



Prosiding Seminar Leksikografi Indonesia (SLI) 2018

Prosiding Seminar Leksikografi Indonesia

LEKSIKOGRAFI DI ERA DIGITAL

1-3 Agustus 2018

Hotel Santika Premiere Slipi, Jakarta

2019

Pusat Pengembangan dan Pelindungan Bahasa dan Sastra
Badan Pengembangan Bahasa dan Perbukuan
Kementerian Pendidikan dan Kebudayaan

ISBN 978-602-437-780-9



9 786024 377809

**PROSIDING
SEMINAR LEKSIKOGRAFI INDONESIA
1—3 AGUSTUS 2018
Hotel Santika Premiere, Slipi, Jakarta**

LEKSIKOGRAFI DI ERA DIGITAL

**BADAN PENGEMBANGAN BAHASA DAN PERBUKUAN
KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
JAKARTA
2019**

PROSIDING SEMINAR LEKSIKOGRAFI INDONESIA

“Leksikografi di Era Digital”

Panitia Pelaksana:

- Pengarah : Dr. Hurip Danu Ismadi, M.Pd.
- Ketua : Dr. Dora Amalia
- Wakil Ketua : Azhari Dasman Darnis, M.Hum.
- Sekretaris : Denda Rinjaya, S.S.
- Permakalahan : Dira Hildayani, S.S.
Dewi Khairiah, S.S.
Vita Luthfia Urfa, M.Hum.
Dina Alfiyanti Fasa, M.Hum.
- Persidangan : Winda Luthfita, S.Hum.
Denny Adrian Nurhuda, S.Pd.
Ambiya Ikrami Adji, S.Hum.
Kunkun Purwati, S.Pd.
Ilham Nuril Huda, S.Kom.
Ibnu Kharish, S.Hum.
M. Rofid Hilmi, S.Kom.
- Kesekretariatan : Selly Rizki Yanita, S.Hum.
Retno Indarwati, S.S.
Nikita Daning Pratami, S.S.
Nadia Almira Sagitta, S.Hum.
Vita Muflihah Fitriyani, S.Hum.
Rizki Maulida Agustin, S.Hum.

Steering Comitee:

Dr. Hurip Danu Ismadi, M.Pd.

Dr. Dora Amalia

Azhari Dasman Darnis, M.Hum.

Deny Arnos Kwary, Ph.D.

Dr. Felicia Nuradi Utorodewo

Totok Suhardijanto, Ph.D.

Apolonius Lase, S.Sos.

Reviewer:

Dr. Dora Amalia

Azhari Dasman Darnis, M.Hum.

Deny Arnos Kwary, Ph.D.

Dr. Felicia Nuradi Utorodewo

Totok Suhardijanto, Ph.D.

Apolonius Lase, S.Sos.

Dra. Cormentyana Sitanggang

Dra. Hari Sulastri, M.Pd.

Dra. Menuk Hardaniwati, M.Pd.

Editor:

Elvi Suzanti, S.S., M.Pd.

Triwulandari, M.Hum.

Dira Hildayani, S.S.

Dewi Khairiah, S.S.

Vita Luthfia Urfa, M.Hum.

Dina Alfiyanti Fasa, M.Hum.

Penerbit:

Badan Pengembangan Bahasa dan Perbukuan

Jl. Daksinapati Barat 4 No.11, RT.11/RW.14

Rawamangun, Kec. Pulo Gadung, Kota Jakarta Timur

Daerah Khusus Ibukota Jakarta 13220

Telp. (021) 4750406

Pos-el badanbahasa@kemdikbud.go.id

Cetakan pertama, September 2019

Hak cipta dilindungi undang-undang.

Dilarang memperbanyak karya tulis ini dalam bentuk apapun

tanpa izin tertulis dari penerbit

BB

499.210 6

PRO

p

Prosiding Seminar Leksikografi Indonesi 1—3 Agustus 2018

Hotel Santika Premiere, Slipi, Jakarta: Leksigrafi di Era Digital /

Hurip Danu Ismadi dkk. (Komite Pelaksana); Elvi Suzanti dkk. (Ed.);

Jakarta: Badan Pengembangan Bahasa dan Perbukuan, 2019. vii, 261 hlm.;
28 cm.

ISBN 987-602-437-339-9

BAHASA INDONESIA – TEMU ILMIAH

KATA PENGANTAR

Perkembangan teknologi informasi yang ikut menandai hadirnya Revolusi IV telah membawa dampak besar dalam berbagai bidang kehidupan. Perkembangan teknologi tersebut turut memengaruhi praktik leksikografi di Indonesia. Dari segi penyusunan, pemanfaatan teknologi dapat terlihat dari pemanfaatan korpus, otomatisasi proses leksikografis (penyuntingan), pengumpulan data melalui metode urun daya (*crowdsourcing*), dan penggunaan aplikasi *dictionary writing system* (DWS). Dari segi penggunaan, kecenderungan penggunaan kamus digital menjadi motivasi utama dalam pengembangan kamus berbasis daring dan luring. Pengembangan kamus tersebut diharapkan dapat mendukung aksesibilitas kamus untuk semua masyarakat.

Seminar Leksikografi Indonesia 2018 mengangkat tema “Leksikografi di Era Digital”. Tujuan utama pemilihan tema ini adalah memperkuat kesadaran masyarakat terhadap pemanfaatan teknologi informasi dalam pekerjaan leksikografis serta perlindungan data digital produk perkamusan. Melalui tema di atas, beberapa subtema diturunkan. Subtema tersebut adalah:

1. Aspek-Aspek Penyusunan Kamus Elektronik;
2. Pemanfaatan Korpus dalam Penyusunan Kamus;
3. Pengembangan Aplikasi untuk Kamus Elektronik;
4. Aspek Hukum dalam Pemanfaatan Data Kamus Elektronik.

Prosiding seminar ini merupakan kumpulan makalah-makalah yang dipaparkan selama Seminar Leksikografi Indonesia. Prosiding ini disusun sebagai bentuk dokumentasi kontribusi intelektual para pekamus dan penggiat bahasa dalam penggunaan teknologi untuk mendukung pengembangan kamus, memberikan pengetahuan leksikografis, dan memotivasi ekosistem kebahasaan, baik pekamus, penggiat bahasa, serta masyarakat luas dalam mengembangkan produk-produk kamus dan praktik leksikografi berkualitas di masa depan.

Ketua Pelaksana Seminar Leksikografi Indonesia 2018

DAFTAR ISI

Kata Pengantar	vi
Daftar Isi	vii
KBBI Daring dan Pengembangan Leksikografi Digital di Indonesia Ian Kamajaya	1
Tatanan Hukum Pemanfaatan Data Kamus Elektronik Dr. Houtman, M.Pd. dan Suryati, S.H., M.H.	29
Persoalan Hukum dan Pemberitahuan Hak Cipta dalam Pemanfaatan Korpus Elektronik Muhammad Fadli Muslimin	37
Analisis Kualitas UX KBBI Luring pada Aplikasi Telepon Genggam: Tinjauan Fungsi dan Penggunaan Kamus Elektronik Satwiko Budiono.....	51
Penerapan Dekomposisi Leksikal dalam Penyusunan Definisi: Studi Kasus KBBI Daring Dessy Irawan, Dien Rovita, dan Totok Suhardijanto	63
Pengembangan Kamus Morfologi dan Aplikasi Konkordansi Daring Bahasa Indonesia dan Malaysia Sumber Terbuka: Malindo Morph dan Malindo Conc David Moeljadi	77
Kompetensi dan Kesiapan Leksikografi dalam Era Digital Rosida Tiurma Manurung	85
Kamus Digital Bahasa Indonesia Masa Depan Mengacu pada Lima Kamus Digital Bahasa Inggris Ternama Almira F. Artha dan Rosita Masfirotul Uyun	95
Pengembangan Aplikasi Kamus Daring Tiga Bahasa Sri Andayani, Sugeng Edy Mulyono, dan Ahmad Izzuddin.....	107
Pembentukan Kamus Bahasa Jawa Berdasarkan Tingkat Bahasa untuk Anak Sekolah (SD, SMP, SMA) Berbasis Korpus Afwin Sulistiawati	117
Perancangan Kamus Saku Register Haji dan Umrah Berbasis Korpus Menggunakan Aplikasi Lexique Pro Muhammad Ridwan, S.S., M.A. dan Azmil Fikri Abdil Haq.....	127

Penyusunan Kamus Sinonim Bahasa Jawa: Penentuan Sinonim Berbasis Korpus dalam Deskripsi Leksikografis Bahasa Jawa Atin Fitriana, Totok Suhardijanto, Dien Rovita.....	141
Pengembangan Kamus Pengucapan Berbasis Gawai untuk Pemelajar Bahasa Indonesia di Tiongkok Hery Yanto The, Latifah, Batari Oja Andini	153
Korpus Data Bahasa Indonesia Dalam Teks Indo-Belanda Namira Choirani Fajri	165
Pemanfaatan Korpus dalam Penyusunan Kamus Konstruksi Preposisional Totok Suhardijanto dan Miel Slager	179
Keragaman Kosakata dalam Korpus Sastra: Sebuah Kontribusi untuk Pengembangan Leksikografi Indonesia Neneng Nurjanah	187
“Si Kamper” Sebagai Upaya Pengenalan Kosakata Baru Bahasa Indonesia kepada Siswa Sekolah Menengah Pertama Ilmatas Sa’diyah, S.Pd., M.Hum.	199
Pemanfaatan <i>Natural Language Processing</i> (NLP) untuk Perancangan Tesaurus Digital Tri Wahyu Retno Ningsih dan Debyo Saptono	211
Rancang Bangun Kamus Html 5 Berbasis Web Ekohariadi, Dodik Arwin Dermawan, dan Setya Chendra Wibawa.....	223
Penyusunan Kamus Digital IELTS Berbasis Android Homsatun Nafiah dan Ahmad Afandi	235
Aplikasi Kamus Elektronik yang Ramah bagi Pengguna Ponsel Pintar Penyandang Tunanetra Aptia Ardiasri.....	249
Algoritma Pelepasan Sisipan pada Proses Stemming Bahasa Indonesia untuk Pembentukan Kata Dasar Yeni Anistyasari, Ekohariadi, dan Ricky Eka Putra	257
Aplikasi Transkripsi Fonetik Bahasa Indonesia Berbasis AI (<i>Artificial Intelligence</i>) untuk Pembelajar BIPA Endang Sartika dan Anggy Eka Pratiwi.....	267

KBBI DARING DAN PENGEMBANGAN LEKSIKOGRAFI DIGITAL DI INDONESIA

Ian Kamajaya

ASTrio Tech Pte Ltd, Singapore

Pos-el: ian.kamajaya@gmail.com

Abstrak

Sejak diluncurkan pada tanggal 28 Oktober 2016, Kamus Besar Bahasa Indonesia Dalam Jaringan (KBBI Daring) telah menjadi ikon pergeseran alur kerja leksikografi di Indonesia dari era manual ke era digital. Dengan dukungan kuat dari Badan Bahasa dan dari masyarakat pengguna/penyumbang usulan bagi Kamus Besar Bahasa Indonesia (KBBI), KBBI Daring telah menjadi platform kerja leksikografi singular bagi pengembangan bahasa Indonesia secara daring di seluruh dunia. KBBI Daring telah menggantikan cara kerja pengumpulan data, pemrosesan data, penyusunan format, pengoreksian, pencetakan, dan penyajian KBBI secara manual – menghilangkan berbagai kelemahan dan kesalahan manusiawi yang diakibatkan oleh cara kerja manual – dan mempercepat pengayaan kosakata bahasa Indonesia – secara khusus, KBBI – secara signifikan melalui pemrosesan usulan penyumbang pengembangan bahasa Indonesia di seluruh dunia.

KBBI Daring telah menjadi kontribusi besar bagi kemajuan leksikografi digital di Indonesia sampai saat ini, tetapi belum secara maksimal. KBBI Daring berpotensi, namun belum, menjadi sebuah platform yang mencakup kebutuhan pengembangan leksikografi digital secara menyeluruh di Indonesia. Sebagai platform yang masih balita, saat ini, KBBI Daring hanya memiliki cakupan yang sangat terbatas: berfokus pada pengembangan dan penyempurnaan KBBI secara digital dan secara daring. Melalui perawatan dan pengembangan secara benar, KBBI Daring memiliki potensi besar agar dapat dimanfaatkan secara lebih maksimal bagi pengembangan leksikografi digital di seluruh Indonesia.

Makalah ini disusun membahas pencapaian KBBI Daring sampai saat ini, usulan rencana pengembangan KBBI Daring ke depan, kendala-kendala yang sedang dan yang mungkin akan dihadapi dalam pengembangan KBBI Daring selanjutnya, serta saran-saran untuk mengatasinya.

Kata Kunci: KBBI, KBBI Daring, Leksikografi Digital, Kamus Daring

I. INTRODUKSI KBBI DARING

KBBI adalah kamus resmi bahasa Indonesia yang dipublikasikan oleh Badan Pengembangan dan Pembinaan Bahasa (atau pendeknya, Badan Bahasa) di bawah naungan Kementerian Pendidikan dan Kebudayaan Republik Indonesia (Kemdikbud RI). KBBI merupakan kamus bahasa Indonesia terlengkap, terakurat, dan yang digunakan sebagai acuan tertinggi penggunaan bahasa Indonesia baku yang pernah diterbitkan di bawah naungan Kemdikbud RI. KBBI edisi kelima (KBBI 5) yang diluncurkan pada tanggal 28 Oktober 2016 merupakan

KBBI edisi termutakhir yang diterbitkan oleh Badan Bahasa dalam tiga format: cetak¹, daring², dan luring³.



© 2016 Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan Republik Indonesia

Gambar 1: Laman Utama KBBI Daring

KBBI Daring pertama kali diluncurkan secara resmi oleh Badan Bahasa pada tahun 2006 berdasarkan data KBBI Edisi Ketiga (KBBI 3) dan untuk kedua kalinya pada bulan Juli 2016 berdasarkan data KBBI edisi Keempat (KBBI 4). Kedua KBBI Daring yang mula-mula diluncurkan tersebut hanya memiliki satu fungsi dasar, yaitu pencarian kata – beserta makna dan contohnya – dalam KBBI secara daring. Pada tanggal 28 Oktober 2016, bertepatan dengan peringatan hari Sumpah Pemuda ke-88, KBBI Daring berbasis data KBBI 5⁴ diluncurkan oleh Badan Bahasa. Gambar 1 menunjukkan laman utama KBBI Daring berbasis data KBBI 5.

Berbeda dengan pendahulu-pendahulunya, KBBI Daring berbasis data KBBI 5 didesain bukan hanya untuk *menyajikan* KBBI secara daring, melainkan juga untuk *menyusun* KBBI secara daring. Di samping memiliki fungsi dasar *pencarian* lema dalam KBBI secara daring, KBBI Daring berbasis data KBBI 5 dilengkapi dengan kekayaan fitur yang secara signifikan mengubah proses *penyusunan* KBBI – yaitu pengumpulan data, pemrosesan data, penyusunan format, pengoreksian, pencetakan, dan penyajian KBBI – dari proses manual ke proses digital dan secara daring, tidak terbatas oleh tempat atau lokasi kerja. Dalam konteks tulisan makalah ini selanjutnya, kecuali diberikan kualifikasi secara khusus, frasa “KBBI Daring” akan dipakai secara eksklusif merujuk pada KBBI Daring yang terakhir diluncurkan. Makalah ini disusun membahas pencapaian KBBI Daring sampai saat ini dalam hubungannya dengan pengembangan leksikografi di Indonesia, potensi pengembangan KBBI Daring ke depan, kendala-kendala yang sedang dan mungkin akan dihadapi dalam pengembangan KBBI Daring selanjutnya, serta saran-saran untuk mengatasinya. Pada bab dua makalah ini,

¹ diterbitkan secara terbatas pada bulan Oktober 2016 dan secara massal pada bulan April 2017

² dimutakhirkan setiap enam bulan sekali, terakhir kali pada bulan April 2018

³ ibid

⁴ dapat diakses melalui alamat laman <https://kbbi.kemdikbud.go.id>

pengembangan digitalisasi leksikografi di Indonesia yang telah dikerjakan melalui KBBI Daring akan dipaparkan. Selanjutnya, bab tiga makalah ini akan membahas potensi, kendala, dan saran dalam mengatasi kendala pengembangan KBBI Daring di masa yang akan datang, terutama dilihat dari sudut pandang kapasitas teknik informasi komputer dan contoh praktik dari kamus daring lain yang sudah ada saat ini. Isi makalah akan ditutup dengan bab empat yang berisikan kesimpulan keseluruhan makalah. Bab lima makalah ini berisikan ucapan terima kasih pada semua perorangan yang telah membantu dalam penyelesaian ini dan bab enam makalah ini berisikan daftar referensi yang digunakan sepanjang makalah ini.

II. DIGITALISASI LEKSIKOGRAFI OLEH KBBI DARING

Pekerjaan editorial penyusunan KBBI mencakup pengumpulan data, pemrosesan data (mencakup pengurutan, penyeleksian, pengecekan – seperti pengecekan akurasi ejaan, definisi, dan contoh –, dsb.), penyusunan format kamus (seperti pengerjaan tata letak, format tulisan, ukuran kertas, dsb.), pengoreksian, pencetakan, dan penyajian kepada masyarakat. Bagaimana KBBI Daring bersumbangsih dalam mengubah proses penyusunan KBBI dari manual ke digital dan secara daring akan dipaparkan secara mendetail dalam bagian selanjutnya pada bab ini.

2.1. Penyusunan KBBI Sebelum KBBI Daring

Sebelum KBBI Daring diluncurkan, semua pekerjaan editorial penyusunan KBBI dikerjakan secara manual dan dibatasi oleh tempat atau lokasi kerja yang berpusat pada (kantor) Badan Bahasa:

- **Pengumpulan data** dilakukan oleh tim redaksi KBBI di Badan Bahasa dengan mengumpulkan sumber data seperti majalah, surat kabar, dan karya tulisan untuk menemukan kata-kata baru yang tidak terdapat dalam KBBI dengan menelusuri daftar kata-kata dalam KBBI cetak. Kata-kata baru yang ditemukan kemudian dituliskan pada potongan kertas/kartu atau slip untuk kemudian diproses lebih jauh. Sementara itu, masyarakat luar yang bukan bagian dari tim redaksi KBBI namun ingin turut berpartisipasi dalam pengumpulan data KBBI mencatat atau mengetik, bahkan terkadang sekedar menyampaikan dalam pertemuan Badan Bahasa dengan publik seperti dalam seminar, kata-kata yang diusulkannya dan mengirimkannya kepada Badan Bahasa melalui pos atau pos-el. Sebagian data kiriman ini tidak jarang tidak sampai kepada, atau lupa/tidak direkam oleh Badan Bahasa maupun “hilang dalam lautan pos-el” yang diterima oleh Badan Bahasa. Sejak diluncurkannya KBBI Daring berbasis data KBBI 3, pengecekan entri sudah mulai dapat dilakukan secara digital dengan memakai KBBI Daring tersebut. Akan tetapi, pengecekan tersebut masih belum maksimal karena KBBI Daring berbasis data KBBI 3 tersebut tidak memiliki pangkalan data dengan struktur/format yang sesuai dengan jenis data dalam KBBI dan tidak memiliki fitur pencarian lanjut (atau penyaringan data) selain dengan memasukkan dan mendapatkan kata yang dicari. Pengumpulan data juga mulai dilakukan tim redaksi menggunakan media pencarian digital, seperti *Google*, akan tetapi belum benar-benar ada platform digital resmi dan sesuai untuk digunakan

mengusulkan data berupa entri/makna/ccontoh untuk masuk ke dalam KBBI. Hasil pengumpulan data yang dilakukan dimasukkan ke dalam suatu fail komputer pada suatu komputer tertentu yang hanya dapat diakses oleh orang-orang yang berada pada lokasi di mana komputer tersebut berada. Data tidak dikumpulkan secara tersentralisasi dalam bentuk usulan digital pada pangkalan data pada sebuah peladen (*server*) yang dapat ditambahkan atau diubah secara daring melalui proses editorial yang sah yang dilakukan secara daring.

- **Pemrosesan data** mencakup penyortiran/pengurutan data⁵, penyeleksian, dan pengecekan kualitas, memastikan kata-kata baru tersebut memenuhi kualifikasi untuk dimasukkan ke dalam KBBI. Pemrosesan data ini biasa dikerjakan melalui pembagian tugas dan “tahap pemrosesan” melalui seorang yang telah dipercaya melakukan suatu “tahap pemrosesan” data biasanya akan bertanggung jawab penuh terhadap proses yang dilakukannya. Misalnya, seorang dipercayakan suatu “tahap pemrosesan” pada kata-kata dengan huruf depan A sampai E untuk menentukan apakah semua kata tersebut sudah memiliki pemenggalan yang benar. Atau seorang yang lain, mungkin dipercayakan untuk menyunting kumpulan kata yang berasal dari bidang ilmu Biologi karena kelebihan tambahan yang dimiliki oleh orang tersebut sebagai penggemar biologi di samping ia adalah seorang ahli bahasa. Perlu dicatat bahwa cara pemrosesan seperti ini memerlukan ketelitian yang sangat besar dari editor yang melakukan pemrosesan tersebut karena terbatasnya akses orang lain selain dirinya dalam memproses data tersebut. Misalnya, definisi dari sebuah istilah baru yang sedang diproses oleh seorang editor akan hanya dapat diakses oleh editor tersebut – atau terkadang ditambah segelintir orang lain yang memiliki akses ke catatan tersebut –, bukan dapat diakses oleh setiap editor.
- **Penyusunan format kamus** kemudian dilakukan pada kata-kata yang telah lulus proses pemrosesan data (seleksi editor). Kata-kata tersebut diketik di dalam fail khusus di sebuah komputer di Badan Bahasa – dalam bentuk *Microsoft Excel* (.xls/.xlsx) dan *Word* (.doc/.docx) – yang mencatat semua data KBBI dengan “format kamus” yang sesuai bagi standar KBBI. Sekali lagi, hal ini biasa dikerjakan dengan pembagian tugas karena banyaknya kata-kata yang perlu disusun dalam “format kamus” yang sesuai untuk KBBI. Penyusunan format ini menyangkut banyak hal seperti ukuran huruf yang digunakan, margin halaman, penggunaan huruf tebal dan/atau huruf miring, urutan makna, pemotongan kata pada akhir baris, urutan peletakan label pada makna, pemberian nomor pada entri yang memiliki lebih dari satu makna, penempatan contoh dari sebuah makna, dan sebagainya. Penyusunan “format kamus” semacam ini harus dilakukan per entri dan per makna oleh semua yang melakukan penyusunan format secara konsisten. Jika ada satu orang saja yang salah dalam menyusun format, maka kesalahan tersebut akan sangat sulit ditelusuri. Gambar 2 menampilkan contoh kata “Indonesia” dan kata-kata turunannya yang dicetak dalam “format kamus”.

⁵ pada pembuatan KBBI edisi pertama dan kedua, misalnya, kumpulan potongan kertas/kartu tersebut kemudian diurutkan secara alfabetis, dan ditaruh di dalam lemari katalog khusus

In.do.ne.sia /indonesia/ *n* **1** nama negara kepulauan di Asia Tenggara yg terletak di antara benua Asia dan benua Australia; **2** bangsa, budaya, bahasa yg ada di negara Indonesia;
meng.in.do.ne.si.a.kan *v* **1** menjadikan Indonesia (tt kata): ~ *kata-kata asing*; **2** menempelkan ke dalam bahasa Indonesia;
peng.in.do.ne.si.a.an *n* proses, cara, perbuatan mengindonesiakan; indonesianisasi;
ke.in.do.ne.si.a.an *n* perihal Indonesia; yg bersangkutan paut dng Indonesia: *bagaimana mengembangkan kebudayaan yg berakar dan berada dalam jiwa* ~

Gambar 2 Contoh Kata yang Dicitak dalam “Format Kamus”

- **Pengoreksian** dilakukan oleh tim redaksi untuk memastikan bahwa data di dalam fail berisikan data KBBI tersebut benar, berada dalam “format kamus” yang benar, dan tidak ada kata-kata yang sebenarnya udah layak masuk KBBI, namun “tertinggal” (tidak dimasukkan dalam fail tersebut). Setelah melalui pengoreksian ini, jika masih ada kata-kata yang memiliki kesalahan definisi, kesalahan format, atau “tertinggal” (lupa diproses untuk dimasukkan ke dalam KBBI), kata-kata tersebut akan tidak dapat lagi dimasukkan/dikoreksi dalam KBBI yang telah dicetak selama bertahun-tahun sampai penerbitan KBBI edisi berikutnya⁶. Proses penyusunan “format kamus” dan pengoreksian yang dilakukan halaman per halaman dan baris per baris oleh tim redaksi ini tidak jarang memerlukan waktu kerja yang padat selama beberapa bulan. Dapat dilihat bagaimana hal ini sangat memberatkan pekerjaan tim redaksi yang melakukan semuanya ini secara manual, tanpa bantuan mesin.



Gambar 3 Kamus Besar Bahasa Indonesia Edisi Kelima Versi Cetak

- **Pencetakan** KBBI kemudian dilakukan sesudah tahap pengoreksian selesai. Fail di mana semua data kamus tercatat tersebut akan dibenahi (atau terkadang diubah) sekali lagi dan dipastikan berada dalam format yang siap cetak⁷ – bukan hanya dalam “format kamus” yang sesuai dengan KBBI, tetapi juga dalam “format cetak” – kemudian barulah fail tersebut dikirimkan kepada percetakan untuk dicetak. Gambar 3 menampilkan KBBI edisi kelima versi cetak.

⁶ Badan Bahasa dapat menerbitkan buku suplemen berupa daftar kata-kata dan makna-makna yang dikoreksi, namun tidak dapat mengoreksi KBBI yang telah dicetak itu sendiri

⁷ Tergantung dari permintaan badan percetakan, fail dalam format yang siap cetak ini memiliki beberapa ketentuan tersendiri yang lebih spesifik. Satu percetakan, misalnya, dapat meminta bentuk fail komputer yang sangat spesifik – yaitu *INDesign Document (INDD)*, bentuk fail komputer yang dikenali oleh program komputer *Adobe InDesign* – untuk menggunakan jasa percetakannya, sedangkan satu percetakan lain dapat menawarkan jasa pembuatan format siap cetak dari bentuk naskah cetak yang relatif “lebih mentah” seperti *Word* atau *PDF*.

- **Penyajian** KBBI sebelum KBBI Daring berbasiskan data KBBI 3 hanya dilakukan dalam bentuk cetak, namun penerbitan KBBI versi cetak ini biasanya tidak langsung diikuti oleh penerbitan KBBI Daring berbasiskan data KBBI versi cetak tersebut⁸, sehingga selama masa-masa ini, pengguna KBBI harus bergantung pada ketersediaan versi cetaknya. KBBI cetak yang diterbitkan dijual di toko buku atau dikirimkan (terkadang secara gratis) kepada badan-badan pemerintah, sekolah-sekolah, maupun organisasi-organisasi khusus lainnya sesuai dengan kebijakan Badan Bahasa. Dengan metode penyajian seperti ini, tidak jarang diperlukan waktu yang cukup panjang untuk sebuah KBBI cetak sampai ke tangan pengguna yang memerlukan, sebagaimana proses pengiriman dan penjualan barang biasa terjadi. Selain itu, masyarakat yang dapat menggunakan KBBI sangat terbatas pada mereka yang mempunyai KBBI cetaknya saja atau yang mampu membelinya.

Dapat dilihat bahwa dalam setiap tahap proses penyusunan KBBI di atas, kesalahan manual dapat terjadi. Apakah itu berupa hilangnya usulan kata untuk masuk ke KBBI dalam proses pengumpulan data, salah diurutkan pada waktu pemrosesan, memiliki format yang tidak konsisten pada waktu disusun dalam “format kamus”, kurang teliti dalam pengoreksian, tidak konsisten dalam bentuk format siap cetaknya, maupun mengalami kerusakan atau hilang dalam perjalanan pada waktu disajikan dalam bentuk cetak, semuanya memberikan dampak pada kualitas KBBI yang disajikan. Selain itu, dapat dilihat bahwa proses penyusunan KBBI yang dikerjakan secara manual tersebut membutuhkan tenaga kerja (tim redaksi) dengan dedikasi dan ketelitian yang ekstra besar untuk menjaga kualitas di dalam setiap tahapan proses penyusunan KBBI yang sebenarnya hampir semua dapat dibantu, atau bahkan dikerjakan dengan lebih baik, oleh mesin. Semua proses tersebut juga dapat dilihat berpusat secara signifikan pada Badan Bahasa. Masyarakat umum memiliki akses yang sangat terbatas untuk berkontribusi dalam menyusun KBBI, bahkan di dalam bagian yang paling kecil seperti mengusulkan kata.

2.2. Penyusunan KBBI Menggunakan KBBI Daring

KBBI Daring didesain untuk membantu dan/atau melenyapkan kelemahan manual dalam proses penyusunan KBBI sebagaimana disajikan dalam bagian awal makalah ini. Fitur KBBI Daring telah dijelaskan secara terperinci dalam (Kamajaya, Moeljadi, & Amalia, 2017). Pada makalah ini, bagaimana fitur-fitur tersebut digunakan untuk membantu dan/atau melenyapkan kelemahan manual dan keterbatasan lokasi dalam proses penyusunan KBBI akan dipaparkan:

⁸ KBBI 4 versi cetak diterbitkan pada tahun 2008, tetapi KBBI Daring berbasiskan data KBBI 4 baru diluncurkan pada bulan Juli 2016, berselang 8 tahun sesudah KBBI 4 dicetak. Demikian juga KBBI 3 versi cetak diterbitkan pada tahun 2005, namun versi daringnya baru tersedia pada tahun 2006.



Ubah Data (Entri)

Jenis Pilihan Lanjut

Jenis Entri Entri Dasar

Entri * daring

Varian Contoh: api; bisa (2) (untuk homonim)

Pemenggalan da.ring

Lafal Contoh: mengéja

Makna/Rujuk Makna

Jumlah Makna 1

+ Tambah - Kurang

Makna #1

Pilihan Lanjut

Makna * dalam jaringan, terhubung melalui jejaring komputer, internet, dsb

Bantuan Pencarian Usulan

Frasa Contoh: jelai, bujang (1) (untuk homonim)

Jenis Entri

Cari

Informasi

Gunakan [Bantuan Pencarian Usulan] di atas untuk memastikan bahwa frasa yang Anda masukkan sebagai [Entri], [Induk Kata], atau [Entri Rujuk] sungguh-sungguh telah/belum terdapat dalam KBBI.

Penjelasan

Tidak seperti [pencarian biasa](#), [Bantuan Pencarian (bagi) Usulan] memiliki aturan yang **lebih ketat** dalam menentukan hasil pencariannya karena ditujukan untuk membantu memastikan usulan yang diberikan memiliki format yang tepat:

1. pencarian frasa [a la carte] (huruf **a** ditulis tanpa diakritik) pada [pencarian biasa](#) akan memunculkan hasil [â la carte] (huruf **â** memiliki diakritik), namun pencarian frasa [a la carte] pada [Bantuan Pencarian Usulan] tidak akan memunculkan hasil [â la carte].
2. pencarian frasa [nya] (tanpa diawali dengan tanda hubung [-]) pada [pencarian biasa](#) akan memunculkan hasil [-nya] (memiliki tanda hubung [-]), namun pencarian frasa [nya] pada [Bantuan Pencarian Usulan] tidak akan memunculkan hasil [-nya].

Gambar 4 Laman Usulan KBBI Daring

Gambar 5 "Meja Redaksi" KBBI Daring

- **Pengumpulan data** dalam KBBI Daring dilakukan dalam bentuk usulan digital, secara daring, dengan format yang terstandarisasi, dan dapat diberikan oleh setiap pengguna terdaftar pada KBBI Daring. Gambar 4 menampilkan laman usulan pada KBBI Daring. KBBI Daring membuat proses pengumpulan data dapat dilakukan secara urun daya oleh siapa saja yang terdaftar dalam laman aplikasi, tidak hanya terbatas dilakukan oleh tim redaksi KBBI yang berada di Badan Bahasa. Pengumpulan data dengan cara ini juga mencegah usulan hilang atau rusak, mengeliminasi variasi format usulan (seorang mungkin mengusulkan dengan format yang satu sedangkan seorang mengusulkan dengan format yang lain), memastikan usulan yang diberikan langsung masuk ke dalam "Meja Redaksi", serta dapat diberikan oleh setiap masyarakat penutur bahasa Indonesia, bukan hanya di Indonesia,

melainkan juga di seluruh dunia secara daring untuk berkontribusi mengembangkan KBBI. Gambar 5 menampilkan fitur “Meja Redaksi” pada KBBI Daring. Pengumpulan data dengan menggunakan format usulan digital semacam ini juga membuat usulan yang diterima tersentralisasi pada pangkalan data KBBI yang dapat diakses langsung oleh tim redaksi melalui “Meja Redaksi”, tidak terpisah-pisah dalam fail komputer masing-masing pembuat dan/atau penerima usulan. Berkat fitur pengusulan dan “Meja Redaksi” secara daring ini, jumlah usulan yang diterima Badan Bahasa telah meningkat drastis semenjak diluncurkannya KBBI Daring.

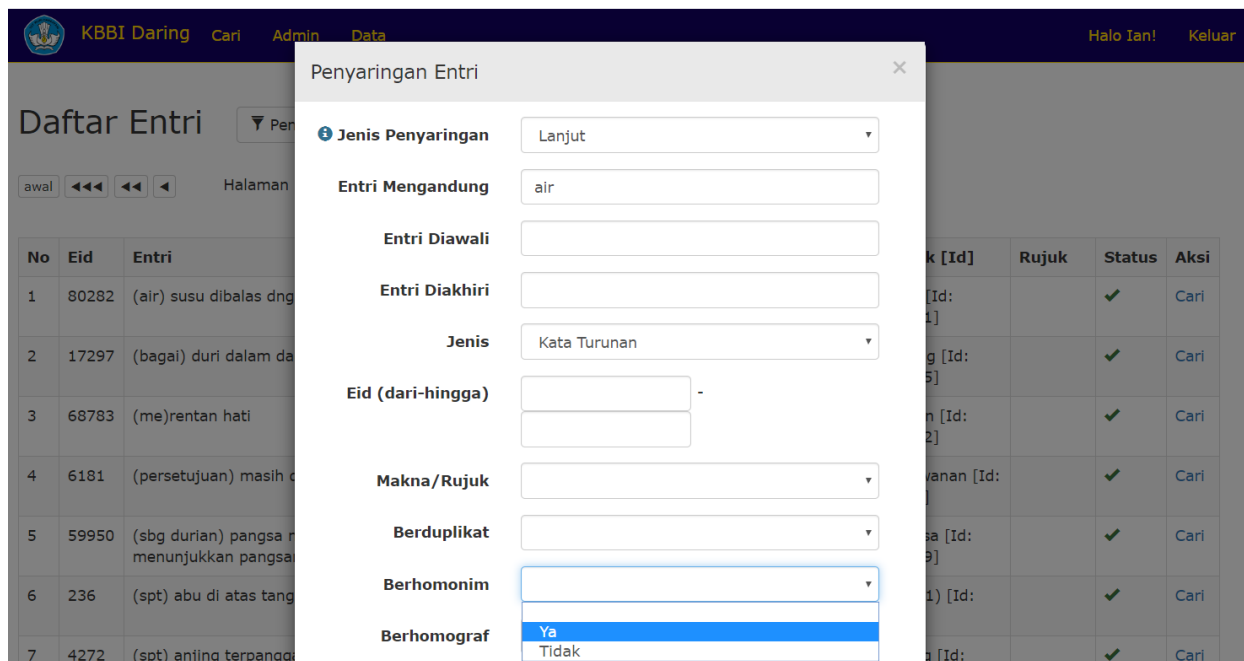
The screenshot shows the 'Daftar Usulan Bagi Editor' (Proposal List for Editors) page on the KBBI Daring website. The page includes a navigation bar with 'KBBI Daring', 'Cari', 'Admin', and 'Data' buttons, and a user greeting 'Halo developer! Keluar'. Below the navigation bar, there is a search filter 'Penyaringan' and a pagination control showing 'Halaman 21 / 1013' and '(Usulan [401] Aroha - [420] arungpitue dari 20256 usulan)'. The main content is a table with 12 columns: No, Entri, Jenis, Pid & Id Tabel, Jumlah, Pengusul & Editor, Redaktur & Validator, Komentar Pengusul, Status, Divalidasi, Aksi, and Ekstra. Three proposals are visible in the table.

No	Entri	Jenis	Pid & Id Tabel	Jumlah	Pengusul & Editor	Redaktur & Validator	Komentar Pengusul	Status	Divalidasi	Aksi	Ekstra
401	Aroha Eid: 135693	+ [icon]	0000041100	1 1 0	Candra Henaulu (2017-11-06 12:47:12.140) (Tidak tersedia)	(Tidak tersedia) (Tidak tersedia)	Rumah adat asal pula...	Diproses (Editor)		Detail Validasi Terima Kembalikan Alihkan/Ubah Tolak Gugurkan Arsipkan Hapus	-
402	aron mulud Eid: 140682	+ [icon]	0000047140	1 1 0	Sunarsih (2018-02-27 10:39:04.960) (Tidak tersedia)	(Tidak tersedia) (Tidak tersedia)	BB Jabar 2018 Sumber...	Diproses (Editor)		Detail Validasi Terima Kembalikan Alihkan/Ubah Tolak Gugurkan Arsipkan Hapus	-
403	arorroq Eid: 118553	+ [icon]	0000022456	1 1 0	Jerniati (2017-10-06 11:33:01.265) (Tidak tersedia)	(Tidak tersedia) (Tidak tersedia)	[bb sulsel] bahasa s...	Diproses (Editor)		Detail Validasi Terima Kembalikan Alihkan/Ubah Tolak Gugurkan Arsipkan Hapus	-

Gambar 6 Laman Daftar Usulan Bagi Editor

- **Pemrosesan data** menggunakan KBBI Daring dikerjakan dalam “Meja Redaksi” digital yang setiap usulannya dapat ditemukan. Gambar 6 menampilkan contoh daftar usulan yang dapat ditemukan dalam “Meja Redaksi”. Pengguna KBBI Daring dengan hak akses editor atau lebih tinggi diberikan hak untuk melihat dan memproses usulan dalam “Meja Redaksi” sesuai dengan layaknya usulan tersebut diproses (misalnya, (i) meneruskan usulan yang baik ke tingkat yang lebih tinggi, (ii) menerima usulan yang sudah diseleksi, atau (iii) menolak usulan yang tidak memenuhi standar kualitas KBBI). Untuk membantu tim redaksi yang pemrosesan usulan, “Meja Redaksi” dilengkapi dengan fitur penyaringan yang kaya akan kriteria pencarian guna membantu tim redaksi menemukan setiap usulan yang memenuhi kriteria pencarian tersebut. Pemrosesan usulan dalam “Meja Redaksi” juga dikerjakan dengan transparan: setiap editor dapat melihat status setiap usulan (kapan usulan tersebut dibuat dan diproses, siapa yang membuat dan memproses usulan tersebut, bagaimana status usulan tersebut saat ini, dsb.), memungkinkan pembagian tugas dan saling-mengawasi di antara tim redaksi untuk dapat dilakukan dengan lebih mudah.
- **Penyusunan format kamus** pada KBBI Daring dilakukan dengan memindahkan data KBBI dari fail dengan format *.doc* dan *.xls* ke pangkalan data. Proses pembuatan pangkalan data KBBI Daring dengan lebih mendetail telah dijelaskan dalam (Moeljadi, Kamajaya, & Amalia, 2017). Dengan menyimpan data KBBI Daring di

dalam pangkalan data – dan bukan di dalam fail komputer khusus –, seorang tidak perlu lagi secara manual dan konsisten menulis data KBBI dalam “format kamus” yang sesuai pada fail dalam komputer mereka untuk menyimpan data KBBI, struktur pangkalan data KBBI Daring-lah yang menjamin setiap data disimpan dalam “format kamus” yang konsisten (sebab setiap data yang dapat dimasukkan ke dalam pangkalan data harus juga telah mengikuti struktur dan jenis data yang sesuai dengan kolom-kolom dan tabel-tabel pada pangkalan data tersebut secara konsisten). Bentuk format kamus cetak (yang mencakup, misalnya, ukuran huruf yang digunakan, margin halaman, penggunaan huruf tebal dan/atau huruf miring, urutan peletakan label pada makna, dsb.) tidak perlu dibuat di dalam tahapan ini, melainkan semua data disimpan dalam bentuk pangkalan data digital yang berstruktur dan cocok dengan sifat penggunaan data KBBI. Pada tahap **pencetakan**, barulah semua data pada pangkalan data digital ini diambil dan diproduksi dalam bentuk fail berisikan naskah dasar yang memiliki format kamus yang konsisten.



Gambar 7 Fitur Penyaringan Entri Secara Lanjut pada (Meja Redaksi) KBBI Daring



Daftar Diagnosis Massal

Penyaringan

awal <<<

Halaman 4 / 14

>>> akhir

(Diagnosis Massal [121] - [160] dari 530 diagnosis massal)

Keterangan Diagnosis

Waktu Mulai: 2018-07-04 12:21:02.130, Waktu Berakhir: 2018-07-04 14:45:07.037, Eid Awal: Terawal, Jumlah Elemen: Semua

No	Id	Entri	Hasil Diagnosis Massal Terakhir	Aksi
121	121	parataksis (Eid: 60429)	<p>Contoh #1-1 (Cid: 16267) Entri ini:</p> <ul style="list-style-type: none"> • besar kecil, tua muda <p>memiliki kesalahan format berikut:</p> <ol style="list-style-type: none"> 1. Contoh ini bukan berasal dari entri yang bersifat afiks, namun tidak memiliki potongan frasa/kata/kalimat yang berasal dari [entri] awalnya, yaitu [parataksis] di dalam kurung siku [...] 2. Contoh ini tidak memiliki potongan frasa/kata/kalimat yang menyerupai [entri] awalnya maupun [varian]nya, yaitu [parataksis] 	Detail Cari
122	122	pasar (1) (Eid: 60679)	<p>Makna #2 (Mid: 71480) Entri ini:</p> <ul style="list-style-type: none"> • Ek kekuatan penawaran dan permintaan, tempat penjual yg ingin menukar barang atau jasa dng uang, dan pembeli yg ingin menukar uang dng barang atau jasa <p>memiliki kesalahan format berikut:</p> <ol style="list-style-type: none"> 1. Makna ini berasal dari entri berjenis [dasar] atau [turunan], yaitu [pasar (1), jenis: dasar], namun tidak memiliki [kelas kata] 	Detail Cari

Gambar 8 Laman Hasil Diagnosis Massal KBBI Daring

Halaman Admin (Fasilitas)

Cetak

Cetak

Hasil

Pengaturan

Pulihkan

Diagnosis & Pemulihan

Bagi Entri

Otomatisasi Lampiran

Buat Id Pencarian

Urutkan Ulang Id

Perbaiki Cid (0 | 0)

Mutakhirkan Entri Populer

Buat Data Cadangan

Diagnosis Massal

Beri Lafal é

Hapus Kelas Gabungan

Singkat Kata Panjang

Panjangkan Kata Singkat

Pemutakhiran KBBI

Mutakhirkan KBBI

Render KBBI

Tampilkan Hasil Render

Gambar 9 Halaman Fasilitas KBBI Daring



Panjangkan Kata Singkat

❓ Fungsi apakah ini?

Fungsi ini digunakan untuk:

- memanjangkan kata-kata "singkat" menjadi bentuk "panjang"nya, sebagaimana lazimnya digunakan dalam KBBI, yaitu:
 - [di] diganti menjadi [dalam]
 - [shg] diganti menjadi [sehingga]

❓ Kapan sebaiknya fungsi ini digunakan?

Gunakan fungsi ini sesekali saja, sesaat ketika Anda hendak membuat naskah mentah KBBI cetak menggunakan aplikasi cetak KBBI Daring di mana pangkalan data sudah dibersihkan sebelumnya.

⚠ Selama sedang berjalan, fungsi ini memakan porsi yang cukup signifikan dari sumber daya perangkat keras peladen KBBI daring. Penggunaan fungsi ini pada jam-jam di mana pemakaian KBBI daring sedang padat dapat mengakibatkan melambatnya respon peladen dalam memenuhi permintaan (pencarian atau pengusulan) penggunaannya. Untuk menjaga performa peladen yang maksimal, jalankan fungsi ini hanya pada jam-jam di mana pemakaian KBBI daring sedang tidak padat.

Gambar 10 Contoh Fasilitas Pemulihan Massal: Pemanjangan Kata yang Disingkat dalam KBBI

- **Pengoreksian** melalui KBBI Daring dibantu oleh fitur penyaringan lanjut, fitur diagnosis otomatis, dan fitur pemulihan otomatis dalam KBBI Daring. Melalui fitur penyaringan lanjut, tim redaksi dapat menemukan kesalahan-kesalahan format yang memiliki pola tertentu, misalnya menemukan entri yang berduplikat, entri berupa kata gabungan yang tidak memiliki induk kata, makna yang tidak memiliki kelas kata, entri yang tidak memiliki pemenggalan kata, definisi yang memiliki kata-kata tidak ditulis dalam bentuk kependekannya, dsb.). Melalui fitur diagnosis otomatis, tim redaksi dapat langsung menemukan kesalahan format yang diidentifikasi oleh sistem tanpa perlu melakukan pengecekan melalui mata satu per satu. Gambar 7 menampilkan penyaringan entri secara lanjut yang disediakan dalam “Meja Redaksi” KBBI Daring. Fitur ini dapat digunakan oleh editor untuk menemukan kumpulan entri yang hendak dicari dengan jauh lebih mudah. Gambar 8 menampilkan laman hasil diagnosis massal pada KBBI Daring. Melalui fitur pemulihan otomatis, data pada KBBI Daring yang memiliki kesalahan format yang bersolusi tunggal⁹ dapat langsung dikoreksi tanpa melibatkan kerja manual tim redaksi. Gambar 9 menampilkan halaman fasilitas yang dapat diakses oleh akun dengan hak “Admin Utama” pada KBBI Daring. Halaman tersebut memuat fitur “Diagnosis Massal” dan beberapa fitur pemulihan massal KBBI Daring seperti “Beri Lafal é”, “Hapus Kelas (kata dalam kata) Gabungan”, “Singkat(kan) Kata (yang ditulis dalam bentuk) Panjang(nya)”, dan “Panjangkan Kata (yang ditulis dalam bentuk) Singkat(nya)”. Gambar 10 menampilkan contoh fasilitas pemulihan massal yang dimiliki oleh KBBI Daring, yaitu untuk memanjangkan kata yang terkadang salah ditulis dalam bentuk singkatannya di dalam KBBI secara massal.

⁹ Data yang memiliki kesalahan format dan *tidak* bersolusi tunggal dapat dideteksi oleh fitur diagnosis otomatis KBBI Daring, tetapi tidak dapat diperbaiki langsung oleh fitur pemulihan otomatis karena kemungkinan solusi dari kesalahan formatnya yang tidak unik

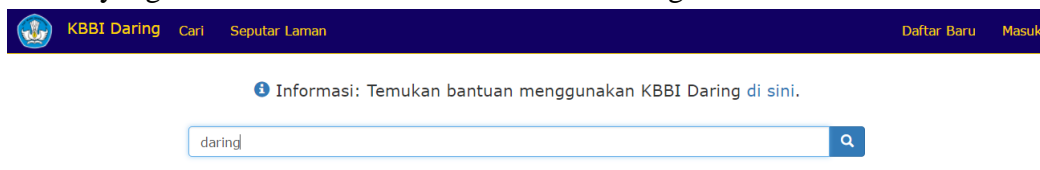
2 ²Abai • abdomen

orang itu ~ anaknya dan tidak mengurusnya lagi; 7 tidak memegang teguh (adat istiadat, aturan, janji); negara itu ~ ketentuan yg telah disepakati bersama;
peng.a.bai *n* orang yg suka mengabaikan (tidak memedulikan); pelengah;
peng.a.bai.an *n* 1 proses, cara, perbuatan mengabaikan (tidak memedulikan, melalaikan); 2 *Huk* tindak kejahatan yg terjadi km tidak melakukan sesuatu yg diharuskan;
abai.an *n* *Sas* penghilangan atau perubahan bagian naskah yg tidak dipahami lagi oleh penyalin;
ter.a.bai *v* tersia-sia; terlalaikan; tidak dipedulikan;
ter.a.bai.kan *v* 1 tidak disengaja diabaikan: *sering hal-hal yg kecil ~*; 2 dapat diabaikan: *bagaimana pun jasa-jasanya tidak ~*
²**Abai** *n* 1 suku bangsa di Provinsi Kalimantan Timur; 2 bahasa yg dituturkan oleh suku Abai
abai.ma.na *n* *ark* 1 lubang pelepasan; dubur; 2 kemaluan
aba.ka *n* tumbuhan yg seratnya dibuat tali belati; *Musa textilis*
ab.ak.si.al *a* *Bio* berada jauh atau menjauhi sumbu panjang tempat melekatnya
abak.ti.nal *a* *Zool* berkenaan dng sisi tubuh yg tidak mengandung mulut
¹**aba.kus** *n* dekak-dekak; swipoa
²**aba.kus** *n* *Ars* lempeng datar di atas kepala tiang dng pinggiran cekung atau beralur

untuk mengurangi (menghilangkan) bunyi (benturan);
meng.a.bar *v* 1 mengurangi (ketegangan, penderitaan): *obat ini dapat ~ penderitaan si sakit*; 2 menghambat; menahan; mengerem;
abar.an *n* *Psi* hambatan kejiwaan;
ter.a.bar *v* terhambat; tertahan: *tadi saya ~ oleh kemacetan lalu lintas*;
ber.a.bar.kan *v* mempunyai dinding penyekat: *bilik itu ~ tepas*
aba.sia *n* *Psi* gangguan koordinasi motorik yg menyebabkan kesulitan di berjalan
abat *n* ikan laut yg berbadan lebar, panjang mencapai 70 cm, hidup di dasar perairan tropis dng kedalaman 150 m, tersebar di perairan Laut Merah dan Afrika Timur sampai Kepulauan Indo-Malaya dan Australia Utara; *Argyrops spinifer*
aba.te /*abaté/* atau /*abate/* *n* serbuk obat pembasmi larva nyamuk
aba.ti.sa.si *n* pembasmi larva nyamuk dng menggunakan abate
aba.to.ar *n* tempat pemotongan hewan untuk umum; pejagalan
abau *n* kura-kura yg hidup di rawa; *Callogau pieta*
Abau *n* 1 suku bangsa yg mendiami daerah sekitar perbatasan Indonesia dan Papua Nugini di Kabupaten Jayapura, Provinsi Papua; 2 bahasa yg dituturkan oleh suku Abau

Gambar 11 Contoh Hasil Cetak KBBI Daring

- **Pencetakan** dilakukan menggunakan fitur cetak KBBI Daring yang dapat diakses oleh akun dengan hak akses “Admin Utama” KBBI Daring. Fitur cetak KBBI Daring menjamin format yang konsisten di seluruh KBBI, menghilangkan kemungkinan format yang tidak konsisten karena dikerjakan halaman per halaman, baris per baris, oleh tim redaksi secara manual dengan mengandalkan ketelitian mata dan pembagian tugas. Fitur cetak KBBI Daring juga mengurangi secara drastis waktu persiapan yang diperlukan untuk mendapatkan naskah siap cetak KBBI dengan menyediakan naskah cetak dasar (yang hanya memerlukan usaha minimum untuk disempurnakan oleh tim redaksi) dengan format kamus yang konsisten hanya dalam waktu beberapa jam setelah fitur cetak dijalankan¹⁰. Gambar 11 menampilkan contoh hasil naskah cetak dasar yang dihasilkan oleh fitur cetak KBBI Daring.



da.ring

n akr dalam jaringan, terhubung melalui jejaring komputer, internet, dan sebagainya

Gambar 12 Contoh Hasil Pencarian Kata pada KBBI Daring

- **Penyajian** KBBI dilakukan dalam bentuk daring, melalui jaringan internet. Dengan demikian, KBBI dapat diakses kapan saja, di mana saja, dan dengan segera oleh setiap orang yang terhubung dengan jaringan internet. Hal ini mengeliminasi kendala yang diakibatkan oleh biaya, waktu, dan risiko barang rusak/berkurang kualitas yang biasa diakibatkan oleh pengiriman barang secara fisik. Satu-satunya keterbatasan yang

¹⁰ naskah cetak dasar KBBI 5 versi cetak dibuat oleh fitur cetak KBBI Daring dalam waktu 2 jam, tetapi tim redaksi biasanya memerlukan waktu beberapa minggu untuk menyempurnakan naskah cetak dasar tersebut menjadi versi siap cetak

diakibatkan oleh hal ini adalah seorang harus terhubung pada jaringan internet untuk dapat mengakses KBBI Daring. Sebagai solusi dari limitasi ini, Badan Bahasa telah menyediakan versi luring dari KBBI untuk sistem operasi *Android*¹¹ dan *iOS*¹². Gambar 12 menampilkan contoh hasil pencarian kata pada KBBI Daring. KBBI Daring (saat ini berbasis data KBBI 5) juga dapat digunakan dengan lebih tetap dibandingkan KBBI Daring versi-versi sebelumnya, karena kelengkapan fiturnya dalam setiap tahap penyusunan KBBI. Penggunaan KBBI Daring secara lebih tetap ini membuat KBBI Daring menjadi versi KBBI yang termutakhir, tidak seperti KBBI Daring versi-versi sebelumnya¹³.

2.3. Pencapaian dan Penggunaan KBBI Daring Sampai Saat Ini

Sejak peluncurannya, KBBI Daring telah menjadi ikon pergeseran alur kerja leksikografi di Indonesia dari era manual ke era digital. KBBI Daring menyatakan keunggulannya sebagai aplikasi kamus digital dalam hal mendigitalisasikan dan mendinginkan seluruh proses *pembuatan, penyuntingan, pengoreksian, dan pencetakan* kamus, bukan hanya melayani *pencarian* kata melalui suatu aplikasi kamus digital. Pada bulan Juli 2018, KBBI Daring telah merekam lebih dari 20 juta pencarian dengan rata-rata 32 ribu pencarian per hari, menerima lebih dari 40 ribu usulan, diakses oleh lebih dari 800 ribu IP unik dari seluruh dunia, dan memiliki lebih dari 24 ribu pengguna terdaftar. Pada bulan April—Mei 2018, laman *Alexa*¹⁴ (dimiliki oleh *Amazon Inc.*) secara konsisten mencatat domain *kemdikbud.go.id*¹⁵ sebagai salah satu di antara 2000 laman paling banyak diakses di seluruh dunia, berada di peringkat 1700-2000 (Sebagai perbandingan, domain *oxforddictionaries.com* berada di posisi 1500-1600 dunia pada bulan-bulan yang sama pada tahun 2018), meningkat pesat dari peringkat 7000-8000 dunia sebelum Oktober 2016. Fitur cetak KBBI Daring telah digunakan untuk mencetak KBBI 5 versi cetak dengan hasil yang memuaskan¹⁶.

KBBI Daring juga telah menjadi platform kerja resmi dan singular penyusunan KBBI. KBBI Daring bukan saja hanya telah digunakan oleh Badan Bahasa, melainkan juga telah diperkenalkan kepada masyarakat umum, disebarluaskan kepada para ahli bahasa Indonesia, dan digunakan oleh Kantor/Balai Bahasa di seluruh Indonesia untuk menyusun KBBI melalui pengusulan dan pemrosesan data untuk pembuatan KBBI. KBBI Daring dimutakhirkan setiap enam bulan sekali. Sampai saat waktu makalah ini ditulis, pemutakhiran KBBI Daring yang terakhir dilakukan pada bulan April 2018. Setiap kali KBBI Daring akan dimutakhirkan, selalu diadakan lokakarya pemutakhiran untuk mendapat masukan dari masyarakat sebelum data KBBI dimutakhirkan. Untuk mengundang partisipasi masyarakat dalam menyumbangkan kosakata ke dalam KBBI, Badan Bahasa juga mengadakan Diseminasi Program Pengayaan Kosakata sejak tahun 2016 di seluruh provinsi di Indonesia dan beberapa universitas dan

¹¹ <https://play.google.com/store/apps/details?id=yuku.kbbi5>

¹² <https://itunes.apple.com/us/app/kamus-besar-bahasa-indonesia/id1173573777?mt=8>

¹³ atau lebih tepatnya, persis terbalik dengan penyajian-penyajian KBBI sebelumnya di mana KBBI Daring *mengikuti* KBBI versi cetaknya, KBBI Daring (yang termutakhir) *mendahului* versi cetaknya

¹⁴ <https://www.alexa.com>

¹⁵ domain di mana laman KBBI Daring, <https://kbbi.kemdikbud.go.id>, dipasang

¹⁶ baik untuk versi cetak terbatas pada bulan Oktober 2016, maupun untuk versi cetak massal pada bulan April 2017

instansi/lembaga. Selain itu, Bengkel Leksikografi menggunakan KBBI Daring diadakan setiap tahun dengan target khususnya adalah editor KBBI untuk meningkatkan kompetensi mereka. KBBI Daring telah menjadi platform kerja resmi dan singular yang digunakan dan disebarluaskan kepada semua golongan yang terlibat dalam penyusunan KBBI: tim redaksi KBBI, para ahli bahasa Indonesia, dan masyarakat luas pengguna/penyumbang pengembangan KBBI. Jelas terlihat bahwa KBBI Daring telah memberikan kontribusi besar bagi pengembangan digitalisasi leksikografi di Indonesia sampai saat ini melalui pengembangan dan penyempurnaan KBBI secara digital dan secara daring, yang memungkinkan pengembangan secara urun daya.

III. POTENSI, KENDALA, DAN SARAN PENGEMBANGAN KBBI DARING

Pada bab II makalah ini, telah dijelaskan bagaimana KBBI Daring telah digunakan sampai saat ini dan telah menjadi ikon pergeseran alur kerja leksikografi di Indonesia dari era manual menuju ke era digital. Telah dijelaskan juga bagaimana KBBI Daring telah membantu, mencegah kesalahan, mempercepat, dan meningkatkan kualitas setiap tahapan pekerjaan editorial penyusunan KBBI di Indonesia, serta bagaimana KBBI Daring telah menjadi platform kerja resmi dan singular untuk tujuan tersebut (penyusunan KBBI). Dapat dilihat dengan jelas bahwa KBBI Daring telah memenuhi tujuan diciptakannya, yaitu untuk mengembangkan dan menyempurnakan KBBI secara digital.

Ditinjau dari sudut pandang kapasitas teknologi informasi komputer yang kita miliki pada zaman ini dan contoh praktik dari kamus daring lain¹⁷, KBBI Daring sesungguhnya masih dapat dikembangkan lebih jauh untuk kemajuan leksikografi digital di Indonesia – melampaui cakupan dasarnya untuk mengembangkan dan menyempurnakan KBBI secara digital. Pada bab ini akan dipaparkan beberapa usulan arah pengembangan KBBI Daring di masa yang akan datang, hanya dengan meninjaunya sesuai dengan kapasitas teknologi informasi komputer saat ini (dengan kata lain, ditinjau secara teori dan dengan mengasumsikan tersedianya sumber daya yang cukup, pengembangan ini secara pasti dapat dilakukan tanpa menunggu ketersediaan teknologi yang baru) dan dengan melihat contoh praktik dari kamus daring lain yang sudah ada.

3.1. Pelayanan dan Lisensi Bagi Peneliti Bahasa dan Pengembang Teknik Informasi Komputer

Salah satu aspek leksikografi adalah aspek penggunaan – bukan hanya penyusunan – kamus. Digitalisasi kamus telah membuka suatu wilayah baru yang sangat luas bagi penggunaan kamus. Salah satu contoh yang paling jelas terlihat dalam wilayah baru ini adalah penggunaan dan analisa menggunakan data kamus secara digital pula. Tidak lagi sebuah kamus semata memiliki data dalam bentuk tertulis yang hanya dapat “dikonsumsi” oleh manusia dengan menggunakan mata untuk membaca dan otak untuk mengerti, tetapi juga

¹⁷ seperti: <https://www.oxforddictionaries.com/>, <https://www.merriam-webster.com/>, atau <https://www.lexilogos.com/>

memiliki data digital yang dapat “dikonsumsi” oleh komputer dengan menggunakan program untuk “membaca” dan “menggunakan” data tersebut.

Pada zaman ini, peneliti bahasa tidak lagi dibatasi dalam proses penelitian mereka dengan hanya menggunakan sumber daya dan alat bantu manual (non-digital). Peneliti bahasa zaman ini dapat menggunakan komputer untuk membantu proses penelitiannya, termasuk juga di dalamnya membuat program bagi penelitiannya. Jika seandainya mereka sendiri tidak mampu membuat program bagi penelitiannya, mereka dapat bekerja sama dengan pengembang teknik informasi komputer yang dapat membantu mereka membuat program untuk membantu mereka melakukan penelitiannya. Saat ini, di Indonesia sudah banyak mahasiswa, peneliti bahasa, dan pengembang aplikasi yang mengajukan permintaan kepada Badan Bahasa untuk mendapatkan akses pada data digital KBBI 5 untuk keperluan penelitian dan/atau pengembangan aplikasi mereka. Hal ini memberikan indikasi kuat bahwa penelitian bahasa menggunakan alat bantu program juga tengah berkembang di Indonesia. Wilayah baru ini, jika diolah dengan benar, dapat berpotensi memicu pengembangan leksikografi di Indonesia secara signifikan.

KBBI Daring dapat memberikan pelayanan kepada kategori masyarakat ini dengan memberikan *Application Programming Interface* (API) dan lisensi penggunaan data bagi yang membutuhkan data digital KBBI bagi penelitian atau pengembangan mereka. Hal ini dapat dengan mudah dikerjakan selama ada aturan yang jelas yang ditetapkan sehubungan dengan cakupan dan aturan penggunaan data yang diberikan. Hal yang sama telah dilakukan oleh kamus daring lain seperti *Oxford*¹⁸ dan *Merriam-Webster*¹⁹. Dengan memberikan API dan lisensi, peneliti bahasa dan pengembang aplikasi dapat memiliki akses pada data KBBI sebagaimana diperlukan untuk materi dan durasi penelitian mereka. Dengan demikian, mereka tidak lagi hanya dapat melakukan penelitian menggunakan sumber daya dan alat bantu manual, tetapi mereka dapat bekerja sama dengan pengembang, memakai program yang telah tersedia, atau memakai komputer dan membuat sendiri program yang mereka inginkan untuk memproses data KBBI yang mereka teliti guna mempercepat dan meningkatkan kualitas penelitian mereka. Pemberian pelayanan dan lisensi bagi peneliti bahasa dan pengembang teknik informasi komputer secara tepat ini pada akhirnya juga akan meningkatkan kualitas leksikografi digital di Indonesia.

3.1.1 Kendala

Kendala utama dalam mengeksekusi rencana pengembangan ini pada saat ini adalah masih lemahnya aturan hukum yang jelas mengenai kepemilikan dan penggunaan data digital KBBI dan ketegasan dalam menindak mereka yang melanggar aturan hukum tersebut di Indonesia. Tanpa aturan hukum yang jelas – siapa saja yang diizinkan untuk menggunakan cakupan data yang mana, untuk tujuan apa, selama berapa lama, dan apa konsekuensinya jika melanggar – dan ketegasan dalam menindak mereka yang melanggar, penyediaan API dan lisensi demi keperluan penelitian dapat dengan sangat mudah disalahgunakan (misalnya, tidak lagi data

¹⁸ <https://www.oxforddictionaries.com/>

¹⁹ <https://www.merriam-webster.com/>

digunakan untuk keperluan penelitian, melainkan untuk keperluan komersial tanpa perizinan yang sah dari Badan Bahasa).

Penyalahgunaan data digital KBBI pernah secara nyata terjadi pada KBBI “daring” berbasis data KBBI 3. Hanya beberapa minggu setelah KBBI “daring” berbasis data KBBI 3 diluncurkan, bermunculan berbagai macam laman/aplikasi yang menggunakan *seluruh* data KBBI 3 tanpa izin dan persetujuan Badan Bahasa. Saat ini, jika kita mencari aplikasi dengan kata kunci “KBBI” untuk *Android* pada *Playstore* atau untuk *iOS* pada *Appstore*, kita dapat langsung mendapati bahwa sebagian besar dari aplikasi tersebut menggunakan data KBBI 3 (sebagian juga menggunakan lambang Tut Wuri Handayani, sebagaimana dipakai oleh Badan Bahasa dalam aplikasi luring resminya berbasis data KBBI 5). KBBI “daring” berbasis data KBBI 3 tidak memiliki fitur pengamanan terhadap pengambilan data – tidak seperti KBBI Daring berbasis data KBBI 5 –, melainkan hanya memiliki fitur pencarian. Sehingga, seorang peretas/perangkak dapat dengan mudah merangkaki laman KBBI “daring” tersebut dan mengambil seluruh data digital KBBI 3 dalam waktu singkat. Ada di antaranya yang bahkan menjual data KBBI 3 yang mereka dapatkan untuk keuntungan komersial. Tetapi karena tidak ada aturan hukum yang jelas maupun ketegasan dalam menindak mereka yang melanggar aturan hukum tersebut, apa yang dikerjakan oleh golongan peretas ini tidak dapat dijerat oleh ketentuan hukum.

Dapat dilihat bahwa aturan hukum dan ketegasan dalam menindak mereka yang melanggar menjadi sangat penting di sini. Tanpanya, pemberian data digital KBBI kepada perorangan/instansi yang lain akan dapat mudah menyimpang menjadi penyalahgunaan data. Pengambilan data tanpa izin memang dapat diatasi dengan menggunakan fitur keamanan sebagaimana diimplementasikan pada KBBI Daring (berbasis data KBBI 5), akan tetapi penindaklanjutan terhadap penyalahgunaan data yang diambil (baik *dengan* maupun *tanpa* izin) hanya dapat dilakukan jika ada aturan hukum yang jelas. Penyediaan API dan lisensi adalah pemberian data dengan izin resmi. Aturan hukum yang jelas dan ketegasan dalam menindak mereka yang melanggar sangat diperlukan untuk menjamin pembagian data digital KBBI boleh sungguh digunakan dengan cara dan untuk tujuan yang benar.

3.1.2.Saran

Sangat disayangkan bahwa secara fakta, di satu sisi, pembuatan aturan hukum yang jelas bagi penggunaan data digital secara umum, bukan hanya bagi penggunaan data digital KBBI, tidak secepat dan semudah itu dapat diselesaikan di Indonesia. Hal ini menjadi lebih rumit dikarenakan tingkat kesadaran masyarakat Indonesia tentang kepemilikan digital yang secara umum masih sangat lemah²⁰. Seandainya aturan hukum yang kuat dan dengan standar yang terbaik sungguh-sungguh dibuat, maka penegasan aturan hukum tersebut juga akan menjadi sangat berat untuk dilakukan karena banyaknya pelanggaran yang akan ditemukan. Di samping itu, juga akan ada reaksi negatif yang kuat dari masyarakat dengan tingkat kesadaran

²⁰ hal ini dapat terlihat misalnya dalam kebiasaan yang ada secara umum di Indonesia: mengonsumsi barang digital berbayar bajakan – tanpa mendapat izin dari perorangan/instansi yang mempunyai hak atas barang digital tersebut dan tanpa membayar perorangan/instansi tersebut –, tanpa merasa bahwa hal tersebut adalah tindakan kriminal

semacam ini terhadap kebijakan yang akan diambil yang juga memerlukan penegasannya sendiri.

Namun di sisi yang lain, pengaturan pembagian data digital KBBI kepada golongan yang sungguh memerlukannya demi kepentingan pengembangan bahasa Indonesia – sebagaimana diindikasikan dengan permintaan untuk mendapatkan akses ke data digital KBBI 5 bagi keperluan penelitian yang terus berlangsung dan terus bertambah – juga diperlukan dengan sangat mendesak. Jika aturan pembagian dan penggunaan data secara resmi tidak segera dibuat, leksikografi digital di Indonesia secara keseluruhanlah yang akan menjadi korbannya. Menimbang dilema yang ada, saran penulis makalah dalam hal ini adalah agar Kemdikbud, secara khusus Badan Bahasa, menempatkan prioritas dalam menangani masalah ini melalui bekerja sama dengan ahli hukum digital sehingga dapat – setidaknya – mempercepat proses penyajian pelayanan kepada golongan masyarakat yang layak dipercayakan menggunakan data digital KBBI ini, yaitu dengan bertanggung jawab dan untuk tujuan pengembangan bahasa Indonesia.

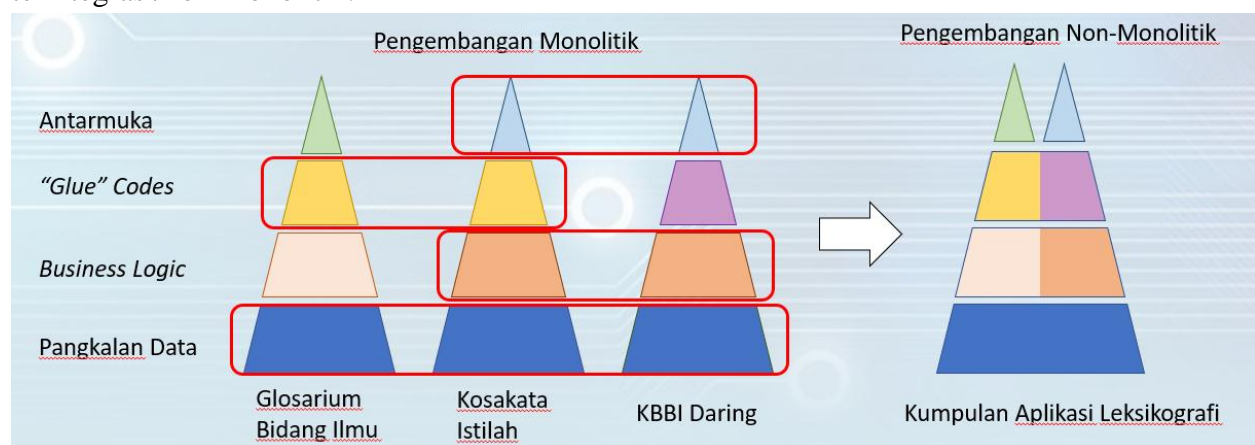
Penulis mengakui, bahwa dilema dalam hal ini tidak mudah untuk diselesaikan. Di satu sisi, aturan hukum yang terlalu berat akan sulit dijalankan karena banyaknya pelanggaran yang mungkin akan ditemukan. Di sisi yang lain, absennya hukum yang mengatur kepemilikan data digital – secara khusus data KBBI – juga yang tidak dapat terus dibiarkan karena hal tersebut akan berdampak juga pada kepentingan golongan yang sungguh memerlukan data digital KBBI bagi pengembangan bahasa Indonesia.

Mengenai permasalahan yang disajikan secara khusus dalam dilema ini, penulis berpendapat bahwa aturan hukum yang tidak terlalu memberatkan baik penegak hukum maupun pelanggar hukum (misalnya, (i) dengan penetapan denda yang progresif untuk pelanggaran yang diulangi, namun tidak terlalu besar bagi pelanggaran yang pertama, atau (ii) dengan memberikan aturan yang memperbolehkan beberapa bentuk “pelanggaran” namun yang masih dapat “ditoleransi”, seperti misalnya kebebasan mengambil, memiliki, dan menggunakan sampai dengan 10% data digital KBBI tanpa izin dari Badan Bahasa), namun dengan segera dan dapat secara konsisten dijalankan (yaitu, semua yang melanggar harus dikenakan sanksi tanpa terkecuali), akan “lebih baik” untuk dimiliki saat ini – bukan untuk dimiliki dalam jangka panjang – dibandingkan dengan aturan hukum yang “terlalu kuat dan tegas”. Aturan hukum yang *tidak* “terlalu kuat dan tegas”, tetapi yang dapat dengan lebih cepat dibuat dan dijalankan tersebut, dianggap “lebih baik” oleh penulis karena aturan hukum semacam ini akan langsung dapat memberikan “lampu hijau” kepada golongan masyarakat yang memerlukan data digital KBBI untuk memakainya secara benar, namun tidak sampai harus membuat penegak hukum menangani semua masyarakat yang melakukan pelanggaran-pelanggaran “ringan”, sehingga tidak terlalu memberatkan pelaksanaan penegakan aturan hukum itu sendiri. Untuk dimiliki dalam jangka panjang, penulis tetap berpendapat bahwa aturan hukum yang lebih benar dan lebih adil tetap harus dibuat, sebagaimana aturan hukum seharusnya.

3.2. Penyatuan Berbagai Platform/Laman Leksikografi Daring ke dalam Platform KBBI Daring

Saat ini, Badan Bahasa – baik secara langsung maupun dengan berkerja sama dengan instansi/perorangan terkait – tengah mengembangkan berbagai platform/laman daring selain KBBI Daring terkait leksikografi (seperti: tesaurus, korpus, kamus istilah, dsb.). Akan tetapi, sebagaimana dengan KBBI Daring, pengembangan setiap platform/laman daring tersebut dilakukan secara *monolitik* – memiliki “bangunan” teknologi pengembangannya sendiri –, tidak terkait satu sama lain. Setiap platform/laman daring tersebut memerlukan pangkalan data, perangkat keras, maupun desain dan pengembangan aplikasinya sendiri.

Di satu sisi, pengembangan monolitik seperti ini lebih mudah untuk dilakukan karena sifat pembagian kerjanya yang sederhana – tidak melibatkan banyak koordinasi kerja dan tidak memerlukan tim “pengembang utama” (atau, tim penyatu *teknis*) untuk melakukan standardisasi dan menyatukan kumpulan “bangunan” teknologi yang digunakan. Di sisi yang lain, hal ini menghasilkan pekerjaan yang redundan dari setiap pengembang platform/laman dan sumber data yang tidak menyatu (data yang ditambahkan, dihapuskan, atau diganti di salah satu platform tidak tercermin pada platform yang lain). Untuk kumpulan platform/laman daring terkait hal yang sama – yaitu, dalam hal ini, leksikografi – pada umumnya terdapat persamaan subset data dan keterkaitan penggunaan yang erat. Gambar 13 menampilkan ilustrasi redundansi kerja yang terjadi akibat pengembangan berbagai macam aplikasi secara monolitik dibandingkan dengan pengembangan yang dilakukan secara terintegrasi/non-monolitik.



Gambar 13 Ilustrasi Redundansi Pengembangan Monolitik vs Pengembangan Non-Monolitik
Misalnya, seorang yang mencari arti kata melalui KBBI Daring, mungkin pada saat bersamaan ingin menemukan sinonim dan antonimnya melalui Tesaurus Daring. Dalam contoh ini, dapat terlihat bahwa terdapat subset data yang sama-sama digunakan pada aplikasi KBBI Daring dan Tesaurus Daring, yaitu kata yang dicari itu sendiri. Cara pengembangan yang monolitik akan menghasilkan sumber data yang terpisah untuk data yang sama persis. Sumber data yang terpisah dalam contoh ini adalah daftar kata yang dapat dicari: satu daftar kata dimiliki oleh KBBI Daring dan satu daftar kata yang lain dimiliki oleh Tesaurus Daring. Hal ini berarti perubahan yang terjadi di salah satu laman tidak akan tercermin pada laman lainnya: jika pada KBBI Daring muncul kata baru yang memiliki sinonim dengan kata-kata lain yang sudah ada, hal tersebut tidak dapat langsung tercermin pada Tesaurus Daring

kecuali kata yang sama ditambahkan juga pada daftar kata yang dimiliki oleh laman Tesaurus Daring. Selain itu, pengembangan secara monolitik juga menghasilkan redundansi usaha pengembangan aplikasi. Diperlukan usaha pengembangan untuk melakukan pencarian pada laman Tesaurus Daring dan usaha pengembangan yang lain untuk melakukan pencarian pada laman KBBI Daring. Jika platform/laman daring tersebut disatukan di bawah naungan yang sama (tidak monolitik), maka pemisahan sumber data dan usaha pengembangan yang redundan dapat dihindarkan.

Secara teknis, KBBI Daring tidak harus menjadi platform yang menaungi platform/laman lainnya. Akan tetapi, oleh karena status ikon, cakupan penggunaan, kelengkapan data, serta kekayaan fiturnya, KBBI Daring merupakan kandidat platform yang terbaik untuk menaungkan platform/laman lainnya.

3.2.1. Kendala

Sampai saat ini, belum ada suatu “tim penyatu *teknis*” yang secara khusus bekerja sama – menyumbangkan tenaga dan keahlian *teknis* – dengan Badan Bahasa untuk tujuan *penyatuan* ini. Bukan berarti bahwa Badan Bahasa tidak memiliki tim yang bekerja dan menyusun dengan baik rencana jangka panjang pengembangan digitalisasi leksikografi yang ada, tetapi berarti bahwa ada kelompok vital yang saat ini masih absen dalam tim tersebut, yaitu tim penyatu *teknis*. Tim yang bekerja “menyatukan” berbagai laman/aplikasi yang ada saat ini *belum* bekerja menyatukan program pengembangan leksikografi digital Badan Bahasa di ranah ontologis – yaitu desain dan teknis yang digunakan untuk menyatukan laman/aplikasi tersebut –, melainkan di ranah ekonomis – berpusat pada manfaat akhir bagi penggunanya, bukan bagi kemudahan dan kesatuan kerja para pengembangnya.

Di samping itu, sudah ada beberapa laman/aplikasi yang telah selesai dibuat (misalnya, aplikasi Pengayaan Kosakata, Tesaurus Tematis Bahasa Indonesia, Ensiklopedia Sastra, dan juga Glosarium Bidang Ilmu²¹) secara monolitik (tanpa memasukkan pertimbangan penyatuan platform di dalam desainnya) dan beberapa laman/aplikasi lain yang sedang digarap – juga secara monolitik. Hal ini akan membuat pekerjaan penyatuan tersebut lebih rumit dibandingkan jika penyatuan tersebut telah *juga* direncanakan dengan matang *secara teknis* sedari mulanya. Akibat lain dari penyatuan *teknis* yang datang terlambat adalah kemungkinan menyebabkan laman/aplikasi yang dibuat lebih awal *kedaluwarsa* – tidak digunakan dalam platform tunggal akhir (hasil akhir penyatuan platform) karena tidak memiliki desain/fitur yang cocok disatukan dengan yang lain atau karena seluruh desain/fiturnya dapat digantikan secara penuh – atau bahkan dengan lebih baik – oleh laman/aplikasi yang lain yang dibuat sesudahnya.

Contoh nyata dari hal ini adalah aplikasi Pengayaan Kosakata istilah yang diluncurkan oleh Badan Bahasa hanya beberapa bulan sebelum KBBI Daring diluncurkan. Aplikasi tersebut dikembangkan dengan tujuan menjadikannya saluran digital bagi masyarakat untuk mengusulkan kosakata baru bagi KBBI, namun setelah KBBI Daring diluncurkan, aplikasi Pengayaan Kosakata tersebut tidak lagi aktif digunakan. Hal ini dikarenakan platform KBBI Daring sendiri telah menyatukan proses pencarian dan pengusulan kosakata secara lebih

²¹ alamat aplikasi-aplikasi ini ditautkan di laman Badan Bahasa (badan.bahasa@kemdikbud.go.id)

mudah dan lebih efisien dibandingkan dengan aplikasi Pengayaan Kosakata. Usulan yang diberikan melalui formulir pada KBBI Daring didesain sesuai dengan struktur pangkalan data KBBI sehingga dapat langsung dimasukkan pada pangkalan data KBBI segera sesudah usulan tersebut disetujui untuk masuk dalam KBBI oleh rapat redaksi. Tetapi tidak demikian halnya dengan formulir pengusulan pada aplikasi Pengayaan Kosakata. Formulir pengusulan aplikasi tersebut didesain secara monolitik, terpisah dari pengembangan KBBI Daring. Maka, secara alamiah, formulir pengusulan tersebut didesain tanpa mempertimbangkan struktur pangkalan data KBBI – karena memang tidak tersedia pada saat itu. Konsekuensinya, usulan yang diberikan melalui aplikasi pengayaan tersebut akhirnya tidak dapat langsung dimasukkan ke dalam pangkalan data KBBI yang muncul sesudahnya, melainkan harus dibuat ulang untuk boleh masuk ke dalam pangkalan data KBBI. Dengan munculnya KBBI Daring yang menyediakan formulir pengusulan yang lebih “siap saji” dibandingkan aplikasi Pengayaan Kosakata, para penyumbang kosakata bahasa Indonesia pun tidak memiliki insentif untuk memakai aplikasi Pengayaan Kosakata tersebut. Selain itu, ditinjau dari sisi tim redaksi yang memproses usulan, tidak seperti KBBI Daring, aplikasi Pengayaan Kosakata tersebut tidak menyediakan platform “Meja Redaksi” sebagaimana dimiliki oleh KBBI Daring untuk memproses usulan yang dikumpulkannya. Hal ini menyebabkan usulan yang diterima oleh aplikasi Pengayaan Kosakata lebih sulit diproses secara digital oleh tim redaksi. Pada akhirnya, hal ini membuat tim redaksi – bukan hanya penyumbang data – juga lebih memilih menggunakan KBBI Daring ketimbang aplikasi Pengayaan Kosakata yang diluncurkan hanya beberapa bulan sebelumnya. Setelah KBBI Daring diluncurkan, aplikasi Pengayaan Kosakata segera menjadi *kedaluwarsa*. Jelas terlihat bahwa tanpa perencanaan penyatuan yang baik, akan mungkin bagi suatu laman/aplikasi yang pada awalnya dikembangkan untuk dengan segera menjadi *kedaluwarsa*.

3.2.2. Saran

Tim penyatu teknis yang berkualitas – secara nyaris mutlak – diperlukan jika penyatuan platform ini ingin dikerjakan dengan baik. Tim ini setidaknya perlu memiliki seorang atau kumpulan pengembang teknik informasi komputer yang memiliki pengertian yang baik tentang pengembangan platform daring secara menyeluruh, mulai dari aspek belakang (*back-end*) aplikasi seperti membangun pangkalan data dan konfigurasi *server*, sampai pada aspek depan (*front-end*) aplikasi seperti penyajian data kepada masyarakat secara luas melalui laman daring. Dengan kata lain: seorang *full-stack developer*, seorang arsitek sistem perangkat lunak (*software system architect*), atau sekumpulan orang yang secara keseluruhan memiliki kapasitas sebanding. Jika tim ini telah terbiasa dengan isu-isu teknis yang spesifik mengenai pengembangan leksikografi secara daring, hal ini merupakan bonus bagi tim tersebut, tetapi tidak harus dimiliki oleh mereka. Tim penyatu teknis ini kemudian perlu bekerja sama secara dekat dengan Badan Bahasa untuk menyumbangkan pekerjaan *teknis* yang diperlukan untuk penyatuan platform ini dan untuk menjaga platform ini sendiri agar tetap boleh digunakan untuk berbagai macam perkembangan program pendaringan leksikografi selanjutnya. Idealnya, tim ini adalah bagian dari tim pengembang Badan Bahasa sendiri yang bekerja penuh waktu di dalam Badan Bahasa.

Jika penyatuan platform ini ingin dikerjakan, pembentukan tim penyatu teknis ini juga perlu dilakukan sedini mungkin. Dengan bertambahnya laman/aplikasi monolitik yang hendak disatukan, akan makin sulit pula pekerjaan tim penyatu teknis tersebut – di samping akan makin besar kemungkinan laman/aplikasi yang lebih awal dibuat untuk tidak lagi menjadi bagian dari platform tunggal akhir yang digunakan (*kedaluwarsa*). Memulai pekerjaan ini makin dini juga akan memberikan lebih banyak waktu bagi tim penyatu teknis untuk mengerjakan penyatuan platform tersebut dengan baik. Hal ini diperlukan secara khusus bagi pengembangan ke arah penyatuan platform ini oleh karena penyatuan platform ini mungkin merupakan pengembangan yang memerlukan paling banyak pekerjaan yang bersifat teknis dibandingkan semua kemungkinan pengembangan lain yang dipaparkan dalam makalah ini.

3.3. Tautan, Pranala Luar, dan/atau Integrasi dengan Aplikasi/Laman Leksikografi Daring Internasional

Saat ini, terdapat fitur tautan pencarian melalui *Google* (bagi pengguna terdaftar) yang dimiliki oleh KBBI Daring. *Google* merupakan salah satu contoh aplikasi/laman internasional yang secara luas digunakan untuk pencarian data secara umum, namun telah terdapat juga aplikasi/laman lainnya yang, secara lebih khusus, terkait leksikografi dan yang telah digunakan dalam ranah internasional oleh peneliti dan pengembang bahasa. Aplikasi/laman yang demikian merupakan aplikasi/laman yang siap digunakan untuk menyajikan informasi yang biasa digunakan oleh peneliti bahasa dan leksikograf dalam pekerjaan mereka.



Gambar 14 Laman WordNet Bahasa

*WordNet*²² adalah salah satu contoh dari laman semacam ini. *WordNet* adalah sebuah (laman) pangkalan data leksikal elektronik (digital) berbasiskan bahasa Inggeris (Miller, 1995).

²² Laman *WordNet* dapat diakses melalui <https://wordnet.princeton.edu/>

WordNet berisi data yang komprehensif tentang kata, definisi, dan contoh yang kata-katanya yang memiliki makna atau konsep yang sama dikelompokkan ke dalam himpunan sinonim (*synonym set* atau *synset*) yang saling berhubungan satu sama lain, misalnya melalui hubungan semantik hipernim, hiponim, holonim, troponim, dan sebagainya. Setiap himpunan sinonim memiliki ID pencarian yang unik. *WordNet Bahasa*²³ (Noor, Sapuan, & Bond, 2011; Bond, Lim, Tang, & Riza, 2014) adalah *WordNet* untuk bahasa Melayu, yang saat ini berisi data bahasa Indonesia dan bahasa Malaysia. Gambar 14 menampilkan laman *WordNet Bahasa*.

KBBI Daring dapat dikembangkan lebih jauh melalui pengembangan pangkalan data yang tersedia saat ini, misalnya dengan mengembangkan tabel entri untuk mencakup juga kumpulan ID himpunan sinonim dari *WordNet Bahasa* untuk kata (entri) tersebut – atau dengan mengembangkan tabel makna untuk mencakup juga ID himpunan sinonim dari *WordNet* yang merupakan konsep yang serupa dengan makna tersebut – dan menambahkan tautan rujukan ke *WordNet Bahasa* pada hasil pencarian kata laman KBBI Daring. Hal yang serupa juga dapat dilakukan untuk aplikasi/laman lain yang serupa, misalnya *BabelNet* (Navigli, 2012) dan *FrameNet* (Baker, Fillmore, & Lowe, 1998).

Dengan rujukan berupa tautan/pranala pada – atau integrasi dengan – aplikasi/laman sedemikian, KBBI Daring dapat juga membantu pekerjaan peneliti bahasa dan leksikograf yang memerlukan informasi dari aplikasi/laman tersebut. Pengembangan ini juga akan mengakibatkan KBBI lebih mudah diakses dan lebih dikenali secara internasional.

3.3.1. Kendala dan Saran

Tingkat kesulitan dalam mengerjakan hal ini tidak menentu. Tergantung aplikasi/laman apa yang ingin ditautkan/diintegrasikan dengan laman KBBI Daring, tingkat kesulitan dan keperluan yang diperlukan untuk membangun kapasitas ini pada KBBI Daring bisa berbeda-beda. Satu aplikasi/laman mungkin memerlukan penambahan atau perombakan yang signifikan sedang satu aplikasi/laman lainnya tidak. Satu hal yang jelas, pengembangan ke arah ini tidak hanya memerlukan tenaga pengembang teknik informasi komputer, tetapi juga – terutama – memerlukan tenaga ahli bahasa yang memiliki pengetahuan tentang aplikasi-aplikasi/laman-laman internasional yang biasa dipakai dalam melakukan penelitian tersebut.

Dalam jenis pengembangan ini, sumbangsih ahli bahasa tersebut adalah hal yang paling diperlukan. Tanpa memiliki seorang atau kumpulan ahli bahasa yang terbiasa memakai aplikasi/laman semacam demikian, pengembang teknik informasi komputer sendiri tidak akan mengetahui bagaimana harus membangun kapasitas KBBI Daring untuk dapat diintegrasikan dengan (dapat mengakomodasi) aplikasi/laman lain tersebut. Jika seorang atau kumpulan ahli bahasa semacam ini ada, mereka dapat bekerja sama dengan Badan Bahasa dan pengembang teknik informasi komputer untuk menjalankan pengembangan ini. Badan Bahasa dapat menawarkan jenis kerja sama seperti ini kepada ahli bahasa sedemikian jika Badan Bahasa juga telah memiliki komitmen pengembangan ke arah ini dan tenaga kerja teknis (pengembang teknis informasi komputer) untuk melakukannya. Sekali lagi, dalam hal ini, idealnya adalah Badan Bahasa memiliki pengembang teknik informasi komputer sebagai

²³ <http://wn-msa.sourceforge.net/>

bagian dari tim internal Badan Bahasa sendiri untuk bekerja sama dengan ahli-ahli bahasa (internal maupun eksternal) dalam mengintegrasikan aplikasi KBBI Daring dengan aplikasi/laman luar tersebut.

3.4. Penggunaan Platform KBBI Daring untuk Penyusunan Kamus Bahasa Daerah

Platform yang sama yang digunakan untuk penyusunan KBBI pada KBBI Daring juga dapat diperluas cakupannya dan digunakan untuk penyusunan kamus bahasa daerah. Platform KBBI Daring telah digunakan secara luas dan telah teruji dengan baik dalam seluruh tahapan penyusunan KBBI secara digital. Dengan mengasumsikan bahwa penyusunan kamus bahasa daerah saat ini menggunakan sistem pengumpulan data, pemrosesan data, penyusunan format kamus, pengoreksian, pencetakan, dan penyajian yang serupa dengan penyusunan KBBI, platform KBBI Daring dapat juga diadopsi sebagai dasar pengembangan platform kamus bahasa daerah daring.

Untuk saat ini, Badan Bahasa menyosialisasikan penggunaan *Lexique Pro* untuk penyusunan kamus bahasa daerah karena *Lexique Pro* merupakan aplikasi yang gratis. Sampai saat makalah ini ditulis, aplikasi *Lexique Pro* sudah digunakan untuk proyek pembuatan lima kamus bidang ilmu: Kamus Filologi, Kamus Budaya Gayo, Kamus Agroteknologi, Kamus Oseanografi Fisika, dan Kamus Ilmu Tafsir Alquran.

3.4.1. Kendala dan Saran

Pengadopsian ini akan memerlukan ketersediaan sumber daya perangkat keras, data digital pada pangkalan data, dan struktur pangkalan data yang serupa bagi kamus bahasa daerah yang ingin dikembangkan sebagaimana data KBBI didigitalisasikan. Jika hal-hal tersebut tersedia, barulah platform yang serupa KBBI Daring dapat diadopsi – dengan penyesuaian yang minimal – untuk menggantikan penyusunan kamus bahasa daerah secara manual menjadi digital.

Dengan demikian, jika pengembangan ke arah ini hendak dikerjakan, digitalisasi data kamus-kamus bahasa daerah menjadi pangkalan data diperlukan. Hal ini memerlukan suatu usaha dan komitmen yang tinggi dari para pekamus dan penyedia data bahasa daerah. Selain melakukan pengumpulan data bahasa daerah, mereka perlu dilengkapi juga dengan kemampuan dasar dalam mengenali dan membuat struktur data digital yang ramah “dibaca” dan diproses lebih lanjut secara digital, yaitu berupa pangkalan data digital. Idealnya adalah jika digitalisasi ini dapat dilakukan dan menghasilkan struktur data yang serupa dengan yang sekarang digunakan dalam KBBI Daring, sehingga platform KBBI Daring pun dapat langsung digunakan untuk melayani kebutuhan pengembangan dan penyempurnaan bahasa daerah, namun jika ternyata didapati bahwa digitalisasi data kamus-kamus bahasa daerah ini tidak dapat dilakukan sepenuhnya mengikuti struktur pangkalan data yang ada pada KBBI Daring saat ini, maka salah satu dari dua pendekatan ini dapat diambil: pertama, pengembangan yang lebih spesifik secara khusus sebagai proyek terpisah untuk keperluan kamus bahasa daerah tersebut dengan memakai platform KBBI Daring “hanya” sebagai dasar (tanpa mengubah platform KBBI Daring itu sendiri), atau kedua, perluasan kapasitas dasar platform KBBI Daring sehingga dapat mencakup juga penyajian data dan pengembangan

kamus daerah secara daring (yaitu dengan mengubah dan memperluas kapasitas platform KBBI Daring itu sendiri).

Di antara dua pendekatan yang disajikan, penulis makalah berpendapat bahasa pendekatan kedua adalah pendekatan yang lebih sesuai dengan kemungkinan orientasi pengembangan kamus bahasa daerah maupun perkembangan KBBI Daring di masa depan. Misalnya, saat ini, KBBI adalah kamus ekabahasa. Dengan demikian pangkalan data yang dimilikinya adalah pangkalan data yang sesuai untuk kamus ekabahasa pula. Akan tetapi pengembangan kamus daerah *sangat mungkin* memiliki orientasi ke arah pengembangan dwibahasa – dari bahasa daerah ke bahasa Indonesia atau sebaliknya – walaupun tidak harus demikian, oleh sebab pengembangan kamus bahasa daerah dwibahasa akan dapat membawa manfaat bagi lebih banyak orang (penutur bahasa daerah tersebut *dan* penutur bahasa Indonesia yang ingin mempelajari bahasa daerah tersebut). Selain itu, ada kemungkinan juga bagi KBBI Daring (atau produk turunannya) di masa depan untuk dikembangkan dan menjadi bersifat dwibahasa – atau bahkan multibahasa, sebagaimana yang dilakukan oleh kamus-kamus daring yang awalnya adalah ekabahasa lain seperti *Oxford* dan *Merriam Webster*. Dengan demikian, adalah lebih baik jika struktur pangkalan data yang terdapat pada KBBI Daring sekarang diperluas sehingga dapat mencakup penyusunan kamus dwibahasa ketimbang setiap pengembangan kamus bahasa daerah perlu melakukan pengembangan spesifik pada struktur pangkalan datanya berdasarkan platform KBBI Daring dasar yang tidak dikembangkan (hanya diperuntukkan untuk kamus ekabahasa) demi mendukung penyajian kamus dwibahasa.

Pengembangan ke arah ini dapat mengantisipasi kebutuhan pengembangan bahasa daerah secara digital yang mungkin akan datang dalam jangka waktu beberapa tahun ke depan. Jika rencana pengembangan ini dimulai dari sekarang, yaitu sebelum kebutuhan akan hal ini makin mendesak, maka metode pengembangan platform yang monolitik untuk setiap laman/aplikasi daring per bahasa daerah dapat lebih mungkin dihindarkan. Jika menunggu sampai kebutuhan pendaring kamus bahasa daerah ini lebih mendesak di masa depan, pengembangan ke arah ini dapat menjadi lebih rumit sebagaimana penyatuan platform yang dibahas dalam bagian 3.2.

3.5. Penambahan Fitur/Aplikasi/Laman Artikel sebagai Pendukung Penggunaan atau Perluasan KBBI Daring

KBBI Daring dapat pula dikembangkan dengan menambahkan fitur/aplikasi/laman artikel pendukung. Fitur/aplikasi pendukung dapat berupa fitur/aplikasi tambahan dalam menggunakan KBBI Daring, sedang laman artikel dapat berisikan tulisan pengayaan penggunaan kamus atau penggunaan bahasa Indonesia.

Salah satu contoh fitur/aplikasi pendukung yang dapat ditambahkan pada KBBI Daring adalah pengecekan kata baku atau tata bahasa baku suatu frasa/kalimat. Dengan keunggulan kelengkapan datanya, KBBI Daring adalah kandidat platform yang ideal untuk penambahan fitur semacam ini secara daring. Salah satu contoh penggunaan aplikasi semacam ini adalah untuk pengecekan tata bahasa tugas akhir mahasiswa.

Contoh lain ialah fitur/aplikasi penerjemahan ke bahasa daerah dan penambahan laman artikel yang mendukung penggunaan kamus tersebut dengan menjelaskan secara lebih

mendetail informasi dari kata yang dicari. Hal ini bukan sesuatu yang tidak pernah dibuat pada kamus daring lainnya. Kamus daring *Oxford*, misalnya, memiliki bukan hanya fitur penerjemahan dari Bahasa Inggris ke bahasa lain, tetapi juga memiliki kamus *Oxford* berbasis bahasa lain tersebut. Misalnya, *Oxford* memiliki kamus daring untuk bahasa Indonesia²⁴. Kamus daring *Oxford* berbasis bahasa Indonesia ini bukan hanya menyediakan fitur pencarian kata, melainkan juga menyediakan berbagai artikel mencakup tips penggunaan bahasa, penjelasan tata bahasa, dan pengetahuan praktis bahasa Indonesia (seperti filosofi dari suatu kata, misalnya *ketupat*) untuk memberikan informasi tambahan terhadap kata yang dicari.

KBBI Daring dapat pula dikembangkan dengan mencakup fitur-fitur yang serupa maupun laman artikel lain yang dianggap dapat membantu penggunaan bahasa Indonesia lainnya (misalnya: laman berisi daftar penulisan kata yang benar dibandingkan yang tidak benar tetapi sering digunakan, laman berisi daftar padanan kata dari istilah asing yang beredar di masyarakat, dsb.). Dapat dilihat bahwa poin utama di sini adalah bahwa platform KBBI Daring masih dapat banyak diperkaya dengan berbagai penyediaan fitur/aplikasi/laman artikel pendukung penggunaannya.

Saat ini, berbagai macam artikel pendukung penggunaan KBBI semacam ini (misalnya, info singkat tentang padanan kata atau artikel pendek mengenai penggunaan ejaan yang benar) dibuat oleh Badan Bahasa dan dipos lewat *Instagram*²⁵, *Facebook*²⁶, atau *Twitter*²⁷, tetapi tidak dapat ditemukan pada laman KBBI Daring sendiri.

3.5.1. Kendala dan Saran

Jenis sumber daya manusia yang diperlukan untuk menjalankan usulan pengembangan ini cukup beragam. Satu usulan pengembangan dapat dilaksanakan terutama berdasarkan ketersediaan ahli bahasa sedangkan usulan yang lain memerlukan ketersediaan pengembang teknik informasi komputer. Misalnya, usulan untuk menambahkan artikel melengkapi kata-kata yang dapat dicari pada KBBI Daring dapat dikerjakan hampir sepenuhnya oleh seorang ahli bahasa (sebagaimana yang telah dilakukan, misalnya dengan mengepos artikel pendek atau gambar lewat *Instagram*, *Facebook*, dan *Twitter* tentang padanan kata), sedangkan pengembang teknik informasi komputer hanya perlu menautkan artikel tersebut pada KBBI Daring. Sedangkan, usulan untuk membuat fitur pengecekan kosakata atau tata bahasa baku akan memerlukan baik ahli bahasa maupun pengembang teknik informasi komputer – atau jika tidak, seorang ahli bahasa komputasi (*computational linguistics expert*), yaitu seorang yang memiliki keahlian gabungan dalam dua bidang tersebut.

Keunggulan dari pengembangan ke arah ini adalah pengembangan semacam ini dapat diadopsi secara fleksibel, tergantung ketersediaan sumber daya manusia. Misalnya, jika lebih banyak terdapat sumber daya ahli bahasa, maka usulan penambahan laman-laman artikel dapat dikerjakan. Akan tetapi, jika terdapat lebih banyak sumber daya pengembang teknik informasi komputer, pembuatan fitur yang lebih banyak memerlukan pengembangan

²⁴ dapat diakses di <https://id.oxforddictionaries.com/>

²⁵ <https://www.instagram.com/badanbahasakemendikbud/>

²⁶ <https://id-id.facebook.com/Badan.Bahasa/>

²⁷ <https://twitter.com/badanbahasa>

perangkat lunak seperti pembuatan fitur pengecekan kosakata dapat dikerjakan. Kekurangan pengembangan semacam ini mungkin terletak pada sifatnya yang lebih merupakan tambahan “barang mewah” dari apa yang sudah dimiliki dan bukan merupakan “barang kebutuhan pokok” yang seharusnya ada. Berbagai macam saran pengembangan pada bagian ini baik jika ada, tetapi tidak harus ada atau belum mendesak kebutuhannya.

Saran dalam mempertimbangkan mengambil pengembangan ke arah ini adalah dengan melihat sumber daya yang lebih tersedia dan menimbang seberapa besar manfaat “barang mewah” tersebut dikembangkan saat ini dibandingkan jika sumber daya pengembangan yang dimiliki digunakan untuk hal yang lain. Pengembangan ke arah penyatuan platform leksikografi yang ada, misalnya, merupakan pengembangan yang lebih mendesak dan yang makin cepat dikerjakan makin sedikit tingkat kesulitannya dibandingkan dengan, misalnya, penambahan laman artikel penjelasan pemakaian kata yang tingkat kesulitan implementasinya cenderung tidak berubah secara drastis dalam waktu dekat. Jika sumber daya yang tersedia benar-benar terbatas, pengembangan ke arah ini juga tidak harus dikerjakan.

4. KESIMPULAN

Makalah ini telah menyajikan latar belakang KBBI Daring dan kontribusi yang telah dihasilkan oleh KBBI Daring bagi pengembangan leksikografi di Indonesia sampai saat ini. Telah dijelaskan bagaimana KBBI Daring telah mengubah proses kerja penyusunan KBBI secara manual yang mencakup pengumpulan data, pemrosesan data, penyusunan format, pengoreksian, pencetakan, dan penyajian KBBI menjadi proses penyusunan secara digital. Makalah ini juga telah memaparkan bagaimana KBBI Daring telah digunakan dan disebarluaskan sampai saat ini dan mengelaborasi bagaimana sesungguhnya platform KBBI Daring dapat dikembangkan lebih jauh hanya dengan meninjaunya dari kapasitas teknologi informasi komputer yang sudah tersedia saat ini dan dari praktik-praktik kamus daring lain. Secara khusus, penulis makalah menyumbangkan beberapa usulan berikut bagi pengembangan KBBI Daring:

- Pelayanan dan Lisensi Bagi Peneliti Bahasa dan Pengembang Teknik Informasi Komputer
- Penyatuan Berbagai Platform/Laman Leksikografi Daring ke Bawah Naungan Platform KBBI Daring
- Tautan, Pranala Luar, dan/atau Integrasi dengan Aplikasi/Laman Leksikografi Daring Internasional
- Penggunaan Platform KBBI Daring untuk Penyusunan Kamus Bahasa Daerah
- Penambahan Fitur/Aplikasi/Laman Artikel Sebagai Pendukung Penggunaan atau Perluasan KBBI Daring

Beserta dengan usulan yang disumbangkan, penulis makalah juga melakukan kajian terhadap kendala-kendala yang sedang dan mungkin akan dihadapi beserta saran-saran dalam mengatasinya sejalan dengan usulan-usulan pengembangan tersebut. Masih banyak lagi kemungkinan pengembangan yang dapat dilakukan melalui KBBI Daring bagi pengembangan leksikografi digital di Indonesia. Secara khusus, pengembangan leksikografi

digital di Indonesia dalam aspek penelitian bahasa dan perluasan wilayah pemakaian kamus ke arah digital – baik bahasa Indonesia, maupun bahasa-bahasa daerah di Indonesia – mendapat tekanan di dalam usulan-usulan pengembangan yang dituliskan dalam makalah ini. Melalui makalah ini, penulis berharap ia dapat memberikan sumbangsih bagi pengembangan leksikografi di Indonesia ke arah digital dengan memberikan analisis dan saran dari sudut pandang teknologi informasi komputer dan praktik-praktik kamus daring lain yang sudah ada saat ini dan dengan menggunakan KBBI Daring sebagai platform dasar bagi pengembangan tersebut.

5. UCAPAN TERIMA KASIH

Penulis menghaturkan terima kasih kepada Ibu Dora Amalia yang membantu mengecek keakuratan fakta-fakta yang ditulis dalam makalah ini, khususnya sehubungan dengan sejarah KBBI dan cara kerja tim redaksi di dalam Badan Bahasa dalam menyusun KBBI. Penulis juga mengucapkan terima kasih kepada David Moeljadi yang membantu mengecek, memberikan rujukan, dan membenahi deskripsi *WordNet*, *WordNet Bahasa*, *BabelNet* dan *FrameNet* yang dituliskan dalam makalah ini.

6. REFERENSI

- Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet Project. *Proceedings of the 17th international conference on Computational linguistics, Volume 1* (pp. 86-90). Association for Computational Linguistics.
- Bond, F., Lim, L. T., Tang, E. K., & Riza, H. (2014). The combined Wordnet Bahasa. *NUSA: Linguistic studies of languages in and around Indonesia, 57*, 83-100.
- Kamajaya, I., Moeljadi, D., & Amalia, D. (2017). KBBI Daring: A Revolution in The Indonesian. *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, (pp. 513-530). Leiden.
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM Vol. 38, No. 11*, 39-41.
- Moeljadi, D., Kamajaya, I., & Amalia, D. (2017). Building the Kamus Besar Bahasa Indonesia (KBBI) Database and Its Applications. *In Proceedings of The 11th International Conference of the Asian Association for Lexicography*, (pp. 64-80). Guangzhou.
- Navigli, R. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence, 193*, 217-250.
- Noor, N. H., Sapuan, S., & Bond, F. (2011). Creating the Open Wordnet Bahasa. *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation (PACLIC 25)*, (pp. 258-267). Singapore.

TATANAN HUKUM PEMANFAATAN DATA KAMUS ELEKTRONIK

Dr. Houtman, M.Pd.¹, Suryati, S.H., M.H.²

Universitas PGRI Palembang

Abstrak

Perkembangan teknologi informasi dan komunikasi menyebabkan hubungan dunia menjadi tanpa batas (*borderless*) dan menyebabkan perubahan sosial, ekonomi, dan budaya secara signifikan dan berlangsung demikian cepat. Pemanfaatan teknologi informasi dalam suatu sistem elektronik adalah penggunaan sistem komputer secara luas yang mencakup perangkat keras, perangkat lunak, jaringan komunikasi, serta data elektronik. Pemanfaatan teknologi juga digunakan untuk pengelolaan data perkamusan. Hal ini berkaitan dengan karakteristik data keluasan diksi itu sendiri yang bersifat multiguna yang terkait dengan pelbagai kebutuhan dalam upaya pencarian makna dari sebuah istilah atau pemahaman terhadap konstruksi suatu kosakata. Pengelolaan kamus elektronik ini untuk memenuhi tuntutan masyarakat yang semakin meningkat untuk mewujudkan pemerintahan yang bersih yang akhirnya akan berkaitan dengan keterbukaan informasi untuk masyarakat sebagai sarana pertukaran informasi.

Dalam kebimbangan pemanfaatan dan kebebasan penggunaan teknologi informasi dan komunikasi tersebut, sebenarnya telah tersedia Undang-undang ITE disahkan tahun 2008. Undang-undang ini dapat dijadikan sebagai landasan dalam mendukung pembangunan pangkalan data perkamusan dan sekaligus menyangkut tata aturan hukum yang berkaitan dengan upaya pemanfaatan kamus elektronik oleh masyarakat luas. Pengaturan mengenai informasi dan transaksi elektronik mengacu pada beberapa instrumen internasional, seperti *UNCITRAL Model Law on e-Commerce* dan *UNCITRAL Model Law on e-Signature*. Bagian ini dimaksudkan untuk mengakomodasi kebutuhan para pelaku bisnis di internet dan masyarakat pada umumnya guna mendapatkan kepastian hukum dalam melakukan transaksi elektronik, termasuk penggunaan kamus elektronik.

Beberapa materi yang diatur itu antara lain: pengakuan informasi atau dokumen elektronik sebagai alat bukti hukum yang sah (Pasal 5 & Pasal 6 UU ITE); penyelenggaraan sistem elektronik (Pasal 15 & Pasal 16 UU ITE); dan perbuatan yang dilarang (*cybercrimes*). Beberapa *cybercrimes* yang diatur dalam UU ITE, antara lain: a) akses ilegal (Pasal 30); b) intersepsi ilegal (Pasal 31); c) gangguan terhadap sistem (*system interference*, Pasal 33 UU ITE); d) penyalahgunaan alat dan perangkat (*misuse of device*, Pasal 34 UU ITE). Sementara itu, aspek hukum dalam penggunaan internet terbagi menjadi: Aspek hak milik intelektual, yaitu yang memberikan perlindungan hukum bagi pembuat karya, seperti hak cipta dan hak paten, yurisdiksi hukum, dan aspek-aspek terkait.

Komponen ini menganalisis dan menentukan keberlakuan hukum yang diterapkan di dalam dunia maya itu. Sebagai penguat, Pasal 26 UU ITE menyebutkan: 1) Penggunaan setiap informasi melalui media elektronik yang menyangkut data pribadi seseorang harus dilakukan atas persetujuan orang yang bersangkutan; 2) Setiap orang yang dilanggar haknya sebagaimana dimaksud pada ayat (1) dapat mengajukan gugatan atas kerugian yang ditimbulkan berdasarkan Undang-undang ini.

I. PENDAHULUAN

Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE) pada dasarnya tidak dapat menjangkau semua aspek atau tatanan hukum dalam kegiatan atau perbuatan hukum yang dilakukan dalam internet, tetapi dapat didukung oleh peraturan perundang-undangan lainnya sehingga tidak akan terjadi kekosongan hukum dalam setiap peristiwa hukum yang terjadi sebagai jalan keluar dalam penegakan hukumnya.

Dalam penjelasan UU ITE disebutkan bahwa kegiatan melalui media sistem elektronik yang disebut juga ruang siber (*cyber space*), meskipun bersifat virtual dapat dikategorikan sebagai tindakan atau perbuatan hukum yang nyata. Secara yuridis kegiatan pada ruang siber tidak dapat didekati dengan ukuran dan kualifikasi hukum konvensional saja. Hal ini sebab jika cara tersebut yang ditempuh akan terlalu banyak kesulitan dan hal yang lolos dari pemberlakuan hukum.

Internet saat ini telah menghubungkan jaringan komputer lebih dari 300 ribu jumlahnya (*networks of networks*) yang menjangkau sekitar lebih dari seratus negara di dunia. Dalam setiap hitungan menit muncul jaringan tambahan lagi, ratusan halaman informasi (*web pages*) yang baru tersajikan setiap menitnya sehingga memperkaya khazanah yang telah ada. Seiring dengan perkembangan komputer, internet juga telah menawarkan sejumlah layanan bagi kehidupan manusia, mulai dari kegiatan kesehatan (*e-medicine*), bisnis (*e-business*), pendidikan (*e-education*), pemerintahan (*e-government*) dan lain sebagainya. Kemajuan teknologi informasi khususnya media internet, dirasakan banyak memberikan manfaat seperti dari segi keamanan, kecepatan serta kenyamanan.²⁸

Dalam Pasal 3 UU ITE disebutkan bahwa internet sebagai sarana informasi yang memiliki asas dan tujuan dalam pemanfaatannya, yaitu pemanfaatan Teknologi Informasi dan Transaksi Elektronik dilaksanakan berdasarkan asas kepastian hukum manfaat, kehati-hatian, iktikad baik, dan kebebasan memilih teknologi atau netral teknologi.

Dari sisi pandang teori sistem, informasi memiliki ruang kebebasan beraksi, mengendalikan pengeluaran, mengefisienkan pengalokasian sumber daya dan waktu. Sirkulasi informasi yang terbuka dan bebas merupakan kondisi yang optimal untuk pemanfaatan informasi.²⁹ Secara pesat, teknologi ini mengubah cara hidup masyarakat, di mana batas ruang dan waktu sudah tidak menjadi kendala besar (*borderless*). Bahkan kehadiran internet yang sangat fenomenal ini semakin mengukuhkan pendapat bahwa

²⁸ Rakha Permana, <http://rakhapermanasblog.blogspot.com/2013/05/aspek-hukum-dan-etika-internet.html>, diunduh pada tanggal 3 Juli 2018.

²⁹ I Made Wiryana, <http://nakula.rvs.uni-bielefeld.de/~made/?id=31>, diunduh pada tanggal 25 Juni 2018.

teknologi informasi dan komunikasi telah menjadi budaya umum pada masyarakat dunia saat ini.³⁰

Oleh karenanya, pengkajian hukum di bidang hukum informasi dan teknologi informasi bagi pelaksanaan UU ITE masih dibutuhkan sebagai tatanan hukum dalam pemanfaatan data elektronik. Salah satu yang masih dapat diperdebatkan (*debatable*) dalam UU ITE tersebut misalnya pihak mana atau siapa yang menjadi penanggungjawab penuh atas penutupan maupun pengawasan isi (*content*) dari sebuah laman (situs) maupun kamus elektronik.

Menurut mantan ketua umum APJII (Asosiasi Penyelenggara Jasa Internet), Sylvia W. Sumarlin, kebebasan berkespresi di ruang internet tidak bisa mengadopsi kebebasan seperti yang diterapkan di Amerika Serikat, yang berbeda budaya dan etika dengan orang Indonesia. Padahal lembaga yang mengurus tata internet di Indonesia belum banyak seperti di Cina, di mana pemerintahannya memiliki jumlah SDM tidak kurang dari 30.000 orang untuk mengurus aturan bisnis dan konten internet.³¹

Walaupun masih sering diperdebatkan karena dinilai banyak pasal yang dianggap “pasal karet” dan dapat dipakai untuk menyerang pihak-pihak tertentu, kehadiran Undang-undang No. 11 Tahun 2008 tentang Informasi dan Transaksi Ekonomi patut diapresiasi. Sebagai payung hukum atas peraturan di bawahnya, undang-undang ini tentu telah memberi paling tidak kepastian kepada pelaku dan pengguna yang terkait dengan Informasi dan Transaksi Elektronik. Tatanan hukum yang mengatur dan melindungi terutama terkait Dokumen Elektronik khususnya Kamus Elektronik setidaknya telah tersedia melalui Undang-undang No. 11 Tahun 2008. Setelahnya jika terasa belum mencukupi, maka perlu diturunkan kembali dalam bentuk peraturan yang lebih detail.

II. PEMBAHASAN

2.1. Kamus Elektronik Sebagai Alat Bukti yang Sah

Untuk membahas mengenai tatanan hukum dari kamus elektronik, sebaiknya perlu disepakati terlebih dahulu bahwa kamus elektronik adalah bagian dari sistem elektronik yang dapat disebut sebagai dokumen elektronik. Terkait penggunaan dokumen elektronik pada sistem elektronik, rujukan Undang-undang yang dapat dipakai salah satunya adalah Undang-undang Republik Indonesia Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik. Dalam penjelasannya, dokumen elektronik³² yang dimaksud adalah setiap informasi elektronik yang dibuat, diteruskan, dikirimkan, diterima atau disimpan dalam bentuk analog, digital, elektromagnetik, optikal, atau sejenisnya, yang dapat dilihat, ditampilkan, dan/atau didengar melalui komputer atau sistem elektronik. Semua itu termasuk tetapi tidak terbatas pada tulisan, suara, gambar, peta, rancangan, foto atau sejenisnya, huruf, tanda, angka, kode akses, simbol atau perforasi yang memiliki makna serta dapat dipahami oleh orang yang mampu memahaminya.

³⁰ Jimly Asshiddiqie, *Hukum Tata Negara dan Pilar-pilar Demokrasi*. Jakarta: Konpress, 2005, hal.234.

³¹ Sylvia W. Sumarlin dalam “Indonesia Butuh Lembaga Konten Internet”, Coverage Monthly Telematic Magazine, Jakarta, Edisi 6 Vol. 02/ Mei 2008 hal. 13.

³² Undang-undang No. 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (ITE) Pasal 1, Ayat (4).

Adapun penjelasan mengenai keabsahan dokumen elektronik yang dalam hal ini kamus elektronik jika nantinya digunakan sebagai bukti hukum telah dijelaskan dalam pasal 5 yakni:

- (1) Informasi Elektronik dan/atau Dokumen Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah.
- (2) Informasi Elektronik dan/atau Dokumen Elektronik dan/atau hasil cetaknya sebagaimana dimaksud pada ayat (1) merupakan perluasan dari alat bukti yang sah sesuai dengan Hukum Acara yang berlaku di Indonesia.
- (3) Informasi Elektronik dan/atau Dokumen Elektronik dinyatakan sah apabila menggunakan Sistem Elektronik sesuai dengan ketentuan yang diatur dalam Undang-Undang ini.
- (4) Ketentuan mengenai Informasi Elektronik dan/atau Dokumen Elektronik sebagaimana dimaksud pada ayat (1) tidak berlaku untuk:
 - a. Surat yang menurut Undang-undang harus dibuat dalam bentuk tertulis; dan
 - b. Surat beserta dokumennya yang menurut Undang-undang harus dibuat dalam bentuk akta notaris atau akta yang dibuat oleh pejabat pembuat akta.³³

Jika merujuk penjelasan pasal ini maka sebetulnya dokumen elektronik yang dalam hal ini Kamus Elektronik merupakan bukti hukum/alat bukti yang sah sesuai dengan Hukum Acara yang berlaku di Indonesia. Walaupun penggunaannya juga masih dapat diperdebatkan terkait apakah Kamus Elektronik juga harus terdapat akta untuk memperkuat kedudukannya sebagai bukti hukum. Terkait dengan penjelasan yang lebih detail terkait alat bukti ini sebenarnya diatur pada pasal 6 yaitu:

“Dalam hal terdapat ketentuan lain selain yang diatur dalam Pasal 5 ayat (4) yang mensyaratkan bahwa suatu informasi harus berbentuk tertulis atau asli, Informasi Elektronik dan/atau Dokumen Elektronik dianggap sah sepanjang informasi yang tercantum di dalamnya dapat diakses, ditampilkan, dijamin keutuhannya, dan dapat dipertanggungjawabkan sehingga menerangkan suatu keadaan.³⁴”

Jika merujuk dari penjelasan pasal 6 tersebut, maka Dokumen Elektronik dalam hal ini kamus Elektronik jika ingin dijadikan bukti sah di pengadilan maka harus berupa dokumen asli. Dengan kata lain pihak yang memungkinkan untuk menggugat adalah pihak yang merasa dirugikan yang dalam hal ini penulis kamus elektronik.

2.2. Penyelenggaraan Sistem Elektronik

Pada dasarnya, penyelenggaraan sistem apa pun baik yang sifatnya konvensional maupun penyelenggaraan dengan sistem elektronik harus berpedoman pada kaidah-kaidah baik secara teknis maupun secara etis yang berlaku. Secara teknis, penyelenggaraan tersebut berkaitan dengan sistem operasi dan prosedur untuk menjalankan sistem tersebut. Sedangkan secara etis, sistem elektronik harus dilakukan secara bertanggung jawab, aman serta tetap

³³ *Ibid.*, pasal 5

³⁴ *Ibid.*, pasal 6

menjaga privasi dari pengguna. Jika merujuk pada peraturan hukum yang berlaku, hal-hal ini telah diatur dalam Undang-undang No. 11 Tahun 2011 pasal 15 yakni :

- (1) Setiap Penyelenggara Sistem Elektronik harus menyelenggarakan Sistem Elektronik secara andal dan aman serta bertanggung jawab terhadap beroperasinya Sistem Elektronik sebagaimana mestinya.
- (2) Penyelenggara Sistem Elektronik bertanggung jawab terhadap Penyelenggaraan Sistem Elektroniknya.
- (3) Ketentuan sebagaimana dimaksud pada ayat (2) tidak berlaku dalam hal dapat dibuktikan terjadinya keadaan memaksa, kesalahan, dan/atau kelalaian pihak pengguna Sistem Elektronik³⁵.

Jika melihat secara rinci memang aturan terkait pasal 15 Undang-undang ITE terasa sangat luas secara pengertian dan tidak bersifat operasional. Untuk mengatasi hal tersebut, kemudian dijelaskan kembali pada pasal 16 sebagai syarat minimum dalam penyelenggaraan sistem elektronik :

- (1) Sepanjang tidak ditentukan lain oleh undang-undang tersendiri, setiap Penyelenggara Sistem Elektronik wajib mengoperasikan Sistem Elektronik yang memenuhi persyaratan minimum sebagai berikut:
 - a. dapat menampilkan kembali Informasi Elektronik dan/atau Dokumen Elektronik secara utuh sesuai dengan masa retensi yang ditetapkan dengan Peraturan Perundang-undangan;
 - b. dapat melindungi ketersediaan, keutuhan, keotentikan, kerahasiaan, dan keteraksesan Informasi Elektronik dalam Penyelenggaraan Sistem Elektronik tersebut;
 - c. dapat beroperasi sesuai dengan prosedur atau petunjuk dalam Penyelenggaraan Sistem Elektronik tersebut;
 - d. dilengkapi dengan prosedur atau petunjuk yang diumumkan dengan bahasa, informasi, atau simbol yang dapat dipahami oleh pihak yang bersangkutan dengan Penyelenggaraan Sistem Elektronik tersebut; dan
 - e. memiliki mekanisme yang berkelanjutan untuk menjaga kebaruan, kejelasan, dan keberanggungjawaban prosedur atau petunjuk.
- (2) Ketentuan lebih lanjut tentang Penyelenggaraan Sistem Elektronik sebagaimana dimaksud pada ayat (1) diatur dengan Peraturan Pemerintah.³⁶

2.3. Perbuatan yang Dilarang (*Cybercrimes*)

Jika ingin melihat tatanan hukum yang berlaku terkait dokumen elektronik khususnya tentang kamus elektronik maka kita harus paham tentang hal-hal yang dilarang (*cybercrimes*). Dalam penggunaan dokumen elektronik terdapat beberapa hal yang dilarang yakni: (a) akses ilegal; (b) intersepsi ilegal; (c) gangguan pada sistem; dan (d) penyalahgunaan alat

³⁵ *Ibid.*, pasal 15

³⁶ *Ibid.*, pasal 16

(perangkat). Hal-hal tersebut sesuai dengan yang dijelaskan dalam Undang-undang ITE sebagai berikut:

a. Akses Ilegal

Dalam Undang-Undang No. 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (ITE) khususnya pasal 30 telah diatur terkait larangan akses secara ilegal. Secara jelas pasal ini berbunyi sebagai berikut:

- (1) Setiap Orang dengan sengaja dan tanpa hak atau melawan hukum mengakses Komputer dan/atau Sistem Elektronik milik Orang lain dengan cara apa pun.
- (2) Setiap Orang dengan sengaja dan tanpa hak atau melawan hukum mengakses Komputer dan/atau Sistem Elektronik dengan cara apa pun dengan tujuan untuk memperoleh Informasi Elektronik dan/atau Dokumen Elektronik.
- (3) Setiap Orang dengan sengaja dan tanpa hak atau melawan hukum mengakses Komputer dan/atau Sistem Elektronik dengan cara apa pun dengan melanggar, menerobos, melampaui, atau menjebol sistem pengamanan.

Merujuk dari penjelasan pasal 30, maka diketahui bahwa terkait akses dokumen elektronik dapat terkategori dilarang apabila diakses melawan hukum dengan mengakses tanpa izin sistem elektronik milik orang lain dengan tujuan untuk memperoleh informasi khususnya terkait dokumen elektronik (kamus elektronik) dengan menerobos dan menjebol sistem milik orang lain dapat dijerat dengan Undang-Undang ITE.

b. Intersepsi Ilegal

Intersepsi atau penyadapan juga termasuk dilarang jika dikaitkan dengan penggunaan dokumen elektronik (kamus elektronik). Intersepsi ilegal diatur dalam pasal 31 Undang-Undang ITE sebagai berikut :

- (1) Setiap Orang dengan sengaja dan tanpa hak atau melawan hukum melakukan intersepsi atau penyadapan atas Informasi Elektronik dan/atau Dokumen Elektronik dalam suatu Komputer dan/atau Sistem Elektronik tertentu milik Orang lain.
- (2) Setiap Orang dengan sengaja dan tanpa hak atau melawan hukum melakukan intersepsi atas transmisi Informasi Elektronik dan/atau Dokumen Elektronik yang tidak bersifat publik dari, ke, dan di dalam suatu Komputer dan/ atau Sistem Elektronik tertentu milik Orang lain, baik yang tidak menyebabkan perubahan apa pun maupun yang menyebabkan adanya perubahan, penghilangan, dan/atau penghentian Informasi Elektronik dan/atau Dokumen Elektronik yang sedang ditransmisikan.
- (3) Kecuali intersepsi sebagaimana dimaksud pada ayat (1) dan ayat (2), intersepsi yang dilakukan dalam rangka penegakan hukum atas permintaan kepolisian, kejaksaan, dan/atau institusi penegak hukum lainnya yang ditetapkan berdasarkan undang-undang.
- (4) Ketentuan lebih lanjut mengenai tata cara intersepsi sebagaimana dimaksud pada ayat (3) diatur dengan Peraturan Pemerintah.

Dalam pasal 31 tersebut, walaupun telah dijelaskan ada beberapa jenis penyadapan yang dikecualikan oleh penegak hukum, namun dinyatakan bahwa intersepsi (penyadapan) terhadap dokumen elektronik (kamus elektronik) dinyatakan dilarang secara hukum. Dalam Undang-undang ITE ini, kegiatan penyadapan yang dilarang terkait dengan perubahan, penghilangan dan/atau penghentian informasi pada Dokumen Elektronik (Kamus Elektronik).

c. Gangguan pada Sistem

Berbicara mengenai Kamus Elektronik sejatinya tidak hanya berbicara konten (isi) pada kamus, melainkan juga berbicara sistem secara keseluruhan. Selain mengganggu isi Kamus Elektronik, mengganggu sistem yang berjalan juga telah diatur dalam UU ITE pasal 33 sehingga dapat dijerat secara hukum seperti dijelaskan sebagai berikut:

“Setiap Orang dengan sengaja dan tanpa hak atau melawan hukum melakukan tindakan apa pun yang berakibat terganggunya Sistem Elektronik dan/atau mengakibatkan Sistem Elektronik menjadi tidak bekerja sebagaimana mestinya.”

d. Penyalahgunaan Alat atau Perangkat (*misuse of device*)

Seperti telah dijelaskan sebelumnya, berjalannya sebuah sistem adalah hal mutlak yang harus mendapat perlindungan. Perangkat atau alat merupakan bagian yang juga dilindungi dalam UU ITE. Hal ini dijelaskan dalam pasal 34 sebagai berikut:

- (1) Setiap Orang dengan sengaja dan tanpa hak atau melawan hukum memproduksi, menjual, mengadakan untuk digunakan, mengimpor, mendistribusikan, menyediakan, atau memiliki:
 - a. perangkat keras atau perangkat lunak Komputer yang dirancang atau secara khusus dikembangkan untuk memfasilitasi perbuatan sebagaimana dimaksud dalam Pasal 27 sampai dengan Pasal 33;
 - b. sandi lewat Komputer, Kode Akses, atau hal yang sejenis dengan itu yang ditujukan agar Sistem Elektronik menjadi dapat diakses dengan tujuan memfasilitasi perbuatan sebagaimana dimaksud dalam Pasal 27 sampai dengan Pasal 33.
- (2) Tindakan sebagaimana dimaksud pada ayat (1) bukan tindak pidana jika ditujukan untuk melakukan kegiatan penelitian, pengujian Sistem Elektronik, untuk perlindungan Sistem Elektronik itu sendiri secara sah dan tidak melawan hukum.

Kegiatan penyalahgunaan alat atau perangkat ini dilarang secara hukum kecuali jika digunakan pada kegiatan-kegiatan tertentu misalnya penelitian, pengujian sistem dan untuk perlindungan sistem itu sendiri.

Untuk memperkuat kedudukan hukum dari terkait pelanggaran dalam penggunaan Dokumen Elektronik khususnya Kamus Elektronik maka secara khusus dijelaskan sebagai penguat ketentuan yang telah ada, dijelaskan pada pasal 26 UU ITE sebagai berikut :

- (1) Kecuali ditentukan lain oleh Peraturan Perundang-undangan, penggunaan setiap informasi melalui media elektronik yang menyangkut data pribadi seseorang harus dilakukan atas persetujuan Orang yang bersangkutan.
- (2) Setiap Orang yang melanggar haknya sebagaimana dimaksud pada ayat (1) dapat mengajukan gugatan atas kerugian yang ditimbulkan berdasarkan Undang-Undang ini.

Jadi jelas bahwa terkait Kamus Elektronik, pihak bersangkutan yang dalam hal ini penulis dari Kamus Elektronik dapat mengajukan secara hukum apabila merasa haknya dilanggar sebagai pemilik atau penulis dari Kamus Elektronik tersebut sesuai dengan Undang-undang yang berlaku.

III. PENUTUP

Dalam pertimbangan pemanfaatan dan kebebasan penggunaan teknologi informasi dan komunikasi tersebut, sebenarnya telah tersedia Undang-undang ITE disahkan tahun 2008, yang dapat dijadikan sebagai landasan dalam mendukung pembangunan pangkalan data perkamusan dan sekaligus menyangkut tata aturan hukum yang berkaitan dengan upaya pemanfaatan kamus elektronik oleh masyarakat luas. Pengaturan mengenai informasi dan transaksi elektronik mengacu pada beberapa instrumen internasional, seperti UNCITRAL *Model Law on eCommerce* dan UNCITRAL *Model Law on eSignature*. Bagian ini dimaksudkan untuk mengakomodir kebutuhan Para pelaku bisnis di internet dan masyarakat umumnya guna mendapatkan kepastian hukum dalam melakukan transaksi elektronik, termasuk penggunaan kamus elektronik. Beberapa materi yang diatur, antara lain: pengakuan informasi/dokumen elektronik sebagai alat bukti hukum yang sah (Pasal 5 & Pasal 6 UU ITE); penyelenggaraan sistem elektronik (Pasal 15 & Pasal 16 UU ITE); dan perbuatan yang dilarang (*cybercrimes*). Beberapa *cybercrimes* yang diatur dalam UU ITE, antara lain: a) akses ilegal (Pasal 30); b) intersepsi ilegal (Pasal 31); c) gangguan terhadap sistem (system interference, Pasal 33 UU ITE); d) penyalahgunaan alat dan perangkat (misuse of device, Pasal 34 UU ITE); Aspek Hukum dalam penggunaan internet terbagi menjadi: Aspek hak milik intelektual, yaitu yang memberikan perlindungan hukum bagi pembuat karya. Contohnya : Hak Cipta dan Hak Paten. Yurisdiksi hukum dan aspek-aspek terkait. Komponen ini menganalisa dan menentukan keberlakuan hukum yang berlaku dan diterapkan di dalam dunia maya itu. Sebagai penguat, Pasal 26 UU ITE menyebutkan: 1) Penggunaan setiap informasi melalui media elektronik yang menyangkut data pribadi seseorang harus dilakukan atas persetujuan Orang yang bersangkutan. 2) Setiap Orang yang dilanggar haknya sebagaimana dimaksud pada ayat (1) dapat mengajukan gugatan atas kerugian yang ditimbulkan berdasarkan Undang-Undang ini.

Daftar Pustaka

- Asshiddiqie, Jimly. 2005. *Hukum Tata Negara dan Pilar-pilar Demokrasi*. Jakarta: Konpress.
- Sumarlin, Sylvia W. *Indonesia Butuh Lembaga Konten Internet*. Coverage Monthly Telematic Magazine, Jakarta, Edisi 6 Vol. 02/ Mei 2008 hal. 13.
- Rakha Permana, <http://rakhapermanasblog.blogspot.com/2013/05/aspek-hukum-dan-etika-internet.html>.
- I Made Wiryana, <http://nakula.rvs.uni-bielefeld.de/~made/?id=31>.
- Undang-undang No. 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (ITE)

**PERSOALAN HUKUM DAN PEMBERITAHUAN HAK CIPTA
DALAM PEMANFAATAN KORPUS ELEKTRONIK**
*(LEGAL ISSUES AND COPYRIGHT NOTICES
IN THE UTILIZATION OF ELECTRONIC CORPUS)*

Muhammad Fadli Muslimin
Ilmu Sastra UGM
fadlimuslimin@gmail.com

Abstrak

Keterbukaan informasi di era digital menyebabkan akses terhadap data semakin akomodatif, begitu juga dengan korpus. Masalah etika terkait penggunaan korpus data yang diperoleh dari responden, izin pendistribusian, dan pengguna korpus data untuk berbagai tujuan menjadi isu penting. Selain itu, persoalan hukum yang ditimbulkan pun dapat beragam dari masalah etika tersebut. Pertama, masalah etika penggunaan korpus data lisan dan tulis terutama terkait privasi data yang diperoleh; kedua, masalah etika pendistribusian korpus, dalam hal ini penerbit korpus; ketiga, pengguna korpus; keempat, hukum positif terkait pengumpulan dan pendistribusian data korpus berbasis daring. Pergeseran paradigma pengolahan data korpus berbasis teks ke daring menempatkan hak cipta berlaku sebagaimana yang terjadi pada data korpus cetak. Hak cipta digital menjadi ujung tombak untuk membentuk keseimbangan dalam pengaturan tradisional yang ada pada hukum hak cipta. Lisensi digital hak cipta yang dimaksud adalah *creative commons* yang bertujuan untuk memberikan setiap orang, dari pencipta individu sampai dengan perusahaan, dan Lembaga besar hak cipta terhadap ciptaan mereka. Setiap jenis hak cipta memungkinkan pemegang hak cipta untuk mengizinkan orang lain untuk menyalin, menyebarkan, dan menggunakan ciptaan mereka. Korpus data di era digital terutama teks-teks elektronik yang tersebar luas di internet yang selanjutnya dipergunakan sebagai korpus perlu diperhatikan penegasan hak ciptanya dari pencipta ataupun penggunaannya. Lisensi digital *creative commons* memberikan kemudahan bagi pencipta dengan tiga “lapis” lisensi. Lisensi ini tidak saja menunjukkan konsep hukum yang berlaku secara konvensional tetapi sesuatu yang dapat dipahami bersama oleh pemegang hak cipta, pengguna ciptaan yang memanfaatkan data web, ataupun data korpus. Perlindungan hak cipta ini selaras dengan Undang-undang Republik Indonesia Nomor 28 Tahun 2014 tentang Hak Cipta, bahwa hak cipta merupakan kekayaan intelektual di bidang ilmu pengetahuan, seni, dan sastra.

Kata kunci : korpus, hak cipta, hukum, lisensi, *creative common*

Abstract

Disclosure of information in the digital age causes access to data more accommodative as well as with the corpus. Ethical issues related to the use of corpus which data obtained from respondents, distribution permits, and users of data corpus for various purposes become an important issue. In addition, legal issues can also vary from ethical issues. First, the ethical issue of the use of oral and written data corpuses primarily concerns the privacy of the data obtained; second, the ethical problem of distributing the corpus in this case the issuer of the corpus; third, the user of the corpus; fourth, positive law related to the collection and distribution of online-based corpus data. The paradigm shift of text-based text-based data processing paradigm puts copyright in effect as occurs in printed corpus data. Digital copyrights are at the forefront of forming a balance in the traditional settings that exist under copyright law. The copyright digital license in question is the Creative Commons which aims to provide everyone, from individual creators to corporations, and large institutions of copyright to their creations. Every type of copyright allows copyright holders to allow others to copy, distribute and use their creations. The data corpus in the digital era, especially the electronic text-text that is widely spread on the internet which then is used as the corpus need to be considered the affirmation of the copyright from the creator or the user. Creative Commons digital licenses make it easy for creators with three "layers" of licenses. The license is not only shows the conventional concept of law but something that can be understood Together by the copyright holder, the user of creation which utilize web data and corpus data. This copyright protection is in line with the Law of the Republic of Indonesia Number 28 Year 2014 on Copyright, that copyright is an intellectual property in the field of science, art, and literature.

Keywords: *corpus, copyright, law, license, creative common*

I. PENDAHULUAN

1.1. Latar Belakang

Keterbukaan informasi pada era *Digital Revolution 4.0* membuka akses informasi dari berbagai bidang ilmu. Hal tersebut memungkinkan setiap pengguna menggunakan data-data yang terdapat di *The World Wide Web*¹ berupa teks-teks yang tersebar di berbagai situs. Data tersebut diperoleh dari berbagai situs yang selanjutnya dapat membangun data korpus. Proses untuk membangun data korpus berbasis data web dapat menggunakan alat tertentu.

Intuitive tool merupakan alat yang dapat digunakan untuk membangun data korpus yang menggunakan teks-teks daring yang berada di basis data web. *Intuitive tool* dapat diperoleh dari berbagai situs yang menyediakan layanan tersebut, misalnya www.sketchengine.eu yang menyediakan layanan *WebBootCat* atau *Common Crawl Foundation*. Mesin tersebut memungkinkan untuk mengunduh secara otomatis teks-teks yang sesuai dengan kata kunci yang dimasukkan pada saat ingin membangun data korpus.

Alat pencari menemukan laman situs yang relevan dan menampilkan teks-teks di dalam halaman situs tersebut. Selanjutnya, memulai unduh dan diproses menjadi korpus data. Mesin pencari tersebut juga dapat digunakan secara manual pada dokumen-dokumen fisik. Caranya yaitu diunggah ke mesin pencari dan selanjutnya akan diproses untuk mendapatkan korpus data. Meskipun, alat lain memiliki cara yang berbeda dalam prosesnya, tetapi tujuan akhirnya mengambil teks-teks dari web menjadi data korpus.

Tidak dapat diingkari bahwa proses pengambilan data dalam data web tersebut bersifat acak. Contohnya pada mesin *BootCat*, pada dasarnya dalam membangun korpus, mesin ini secara otomatis mencari data di google ataupun sumber lainnya dan mengekstraksi korpus baru berdasarkan serangkaian pertanyaan otomatis google. Pada mesin *Wacky*, data dikonstruksi melalui empat langkah, yaitu *Selecting seed URL*, *Crawling*, *Cleaning The Data*, and *Annotating the Data* (Kristoffersen, 2017:9). Pada mesin *SpiderLing*, *COW*, dan *WebCorpus*, proses membangun data korpus dari teks web memiliki kesamaan, yaitu mengumpulkan, mengelola, membersihkan, dan menyusun korpus.

Permasalahan pada konstruksi data korpus adalah *crawling*. Menurut Kristoffersen (2017:11) *A crawler is programmed to visit web pages and follow links from those web pages to other web pages, and then repeat the procedure recursively. Crawler* melakukan penelusuran langsung mengikuti tautan-tautan pada web sesuai dengan kata kunci yang dimasukkan oleh pengguna. Dalam proses koleksi data, mesin tersebut menggunakan aplikasi pihak kedua, misalnya *DIY crawl*, *Sedd URL*, *Host Bias*, dan *Common Crawl*. *Common Crawl* memiliki tiga langkah dalam menghasilkan data korpus, yaitu *Raw crawl data*, *Metadata*, dan *Extracted text data*. Semua data yang dikumpulkan melalui alat tersebut diperoleh dari seleksi acak. Artinya, data web diperoleh dari beragam situs yang memiliki relevansi sesuai pencarian. *Only the entries for which there exists content that it is possible to extract text from is in these files. This means that the request and metadata entries are stripped, as well as content where the content-type of the response is considered to not be textual* (Kristoffersen, 2017). Hanya data-data sesuai konten yang dapat diekstrak. Artinya, data-data lainnya yang tidak relevan diabaikan termasuk metadata.

Dengan kata lain, pengambilan data web melalui berbagai situs dilakukan tanpa seizin atau sepengetahuan pemegang hak cipta. Pemegang hak cipta tentunya memiliki hak siapa yang berhak dan tidak berhak untuk menggunakan data web tersebut. Dalam hal ini, proteksi terhadap data tersebut menjadi isu yang diperhatikan, tetapi terdapat hal yang perlu dipertimbangkan juga oleh pemegang hak cipta, yaitu apakah data web tersebut telah menyematkan pemberitahuan hak cipta di situsnya atau tidak. Hal tersebut berguna bagi pemilik dan pengguna data korpus agar setiap bentuk pengambilan data dapat dilakukan tanpa menimbulkan permasalahan yang tidak perlu.

Meskipun pemegang hak cipta yang tidak memiliki kesepakatan untuk proyek korpus, hal tersebut tidak menjadi masalah jika pengambilan data dilakukan untuk keperluan penelitian non-komersial. Tetapi, jika hal tersebut memiliki tujuan komersial, hal tersebut menjadi ilegal dan tidak beretika jika menggunakan data-data dari pemegang hak cipta untuk menghasilkan uang untuk dirinya sendiri ataupun kelompok.

Pemegang hak cipta dapat saja mempermasalahkan hal tersebut meskipun pengguna data korpus tidak bermaksud untuk menjual data korpus tersebut. *Copyright holders may still take you to the court. They may, for example, suffer a loss of profit because your use of their*

material diminishes their ability to sell it: why buy a book when you can read it for free in a corpus (cf. also Amsler 2002) dalam (Indurkha & Damerau, 2010:154). Data-data korpus yang tersedia secara daring, dalam hal ini, menjadi rentan untuk dipersoalkan karena berkaitan dengan royalti yang hilang. Kehilangan potensi ekonomi ini menimbulkan persoalan tersendiri terkait pemanfaatan hak ekonomi terhadap ciptaan tanpa pemberitahuan terlebih dahulu.

Pada hakikatnya, undang-undang hak cipta hadir untuk mengatur pemanfaatan ciptaan, dalam hal ini adalah data web dan data korpus. Meskipun demikian, peraturan hak cipta yang sifatnya mengikat terhadap data web dan data korpus masih menjadi perbincangan. Hak cipta digital dapat dijadikan sebagai solusi agar terhindar dari pelanggaran hak. Hak cipta digital menjadi sebuah langkah bagi pemegang hak cipta dan pengguna data agar terhindar dari hal-hal tersebut. Dengan menyertakan pemberitahuan hak cipta digital pada data web dapat memberikan jaminan proteksi terhadap pemegang hak cipta ataupun pengguna data dalam pemanfaatannya untuk tujuan penelitian atau komersial dengan memperhatikan hukum yang berlaku.

Persoalan hukum dari penggunaan korpus juga menjadi isu yang tidak dapat diabaikan begitu saja di era keterbukaan data saat ini karena penggunaan data korpus beragam dan dalam berbagai tujuan. Data-data web diyakini masih banyak yang belum terlindungi oleh hak cipta sehingga penting untuk memahami bagaimana hak cipta digital dapat menjadi langkah positif bagi pemegang dan pengguna data. Perbedaan status hak cipta yang disertakan terhadap data web dan data korpus memerlukan sebuah lisensi konten terbuka untuk mempertegas status data.

1.2. Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka dapat dirumuskan permasalahan sebagai berikut:

1. Apa persoalan hukum terkait data korpus elektronik?
2. Bagaimana hak cipta digital dapat melindungi pemegang hak cipta dan pengguna korpus data elektronik?

1.3. Tujuan

Berdasarkan latar belakang penelitian dan masalah yang telah diungkapkan, penelitian ini memiliki tujuan sebagai berikut:

1. Mengetahui persoalan hukum terkait data korpus elektronik.
2. Mengetahui konsep hak cipta digital dalam pemanfaatan korpus elektronik.

1.4. Tinjauan Pustaka

Secara spesifik, peneliti belum menemukan penelitian yang membahas secara komprehensif mengenai persoalan hukum terkait dengan data korpus elektronik. Namun, penelitian yang dilakukan oleh Maarten Tryuens dan Patrick Van Eecke yang berjudul *Legal Aspects of Text Mining* dari *University of Antwerp* membuka cakrawala pengetahuan

mengenai *data mining* atau penambangan data, dalam istilah sederhananya disebut pengambilan data, khususnya *text mining*. Penelitian tersebut mengungkapkan bahwa *text mining* mendapatkan perhatian yang terbatas yang berhubungan pada ketentuan hukumnya. Terdapat hambatan hukum yang menarik antara ketatnya aturan hak cipta dan hukum basis data. Konflik dapat saja terjadi jika data diambil dari pangkalan data komersial dan dipergunakan bahkan pada penelitian non-komersial. Perizinan menjadi hal yang sulit untuk dilakukan dan memakan waktu untuk mendapatkan izin penggunaan data untuk membangun data korpus. (Truyens & Van Eecke, 2014).

II. KERANGKA TEORI DAN METODE PENELITIAN

Menurut World Intellectual Property Organization (WIPO, 2008:40-41) *Copyright deals with the rights of intellectual creators in their creation*. Hal ini berkaitan dengan hak cipta dari pencipta karya; *Copyright law is a branch of that part of the law which deals with the rights of intellectual creators*. Hal ini berkaitan dengan undang-undang yang mengatur tentang hak-hak pencipta. *International solutions to the legal and administrative problems posed by digital technology, especially the Internet, to the traditional notions and practices of intellectual property*. Perkembangan teknologi memberikan dampak terhadap permasalahan administratif dan hukum terhadap gagasan hak cipta tradisional sehingga perlu untuk dilakukan analisis terhadap permasalahan ini. *Open content licenses enables your work to circulate in a much wider manner than if there were restrictions on the work. Apart from making you more visible, this model of distribution also enables you to either obtain more shows, more work, sponsorship etc* (Liang, 2004:34). Lisensi konten terbuka memungkinkan pemegang hak cipta memberikan keleluasaan terhadap pengguna datanya lebih leluasa.

Metode penelitian ini adalah deskriptif analisis untuk menemukan persoalan hukum yang terjadi pada penggunaan data beserta ketentuan hukum hak ciptanya dalam membangun data korpus yang dianalisis dalam perspektif hak cipta ‘tradisional’. Selanjutnya, menemukan solusi alternatif terhadap hal tersebut di era keterbukaan informasi dengan menawarkan konsep lisensi konten terbuka yang menggunakan hak cipta digital, yaitu *creative common*.

III. ANALISIS PENELITIAN

3.1. Persoalan Hukum Terkait Pemanfaatan Data Web dan Data Korpus

Proteksi terhadap data merupakan aspek privasi yang perlu diperhatikan. Sebagai contoh, proyek korpus lisan yang dalam prosesnya meliputi *collection, storage, modification, transmission, blocking, deletion and other uses* (DFG Review Board on Linguistics, 2017:6) membutuhkan persetujuan dan kerja sama dari orang-orang yang terlibat dalam perekaman. Hal ini membutuhkan kepastian hukum terhadap proteksi data dari orang-orang yang menjadi subjek pengambilan data yang tentunya sejalan dengan kepentingan komunitas peneliti. Tujuan dari proteksi data tersebut merupakan implementasi terhadap undang-undang data proteksi dan regulasi umum yang melindungi individu terhadap data yang diolah dan digunakan oleh otoritas lain.

Perlindungan terhadap data tersebut dalam undang-undang nasional Jerman disebut *Bundesdatenschutzgesetz BDSG*. Undang-undang tersebut menyatakan bahwa “*The purpose of this Act is to protect the individual against his/her right to privacy being impaired through the handling of his/her personal data.*” (DFG Review Board on Linguistics, 2017:). Ketentuan hukum berlaku terhadap korpus lisan yang meliputi audio dan rekaman video, transkrip, dan metadata mengenai pembicara. Proses persetujuan pun dilakukan melalui hukum yang berlaku dengan menandatangani persetujuan dari setiap subjek rekaman. Persetujuan dilakukan secara tertulis, deklarasi persetujuan oral harus dituliskan, dan deklarasi persetujuan elektronik harus didokumentasikan.

Pengarsipan dan publikasi data setelah ataupun sebelum proyek korpus harus telah dinyatakan secara eksplisit dalam deklarasi. Aspek hak cipta terhadap korpus lisan berkaitan dengan rekaman audio dan video, material tertulis, gambar, grafik, power point, dan lain-lain, merupakan persoalan yang berhubungan dengan penggunaannya di dalam program penelitian. Data-data yang telah diunggah ke situs-situs secara daring memberikan kemungkinan pengambilan data menggunakan *Intuitive tool*. Mengacu kepada ketentuan penggunaan data yang memiliki hak cipta, tentunya memerlukan persetujuan legal secara tertulis daripemegang hak cipta diperlukan untuk melakukan penelitian ataupun untuk kepentingan lainnya.

Hal ini tentunya menjadi persoalan yang kompleks bagi peneliti ataupun pengguna data korpus yang memfokuskan pada penelitian tertentu yang memanfaatkan data-data web untuk membangun korpus tersebut di era keterbukaan informasi. Dengan data-data web, memungkinkan penggunanya dapat menggunakan data tersebut untuk berbagai kepentingan. Undang-undang hak cipta ‘tradisional’ memberikan jaminan hukum terhadap data-data pemegang hak cipta, tetapi disisi lain memandang bahwa kemungkinan persoalan hukum yang diakibatkan oleh pengambilan data tanpa izin dapat menyebabkan berlakunya sanksi-sanksi berdasarkan hukum hak cipta. Namun demikian, data web dalam pemanfaatannya di era keterbukaan informasi memungkinkan data-data tersebut dapat saja diperoleh tanpa persetujuan langsung dari pemegang hak cipta. Lisensi konten terbuka dapat menjadi jembatan sebagai alternatif lisensi konten terbuka untuk menjembatani hal tersebut dengan tetap memberikan ketentuan-ketentuan hukum yang berlaku.

3.2. Penambahan Teks Web

Keterbukaan informasi menyebabkan pengambilan data secara masif melalui mesin-mesin tertentu menyebabkan masalah privasi, apalagi jika dikaitkan dengan hak cipta terhadap data web yang dimiliki oleh pemegang hak cipta. Hal yang perlu dipahami terkait hal ini adalah masalah hukum yang disebabkan oleh pengambilan data web dalam jumlah yang besar untuk kepentingan data korpus. Istilah ini dikenal dengan *web mining*. *Web mining is a concept that gathers all techniques, methods and algorithms used to extract information and knowledge from data originating on the web (web data). Thus, it could be said that this is an application of data mining theory onto web data* (Velásquez 2013:5229). Data web yang dimaksud dalam hal ini adalah data-data yang bersifat pribadi, misalnya nama, alamat rumah, *email* untuk kepentingan komersial. Dalam sudut pandang hukum, *web mining* telah mendapatkan perhatian hukum terkait proteksi data penggunaannya, tetapi

berbanding terbalik terhadap *text mining*. Diskusi mengenai *text mining* kurang mendapatkan perhatian, hal ini disebabkan *less (direct) privacy impact. Even so, the European Commission recently acknowledged the importance of text mining, and wants to promote its use for scientific research purposes. While attention for text mining is slowly rising, legal discussions remain scarce.* (Truyens & Van Eecke, 2014:2182). Isu mengenai penggunaan data berbasis teks belum menjadi perhatian utama. Belum terdapatnya hukum khusus yang menangani *teks mining* menyebabkan perangkat lunak, algoritma dan metode bisnis tertentu dapat memanen data berbasis teks untuk keperluan tertentu.

Dikutip dari laman <https://corpus.byu.edu>, menyatakan bahwa penggunaan teks-teks yang dipanen melalui berbagai situs telah mengikuti ketentuan kebijakan hukum yang berlaku di Amerika yang disebut *Fair Use Law*. Dikutip dari laman <https://fairuse.stanford.edu/overview/fair-use/>, *Fair use is a copyright principle based on the belief that the public is entitled to freely use portions of copyrighted materials for purposes of commentary and criticism.* Kebijakan ini tentunya berbeda di setiap negara,

Di Indonesia, pemerintah belum memiliki peraturan khusus untuk melindungi data yang diperoleh melalui *web mining* ataupun *teks mining*. Dikutip dari merdeka.com, ‘Pemerintah Indonesia hingga kini belum mempunyai Undang-Undang khusus untuk mengatur dan menjamin perlindungan data privasi, baik yang dikelola oleh pemerintah maupun swasta. Padahal, trend penggunaan internet kian meningkat, mulai dari sosialita, bisnis, hingga politik. Pengamanan berupa landasan hukum perlu dibentuk untuk mencegah kejahatan dan menyelesaikan masalah yang terjadi khususnya di dalam era Big Data’ (Putra, 2017). Artinya mesin-mesin pencari tersebut memiliki kebebasan untuk memanen data berupa teks di berbagai situs di Indonesia terutama dalam penyusunan data korpus berbahasa Indonesia. Ketersediaan data web memungkinkan data diambil tanpa perlu mengajukan izin atau pemberitahuan kepada pemegang hak cipta.

Pertanyaan yang mengemuka terkait *text mining* yang dilakukan mesin pencari adalah apakah hal tersebut merupakan sebuah bentuk pelanggaran hak cipta terhadap pengambilan data web tanpa seizin pemilik yang berupa kombinasi kata-kata atau sekedar kata-kata. Meskipun demikian, menurut Maarten Tryens setelah mengevaluasi keputusan Pengadilan Hukum Uni Eropa, dia menyatakan bahwa hal tersebut tergantung dari isinya, sebuah fragmen teks sejumlah sebelas kata bisa cukup asli, dan dengan demikian dilindungi.

Pada gilirannya, undang-undang hak cipta berada di garis terdepan dalam memberikan perhatian terhadap properti intelektual dalam konteks ini. Tetapi, definisi dari properti intelektual pun sebenarnya jika dikaitkan dengan data-data web, yaitu teks-teks atau karya tekstual merupakan sebuah perbedabatan apakah undang-undang hak cipta memberikan perlindungan terhadap data-data yang berupa teks atau tidak. Di Indonesia, terkait hal tersebut, UU nomor 28 tahun 2014 Pasal 40 ayat 1 menyatakan bahwa kompilasi ciptaan atau data, baik dalam format yang dapat dibaca dengan program komputer maupun media lainnya mendapatkan perlindungan. Mengacu kepada pengertian tersebut, undang-undang hak cipta Indonesia memberikan proteksi terhadap data web yang terdapat di berbagai situs. Namun demikian, belum terdapatnya peraturan khusus yang mengatur *text mining* masih membutuhkan telaah yang lebih dalam.

3.3. Perkembangan Hak Cipta

Hak cipta secara otomatis memberikan perlindungan terhadap intelektual properti sekaligus menjadi hukum yang mengikat terkait kepemilikan suatu ciptaan. Latar belakang lahirnya hak cipta bermula di Inggris pada tahun 1556 mengenai hak abadi dan eksklusif terhadap penerbit dalam memublikasikan karya-karya tanpa memperhatikan dan berniat memproteksi atau memberikan penghargaan kepada penulisnya. Setelah hampir satu setengah abad pengabaian terhadap hukum yang mengatur hak-hak penulis, tahun 1790, undang-undang federal pertama mengatur hak cipta dan paten akhirnya diterbitkan (Gorman, 2006:2). Hal tersebut terkait tuntutan terhadap pemerintah terkait nasib penulis dan keluarganya yang disebabkan oleh penyalinan tidak sah terhadap buku-buku mereka.

Setelah itu, terjadi perubahan signifikan terhadap perkembangan hak cipta dalam melindungi pemegang hak cipta. Penetapan hak cipta tahun 1909 berkaitan dengan hak-hak eksklusif yang diperoleh oleh pemegang hak cipta setelah karyanya dicetak, disalin, diadaptasi, dijual, dan dipamerkan. Penetapan hak cipta Tahun 1976 meningkatkan status proteksi terhadap publikasi dari pemegang hak cipta menjadi 75 tahun setelah meninggal dunia, setelah sebelumnya hanya 50 tahun. Hak eksklusif lainnya, yaitu hak terhadap karya yang ditampilkan melalui televisi ataupun komputer. Hingga sampai tahun 1998 proteksi terhadap hak cipta menjadi 95 tahun.

Hak cipta menjadi sebuah instrumen untuk memproteksi karya yang dimiliki oleh individu dengan syarat mendaftarkannya terlebih dahulu. Menurut Adya Paramita Prabandari (2013:170) bahwa hak cipta secara otomatis melekat pada suatu hasil karya atau ciptaan pada saat ciptaan tersebut mempunyai bentuk yang nyata, riil, atau pendaftaran hak cipta ke kantor hak cipta hanya untuk memberikan posisi tawar yang lebih kuat bagi si pencipta atau pemegang hak cipta apabila terjadi sengketa hak cipta di kemudian hari. Hak cipta dalam ruang lingkup ini merupakan konsep hak cipta ‘tradisional’ yang berlaku pada karya-karya yang memiliki bentuk nyata.

Pada gilirannya, hak cipta digital berperan terhadap proteksi properti intelektual yang jangkauannya lebih luas dan difungsikan secara digital. Kedudukannya sama dengan hak cipta ‘tradisional’. Perbedaannya terletak pada izin otomatis atau tidak terhadap karya.

Hak Cipta	Lisensi Digital
All right Reserved	Some right reserved

Perbedaannya terletak pada frasa yang tercantum. Frasa *All right reserved* mengandung pengertian bahwa pemegang hak cipta memegang otoritas penuh terhadap karyanya. Sedangkan, *Some Right Reserved* memberikan keleluasaan langsung kepada pengguna untuk menyebarluaskan, mengubah, atau menggubah tanpa izin dari pemilik konten. Kepemilikan lisensi terbuka tersebut dapat menjadi solusi terhadap data web yang bersifat terbuka bagi mesin pencari ataupun bagi pemilik konten.

3.4. Hak Cipta Digital

Hak cipta digital menjadi alat yang penting di era digital keterbukaan data. Pertanyaan-pertanyaan mendasar mengenai hal ini adalah apakah ‘kepemilikan’ yang akan dilindungi

atau ‘pembajakan’? (Lessig, 2004:11) di era digital keterbukaan data. Siapakah yang berperan terhadap dua kutub ‘kepemilikan’ dan ‘pembajakan’, tentu jawabannya adalah internet. Menurut Lawrence Lessig (2004:19) internetlah yang memprovokasi perang ini. Internet memungkinkan persebaran konten yang efisien. Proses berbagi file ke sesama pengguna (*peer-to-peer*, atau selanjutnya disingkat p2p) merupakan cara berbagi file yang paling efisien dari teknologi-teknologi efisien yang disediakan internet. Dengan menggunakan kecerdasan yang tersebar (*distributed intelligent*), sistem p2p memudahkan proses berbagi konten yang tidak pernah terbayang di generasi yang sebelumnya. Aturan-aturan formal dibuat untuk mencegah dampak kebebasan individu ataupun kelompok dalam memproduksi ciptaan berdasarkan ciptaan orang lain untuk dijadikan komoditas berharga dan dipasarkan tanpa sepengetahuan pemilik hak cipta.

Sehubungan dengan hal tersebut, menurut Till Kreutzer, rezim hak cipta sebagai sesuatu yang terlalu membatasi para pengguna ciptaan dan dunia penciptanya sendiri. Oleh karena itu, lisensi konten terbuka hadir mewujudkan kepemilikan digital bersama dan budaya bebas yang didasarkan pada tiga prinsip dasar, yaitu hubungan hukum yang lebih sederhana, pemberian izin yang luas dan bebas royalti untuk menggunakan ciptaan, dan pengurangan ketidakpastian hukum.

Pemegang hak cipta memiliki wewenang, yaitu membuat karya mereka dapat diakses bebas dengan lisensi digital hak cipta. Lisensi digital *Creative Common* dapat menjadi alternatif pada perlindungan hak cipta khususnya dalam bentuk digital. *Creative Common* menghadirkan prinsip lisensi konten terbuka. Lisensi konten terbuka memiliki manfaat bagi banyak pihak terutama bagi pengguna data korpus ataupun data web. Hal ini disebabkan oleh setiap pihak pemegang hak cipta dapat membuat konten datanya dapat digunakan untuk kepentingan tertentu hanya dengan menyematkan jenis lisensi yang diinginkan tanpa perlu memperoleh perizinan langsung, dan bahasa lisensi yang mudah dimengerti penggunaannya. Pengguna data web atau data korpus elektronik dapat secara mudah menjadi seorang pencipta karena lisensi terbuka merupakan alat yang membuat hak cipta kompatibel dengan era digital.

3.5. Lisensi Digital *Creative Common*

Lisensi digital *Creative Common* merupakan lisensi alternatif untuk melindungi hak cipta dalam bentuk digital. Lisensi hak cipta digital memiliki tiga bentuk ringkas yang memungkinkan pihak ketiga yang memiliki ikatan hukum serta kode metadata yang melekat pada ciptaan digital.

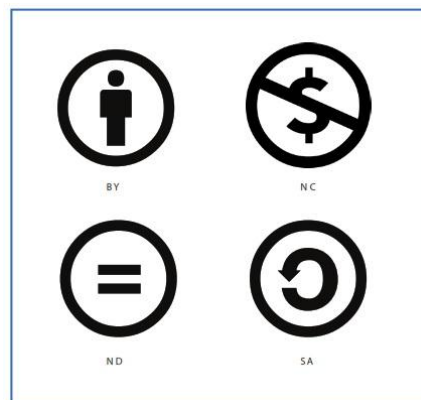
Data web atau data korpus dapat dikategorikan ke dalam properti intelektual yang dapat disematkan lisensi digital. Hal ini mengacu pada ruang lingkup hak-hak properti *intelektual data managemen* yang dikutip dari data.research.cornell.edu. *A builder of a database or other data resource will have an interest in who owns that resource and how others may use it. Someone who may populate that resource with data provided in part by others will want to make sure that all legal, ethical, and professional obligations that one may have to the provider of the data are met. Since the benefits of data sharing are so well known and documented, a researcher may wish to share their database and/or content with others* (Cornell University, n.d.). Meskipun, pernyataan eksplisit mengenai data web ataupun data korpus dapat dikategorikan ke dalam properti intelektual, tetapi fakta bahwa data web

ataupun data korpus merupakan sumber data yang memerlukan kewajiban hukum, etika dan profesionalnya telah terpenuhi.

Secara umum, lisensi konten terbuka dapat dipahami sebagai lisensi yang berguna untuk menyebarluaskan, membuat tersedia untuk publik, dan memproduksi kembali sebuah ciptaan untuk tujuan nonkomersial dengan cara apapun dan melalui media apapun tanpa dikenakan biaya (Kreutzer, 2014:15). Dalam pengertian ini pengguna data web dan pengguna data korpus dengan produk turunannya (jika ada) tidak perlu khawatir terhadap persoalan hukum yang mungkin timbul sebagai akibat penggunaan data tersebut sepanjang mengikuti aturan lisensi yang ditentukan.

Disamping itu, terdapat tiga keuntungan lisensi terbuka *Creative Common*, yaitu (1) penyebarluasan yang lebih luas. Hal ini memiliki efek positif terhadap berbagi konten secara sah tanpa melanggar hukum hak cipta yang telah ada; (2) meningkatkan kepastian hukum dan menyederhanakan hubungan hukum. Narasi dalam bahasa sederhana, terkait ketentuan lisensi memberikan kemudahan bagi pemberi lisensi dan penerima lisensi dalam hal memahami informasi mengenai aturan menggunakan ciptaan. Pada tingkatan ini, menyederhanakan hubungan hukum kedua belah pihak; dan (3) dengan sengaja menyerahkan kendali. Hal ini dimaksudkan memberikan kendali terhadap penggunaan ciptaan seseorang. Saat data telah diunggah ke jaringan, maka sulit untuk membatasi atau mengendalikan data tersebut. Dengan memberikan akses terhadap data tersebut melalui ketentuan lisensi terbuka, maka baik pemilik ataupun pengguna dapat saling menguntungkan.

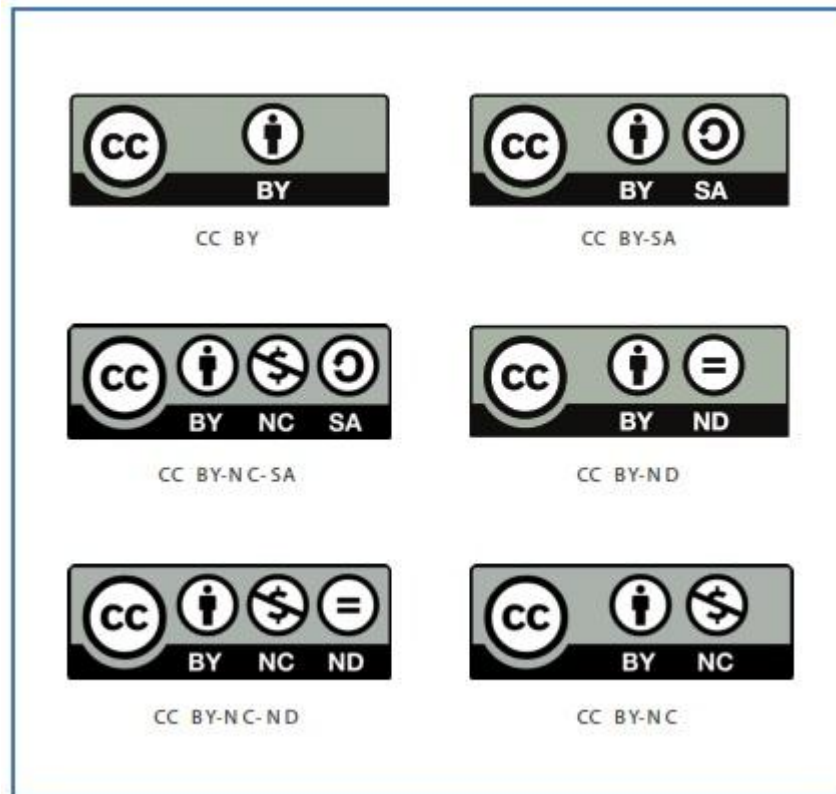
Terdapat enam jenis lisensi kombinasi *Creative Common* untuk memenuhi kebutuhan pemegang hak cipta yang diperoleh dari empat unsur dasar lisensi.



Gambar 1

“BY” adalah singkatan untuk atribusi (kewajiban untuk mencantumkan nama pencipta dan pihak lain yang harus dicantumkan namanya); “NC” adalah NonKomersial (penggunaan untuk tujuan komersial tidak diizinkan oleh lisensi); “ND” adalah Tanpa Turunan (hanya salinan verbatim dari ciptaan yang boleh dibagikan); dan “SA” adalah Berbagi Serupa (contohnya, sebuah ciptaan dapat menjadi dasar pembuatan ciptaan turunan dan ciptaan turunan tersebut hanya dapat disebarluaskan dengan lisensi yang sama dengan lisensi pada ciptaan atau dengan lisensi yang sesuai atau kompatibel) (Kreutzer, 2014:32).

Keempat unsur dasar lisensi tersebut, selanjutnya dapat diturunkan lagi menjadi enam macam lisensi yang memiliki tujuan yang berbeda.



Gambar 2

Penjelasan singkat mengenai enam jenis lisensi tersebut adalah sebagai berikut. CC BY (Atribusi), lisensi ini memberikan lisensi tidak terbatas untuk menggunakan konten yang dilisensikan; CC BY-SA (Atribusi-Berbagi Serupa), lisensi umum yang digunakan Wikipedia, lisensi ini adalah yang paling sering digunakan. Lisensi yang memungkinkan konten mereka untuk diunggah ke Wikipedia, atau ingin menggabungkan konten Wikipedia.; CC BY-ND (Atribusi-Tanpa Turunan), lisensi yang tidak mengizinkan pembuatan ciptaan turunan terhadap ciptaan berlisensi. Untuk melindungi integritas dari ciptaan tersebut, hanya salinan verbatim dari ciptaan tersebut yang dapat disebarluaskan dan dibagikan. Pembatasan tanpa turunan dapat menimbulkan beberapa masalah yang signifikan. (Kreutzer, 2014:34)

CC BY-NC (Atribusi-NonKomersial), bertentangan dengan beberapa lisensi sebelumnya, lisensi ini tidak mengizinkan pengguna konten untuk tujuan komersial; CC BY-NC-SA (Atribusi-NonKomersial-Berbagi Serupa), lisensi yang menggabungkan fitur NonKomersial dengan fitur Berbagi Serupa. Dengan demikian, sebuah ciptaan dapat didaptasi dan bentuk adaptasi tersebut dapat dibagikan di bawah ketentuan tertentu; CC BY-NC-ND (Atribusi-NonKomersial-Tanpa Turunan), yaitu lisensi yang paling terbatas. Meskipun demikian, terdapat akibat hukum menggunakan lisensi konten terbuka, yaitu hilangnya kendali terhadap konten tertentu, keuntungan terhadap konten menjadi terbatas, penggunaan lisensi tidak dapat dibatalkan.

Prinsip dasar konten terbuka memungkinkan pemegang hak cipta untuk menyebarluaskan ciptaan mereka dengan lebih mudah dengan memperbolehkan setiap orang

untuk menggunakan, menyebarluaskan, dan mengubah atau menggubah ciptaan mereka tanpa harus meminta izin dengan ketentuan lisensi CC BY-NC (Atribusi-NonKomersial), dan juga mempertimbangkan konten terbuka mempermudah akses informasi dan pengetahuan serta selaras dengan era digital. Pada dasarnya, selaras dengan Undang-Undang Republik Indonesia Nomor 28 Tahun 2014 tentang Hak Cipta Pasal 1 Nomor 20 tentang Lisensi adalah izin tertulis yang diberikan oleh Pemegang Hak Cipta atau Pemilik Hak Terkait kepada pihak lain untuk melaksanakan hak ekonomi atas Ciptaannya atau produk Hak Terkait dengan syarat tertentu.

IV. SIMPULAN

Persoalan hukum terkait pemanfaat data korpus elektronik ataupun data web dipicu oleh kemampuan lisensi hak cipta untuk mengakomodasi berbagai kepentingan pengguna data tersebut. Tentunya, undang-undang hak cipta “tradisional” memberikan beban terhadap data itu sendiri untuk diolah secara bebas untuk kepentingan nonkomersial terutama bagi peneliti untuk kepentingan akademik. Sementara itu, Undang-undang hak cipta secara global ataupun di Indonesia belum mempunyai regulasi yang spesifik untuk mengatur *text mining* terhadap data web. Di sisi lain, keterbukaan informasi tidak dapat dipungkiri memberikan keleluasaan bagi pemegang hak cipta untuk mengizinkan atau membatasi datanya dengan lisensi digital konten terbuka.

Tentunya, lisensi tersebut dapat memberikan bantuan bagi pemegang hak cipta terhadap data web dan bagi penggunaannya dalam hal ini membangun data korpus. Tetapi, perlu dilakukan telaah lebih jauh mengenai kemampuan lisensi *Creative Common* yang lebih spesifik terhadap data web dan data korpus.

V. DAFTAR PUSTAKA

- Cornell University. (n.d.). *Introduction to intellectual property rights in data management*. Retrieved July 12, 2018, from <https://data.research.cornell.edu/content/intellectual-property>
- DFG Review Board on Linguistics. (2017). *Guidelines for Building Language Corpora Under German Law*. Jerman: CLARIN Legal Issues Committee.
- Gorman, R. A. (2006). *Copyright Law*. Pennsylvania: Federal Judicial Center.
- Indurkha, N., & Damerau, F. J. (2010). *Natural Language Processing* (2nd ed.). United State of America: CRC Press.
- Kreutzer, T. (2014). *Konten Terbuka-Pedoman Praktis Penggunaan Lisensi Creative Commons*. (A. Q. Sudharto, Ed.). Jakarta Pusat: Perkumpulan Wikimedia Indonesia.
- Kristoffersen, K. B. (2017). *Common Crawled Web Corpora*. University of Oslo.
- Lessig, L. (2004). *Budaya Bebas : Bagaimana Media Besar Memakai Teknologi dan Hukum untuk Membatasi Budaya dan Mengontrol Kreativitas*. (F. Thajib, Ed.). Yogyakarta: Kunci Cultural Studies Center.
- Liang, L. (2004). *Guide to Open Content Licenses*. Rotterdam: Piet Zwart Institute.
- Prabandari, A. P. (2013). *Komparasi Pengaturan Hak Cipta Di Indonesia dan Amerika*

- Serikat. Masalah-Masalah Hukum; Masalah-Masalah Hukum Jilid 42, Nomor 2, Tahun 2013* DO - 10.14710/Mmh.42.2.2013.162-171 . Retrieved from <https://ejournal.undip.ac.id/index.php/mmh/article/view/5803>
- Putra, I. R. (2017). *Genjot ekonomi digital, pemerintah diminta buat UU perlindungan data / merdeka.com*. Retrieved July 8, 2018, from <https://www.merdeka.com/uang/genjot-ekonomi-digital-pemerintah-diminta-buat-uu-perlindungan-data.html>
- Truyens, M., & Van Eecke, P. (2014). *Legal aspects of text mining*. In N. Calzolari (Ed.), *LREC 2014, Ninth International Conference on Language Resources and Evaluation* (pp. 153–170). Iceland: the European Language Resources Association.
- Velásquez, J. D. (2013). *Web mining and privacy concerns: Some important legal issues to be consider before applying any data and information extraction technique in web-based environments*. *Expert Systems with Applications*, 40(13), 5228–5239. <https://doi.org/10.1016/J.ESWA.2013.03.008>
- WIPO. (2008). *WIPO Intellectual Property Handbook* (2nd ed.). Switzerland: World Intellectual Property Organization.

ANALISIS KUALITAS UX KBBI LURING PADA APLIKASI TELEPON GENGAM: TINJAUAN FUNGSI DAN PENGGUNAAN KAMUS ELEKTRONIK

(*UX QUALITY ANALYSIS OF OFFLINE KBBI IN MOBILE APPS: A REVIEW OF ELECTRONIC DICTIONARY FUNCTION AND USAGE*)

Satwiko Budiono

Pusat Pengembangan dan Pelindungan, Badan Pengembangan dan Pembinaan Bahasa
satwiko.iko@gmail.com atau satwiko.budiono@kemdikbud.go.id

Abstrak

Kualitas *User Experience* atau biasa disebut UX akan menentukan tingkat kesuksesannya pemenuhan kebutuhan pengguna dalam pengembangan produk. Jika UX dari sebuah produk tidak disesuaikan dengan kebutuhan, maka produk tersebut akan ditinggalkan penggunaannya. Dalam hal ini, KBBI luring sebagai kamus elektronik yang bertujuan memenuhi kebutuhan pengguna bahasa Indonesia dalam mencari arti dan makna kata dituntut memiliki UX yang efektif dan efisien. Hal ini disebabkan KBBI luring merupakan sumber rujukan yang digunakan secara luas sehingga penggunaannya diharuskan mempermudah pengguna. Untuk itu, penelitian ini dilakukan untuk menelusuri sejauh mana kualitas UX KBBI luring memenuhi kebutuhan pengguna. KBBI luring dipilih menjadi objek penelitian karena aplikasi telepon genggam lebih dekat dengan pengguna dibandingkan laman yang membutuhkan jaringan internet. Penelitian ini akan menyajikan survei tertulis pengguna terhadap UX KBBI luring yang disesuaikan dengan daftar pertanyaan dari Anderson (2010). Pengguna yang dipilih sebagai partisipan adalah mahasiswa aktif yang kesehariannya dekat dengan penggunaan KBBI untuk tujuan referensi umum. Temuan penelitian ini akan dapat membantu meningkatkan fungsi dalam penggunaan kamus elektronik secara efektif sehingga masukan dalam penelitian akan dapat menjadi bahan evaluasi untuk pengembangan tampilan KBBI luring.

Kata kunci: leksikografi digital, kamus elektronik, UX aplikasi telepon genggam, dan KBBI luring.

Abstract

User Experience quality or is called UX would determine the success levels of user needs fulfillment in product development. If the UX of a product is not adjusted to the needs, the product will be left behind by the users. In the same case, Offline KBBI—'Indonesian Dictionary' as an electronic dictionary, which is aimed to fulfill the users needs of finding Indonesian word meanings and definitions, is expected to have an effective and efficient UX. Offline KBBI as a widely used reference is required to be user-friendly. Therefore, the

purpose of this research is to explore how far the offline KBBI in the form of mobile application in fulfilling the users needs. Offline KBBI in mobile application is chosen because it is closer to the user than the website version which needs internet connection. This research will present the user written survey towards UX offline KBBI in form of mobile application according to the question list of Anderson (2010). The users chosen as participants are college students who use offline KBBI in their daily lives as a general reference. The research findings are to help developing the function of electronic dictionary usage effectively and to evaluate the development of offline KBBI mobile application features.

Keywords: *digital lexicography, electronic dictionary, UX mobile application, and offline KBBI.*

I. PENDAHULUAN

Perkembangan teknologi semakin lama semakin pesat, canggih, dan memudahkan penggunaannya. Dimulai dari adanya keberadaan komputer yang dilanjutkan dengan munculnya telepon genggam. Bahkan, kecanggihan telepon genggam yang ada sekarang membuat benda tersebut menjelma menjadi telepon pintar atau lebih dikenal dengan *smartphone*. Dengan ditambah koneksi internet, telepon pintar sangat membantu penggunaannya dalam berbagai hal mulai dari berkomunikasi, berselancar, hingga mencari sesuatu hal yang awalnya sulit menjadi mudah. Salah satu contoh dari pencarian yang awalnya sulit menjadi mudah adalah adanya aplikasi Kamus Besar Bahasa Indonesia (KBBI) pada telepon genggam. Hal ini disebabkan pencarian kata pada kamus tidak membutuhkan waktu lama dan tidak pula membutuhkan biaya karena pengguna tidak harus membeli kamus cetaknya terlebih dahulu. Ditambah lagi, aplikasi KBBI pada telepon genggam ini bersifat luring sehingga tidak membutuhkan internet untuk pemakaiannya.

Kemunculan aplikasi KBBI ini sangat berguna untuk memenuhi kebutuhan masyarakat akan kata baku, makna kata, dan sebagainya. Terlebih lagi, pengguna telepon genggam di Indonesia lebih banyak dibandingkan pengguna laptop atau komputer (Google Consumer Barometer, 2018). Hal ini semakin menguatkan bahwa keberadaan KBBI luring ini membantu masyarakat Indonesia perihal penguasaan kosakata bahasa Indonesia. Dengan begitu, aplikasi KBBI luring ini harus memiliki tampilan menarik dan mudah dipahami oleh orang banyak. Hal tersebut sesuai dengan pendapat Levy (2015) yang menyatakan bahwa desain *User Experience* (UX) sangat menentukan nilai atau citra merek dari sebuah produk. Hal yang dimaksud adalah ketika desain UX kurang bagus akan berdampak pada nilai atau citra merek dari produk tersebut juga kurang bagus. Sebaliknya, ketika desain UX bagus akan berdampak pada nilai atau citra merek dari produk juga bagus.

Dalam hal ini, ukuran bagus atau tidaknya desain UX bergantung pada seberapa besar aplikasi tersebut memahami kebutuhan pengguna sehingga pengguna memberikan tanggapan memuaskan dengan cara memakai aplikasi tersebut secara rutin dan diakses oleh orang banyak. Hal tersebut diketahui dari survei yang dilakukan Loop (2018) tentang pemahaman UX di beberapa negara besar, seperti Amerika, Australia, India, Inggris, dan masih banyak lagi. Dari hal tersebut diketahui bahwa selama penggunaannya banyak maka desain UX sudah

dianggap baik karena sesuai dengan kebutuhan pengguna. Namun, jumlah pengguna yang sudah banyak sekalipun tetap harus dipertahankan dan ditambahkan supaya lebih banyak lagi penggunaannya. Bahkan, jumlah pengguna yang banyak juga akan membuat kebutuhan pengguna pun ikut bertambah. Hal ini membuat fitur yang ada pada aplikasi membutuhkan pembaruan supaya kebutuhan pengguna yang bertambah tersebut dapat tercukupi.

Sejalan dengan hal tersebut, Dix (dalam Farouqi, 2018) menjelaskan pula bahwa desain UX pada aplikasi harus selalu dikembangkan dan disesuaikan kebutuhan pengguna dengan cara melakukan evaluasi. Bahkan, Dix menambahkan evaluasi harus selalu dilakukan sebagai bagian integral dari proses perancangan dan harus dilakukan sepanjang siklus hidup desain. Penjelasan Dix tersebut menandakan desain UX harus diperbarui dengan cara mengadakan evaluasi supaya kebutuhan pengguna yang semakin banyak dapat terpenuhi dan pengguna akan semakin setia dengan aplikasi tersebut. Maka dari itu, KBBI sebagai aplikasi luring yang dikembangkan oleh Pusat Pengembangan dan Pelindungan, Badan Pengembangan dan Pembinaan Bahasa juga perlu dilakukan pembaruan desain UX yang disesuaikan dengan hasil evaluasi sesuai kebutuhan pengguna.

Hal ini disebabkan sejak kemunculannya pada tahun 2016, KBBI luring masih belum pernah diperbarui sementara pengguna KBBI ini semakin lama semakin meningkat. Hal tersebut terlihat dari jumlah pengguna yang hingga saat ini sudah lebih dari satu juta pengguna pada Google Play. Pengguna yang banyak tersebut pastinya berhubungan dengan kebutuhan pengguna yang lebih beragam dan bervariasi dari sebelumnya pula. Untuk itu, penelitian tentang kualitas UX KBBI Luring ini perlu dilakukan. Hal ini bertujuan untuk mengetahui sejauh mana pengetahuan dan pengalaman masyarakat saat menggunakan KBBI luring dan mengidentifikasi pemenuhan kebutuhan dari KBBI luring sebagai bahan evaluasi dalam pembaruan desain UX KBBI luring yang lebih efektif, efisien, dan bermanfaat.

Di samping itu, penelitian tentang analisis desain UX pada aplikasi telepon genggam yang sejenis dengan penelitian sudah banyak jumlahnya. Dari banyaknya penelitian tentang analisis desain UX pada aplikasi telepon genggam terdapat empat penelitian yang memiliki tipe penelitian yang mirip, yaitu penelitian Loucky (2010), Ivan (2011), Suzianti, dkk (2017), dan Farouqi, dkk (2018). Keempat penelitian tersebut menjadi tinjauan pustaka pada penelitian ini karena adanya kesamaan topik, objek yang diteliti, atau tujuan penelitian. Loucky (2010) melakukan penelitian tentang kamus elektronik pada mahasiswa Jepang yang belajar bahasa Inggris sebagai bahasa kedua. Pada penelitiannya, Loucky menyajikan survei deskriptif dari mahasiswa Jepang dalam menggunakan kamus elektronik. Dari penelitian tersebut, Loucky memberikan rekomendasi praktis untuk meningkatkan pembelajaran kosakata dan pelatihan strategi sistematis dalam penggunaan elektronik yang efektif. Hal yang membedakan penelitian Loucky dengan penelitian ini adalah penelitian Loucky mengambil kamus elektronik daring dan luring secara bersamaan sebagai objek kajian sementara penelitian ini hanya mengambil kamus elektronik luring sebagai objek kajian.

Di sisi lain, Ivan (2011) membahas tentang analisis kualitas aplikasi luring. Meskipun tidak menggunakan objek data penelitian berupa kamus elektronik, tetapi penelitian Ivan masih memiliki hubungan dengan penelitian ini karena menggunakan aplikasi luring dan menganalisis kualitasnya. Pada penelitian Ivan, karakteristik kualitas didefinisikan dan indikatornya diperhitungkan untuk mengukur kualitasnya. Ada sebelas parameter yang dipertimbangkan dalam analisis aplikasi luring. Parameter tersebut disusun menggunakan

bobot dan melakukan analisis rinci dari sistem bobot. Untuk pengukuran kinerja aplikasi luringnya dilakukan dengan menggunakan indikator agregat berdasarkan pada sistem bobot yang diperoleh. Sementara itu, Suzianti, dkk (2017) melakukan penelitian untuk menganalisis UX pada aplikasi luring. Akan tetapi, aplikasi luring yang dijadikan objek penelitian adalah aplikasi luring pelacak kesehatan. Meskipun demikian, analisis UX yang dilakukan Suzianti, dkk ini memiliki sumbangsih dalam mengembangkan aplikasi supaya aplikasi tersebut sesuai dengan kebutuhan pengguna. Pengetahuan tentang elemen yang dibutuhkan dalam memaksimalkan pengalaman pengguna melalui desain UX pada aplikasi merupakan hal yang dicari pada penelitian Suzianti, dkk yang sama dengan tujuan penelitian ini.

Penelitian terakhir sekaligus penelitian terbaru yang dijadikan tinjauan pustaka adalah penelitian dari Farouqi, dkk (2018). Penelitian Farouqi, dkk membahas tentang evaluasi kegunaan aplikasi ojek daring untuk mengetahui permasalahan dan tingkat kegunaan yang ada pada aplikasi GO-JEK. Cara untuk mendapatkan hasil, Farouqi, dkk mewawancarai pengguna GO-JEK secara langsung dengan memberikan beberapa studi kasus yang akan dilakukan oleh partisipan tersebut. Hasil yang didapatkan pun disajikan secara kuantitatif yang terbagi dalam beberapa kategori, yaitu tingkat kemudahan pengguna, tingkat kecepatan pengguna, tingkat kesalahan pengguna, dan tingkat kepuasan pengguna. Pada akhir penelitian, Farouqi, dkk pun menjabarkan masukan dari partisipan terkait permasalahan yang ada pada aplikasi GO-JEK sebagai bahan evaluasi yang dapat digunakan GO-JEK guna memperbaiki tampilan aplikasinya.

II. KERANGKA TEORI DAN METODE PENELITIAN

Untuk melakukan analisis kualitas UX pada aplikasi KBBI luring, pada penelitian ini menggunakan teori karakteristik kualitas aplikasi luring (Gh. Nocşa dalam Ivan 2011) dan teori analisis pengamatan penelitian UX dari Anderson (2010). Pemakaian kedua teori tersebut disebabkan satu teori saja tidak dapat mencakup ruang lingkup pada penelitian ini. Sesuai dengan judul penelitian untuk analisis kualitas UX pada aplikasi luring, kedua teori digunakan supaya tujuan dari penelitian ini dapat terpenuhi. Dengan penggabungan kedua teori tersebut diharapkan pula penelitian ini memiliki indikator yang jelas dalam tahap analisis. Selain itu, kedua teori yang dijadikan kerangka teori juga menjadi acuan dalam menentukan metode penelitian. Mulai dari partisipan penelitian, daftar pertanyaan, metode yang digunakan, hingga teknik penelitiannya.

Dalam teori karakteristik kualitas aplikasi luring (Gh. Nocşa dalam Ivan 2011), diketahui bahwa ada karakteristik internal dan eksternal. Maksud dari karakteristik internal adalah karakteristik untuk memeriksa proses pengembangan perangkat lunak, sedangkan maksud dari karakteristik eksternal adalah karakteristik untuk memeriksa perangkat lunak pengguna. Karakteristik tersebut dijelaskan dalam lima kategori, yaitu kategori akurasi aplikasi, *interface* aplikasi yang ramah, kelancaran pengguna, portabilitas, dan keamanan. Dari kelima kategori tersebut, kelancaran pengguna mewakili derajat aplikasi dalam hal pengguna memberikan toleransi kesalahan yang ada maupun ketersediaan kebutuhan pada aplikasi. Sementara itu, akurasi merupakan kategori karakteristik kualitas yang paling penting. Berikut penjelasan tentang kategori karakteristik kualitas aplikasi luring.

1	Akurasi aplikasi:	Karakteristik ini melihat kemampuan aplikasi untuk menghasilkan keluaran yang benar untuk data dasar yang dimasukkan dengan benar. Dengan kata lain, karakteristik ini mengecek kesalahan aplikasi jika ditemukan. Komponen yang dilihat dikaitkan dengan komponen individu dan komponen kontekstual dalam konteks tertentu.
2	Interface aplikasi:	Karakteristik ini penting untuk aplikasi luring pada telepon genggam. Hal ini disebabkan aplikasi pada telepon genggam harus lebih dekat dengan <i>interface</i> pengguna supaya pengguna dapat menggunakan aplikasi seintensif mungkin. Aplikasi yang sulit digunakan akan membutuhkan lebih banyak waktu untuk penggunaannya. Akibatnya, pengguna akan berpindah ke aplikasi lain yang lebih mudah digunakan mengingat aplikasi yang ada memiliki kualitas rendah.
3	Kelancaran pengguna:	Karakteristik ini mendukung karakteristik yang disajikan sebelumnya. Elemen-elemen yang ada pada aplikasi harusnya akrab dan mudah dikenal dari aplikasi luring lainnya sehingga pengguna tidak perlu berusaha untuk belajar menggunakan aplikasinya. Jika dalam aplikasi disajikan elemen yang sudah dikenal oleh pengguna, karakteristik lancarnya lebih besar. Dengan demikian, aplikasi dapat digolongkan lebih dekat kepada pengguna.
4	Portabilitas:	Karakteristik ini mengukur aplikasi luring yang berjalan pada telepon genggam. Pada dasarnya, karakteristik ini memastikan kehadiran elemen dan keberlanjutan proses kerja dari aplikasi.
5	Keamanan:	Karakteristik ini melihat kemampuan aplikasi dalam menghemat beberapa elemen ke parameter yang ditentukan oleh pengembang dan pengguna.

Setelah karakteristik kualitas aplikasi luring diketahui, berikutnya aspek analisis pengamatan penelitian UX dari Anderson (2010) akan dijelaskan untuk memperkaya indikator pada penelitian ini. Dalam melakukan analisis pengamatan penelitian UX, Anderson menjelaskan bahwa ada tiga aspek yang perlu diketahui. Ketiga aspek tersebut adalah (1) menemukan persona, (2) mengetahui pengalaman pengguna, dan (3) menggali prioritas pengguna. Dengan menggunakan ketiga aspek tersebut, didapat pola yang digunakan untuk mengidentifikasi dan memperdalam pemahaman tentang target pengguna. Hal ini membantu mengetahui pola umum dalam melihat interaksi, masalah, dan peluang yang ada pada desain UX aplikasinya.

Hal pertama yang dilihat adalah (1) menemukan persona. Dalam hal ini, peneliti pengguna memanfaatkan pengalamannya dalam berbagai teknik untuk mengidentifikasi tren dan kesamaan yang dapat digunakan untuk menyatukan atribut, perilaku, dan pengamatan dari satu set pengguna menjadi satu perwakilan persona. Kemudian, peneliti merancang beberapa persona untuk mewakili keragaman penuh para pengguna dalam bentuk yang telah tersaring dan terfokus. Hal ini dilakukan guna membantu mengembangkan empati dan intuisi yang akurat untuk kebutuhan pengguna dalam perencanaan UX. Lebih lanjut, persona yang dimaksud harus kaya dengan detail yang membantu orang menempatkan diri di dalam pengguna.

Setelah persona ditemukan, langkah berikutnya adalah (2) mengetahui pengalaman pengguna. Dalam hal ini, informasi perlu disampaikan dengan cara menghubungkan satu hal dengan hal lain dan mengungkapkan implikasi temuan untuk memberi tahu cerita lebih detail. Hal tersebut bertujuan untuk mengubah informasi menjadi sesuatu dengan kualitas emotif yang menghasilkan respon yang baik dari para pengguna. Pengalaman pengguna yang dipilih tentunya pengalaman yang memiliki keterkaitan dengan produk sehingga tejalin benang merah. Beberapa hal yang perlu diketahui adalah bagaimana pengguna berusaha mencapai tujuannya, masalah dan kesulitan yang dihadapi pengguna, harapan yang diinginkan pengguna, dan momen dalam kehidupan seorang pengguna dalam menggunakan produk.

Selanjutnya, hal yang perlu dilakukan dan pertimbangkan adalah (3) menggali prioritas pengguna. Hal tersebut bertujuan untuk melihat perspektif yang kuat tentang produk yang seharusnya diunggulkan dengan cara mempelajari perspektif pengguna tentang masalah, peluang, dan kemungkinan tujuan produk. Dalam hal ini, beberapa pendapat tentang prioritas akan terjadi secara langsung oleh pengguna. Di samping itu, pendapat lain juga dapat disimpulkan dengan melihat pola dan tren dalam perilaku dan umpan balik dari pengguna itu sendiri. Hal ini akan membantu menentukan seberapa pentingnya ide yang ada dalam upaya membangun minat dan pemahaman tentang produk.

Aspek analisis pengamatan penelitian UX tersebut juga dijadikan acuan dalam membuat daftar pertanyaan selain menjadi indikator dalam melakukan analisis. Beberapa aspek itu diolah dan disesuaikan dengan kondisi aplikasi KBBI luring. Dengan begitu, penulis membuat daftar pertanyaan menjadi empat kategori, yaitu kategori pemahaman produk, kategori pemahaman tahapan, kategori pemahaman tampilan, dan kategori berbagi pengalaman dalam menggunakan aplikasi KBBI luring. Daftar pertanyaan pada kategori pemahaman produk, tahapan, dan tampilan menggunakan pilihan jawaban dengan disertai isian untuk jawaban tertentu. Sementara itu, daftar pertanyaan pada kategori berbagi

pengalaman dalam menggunakan KBBI luring menggunakan jawaban isian supaya dapat menggali lebih detail dari pengguna.

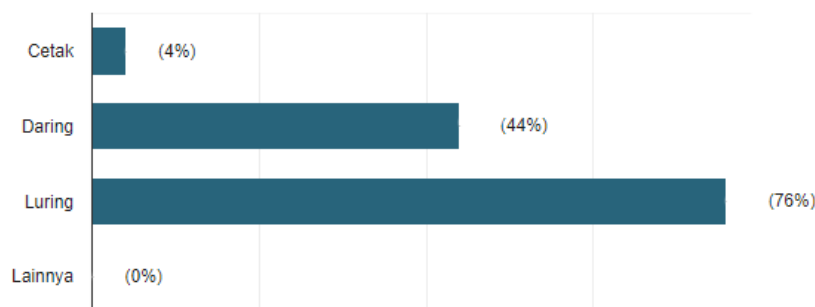
Daftar pertanyaan yang sudah dibuat kemudian diberikan kepada partisipan yang merupakan pengguna KBBI. Maka dari itu, pada penelitian ini partisipan yang dipilih adalah mahasiswa akhir dari program studi Sastra Indonesia di Universitas Indonesia dan Universitas Sebelas Maret yang sedang mengerjakan skripsi. Hal ini disebabkan mahasiswa akhir yang sedang mengerjakan skripsi pastinya bergelut dengan KBBI sehingga partisipan dapat mengetahui hal yang perlu dikembangkan sebagai bahan evaluasi tampilan UX KBBI luring. Partisipan yang diambil dibatasi hanya 25 mahasiswa saja. Hal ini disebabkan penelitian ini dapat dikatakan merupakan penelitian awal untuk nantinya dapat digunakan dan dimodifikasi oleh Subbidang Kosakata lebih dalam lagi.

Di sisi lain, metode yang dipakai pada penelitian ini adalah metode kuantitatif dan kualitatif. Metode kuantitatif digunakan untuk mendapatkan persentase dari daftar pertanyaan yang menggunakan pilihan ganda. Hasil data kuantitatif tersebut kemudian dianalisis dengan menggunakan metode kualitatif ditambah dengan masukan atau saran dari partisipan yang merupakan pengguna KBBI luring. Dengan begitu, hasil analisis ini lebih komprehensif dan objektif sehingga dapat dijadikan pijakan awal untuk mengembangkan penelitian sejenis demi kemajuan KBBI luring bagi kebermanfaatan masyarakat Indonesia.

III. ANALISIS KUALITAS UX KBBI LURING PADA APLIKASI TELEPON GENGAM

Seperti yang telah dijelaskan pada bagian sebelumnya bahwa daftar pertanyaan dibuat berdasarkan acuan dari Anderson (2010) dengan modifikasi yang disesuaikan dengan aplikasi KBBI luring. Daftar pertanyaan pun dibuat dengan beberapa kategori, yaitu kategori pemahaman produk, kategori pemahaman tahapan, kategori pemahaman tampilan, dan kategori berbagi pengalaman dalam menggunakan aplikasi KBBI luring. Berdasarkan daftar pertanyaan tersebut, pada tahap analisis penelitian ini juga dibahas berdasarkan kategori yang ada. Hal ini dimaksudkan untuk dapat lebih memahami temuan pada setiap kategorinya. Semua temuan data pada penelitian ini pun tidak semua ditampilkan mengingat terbatasnya jumlah halaman pada makalah yang harus dibuat. Untuk itu, pada tahap analisis ini pembahasan temuan data pada penelitian dilakukan dengan menampilkan temuan data yang menarik dan mewakili setiap kategorinya saja. Berikut penjelasan dari temuan pada kategori pemahaman produk.

Menurut kamu lebih nyaman menggunakan KBBI versi apa?

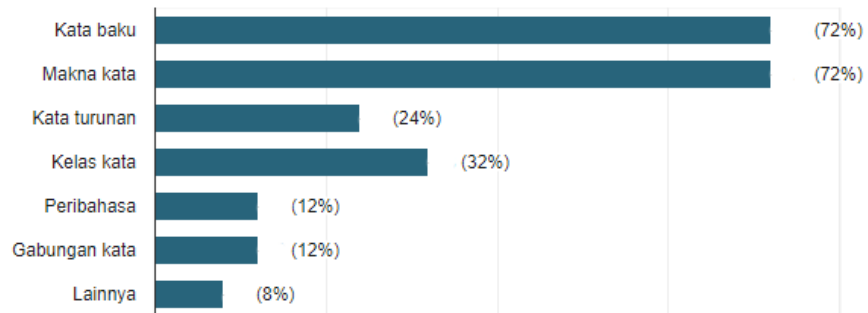


Gambar 1. Perolehan KBBI yang paling banyak digunakan

Dari gambar di atas dapat diketahui bahwa KBBI versi luring menjadi paling nyaman digunakan oleh partisipan. Hal ini juga sejalan dengan pertanyaan lainnya yang menyebutkan KBBI yang paling sering digunakan adalah KBBI versi luring. Hal ini disebabkan KBBI versi luring tidak membutuhkan jaringan internet sehingga pengguna yang memiliki kuota internet terbatas tidak kesulitan mengakses KBBI. Selain itu, KBBI versi luring juga mudah digunakan karena tidak membutuhkan daftar anggota seperti pada KBBI versi daring. Terlebih lagi, pada KBBI versi luring lebih memuat banyak hasil pencarian. Hal tersebut berbeda dengan KBBI versi daring yang harus menjadi anggota terlebih dahulu untuk mendapatkan hasil pencarian yang lengkap dengan kata turunan atau peribahasanya jika tersedia. Perolehan KBBI versi luring yang menjadi KBBI ternyaman juga dapat dijadikan acuan pengembangan KBBI versi luring supaya dapat lebih memenuhi kebutuhan masyarakat sehingga penggunaanya terus bertambah.

Sementara itu, pada kategori pemahaman tahapan daftar pertanyaan yang diajukan lebih pada proses partisipan mencari apa yang diinginkan pada KBBI luring. Pada kategori ini daftar pertanyaan sudah mengarah pada KBBI luring. Hal tersebut berbeda pada kategori sebelumnya tentang pemahaman produk yang daftar pertanyaannya masih secara umum. Hal tersebut dimaksudkan untuk mengetahui sejauh mana partisipan mengenal KBBI. Berikut ditampilkan perolehan persentase dari kategori pemahaman tahapan.

Saat membuka KBBI luring, apa yang sering kamu cari?



Gambar 2. Perolehan pencarian terbanyak dalam KBBI luring

Dari salah satu pertanyaan dari kategori pemahaman tahapan di atas dapat diperoleh informasi bahwa partisipan paling banyak menggunakan KBBI untuk mencari kata baku dan makna kata. Kedua hal tersebut memiliki persentase yang sama banyaknya. Kemudian, pencarian partisipan yang biasa dicari setelah kata baku dan makna kata adalah kelas kata. Hal yang perlu diketahui pada perolehan di atas adalah partisipan dapat memilih lebih dari satu jawaban sehingga persentase tidak bulat 100% jika dijumlahkan. Hal ini disebabkan untuk melihat seberapa banyak pencarian yang dilakukan partisipan di dalam KBBI luring. Temuan di atas juga dapat menandakan bahwa selama beberapa aspek seperti yang tersebut di atas sudah ada pada setiap lema pada KBBI luring, pengembang dapat fokus pada penambahan lema dan konsistensi isi konten setiap lemanya supaya seragam.

Selain itu, pada kategori tahapan juga terdapat temuan yang menarik untuk dilihat lebih jauh. Hal ini disebabkan ada pertanyaan yang mengarah kepada kebermanfaatan fitur yang ada pada KBBI luring. Hal tersebut dijadikan pertanyaan untuk mengetahui sejauh mana partisipan menggunakan KBBI luring selain dari mencari kata baku, makna kata. Terlebih lagi, pada KBBI luring terdapat fitur pengelompokan. Misalnya, pencarian berdasarkan kelas kata, ragam, bahasa, bidang, dan jenis. Berikut perolehan tentang kebermanfaatan fitur pengelompokan tersebut pada KBBI luring.

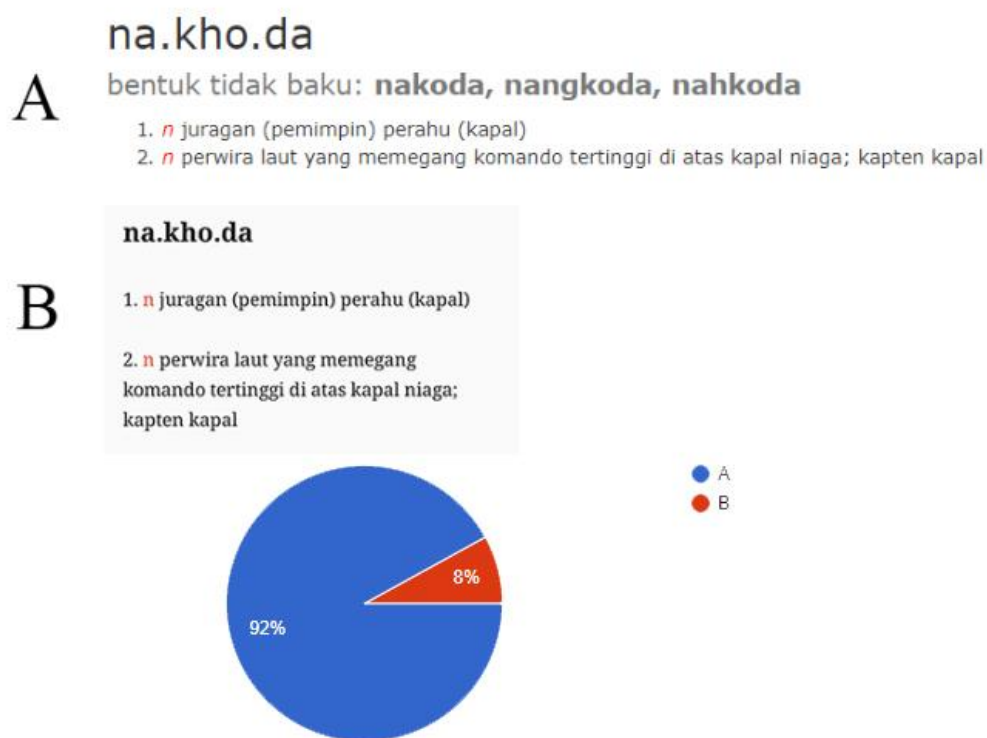
Apakah kamu pernah menggunakan kolom yang dibulatkan saat melakukan pencarian kata?



Gambar 3. Perolehan kebermanfaatan fitur pengelompokan pada KBBI luring

Jika dilihat pada perolehan di atas, dapat diketahui bahwa sebagian besar partisipan tidak pernah melakukan pencarian pada fitur pengelompokan yang ada pada KBBI luring. Hal tersebut dijelaskan pula oleh partisipan karena pengelompokan tersebut jarang dicari. Jika ada partisipan yang mencarinya, fitur tersebut hanya dapat dilakukan secara manual. Hal ini disebabkan pada fitur pengelompokan tersebut tidak ada kolom pencarian khusus pada kelompok tersebut sehingga hanya ada kolom pencarian umum dan jika pencarian umum tersebut digunakan maka pengelompokan tersebut tidak ada lagi atau dapat dikatakan kembali seperti pencarian pada tampilan utama KBBI luring. Hal ini yang sekiranya butuh ditindaklanjuti untuk menambahkan kolom pencarian khusus pada setiap pengelompokan yang telah dibuat atau menghilangkan fitur tersebut karena pengguna juga jarang menggunakannya.

Berikutnya, pada kategori pemahaman tampilan daftar pertanyaan lebih mengarah pada tampilan isi konten yang ada pada KBBI luring. Tampilan isi konten KBBI luring pun diperbandingkan dengan tampilan KBBI daring. Hal ini disebabkan tampilan isi konten KBBI luring dan daring berbeda sehingga pada bagian ini mau melihat keberterimaan isi konten mana yang lebih dimengerti partisipan sebagai pengguna KBBI. Berikut perolehn tentang tampilan isi konten KBBI luring dan daring.



Gambar 4. Perolehan perbandingan tampilan KBBI luring dan daring

Pada gambar di atas, perlu dijelaskan terlebih dahulu bahwa tampilan (A) merupakan tampilan pada KBBI daring dan tampilan (B) merupakan tampilan pada KBBI luring. Dari tampilan keduanya, partisipan lebih menyukai tampilan (A) untuk diaplikasikan pada KBBI luring. Hal ini disebabkan tampilan (A) dirasa lebih jelas dan memuat informasi yang cukup

lengkap. Terlebih lagi, jika kata yang baku merupakan kata yang kurang akrab oleh masyarakat sehingga perlu adanya penambahan bentuk tidak baku supaya masyarakat sebagai pengguna memahami. Hal ini disebabkan pada KBBI luring tidak ada petunjuk penggunaan dan versi luring harus dibuat sedetail dan semudah mungkin untuk dapat dimengerti masyarakat tanpa harus membaca petunjuk penggunaannya.

Selanjutnya, kategori terakhir yang dijadikan daftar pertanyaan adalah kategori berbagi pengalaman dalam menggunakan KBBI luring. Dalam hal ini, partisipan diajak untuk menuliskan isian singkat tentang hal yang dirasakan saat menggunakan KBBI luring. Dari beberapa pertanyaan yang ada, hal yang membingungkan menarik untuk dibahas. Hal ini disebabkan hal yang membingungkan ini jika tidak ditindaklanjuti akan berpengaruh terhadap jumlah pengguna karena pengguna yang merasa bingung akan mencari aplikasi lain yang lebih dimengerti. Berikut

Apakah ada bagian pada KBBI versi luring yang membuatmu bingung?

Tidak (5)

Belum ada (2)

Tidak hapal singkatan yang digunakan sehingga kesulitan untuk menangkap maksudnya, misal singkatan kelas kata (p, par)

Kelas kata yang berupa singkatan. Kadang saya kurang mengerti arti singkatan tersebut.

pelafalan

Fungsi yang ada di pojok kiri atas dan sering sekali tertekan secara tidak sengaja

Halaman awal. Lebih baik langsung ada kolom pencarian seperti pada kbki daring

Bagian kotak kotak yg ada kelas kata, ragam bahasa, dll

Sejauh ini tidak

tidak

Ku tidak suka pakai KBBI versi luring Kemdikbud karena ketika mencari satu kata, "siap" misalnya, tidak bisa langsung tahu kata turunannya

Gambar 5. Perolehan kebingungan partisipan menggunakan KBBI luring

Pada gambar di atas dapat terlihat dengan jelas ada dua hal yang membuat partisipan bingung dalam penggunaan KBBI luring. Hal yang membuat bingung adalah singkatan kelas kata pada setiap lema atau kata dan kolom pengelompokan yang terdapat pada tampilan awal KBBI luring. Perihal singkatan kelas kata yang membuat bingung dapat diatasi dengan tidak menyingkat kelas kata tersebut sehingga partisipan tidak perlu menebak singkatan kelas kata tersebut atau pada kolom pengelompokan khususnya pengelompokan kelas kata diberi keterangan singkatan dari kelas kata. Misalnya, nomina (n), verba (v), dan sebagainya. Untuk itu, diperlukan kajian lebih dalam terkait mana yang lebih efektif dan efisien sehingga pengguna tidak akan kebingungan dalam menggunakan KBBI luring.

IV. SIMPULAN

Dari pembahasan di atas dapat disimpulkan bahwa tampilan UX KBBI luring masih membutuhkan perbaikan untuk diperbarui. Hal ini disebabkan partisipan lebih banyak menggunakan KBBI luring dibandingkan versi lainnya sehingga KBBI luring harus mendapatkan perhatian lebih dan dikembangkan lebih mendalam supaya pengguna bertambah banyak dan kebermanfaatannya dirasakan masyarakat. Selain itu, partisipan sebagai pengguna KBBI luring masih merasa bingung dengan tampilan dan fitur yang ada. Sejalan dengan hal tersebut, tampilan KBBI luring sebaiknya disamakan terlebih dahulu dengan KBBI daring karena partisipan lebih menyukai tampilan isi konten dari KBBI daring dibandingkan KBBI versi luring karena lebih jelas. Selain itu, fitur yang ada pada KBBI luring juga perlu dipertimbangkan lagi keberadaannya dan jika ingin dipertahankan fitur tersebut harus memiliki kreasi supaya kebutuhan masyarakat dapat terpenuhi.

Daftar Pustaka

- Atkins, B.T. dan Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Anderson, Jonathan, etc. 2010. *Effective UI*. California: O'Reilly Media.
- Loucky, John Paul. 2010. "Comparing Electronic Dictionary Functions and Use". *CALICO Journal*. Vol. 28, No. 1, hlm. 156-174.
- Ivan, Ion and Alin Zamfiroiu. 2011. "Quality Analysis of Mobile Applications. *Informatica Economica* Vol. 15 No. 3, Bucharest: Rumania.
- F. Pucillo and G.Cascini. 2013. "Framework for user experience, needs, and affordance," *Design Studies*. hlm. 160-17
- L'Homme, Marie-Claude dan Monique C.Cormier. 2014. "Dictionaries and the Digital Revolution: A Focus on Users and Lexical Databases". *International Journal of Lexicography*, Vol. 27 No. 4, hlm. 331–340.
- Levy, Jaime. 2015. *UX Strategy: How to devise innovate digital products that people want*. California: O'Reilly Media, Inc.,
- Levy, Mike dan Caroline Steel. 2015. "Language learner perspectives on the functionality and use of electronic language dictionaries". *The Journal of the European Association for Computer Assisted Language Learning (ReCALL)*. Volume 27, Issue 2, hlm. 177-196.
- Suzianti, Amalia, Rizky Puti Minanga, dan Felisa Fitriani. 2017. "Analysis of User Experience (UX) on Health-Tracker Mobile Apps". *International Journal of Computer Theory and Engineering* Vol. 9 No.4.
- Farouqi, dkk. 2018. "Evaluasi Usability pada Aplikasi GO-JEK dengan Menggunakan Metode Pengujian Usability". *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* Vol. 2, No.9, hlm. 3110-3117.

**PENERAPAN DEKOMPOSISI LEKSIKAL DALAM PENYUSUNAN
DEFINISI: STUDI KASUS KBBI DARING**
*(THE USE OF LEXICAL DECOMPOSITION IN THE PREPARATION OF
DEFINITIONS: CASE STUDIES OF KBBI DARING)*

Dessy Irawan, Dien Rovita, Totok Suhardijanto
Universitas Indonesia

Abstrak

Pendefinisian menjadi salah satu aspek utama yang perlu diperhatikan dalam proses penyusunan kamus. Pemerian definisi untuk sebuah lema atau sublema dapat menerapkan dekomposisi leksikal. Dekomposisi leksikal adalah istilah yang digunakan Cruse (2000: 235) dalam menganalisis makna. Dalam KBBI Daring, definisi lema dan sublema yang berhiponim memperlihatkan pola yang beragam. Dengan demikian, persoalan mengenai definisi menjadi masalah yang menarik untuk diteliti. Penelitian kualitatif ini bertujuan untuk melihat penerapan dekomposisi leksikal dalam menyusun definisi lema atau sublema. Data penelitian berupa 961 lema dan sublema bidang komputer dalam KBBI Daring. Hasil penelitian menunjukkan bahwa dekomposisi leksikal dapat diterapkan dalam proses penyusunan definisi lema atau sublema. Dengan menerapkan dekomposisi leksikal, genus dalam definisi yang dihasilkan untuk lema dan sublema yang berhiponim akan menjadi beragam. Selain itu, penerapan dekomposisi leksikal juga dapat menghindari definisi dengan sinonim yang mengandung makna melingkar.

Kata Kunci: definisi, dekomposisi leksikal, istilah komputer, KBBI Daring

Abstract

Definition becomes one of the main aspects that need to be considered in the process of compiling the dictionary. In the process of constructing a definition for an entries and subentries can apply lexical decomposition. Lexical decomposition is the term used by Cruse (2000: 235) in analyzing meanings. In KBBI Daring, the pattern of definitions of entries and subentries that are hyponym are very diverse. Thus, the issue of definition becomes an interesting issue to be studied. This qualitative research aims to see the application of lexical decomposition in constructing the definition of an entries and subentries. The research data are 961 entries and subentries of computer field in KBBI Daring. The results show that lexical decomposition can be applied in the process of composing an entries definition or subentries. By applying lexical decomposition, the genus in the resulting definition for entries and subentries that are hyponym will be the same. In addition, the application of lexical decomposition can also avoid definitions with synonyms containing circular meanings.

Keywords: definitions, lexical decomposition, computer terms, KBBI Daring

I. PENDAHULUAN

1.1 Latar Belakang

Kamus adalah buku yang memuat informasi mengenai makna suatu kata. Harimurti Kridalaksana (1985: 12) mendefinisikan kamus sebagai buku referensi yang memuat daftar kata atau gabungan kata dengan keterangan mengenai berbagai segi maknanya dan penggunaannya dalam bahasa; biasanya disusun menurut abjad (dalam tradisi Yunani-Romawi menurut urutan abjad Yunani-Romawi), kemudian menurut abjad bahasa bersangkutan; dalam tradisi Arab menurut jumlah urutan konsonan. Kamus menurut Pierre Labrousse (dalam Chaer, 2008: 180) didefinisikan sebagai buku yang berisi kumpulan kata-kata sebuah bahasa yang disusun secara alfabetis diikuti dengan definisi atau terjemahannya dalam bahasa lain. Sejalan dengan Kridalaksana dan Labrousse, Keraf (2001: 44) juga mendefinisikan kamus sebagai buku referensi yang memuat daftar kosakata yang terdapat dalam sebuah bahasa, yang disusun secara alfabetis disertai keterangan bagaimana menggunakan kata itu.

Atkins dan Rundell (2008: 24—25) mengklasifikasikan kamus berdasarkan delapan hal, di antaranya berdasarkan jumlah bahasa dalam kamus, berdasarkan cakupan isinya, berdasarkan ukurannya, dan berdasarkan medianya. Berdasarkan jumlah bahasa dalam kamus, terdapat tiga tipe kamus, yaitu kamus monolingual, kamus bilingual, dan kamus multilingual. Berdasarkan cakupan isinya, kamus dapat dibedakan atas kamus umum, kamus ensiklopedis, kamus istilah khusus, dan kamus yang berisi penggunaan bahasa secara khusus (seperti kamus kolokasi, peribahasa, dan idiom). Berdasarkan ukurannya, kamus dapat dibedakan atas kamus standar, kamus ringkas, dan kamus saku. Berdasarkan medianya, kamus dapat dibedakan atas kamus cetak, kamus elektronik, dan kamus yang berbasis situs.

Di Indonesia terdapat kamus standar yang diterbitkan oleh Badan Pengembangan dan Pembinaan Bahasa, yaitu *Kamus Besar Bahasa Indonesia* (KBBI). Sampai tahun 2018 ini, Badan Pengembangan dan Pembinaan Bahasa sudah menerbitkan lima edisi KBBI. Berdasarkan klasifikasi tipe kamus menurut Atkins dan Rundell (2008), KBBI dapat digolongkan sebagai kamus monolingual umum yang berukuran standar. KBBI Edisi Kelima diterbitkan dalam tiga versi sekaligus, yaitu versi cetak, daring, dan luring. Oleh karena itu, berbeda dengan edisi sebelumnya, KBBI Edisi Kelima tidak hanya dapat digolongkan sebagai kamus cetak, melainkan juga sebagai kamus elektronik dan kamus yang berbasis situs.

Ketiga versi KBBI Edisi Kelima dimungkinkan terbit secara bersamaan berkat digunakannya aplikasi penyusunan kamus yang mempermudah dan mempercepat proses penyusunan (Tim Penyusun, 2016: xxxv). *Kamus Besar Bahasa Indonesia Daring* (KBBI Daring) dapat diakses melalui laman kbbi.kemendikbud.go.id. *Kamus Besar Bahasa Indonesia Luring* (KBBI Luring) dapat diunduh di aplikasi Playstore atau Appstore di ponsel cerdas. Tujuan KBBI Edisi Kelima diterbitkan dalam tiga versi sekaligus adalah untuk memperluas ketersebaran penggunaan dan memenuhi prinsip keadilan bagi semua masyarakat pengguna (Tim Penyusun, 2016: xxxv).

Bagian kamus yang penting untuk diperhatikan dalam proses penyusunan kamus adalah bagian definisi. Keraf (2001: 56) menyebutkan bahwa inti dari sebuah kamus adalah memberikan batas pengertian atau definisi sebuah kata (Keraf, 2001: 56). Sejalan dengan hal tersebut, Sumarsono (2002: 66) juga menyebutkan bahwa salah satu unsur kamus yang sangat

penting ialah definisi bagi lema yang memerlukan pemerian atau penjelasan. Dalam *Pedoman Penyuntingan Kamus Bahasa Indonesia* (Pusat Bahasa, 1987: 39), juga disebutkan bahwa tahap penyusunan definisi di dalam proses penyusunan kamus adalah tahap yang paling memerlukan ketekunan, ketajaman analisis, kecermatan, dan kelengkapan bahan acuan. Tahap-tahap tersebut yang dapat dikatakan sebagai tahap penentu keberhasilan tingkat mutu kamus yang akan dihasilkan. Dengan demikian, berdasarkan uraian tersebut, penulis menganggap bahwa mutu kamus dapat ditentukan oleh ketepatan pemerian definisinya.

Berbeda dengan versi cetak, dalam KBBI Daring dan Luring pengguna kamus dapat mencari lema dan sublema berdasarkan beberapa kategori, yaitu kategori bahasa, kategori bidang, kategori kelas kata, kategori ragam, dan kategori jenis. Dalam kategori bidang, pengguna kamus dapat menemukan kelompok 83 ragam bidang. Salah satu bidang yang terdapat dalam KBBI Daring dan Luring adalah bidang komputer. Dalam KBBI Daring dan Luring, lema atau sublema yang tergolong ke dalam bidang komputer tersebut diberi label *Komp* yang merupakan bentuk abreviasi dari kata komputer. Berikut adalah contoh pendefinisian lema dan sublema berlabel *Komp* dalam KBBI Daring (2018) yang berada dalam satu kategori semantis.

<p>menu <i>n Komp</i> daftar perintah atau pilihan yang terdapat pada layar monitor komputer</p> <p>menu batang <i>Komp</i> batang persegi panjang yang ditampilkan di layar jendela program aplikasi, sering di bagian atas dan menunya dapat dipilih oleh pengguna</p> <p>menu hierarkis <i>Komp</i> menu yang memiliki satu atau beberapa submenu di bawahnya</p> <p>menu layar penuh <i>Komp</i> kemampuan pemrograman yang memungkinkan menu ditampilkan dalam setiap baris, kolom, atau piksel lokasi pada layar secara penuh</p>

Sublema *menu batang*, *menu hierarkis*, dan *menu layar penuh* merupakan hiponim dari lema *menu*. Akan tetapi, ketiga sublema tersebut memiliki pola definisi yang beragam. Tidak ada keseragaman kata atau frasa pertama yang mengawali tiga definisi sublema yang berhiponim tersebut. Dengan demikian, persoalan mengenai definisi menjadi masalah yang menarik untuk diteliti.

Berdasarkan hal-hal yang sudah dipaparkan di atas, persoalan mengenai definisi menjadi masalah yang menarik untuk diteliti. Lema dan sublema dalam bidang komputer yang memiliki kemiripan leksikal dan yang berhiponim dengan pola pendefinisian yang masih beragam perlu dibenahi, salah satunya dengan menggunakan analisis dekomposisi leksikal dalam menyusun definisi. Berdasarkan hal-hal tersebut, dapat dirumuskan masalah, yaitu bagaimana penerapan dekomposisi leksikal dalam menyusun lema dan sublema bidang komputer dalam KBBI Daring (2018). Merujuk pada masalah yang telah dirumuskan, penelitian ini bertujuan untuk melihat penerapan dekomposisi leksikal dalam menyusun definisi lema atau sublema, khususnya lema atau sublema yang mirip secara leksikal dan yang berhiponim.

Penelitian mengenai pendefinisian dalam KBBI sudah pernah dilakukan. Koen (1996) meninjau pendefinisian terhadap KBBI Edisi Pertama (1988) dengan sistem acu silang.

Dalam penelitian tersebut, Koen (1996) menegaskan bahwa definisi dalam KBBI sebaiknya memberikan informasi faktual yang benar, tidak mengandung kesalahan penulisan, menerangkan lema yang didefinisikan, sederhana dan jelas, mengacu pada etimologi, menggunakan kata-kata yang sudah dicek dengan sistem acu silang, dapat mencantumkan antonim, memberikan ketepatan sinonim, menunjukkan segi pokok, tidak menggunakan pleonasme, memperhatikan lema yang berhomofon atau berhomonim, mengikuti prinsip-prinsip dasar membuat definisi. Selain itu, Koen (1996) juga menegaskan bahwa definisi lema yang berada dalam satu sistem ilmu sebaiknya diseragamkan. Dalam definisi yang menggunakan sinonim, Koen (1996) menyarankan supaya tidak ada kesejajaran yang sulit dipahami dan harus bersesuaian.

Penelitian lain yang berkaitan dengan pendefinisian dalam KBBI dilakukan oleh Sumarsono (2002). Sumarsono (2002) mengupas persoalan klasifikasi dan definisi dalam KBBI Edisi Ketiga (2001) dengan menggunakan analisis komponen. Tulisan tersebut merupakan sebuah tulisan pendahuluan yang ingin mengajak para peneliti bahasa dan leksikograf untuk menelaah berbagai persoalan yang terdapat dalam KBBI. Oleh karena sifatnya yang merupakan penelitian pendahuluan, data kajian Sumarsono (2002) tidak mencakup semua lema alat transportasi yang terdapat di dalam KBBI. Penelitian Sumarsono (2002) diawali dengan mendaftarkan lema yang dijadikan data, kemudian melihat definisi dari berbagai lema, diakhiri dengan rekomendasi yang merupakan salah satu alternatif penyelesaian dalam membuat definisi yang logis dan objektif. Hasil penelitian Sumarsono (2002) tersebut menekankan perihal pentingnya persoalan definisi dalam kamus.

Penelitian yang berfokus pada pendefinisian kamus berikutnya dilakukan oleh Muis (2005). Dalam penelitian tersebut, data yang digunakan adalah lema alat musik dalam KBBI Edisi Ketiga (2001). Sama halnya dengan Sumarsono (2002), Muis juga mengupas data dengan menggunakan analisis komponen. Hasil penelitian Muis (2005) menunjukkan bahwa dalam KBBI (2001) masih banyak sekali definisi lema alat musik yang perlu dibenahi, masih terdapat definisi lema alat musik yang berputar-putar, dan pendefinisian lema dan sublema alat musik perlu menggunakan pendekatan analisis komponen makna.

Berdasarkan penelitian yang sudah dilakukan sebelumnya, penulis melihat penelitian mengenai strategi pendefinisian lema dan sublema bidang komputer dalam KBBI belum dilakukan. Oleh karena itu, dalam penelitian ini penulis menguraikan penerapan dekomposisi leksikal sebagai salah satu strategi pendefinisian lema dan sublema bidang komputer dalam KBBI Daring (2018).

II. KERANGKA TEORI DAN METODE PENELITIAN

2.1 Kerangka Teori

2.1.1 Mikrostruktur Kamus

Mikrostruktur kamus berkaitan dengan susunan informasi dalam entri kamus, termasuk artikel dalam ensiklopedia (Kridalaksana, 2003: xiii—xiv). Mikrostruktur kamus, terdiri atas bagian inti dan pelengkap. Bagian inti memuat lema atau kata kepala dan definisi, sedangkan bagian pelengkap memuat kelas kata, sublema, contoh, derivasi dan infleksi, lafal, variasi historis sosial dan geografis, gaya bahasa, konteks penggunaan bahasa, sinonim dan antonim, serta etimologi.

Salah satu bagian penting dalam mikrostruktur kamus adalah definisi. Definisi adalah karakterisasi arti dari leksem; definisi itu bukanlah penjelasan yang menyeluruh dari acuan yang mungkin (Zgusta, 1971: 252). Terdapat dua konsep umum yang merupakan bagian dari definisi, yaitu *definiendum* dan *definiens* (Svensen, 2009: 218). Dalam kamus, lema mewakili konsep yang akan didefinisikan, sedangkan definisinya adalah deskripsi verbal dari konsep tersebut. Dalam konteks ini, konsep (diwakili oleh lema) disebut *definiendum*. Istilah *definiendum* merupakan bahasa Latin, yang berarti ‘yang harus didefinisikan’. Definisi itu sendiri disebut *definiens* yang juga merupakan bahasa Latin yang berarti ‘yang mendefinisikan’.

2.1.2 Dekomposisi Leksikal

Dekomposisi leksikal adalah istilah yang digunakan Cruse (2000: 235) dalam menganalisis makna. Cruse (2000: 235—238) melihat beberapa hal yang menyebabkan dibutuhkan dekomposisi leksikal, yaitu kata-kata yang memiliki kesamaan parsial, kata-kata yang memiliki korelasi, diskontinuitas, dan paralel kompleks. Pertama, kata-kata yang memiliki kesamaan parsial. Satu kata dengan kata yang lainnya memiliki dua kemungkinan, yaitu memiliki kesamaan makna dan mungkin juga memiliki perbedaan makna (2000: 236). Contoh kesamaan parsial dalam bahasa Inggris, yaitu kata *mare* (kuda betina) dan *stallion* (kuda jantan). Persamaan kedua kata tersebut adalah keduanya mengandung komponen [KUDA]. Akan tetapi, keduanya juga memiliki perbedaan. Jika *mare* mengandung komponen [BETINA], *stallion* tidak memiliki komponen tersebut. *Stallion* mengandung komponen [JANTAN], sedangkan *mare* tidak.

Kedua, kata yang memiliki kesamaan parsial merupakan bukti bahwa adanya korelasi antara satu kata dengan kata lainnya (Cruse, 2000: 236). Contoh korelasi yang terdapat dalam komponen [PEREMPUAN] yaitu *ibu*, *istri*, *wanita*, *tante*, *betina*, dsb. Ketiga, dalam beberapa hal, ada beberapa bukti langsung yang menunjukkan ketidaksenambungan makna atau yang disebut dengan diskontinuitas. Misalnya, dalam *I almost killed her* terdapat ambiguitas, yaitu (1) “saya sedang melakukan aksi, (misalnya menarik pelatuk pistol) yang menyebabkan dia (*her*) meninggal”, dan (2) “saya beraksi sedemikian rupa sehingga menyebabkan dia hampir meninggal”. Kedua ambiguitas kalimat tersebut merupakan anggapan otonomi fungsional untuk komponen [CAUSE] atau *menyebabkan* dan [DIE] atau *mati* di dalam makna *kill* atau *membunuh*. Kedua makna tersebut dapat digambarkan sebagai berikut.

- | |
|--|
| A. almost ([CAUSE] [DIE])
B. [CAUSE] (almost [DIE]) |
|--|

Keempat, dalam banyak kasus, bentuk sederhana gramatikal memiliki ciri semantik yang serupa ataupun sejajar dengan bentuk yang kompleks (Cruse, 2000: 238). Ada banyak kasus, misalnya perbandingan *rise/fall* dengan *lengthen/shorten* dalam bahasa Inggris. *Lengthen* dan *shorten* memiliki hubungan morfologis dengan *long* dan *short*, maknanya pun dapat dianalisis menjadi [BECOME] [MORE] [LONG] dan [BECOME] [MORE] [SHORT]. Mengingat bahwa perbedaan antara *lengthen* dan *shorten* sama dengan perbedaan antara *rise* dan *fall* dan mengingat bahwa relasi makna antara *lengthen* dan *long* sama dengan relasi

makna antara *rise* dan *high*, tentunya komponen dari *rise* dan *fall* dapat menjadi [BECOME] [MORE] [HIGH] dan [BECOME] [MORE] [LOW].

Terdapat lima tujuan dekomposisi leksikal menurut Cruse (2000: 238—250). Tujuan pertama, dekomposisi leksikal digunakan untuk mereduksi kata sehingga menghasilkan komponen makna (Cruse, 2000: 238—244). Cruse memiliki model analisis komponen makna yang berbeda dengan Nida. Berikut adalah contoh analisis komponen makna untuk *rise*, *fall*, *lengthen*, *shorten*.

rise	raise	high
fall	lower	low
lengthen ¹	lengthen ²	long
shorten ¹	shorten ²	short
rise	= [BECOME] [MORE] [HIGH]	
fall	= [BECOME] [MORE] [LOW]	
raise	= [CAUSE] [BECOME] [MORE] [HIGH]	
lower	= [CAUSE] [BECOME] [MORE] [LOW]	
lengthen ¹	= [BECOME] [MORE] [LONG]	
shorten	= [BECOME] [MORE] [SHORT]	
lengthen ¹	= [CAUSE] [BECOME] [MORE] [LONG]	
shorten	= [CAUSE] [BECOME] [MORE] [SHORT]	

Berdasarkan uraian analisis komponen makna di atas, terdapat 7 komponen, yaitu [BECOME], [MORE], [HIGH], [CAUSE], [LOW], [LONG], [SHORT]. Analisis dekomposisi leksikal milik Cruse cenderung berupa definisi kata yang terintegrasi dalam komponen-komponen makna. Berikut adalah contohnya dalam bahasa Indonesia.

kakek	= [MANUSIA] [LAKI-LAKI] [TUA]
nenek	= [MANUSIA] [PEREMPUAN] [TUA]

Tujuan kedua, yaitu dekomposisi leksikal digunakan untuk mengetahui persamaan dan perbedaan leksikal. Analisis komponensial harus dimulai dengan kata yang paling jauh maknanya sampai yang terdekat supaya lebih sistematis. Cruse (2000: 244) mencontohkan analisis komponensial yang memperlihatkan persamaan dan perbedaan leksikal sebagai berikut.

chair	vs. thought [CONCRETE]
	vs. cat [INANIMATE]
	vs. trumpet [FURNITURE]
	vs. table [FOR SITTING]
	vs. sofa [FOR ONE]
	vs. tool [WITH BACK]

Tujuan ketiga, yaitu dekomposisi leksikal digunakan untuk mengetahui adanya relasi dan pertautan leksikal. Cruse (2000: 246) mencontohkan beberapa relasi dan pertautan leksikal sebagai berikut.

<i>stallion</i>	[ANIMAL] [EQUINE] [MALE] hiponim dari
<i>horse</i>	[ANIMAL] [EQUINE]
<i>chair</i>	[CONCRETE] [INANIMATE] [FURNITURE] [FOR SITTING] [FOR ONE] [WITH BACK] hiponim dari
<i>furniture</i>	[CONCRETE] [INANIMATE] [FURNITURE]

Tujuan keempat, dekomposisi leksikal digunakan untuk memprediksi kata yang beranomali. Dalam memprediksi kata-kata yang anomali atau normal, digunakan sistem komponen makna dengan menentukan pembatasan seleksional. Bagian yang menyertai kata-kata tersebut harus memiliki hasil akhir berupa kalimat normal. Pembatasan seleksional dapat diberlakukan untuk sebuah subjek dari kata kerja. Sebagai contoh, terdapat perbedaan arti *expired* dalam kalimat *John expired* dan *my driving licence has expired*. Dalam kalimat *John expired*, kata *expired* bermakna ‘mati’, sedangkan dalam kalimat *my driving licence has expired*, kata *expired* bermakna ‘sudah tidak valid’. Berikut uraian perbedaan komponen makna *expired* untuk dua subjek yang berbeda.

<i>expire</i> = [BECOME] [NOT] [ALIVE] <[HUMAN]> <i>expire</i> = [BECOME] [NOT] [VALID] <[DOCUMENT]>

Contoh kedua untuk memprediksi kata yang beranomali oleh Cruse (2000: 248), yaitu *pregnant*.

<i>pregnant</i> = [WITH CHILD IN WOMB] <[ANIMAL] [FEMALE]>
--

Benar apabila diprediksi bahwa kalimat *my sister is pregnant* adalah normal. Akan tetapi *the chair is pregnant* adalah kalimat yang anomali. Kalimat *my neighbour is pregnant* pun dapat menjadi anomali karena komponen makna *neighbour* dapat mencakup [ANIMAL] dan [HUMAN] yang belum tentu [FEMALE].

Tujuan kelima, dekomposisi leksikal digunakan untuk menjelaskan adanya diskontinuitas. Analisis komponen makna dikontinuitas memberikan penjelasan untuk sifat berlainan variabel kata penghubung, seperti kata lagi (*again*), hampir (*almost*), dan tidak (*not*) di dalam makna unsur leksikal. Contoh ambiguitas dalam kalimat “*I almost killed her*” (aku hampir membunuhnya).

- | |
|--|
| A. “ <i>I was on the point of carrying out an action which would have caused her to die</i> ”.
(Saya berada di titik melaksanakan tindakan yang akan menyebabkan dia mati)
B. “ <i>I acted in such a way as to cause her to be almost dead</i> ”.
(Saya bertindak sedemikian rupa untuk menyebabkan dia menjadi hampir mati). |
|--|

Kedua kalimat di atas menunjukkan otonomi fungsional untuk komponen [CAUSE] dan [DIE] dalam arti [KILL]. Kedua makna tersebut dapat digambarkan sebagai berikut.

- A. almost ([CAUSE] [DIE])
- B. [CAUSE] (almost [DIE])

2.2 Metode Penelitian

Penelitian ini merupakan penelitian yang bersifat deskriptif. Penelitian yang bersifat deskriptif merupakan prosedur pemecahan masalah yang diselidiki dengan menggambarkan atau melukiskan subjek atau objek penelitian pada saat sekarang berdasarkan fakta-fakta yang tampak atau sebagaimana adanya (Hadari, 1985: 63). Dalam penelitian yang bersifat deskriptif ini, penulis akan menganalisis data dengan menggunakan teori dekomposisi leksikal oleh Cruse (2000).

Teknik pengumpulan data yang penulis lakukan adalah teknik baca dan teknik catat. Data yang dijadikan objek penelitian penulis merupakan lema dan sublema bidang komputer dalam KBBI Daring (2018). Lema dan sublema bidang komputer dalam KBBI Daring memiliki label *Komp*. Dalam antarmuka pengguna, lema dan sublema berlabel *Komp* dapat ditemukan di fitur pencarian kelompok bidang komputer. Lema dan sublema berlabel *Komp* dalam KBBI Daring berjumlah 961 per Maret 2018.

Langkah-langkah penelitian ini, yaitu (1) menginventarisasi lema dan sublema bidang komputer dalam KBBI Daring, (2) mengelompokkan data ke dalam beberapa kelompok berdasarkan data yang mirip secara leksikal dan data yang berhiponim, (3) menguraikan proses penerapan dekomposisi leksikal terhadap data, dan (4) merumuskan kesimpulan berdasarkan bagian analisis.

III. ANALISIS PENELITIAN

3.1 Lema dan Sublema Bidang Komputer dalam KBBI Daring (2018) yang Memiliki Kemiripan Leksikal dan yang Berhiponim

Dalam KBBI Daring (2018), terdapat lema dan sublema bidang komputer yang memiliki kemiripan leksikal. Daftar lema dan sublema bidang komputer yang memiliki kemiripan leksikal, di antaranya sebagai berikut.

Tabel 1. Daftar Data Lema dan Sublema Bidang Komputer dalam KBBI Daring yang Memiliki Kemiripan Leksikal

No.	Contoh Data
1	koneksi – jaringan – sambungan
2	nirkoneksi – nirhubung
3	akses berkas – akses ke berkas – akses ke data
4	berkas – data – fail – dokumen

Selain itu, dalam KBBI Daring (2018) juga terdapat lema dan sublema bidang komputer yang berhiponim. Kelompok lema dan sublema bidang komputer yang berhiponim tersebut berjumlah 60. Kelompok data yang berjumlah 60 tersebut terdiri dari 416 lema dan sublema. Genus dalam definisi lema dan sublema yang berhiponim tersebut masih beragam, seperti contoh berikut.

<p>kode <u>n komp</u> aturan transformasi data atau informasi lainnya dari satu bentuk simbolik ke bentuk lainnya</p> <p>kode biner <u>komp</u> kode dalam komputer yang menggunakan sistem angka biner, yaitu angka 1 dan 0</p> <p>kode bita <u>komp</u> bentuk-bentuk instruksi java yang dapat diaplikasikan pada mesin virtual Java</p> <p>kode bita java <u>komp</u> bentuk-bentuk instruksi yang dapat dilakukan oleh mesin virtual Java</p> <p>kode instruksi <u>komp</u> bagian dari bahasa mesin atau instruksi bahasa perakitan yang menentukan jenis instruksi dan struktur data tempat ia beroperasi</p> <p>kode kendali <u>komp</u> deretan satu atau beberapa karakter yang digunakan untuk mengendalikan perangkat keras</p> <p>kode mesin <u>komp</u> bahasa mesin</p>
--

Dalam KBBI Daring (2018) terdapat 60 kelompok data bidang komputer dengan 416 lema dan sublema yang berhiponim seperti contoh di atas. Hal tersebut menunjukkan bahwa terdapat sekitar 416 definisi lema dan sublema bidang komputer yang masih memiliki genus yang beragam.

3.2 Penerapan Dekomposisi Leksikal dalam Menyusun Definisi dan Manfaatnya

Berkaitan dengan lema dan sublema berlabel *Komp* dalam KBBI Daring (2018), dekomposisi leksikal dapat digunakan mengetahui persamaan dan perbedaan leksikal dan mengetahui adanya relasi dan pertautan leksikal. Dalam lema dan sublema berlabel *Komp* dalam KBBI Daring (2018), terdapat beberapa lema dan sublema yang memperlihatkan adanya kemiripan leksikal, misalnya antara lema *berkas*, *data*, *fail*, dan *dokumen*. Lema *berkas* bermakna ‘kumpulan data yang dapat diberi nama berbeda dan disimpan pada tempat penyimpanan sekunder dalam komputer’ (KBBI Daring, 2018). Lema *berkas* memiliki perbedaan leksikal dengan lema *kertas*, *komputer*, *map*, *data*, *fail*, dan *dokumen*. Akan tetapi, lema *berkas* juga memiliki persamaan leksikal dengan lema *komputer*, *map*, *data*, *fail*, dan *dokumen*. Perbedaan dan persamaan leksikal antara lema *berkas* dan yang lainnya tersebut dapat dilihat dalam analisis komponensial berikut ini.

berkas vs kertas [KOMPUTER]
berkas vs komputer [BAGIAN DALAM KOMPUTER]
berkas vs map [YANG DISIMPAN DALAM MAP]
berkas vs data [KUMPULAN DATA]
berkas vs fail [BAGIAN DARI FAIL]
berkas vs dokumen [TIDAK HANYA TEKS]

Analisis komponensial di atas dimulai dengan kata yang paling jauh maknanya dengan lema *berkas* sampai dengan yang maknanya paling dekat dengan lema *berkas*. Lema *berkas* merupakan istilah yang termasuk dalam bidang komputer, sedangkan lema *kertas* tidak. Perbedaannya dengan lema *komputer*, lema *berkas* merupakan bagian dari sebuah sistem komputer. Persamaan lema *berkas* dan lema *komputer*, yaitu keduanya termasuk ke dalam bidang komputer. Lema *berkas* dan lema *map* memiliki perbedaan leksikal, yaitu *berkas* merujuk pada kumpulan data yang akan disimpan, sedangkan *map* merujuk pada tempat penyimpanan berkas. Lema *berkas* dan lema *data* pun masih memiliki perbedaan, yaitu lema *berkas* merujuk pada data-data yang sudah dikumpulkan sehingga disebut dengan kumpulan data. Lema *berkas* dan *fail* sama-sama merujuk pada sebuah kumpulan data, tetapi secara terperinci lema *fail* merujuk pada berkas atau dokumen yang dihasilkan. Kata yang paling dekat maknanya dengan lema *berkas*, yaitu lema *dokumen*. Meskipun keduanya merujuk pada kumpulan data, secara lebih terperinci lema *dokumen* hanya berisi kumpulan data yang berupa teks yang dibuat dengan perangkat lunak pengolah kata.

Berdasarkan uraian di atas, terlihat bahwa dekomposisi leksikal dapat digunakan untuk menunjukkan adanya perbedaan dan persamaan leksikal. Dalam sebuah proses penyusunan kamus, hal tersebut dapat mempermudah penyusun kamus dalam menentukan definisi untuk lema atau sublema yang memiliki kemiripan leksikal. Adanya definisi yang jelas untuk lema dan sublema yang memiliki kemiripan leksikal dapat menghindari definisi dengan sinonim yang mengandung makna melingkar.

Dalam lema dan sublema berlabel *Komp* dalam KBBI Daring (2018) terdapat banyak lema dan sublema yang berhiponim, misalnya sublema *program antivirus* yang merupakan hiponim dari lema *program*. Lema dan sublema yang berhiponim tersebut menandakan adanya relasi dan pertautan leksikal. Dalam hal ini, dekomposisi leksikal juga dapat digunakan untuk menguraikan adanya relasi dan pertautan leksikal. Berikut adalah contoh analisis komponensial yang memperlihatkan adanya relasi dan pertautan leksikal dalam lema dan sublema berlabel *Komp* dalam KBBI Edisi Kelima.

program	= [URUTAN PERINTAH] [KOMPUTER] [UNTUK MEMBUAT FUNGSI/TUGAS]
program antivirus	= [PROGRAM] [KOMPUTER] [MELINDUNGI DARI VIRUS]
program berdesain pesanan	= [PROGRAM] [KOMPUTER] [UNTUK KLIEN]
program faks	= [PROGRAM] [KOMPUTER] [MENGIRIM INFORMASI]
program kendali	= [PROGRAM] [KOMPUTER] [MENGAWASI INSTRUKSI]
program layanan	= [PROGRAM] [KOMPUTER] [MENDUKUNG FUNGSI SISTEM]

Berdasarkan uraian analisis komponensial di atas, terlihat bahwa sublema *program layanan, program antivirus, program berdesain pesanan, program faks, program kendali, program layanan* merupakan hiponim dari lema *program*. Definisi dalam sublema *program layanan, program antivirus, program berdesain pesanan, program faks, program kendali, program layanan* dapat diawali dengan genus sama, yaitu keterangan *program dalam komputer*. Dengan demikian, dapat dikatakan bahwa dengan mengetahui adanya relasi dan pertautan leksikal dalam sebuah daftar lema atau sublema, genus dalam definisi yang dihasilkan untuk lema dan sublema yang berhiponim akan menjadi seragam.

Dalam KBBI Daring (2018), definisi *program, program layanan, program antivirus, program berdesain pesanan, program faks, program kendali, program layanan* adalah sebagai berikut.

program	<i>n komp</i> urutan perintah yang diberikan pada komputer untuk membuat fungsi atau tugas tertentu
program antivirus	<i>komp</i> program yang diciptakan secara khusus untuk melindungi komputer dari serangan virus
program berdesain pesanan	<i>komp</i> jenis program yang dikembangkan untuk klien tertentu atau untuk mengatasi kebutuhan khusus
program faks	<i>komp</i> komputer atau perangkat lunak yang digunakan untuk mengirim informasi
program kendali	<i>komp</i> perangkat lunak yang menjadwalkan dan mengawasi pelaksanaan instruksi-instruksi pada komputer
program layanan	<i>komp</i> program komputer yang digunakan dalam sistem komputer untuk mendukung fungsi sistem, seperti sebagai pustaka atau program utilitas (KBBI Daring, 2018)

Dengan mengetahui adanya relasi dan pertautan leksikal antara *program, program layanan, program antivirus, program berdesain pesanan, program faks, program kendali, program layanan*, usulan perbaikan definisi untuk enam sublema tersebut adalah sebagai berikut.

<p>program <i>n komp</i> urutan perintah yang diberikan pada komputer untuk membuat fungsi atau tugas tertentu</p> <p>program antivirus <i>komp</i> program pada komputer yang diciptakan secara khusus untuk melindungi komputer dari serangan virus</p> <p>program berdesain pesanan <i>komp</i> program pada komputer yang dikembangkan untuk klien tertentu atau untuk mengatasi kebutuhan khusus</p> <p>program faks <i>komp</i> program pada komputer yang digunakan untuk mengirim informasi</p> <p>program kendali <i>komp</i> program pada komputer yang menjadwalkan dan mengawasi pelaksanaan instruksi-instruksi pada komputer</p> <p>program layanan <i>komp</i> program pada komputer yang digunakan dalam sistem komputer untuk mendukung fungsi sistem, seperti sebagai pustaka atau program utilitas (KBBI Daring, 2018)</p>
--

IV. SIMPULAN

Dalam KBBI Daring (2018), terdapat lema dan sublema bidang komputer yang memiliki kemiripan leksikal, seperti *berkas*, *data*, *file*, dan *dokumen*. Selain itu, dalam KBBI Daring (2018) juga terdapat lema dan sublema bidang komputer yang berhiponim. Kelompok lema dan sublema bidang komputer yang berhiponim tersebut berjumlah 60. Genus dalam definisi lema dan sublema yang berhiponim tersebut masih beragam. Berdasarkan hasil penelitian, terlihat bahwa dekomposisi leksikal dapat diterapkan sebagai salah satu strategi pendefinisian lema dan sublema bidang komputer dalam KBBI Daring (2018), khususnya dalam menyusun definisi lema dan sublema yang memiliki kemiripan leksikal dan yang berhiponim.

Dalam sebuah proses penyusunan kamus, pengetahuan adanya persamaan dan perbedaan leksikal dapat mempermudah penyusunan kamus dalam menentukan definisi untuk lema atau sublema yang memiliki kemiripan leksikal. Adanya definisi yang jelas untuk lema dan sublema yang memiliki kemiripan leksikal dapat menghindari definisi dengan sinonim yang mengandung makna melingkar. Selanjutnya, dengan mengetahui adanya relasi dan pertautan leksikal dalam sebuah daftar lema atau sublema, genus dalam definisi yang dihasilkan untuk lema dan sublema yang berhiponim akan menjadi seragam. Dekomposisi leksikal tidak hanya dapat diterapkan dalam menyusun lema dan sublema bidang komputer, tetapi juga bidang-bidang lainnya.

Daftar Pustaka

- Atkins, B. T. dan Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan. 2016. *Pedoman Umum Ejaan Bahasa Indonesia*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.
- Chaer, Abdul. 2007. *Leksikologi dan Leksikografi Indonesia*. Jakarta: Rineka Cipta.

- Cruse, Alan. 2000. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: OUP Oxford
- Keraf, Gorys. 1981. *Diksi dan Gaya Bahasa*. Jakarta: Penerbit Nusa Indah.
- Kridalaksana, Harimurti. 1985. *Tata Bahasa Deskriptif Indonesia: Sintaksis*. Jakarta: Universitas Indonesia.
- _____. 2008. *Kamus Linguistik*. Jakarta: PT. Gramedia Pustaka Utama.
- Koen, Willie. 1996. "Kamus Besar Bahasa Indonesia: Sebuah Tinjauan Kecil dengan Acu Silang". Dalam Purwo, Bambang Kaswanti (ed.). (1996). *Pelba 9 Pertemuan Linguistik Lembaga Bahasa Atma Jaya: Kesembilan; Linguistik Lapangan, Bahasa, dan Politik, Evaluasi Kamus*. Yogyakarta: Kanisius.
- Muis, Muhammad. (2009). *Pendefinisian Lema Alat Musik di Dalam Kamus Besar Bahasa Indonesia*. Jakarta: Pusat Bahasa, Departemen Pendidikan Nasional.
- Nawawi, Hadari. (1985). *Metode Penelitian Bidang Sosial*. Yogyakarta: Gadjah Mada University Press
- Sumarsono. (2002). "Klasifikasi dan Definisi dalam Kamus Besar Bahasa Indonesia". Dalam *Bahasa dan Sastra*, Volume 20, Nomor 1, Edisi Januari—Maret 2002. Hlm. 66—67. Jakarta.
- Sunaryo, Adi dan Hermanu Maulana. 1987. *Pedoman Penyuntingan Kamus Bahasa Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
- Suratminto, Lilie dan Munawar Holil. 2003. *Rintisan Kajian Leksikologi dan Leksikografi*. Jakarta: FIB UI dan Dewan Bahasa dan Pustaka Malaysia.
- Svensen, Bo. 2009. *A Handbook of Lexicography*. New York: Cambridge University Press.
- Tim Penyusun. 2016. *Kamus Besar Bahasa Indonesia Edisi Kelima*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan.
- Tim Penyusun. 2016. "KBBI Daring". <https://kbbi.kemendikbud.go.id> yang diakses pada 3 Februari 2018
- Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

**PENGEMBANGAN KAMUS MORFOLOGI DAN APLIKASI
KONKORDANSI DARING BAHASA INDONESIA
DAN MALAYSIA SUMBER TERBUKA: MALINDO MORPH DAN
MALINDO CONC**

*(DEVELOPMENT OF AN OPEN SOURCE INDONESIAN AND MALAYSIAN
MORPHOLOGICAL DICTIONARY AND AN ONLINE CONCORDANCE
APPLICATION: MALINDO MORPH AND MALINDO CONC)*

David Moeljadi

Universitas Teknologi Nanyang, Singapura

Pos-el: D001@e.ntu.edu.sg, davidmoeljadi@gmail.com

Abstrak

Kerja sama berbagai pihak, antara pihak pemerintah dan banyak lembaga penelitian atau universitas, baik dari dalam maupun luar negeri, diperlukan dalam pengembangan sumber daya dan teknologi bahasa. Kerja sama para pakar dari berbagai bidang, yaitu leksikografi, linguistik, dan teknologi informasi dapat membuka berbagai kemungkinan dan hal baru. Selain itu, akses sumber terbuka dengan lisensi yang jelas sangat menunjang kemajuan penelitian bahasa dan leksikografi karena para peneliti dapat menggunakan hasil penelitian dan alat atau media yang sudah ada dengan leluasa.

Makalah ini membahas dua produk penelitian bahasa Indonesia dan bahasa Malaysia yang berupa kamus morfologi dan aplikasi konkordansi daring, yakni MALINDO Morph (Nomoto, Choi, Moeljadi, dan Bond 2018) dan MALINDO Conc (Nomoto, Akasegawa, dan Shiohara 2018). Kedua produk ini menggunakan berbagai sumber yang sudah ada, yaitu data morfologi atau pembentukan kata dari Kamus Besar Bahasa Indonesia Edisi Kelima (KBBI V), Kamus Dewan (KD), dan korpus Leipzig (Goldhahn, Eckart, dan Quasthoff 2012). Walaupun banyak keterbatasan dalam pembuatannya, seperti akses ke sumber pangkalan data, kedua produk ini dapat dihasilkan berkat kerja sama para pakar dari berbagai bidang ilmu yang berasal dari berbagai universitas dan lembaga mancanegara, yakni Universitas Kajian Asing Tokyo di Jepang, Universitas Teknologi Nanyang di Singapura, dan Institut Bahasa Lago di Jepang.

MALINDO Morph dan MALINDO Conc menunjukkan pentingnya kerja sama para pakar dari berbagai bidang, yaitu pekamus, ahli ilmu bahasa, dan pakar teknologi informasi; kerja sama berbagai universitas dan lembaga mancanegara; dan lisensi sumber terbuka dalam penyusunan dan pengembangan sumber daya dan teknologi bahasa, seperti pangkalan data, korpus, kamus morfologi, dan sistem antarmuka pengguna yang selanjutnya dapat dimanfaatkan oleh pekamus dalam pemutakhiran kamus, ahli ilmu bahasa dalam penelitian bahasa, dan masyarakat penutur bahasa.

Kata kunci: pangkalan data, kamus elektronik, kamus morfologi, aplikasi korpus daring

Abstract

Cooperation between institutions, such as government bodies and research institutes or universities, both inside and outside Indonesia, is needed in the development of language resources and technology. Cooperation between experts from different fields, i.e. lexicography, linguistics, and computer science, can open new possibilities. In addition, open source license supports the

improvement of lexicography and language research because researchers can use the existing tools or media and research results.

This paper discusses two products of Indonesian and Malaysian language research which take forms of a morphological dictionary and an online concordance application: MALINDO Morph (Nomoto, Choi, Moeljadi, and Bond 2018) and MALINDO Conc (Nomoto, Akasegawa, and Shiohara 2018). Both products use the existing sources, i.e. morphological/word formation data derived from Kamus Besar Bahasa Indonesia Fifth Edition (KBBI V), Kamus Dewan (KD), and Leipzig corpus (Goldhahn, Eckart, and Quasthoff 2014). Despite many limitations such as access to the original database source, both products could be made thanks to the cooperation of experts from various fields in international institutions, i.e. Tokyo University of Foreign Studies in Japan, Nanyang Technological University in Singapore, and Lago Institute of Language in Japan.

MALINDO Morph and MALINDO Conc show the importance of cooperation between researchers from many fields, i.e. lexicographers, linguists, and computer scientists; cooperation between international institutions and universities; and open source licenses in the creation and development of language technology and resources, such as a morphological dictionary and an online corpus application that can be subsequently utilized by lexicographers in updating the dictionaries, by linguists in linguistics research, and by the language speakers.

Keywords: *database, electronic dictionary, morphology dictionary, online corpus dictionary*

I. PENDAHULUAN

Sumber-sumber linguistik yang bermutu, seperti pangkalan data kamus elektronik dan korpus, diperlukan dalam berbagai aplikasi pemrosesan bahasa alami atau *natural language processing* (NLP), seperti penerjemahan mesin, analisis sentimen, dan anotasi kelas kata. Dibandingkan bahasa Inggris dan bahasa Jepang, berbagai dialek bahasa Melayu, terutama bahasa Indonesia dan bahasa Malaysia yang banyak jumlah penuturnya, masih kurang dalam jumlah dan mutu sumber-sumber linguistik tersebut. Pangkalan data kamus elektronik dapat dimanfaatkan untuk berbagai aplikasi, misalnya analisis leksikografi dan analisis linguistik (Moeljadi, Kamajaya, dan Amalia 2017). Pangkalan data kamus elektronik yang berisi informasi tentang bentuk dasar dan berbagai bentuk turunan suatu lema dapat dimanfaatkan dalam penyusunan kamus morfologi yang selanjutnya dapat digunakan dalam penyusunan sistem pencarian kata dan kolokasi korpus. Pangkalan data Kamus Besar Bahasa Indonesia Edisi Kelima (KBBI V) dan Kamus Dewan (KD) berisi entri kata yang terdiri atas bentuk dasar atau akar kata dan bentuk turunannya yang berafiks atau bereduplikasi dan dapat dimanfaatkan dalam pengembangan aplikasi anotasi kelas kata dan fitur pencarian secara morfologis di aplikasi korpus daring. Selain itu, Korpus Leipzig atau Leipzig Corpora Collection (Goldhahn, Eckart, and Quasthoff 2012) yang berlisensi sumber terbuka dapat dimanfaatkan sebagai langkah awal pembuatan aplikasi daring konkordansi bahasa Indonesia dan Malaysia. Untuk memfasilitasi pengembangan berbagai sumber linguistik dan aplikasi untuk berbagai dialek bahasa Melayu, Badan Pengembangan dan Pembinaan Bahasa (Badan Bahasa), Dewan Bahasa dan Pustaka (DBP) Malaysia, Dewan Bahasa dan Pustaka (DBP) Brunei Darussalam, dan terutama Majlis Bahasa Brunei-Indonesia-Malaysia (MABBIM) dapat membantu dalam hal sarana dan prasarana yang menunjang penelitian-penelitian leksikografi dan linguistik (komputasional).

Walaupun bahasa Indonesia dan bahasa Malaysia memiliki banyak sumber linguistik seperti yang telah diuraikan, tetapi kebanyakan tidak berlisensi sumber terbuka atau tidak jelas lisensinya, seperti pangkalan data KBBI V dan KD. Kerja sama dan pertukaran informasi antara lembaga pemerintah dan nonpemerintah, baik domestik dan internasional, serta kerja sama antarpeleliti berbagai bidang pun perlu digalakkan. Selain itu, tenaga ahli di bidang linguistik komputasional perlu ditambahkan ke Badan Bahasa, DBP Malaysia, DBP Brunei, dan MABBIM.

Makalah ini bertujuan menyadarkan pentingnya lisensi sumber terbuka untuk sumber-sumber linguistik dan pentingnya kerja sama berbagai pihak dari berbagai bidang untuk peningkatan mutu berbagai sumber linguistik dan aplikasi. Hal tersebut dilakukan dengan memperkenalkan dua produk yang telah dihasilkan oleh peneliti dari berbagai institusi dan bidang ilmu di tengah segala keterbatasan yang ada, yaitu kamus morfologi MALINDO Morph dan aplikasi konkordansi daring MALINDO Conc. Penulis berharap Badan Bahasa, DBP Malaysia, DBP Brunei, dan MABBIM dapat terlibat dalam pengembangan MALINDO Morph dan MALINDO Conc selanjutnya, baik pengembangan jumlah dan kualitas data maupun dalam aspek teknisnya.

II. MALINDO MORPH

Analisis morfologi yang tepat secara linguistik dengan cakupan luas memiliki peran vital dalam aplikasi pemrosesan bahasa alami, seperti anotasi kelas kata dan penguraian sintaks kalimat. Data kamus elektronik yang lengkap menjadi fondasi atau dasar analisis morfologi ini.

MALINDO Morph (Nomoto, Choi, Moeljadi, dan Bond 2018) merupakan kamus morfologi pertama berlisensi sumber terbuka untuk bahasa Indonesia dan bahasa Malaysia yang disusun oleh para ahli ilmu bahasa dan linguistik komputasional Universitas Kajian Asing Tokyo dan Universitas Teknologi Nanyang di Singapura. Kamus ini berlisensi Creative Commons Attribution 4.0 International (CC BY 4.0) dan dapat diunduh di GitHub (https://github.com/matbahasa/MALINDO_Morph). Kamus ini mula-mula disusun dari data pembentukan kata-kata turunan bahasa Indonesia yang disusun penulis berdasarkan data akar kata dan bentuk turunan di KBBI V untuk bahasa Indonesia, serta dari data KD untuk bahasa Malaysia. Data pembentukan kata-kata turunan di KBBI V tersebut pada awalnya disusun dengan tujuan pencetakan KBBI V. Pangkalan data KD tidak berlisensi sumber terbuka. Data KD dimasukkan ke MALINDO Morph satu per satu secara manual.

Selanjutnya, kata-kata yang sering muncul di korpus Leipzig (Goldhahn, Eckart, dan Quasthoff 2014) yang tidak ada, baik di KBBI V maupun di KD, termasuk nama, singkatan, dan salah ketik (saltik) ditambahkan ke dalam kamus ini sehingga kamus ini memiliki jumlah bentuk dasar dan bentuk turunan yang lebih banyak daripada KBBI dan KD. Selain itu, semua bentuk dasar dan turunan yang memiliki kemungkinan berenklitik, seperti *masalah* dan *penanya* ditambahkan secara manual. Bentuk *masalah* dapat dianalisis sebagai bentuk turunan dari bentuk dasar *masa* dan partikel *-lah*, sedangkan bentuk *penanya* dapat dianalisis sebagai bentuk turunan dari bentuk dasar *pena* dan enklitik *-nya*.

Saat ini kamus ini berisi 232.550 baris, berdasarkan 78.750 akar kata, dan setiap barisnya berisi analisis satu bentuk (*case sensitive*). Kamus ini disusun secara semiotomatis.

Bahasa pemrograman Python digunakan dalam proses analisis morfologi yang hasilnya diperiksa secara manual untuk menunjang ketepatan analisis. Kamus ini berisi berbagai informasi, seperti nomor entri, akar kata atau bentuk dasar, bentuk turunan, prefiks/proklitik, sufiks/enklitik, konfiks, dan tiga jenis reduplikasi (reduplikasi penuh atau dwilingga, reduplikasi sebagian atau dwipurwa, dan reduplikasi berubah bunyi atau dwilingga salin suara). Gambar 1 menunjukkan data informasi morfologi bentuk-bentuk turunan yang berasal dari akar kata *alah*.

Saat ini, MALINDO Conc hanya memuat imbuhan dan bentuk reduplikasi produktif dalam bahasa Indonesia dan Malaysia. Semua imbuhan yang berasal dari bahasa asing, seperti *anti-* dan *pra-* belum dimuat, kecuali jika digabung dengan kata asli bahasa Indonesia atau bahasa Malaysia, seperti *anti-pemerintah*. Selain itu, belum ada perbedaan antara sufiks *-nya* (seperti dalam kata keterangan *biasanya*, pembentuk kata benda dari kata kerja *adanya*, dan pembentuk kata seruan dari kata sifat *enaknya*) dan enklitik *-nya* (seperti dalam *bukunya*, *buku merahnya*, dan *dibelinya*). Informasi tambahan seperti ini dapat ditambahkan dalam pengembangan selanjutnya. Selain itu, informasi dialek dan bahasa, seperti bahasa Melayu Malaysia, bahasa Melayu Brunei, dan bahasa Melayu Singapura, dapat juga ditambahkan. Fitur lain yang dapat dikembangkan adalah informasi kelas kata yang dapat diketahui dari informasi kelas kata di KBBI V dan KD dan jenis imbuhan yang ada.

Walaupun masih banyak informasi yang dapat ditambahkan untuk pengembangan MALINDO Morph ini, kamus ini memiliki data morfologi yang cukup untuk dimanfaatkan dalam penyusunan aplikasi konkordansi daring untuk pencarian kata atau kolokasi korpus bahasa Indonesia dan Malaysia, bernama MALINDO Conc (Nomoto, Akasegawa, dan Shiohara 2018).

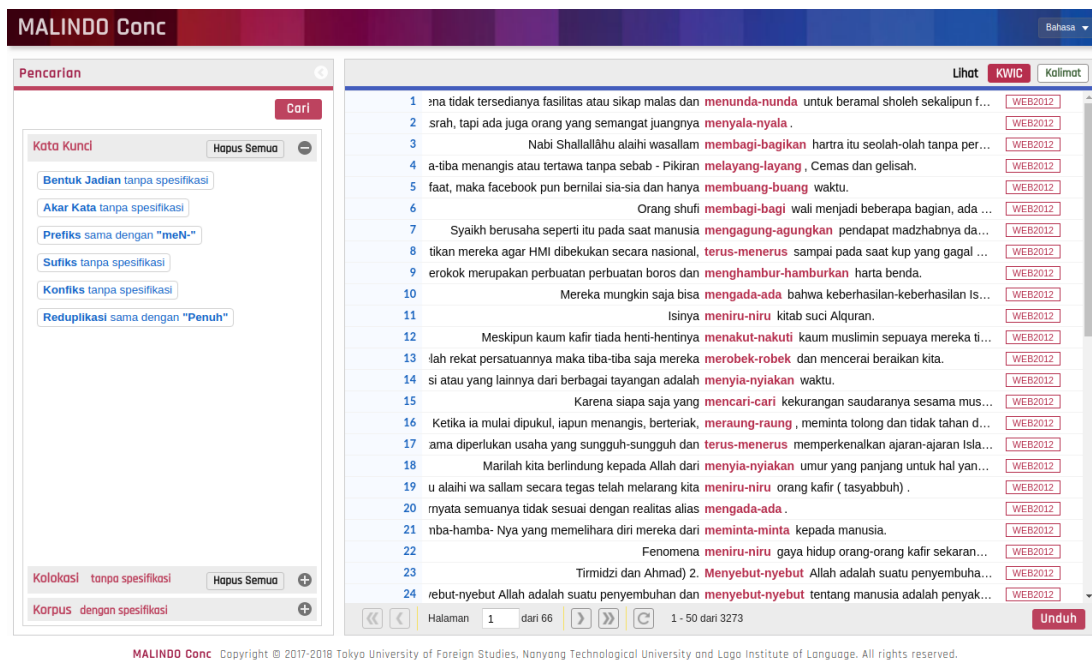
cc-5409	<u>alah</u>	<u>alah</u>		0	0	0	0
cc-5410	<u>alah</u>	<u>alah-mengalahi</u>	<u>meN-</u>	<u>-i</u>		0	<u>R-penuh</u>
cc-5411	<u>alah</u>	<u>alahan</u>		0	<u>-an</u>	0	0
cc-5412	<u>alah</u>	<u>beralah</u>	<u>ber-</u>		0	0	0
cc-5413	<u>alah</u>	<u>beralahan</u>		0	0	<u>ber--an</u>	0
cc-5414	<u>alah</u>	<u>alahan</u>	<u>ber-</u>	<u>-an</u>		0	0
cc-5415	<u>alah</u>	<u>kealahan</u>		0	0	<u>ke--an</u>	0
cc-5416	<u>alah</u>	<u>mengalah</u>	<u>meN-</u>		0	0	0
cc-5417	<u>alah</u>	<u>mengalahi</u>	<u>meN-</u>	<u>-i</u>		0	0
cc-5418	<u>alah</u>	<u>mengalahkan</u>	<u>meN-</u>	<u>-kan</u>		0	0
cc-5419	<u>alah</u>	<u>pengalahan</u>		0	0	<u>peN--an</u>	0
cc-5420	<u>alah</u>	<u>peralahan</u>		0	0	<u>per--an</u>	0
cc-5421	<u>alah</u>	<u>teralahkan</u>	<u>ter-</u>	<u>-kan</u>		0	0

Gambar 1. Tangkapan Layar Informasi Morfologi Bentuk-Bentuk Turunan dari Akar Kata *alah* di MALINDO Morph

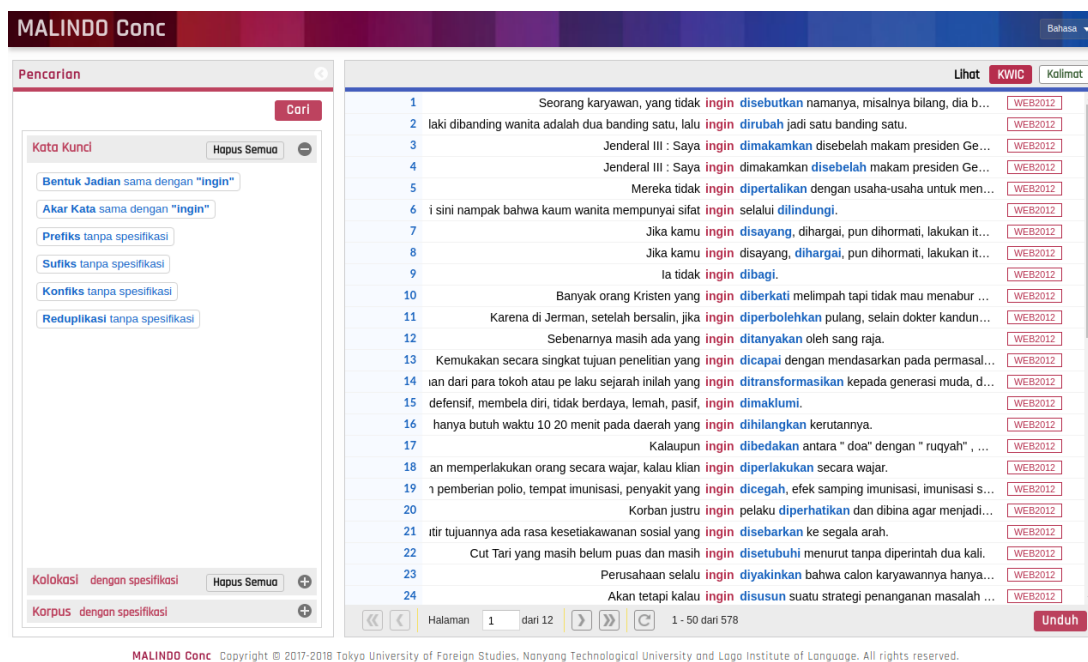
III. MALINDO CONC

MALINDO Conc adalah aplikasi konkordansi daring bahasa Indonesia dan bahasa Malaysia berlisensi sumber terbuka yang dirancang sebagai alat bantu peneliti bahasa Indonesia dan Malaysia yang tak berbayar, mudah digunakan, dan memiliki fitur pencarian mendetail. Aplikasi ini disusun para linguist bahasa Indonesia dan Malaysia dari Universitas Kajian Asing Tokyo dan pakar ilmu komputer dari Institusi Bahasa Lago di Jepang berdasarkan dan memanfaatkan semua fitur yang ada di Malay Concordance Project (<http://mcp.anu.edu.au/>),

aplikasi konkordansi daring untuk bahasa Melayu Klasik, dengan menggunakan informasi morfologi yang ada di MALINDO Morph. Oleh karena itu, MALINDO Conc akan dikembangkan lebih lanjut dengan memasukkan berbagai dialek bahasa Melayu, seperti bahasa Indonesia, bahasa Malaysia, bahasa Melayu Brunei, dan bahasa Melayu di Singapura. Aplikasi konkordansi berlisensi sumber terbuka yang tersedia saat ini, Korpus DBP (<http://sbmb.dbp.gov.my/korpusdbp/SelectUserCat.aspx>) dan SEALang Library Corpora (<http://sealang.net/malay/corpus.htm>, <http://sealang.net/indonesia/corpus.htm>), berisi berbagai dialek bahasa Melayu dari berbagai daerah. MALINDO Conc berisi fitur pencarian secara morfologis sehingga pengguna dapat mencari kata-kata atau bentuk-bentuk turunan yang ada di korpus secara morfologis. Sebagai contoh, pengguna dapat (1) mencari bentuk-bentuk infleksi kata *pikir* dan *pikirkan* dengan mengatur kata kunci sebagai berikut: akar kata sama dengan *pikir*, prefiks sama dengan *meN-* atau *di-*, atau tanpa prefiks, sufiks sama dengan *-kan* atau tanpa sufiks, dan tanpa konfiks; (2) mencari semua kata kerja berimbuhan *ber-...-kan* dengan kata kunci sebagai berikut: akar kata tanpa spesifikasi dan konfiks sama dengan *ber-...-kan*; (3) mencari semua kata kerja berpola *meN-X-X* dan *X-meN-X* seperti *memukul-mukul* dan *pukul-memukul*, dengan kata kunci sebagai berikut: akar kata tanpa spesifikasi, prefiks sama dengan *meN-*, dan reduplikasi sama dengan penuh (lihat Gambar 2); (4) mencari kolokasi kata *ingin* seperti *ingin* + kata kerja berprefiks *di-*, *ingin* + kata (misalnya *untuk*) + kata kerja berprefiks *di-* dengan pengaturan kolokasi sebagai berikut: bentuk permukaan sama dengan *ingin*, cari kolokasi di antara kanan 1 dan kanan 2, akar kata tanpa spesifikasi, dan prefiks sama dengan *di-* (lihat Gambar 3). MALINDO Conc berisi fitur pencarian yang lebih detail, dibandingkan Korpus DBP dan SEALang Library Corpora yang hanya berisi pencarian kata kunci sederhana, sehingga dapat digunakan untuk penelitian morfosintaks. Hasil pencarian MALINDO Conc dapat diunduh pengguna.



Gambar 2. Tangkapan Layar Hasil Pencarian Bentuk Turunan di MALINDO Conc



Gambar 3. Tangkapan Layar Hasil Pencarian Kolokasi di MALINDO Conc

Karena MALINDO Conc menggunakan data morfologi MALINDO Morph, belum ada pilihan pencarian imbuhan dari bahasa asing, enklitik *-nya* dan sufiks *-nya*, dan belum ada anotasi kelas kata sehingga pencarian struktur sintaks secara mendetail belum dapat dilakukan. Meskipun demikian, MALINDO Conc merupakan aplikasi konkordansi bahasa Indonesia dan bahasa Malaysia daring pertama yang memiliki fitur pencarian dan kolokasi morfologis yang memudahkan ahli ilmu bahasa (linguis) mencari data bahasa, menganalisis fenomena bahasa, dan menguji hipotesis analisis linguistik, serta membantu pekamus memutakhirkan kamus dengan data terbaru dari korpus. Pekamus dapat mencari kata-kata baru yang digunakan dalam masyarakat beserta contohnya. Penulis mengundang para ahli ilmu bahasa dan pekamus menggunakan MALINDO Conc untuk penelitian linguistik dan leksikografi. Hasil penelitian tersebut dapat diterbitkan di edisi spesial Jurnal NUSA terbitan Universitas Kajian Asing Tokyo. Hingga makalah ini ditulis (23 Juli 2018), MALINDO Conc dapat diakses di <https://malindoconc.lagoinst.info/concordance/ind/> dengan peladen atau *server* milik Institut Bahasa Lago di Jepang. Oleh karena aplikasi konkordansi ini ditujukan untuk berbagai dialek bahasa Melayu, alangkah baiknya jika MABBIM dapat berkontribusi dalam pengembangan korpus dan menyediakan peladennya.

IV. SIMPULAN

Lisensi sumber terbuka sangat diperlukan untuk pengembangan berbagai aplikasi pemrosesan bahasa alami. Peneliti bahasa dan leksikograf sering terhambat dengan kurangnya berbagai sumber linguistik bermutu dan berlisensi sumber terbuka. Kamus morfologi MALINDO Morph dan aplikasi konkordansi daring MALINDO Conc menunjukkan pentingnya kerja sama pekamus, ahli ilmu bahasa, dan pakar teknologi informasi dalam penyusunan pangkalan data, aplikasi konkordansi daring, dan kamus morfologi berlisensi sumber terbuka yang selanjutnya dapat dimanfaatkan oleh pekamus dalam pemutakhiran kamus, ahli ilmu bahasa

dalam penelitian bahasa, dan masyarakat penutur bahasa pada umumnya. Besar harapan penulis bahwa MALINDO Morph dan MALINDO Conc dapat memajukan penelitian linguistik dan leksikografi bahasa Indonesia dan bahasa Malaysia. Penulis juga menaruh harapan pada Badan Bahasa, DBP Malaysia, DBP Brunei, dan MABBIM untuk memfasilitasi pengembangan aplikasi-aplikasi pemrosesan bahasa alami berbagai dialek bahasa Melayu.

Daftar Pustaka

- Badan Bahasa. *Kamus Besar Bahasa Indonesia*, Edisi kelima. Jakarta: Badan Pengembangan dan Pembinaan Bahasa, 2016.
- Dewan Bahasa dan Pustaka. *Kamus Dewan*, Edisi keempat. Kuala Lumpur: Dewan Bahasa dan Pustaka, 2005.
- Goldhahn, Dirk, Thomas Eckart, dan Uwe Quasthoff. "Building large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages." Dalam *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 759–765. Istanbul, Turki: European Languages Resources Association, 2012.
- Moeljadi, David, Ian Kamajaya, dan Dora Amalia. "Building the Kamus Besar Bahasa Indonesia (KBBI) Database and Its Applications." Dalam Hai Xu (Ed.), *Proceedings of The 11th International Conference of the Asian Association for Lexicography*, 64–80. Guangzhou, Cina: The Asian Association for Lexicography, 2017.
- Nomoto, Hiroki, Hannah Choi, David Moeljadi, dan Francis Bond. MALINDO Morph: Morphological dictionary and analyser for Malay/Indonesian. Dalam Kiyooki Shirai (Ed.), *Proceedings of the LREC 2018 Workshop: The 13th Workshop on Asian Language Resources*, 36–43. Miyazaki, Jepang: Languages Resources and Evaluation Conference, 2018.
- Nomoto, Hiroki, Shiro Akasegawa, dan Asako Shiohara. "Building an Open Concordancer for Malay/Indonesian." Makalah dipresentasikan dalam *The Twenty-Second International Symposium On Malay/Indonesian Linguistics (ISMIL 22)*, 11–12 Mei 2018, Los Angeles, Amerika Serikat.

**KOMPETENSI DAN KESIAPAN LEKSIKOGRAFI
DALAM ERA DIGITAL**
(*COMPETENCE AND READINESS OF LEXICOGRAPHY IN DIGITAL ERA*)

Rosida Tiurma Manurung
Universitas Kristen Maranatha
rosida.tm@psy.maranatha.edu

Abstrak

Era digital merupakan fenomena penyebab pergeseran aktivitas masyarakat dari dunia nyata ke dunia maya. Fenomena ini berkembang pada perubahan pola perkamusan. Kemunculan kamus daring adalah salah satu dampak era digital yang paling diperlukan sebagai perangkat dan fasilitator untuk berbagai aktivitas, baik bidang pendidikan, sosial, ekonomi, maupun politik.

Era digital bisa disebut sebagai era disrupsi, yaitu era yang menganut "harus berkompetisi, untuk bisa menang". Oleh sebab itu, kita harus bisa segera beradaptasi dan mengenali bagaimana keadaan era digital yang penuh dengan perubahan. Pola kehidupan tidak sekadar berubah, melainkan langsung bergeser atau menggantikan yang sudah berdiri sebelumnya dalam waktu yang cepat. Demikian juga di bidang leksikografi, diperlukan kosakata baru untuk menggantikan kosakata usang yang tidak relevan dalam era digital ini.

Kata kunci : kompetensi leksikografi, era digital, kebutuhan akan kosakata baru

Abstract

Digital era is a phenomenon that causes people to shift the activities that were originally done in the real world, to the virtual world. This phenomenon develops in the changing pattern of cultivation. The emergence of an online dictionary is one of the most indispensable impacts of the digital age as a tool and facilitator for various activities, whether educational, social, economic, or political.

The digital era can be called an era of disruption, the era that adheres to the idea of "must compete, to be able to win". Therefore, we must be able to adapt immediately and to recognize how the state of the digital era is full of changes because the pattern of life is not just changing, but directly shifted or replaced the already standing earlier fastly. Likewise in the field of lexicography, new words are needed to replace the irrelevant obsolete words in this digital age.

Keywords: *lexicographic competence, digital age, the need for new vocabulary*

I. PENDAHULUAN

1.1. Latar Belakang Masalah

Sejalan dengan era digital yang perkembangannya bergerak cepat dalam hitungan menit, terjadi perubahan dari teknologi mekanik dan elektronik analogteknologi digital yang telah terjadi sejak tahun 1980 dan berlanjut sampai hari ini. Era digital telah mengubah cara pandang seseorang dalam menjalani kehidupan yang sangat canggih saat ini. Era digital dengan teknologi canggihnya telah membuat perubahan besar kepada seluruh dunia, dari mulai membantu mempermudah segala urusan sampai membuat masalah karena tidak bisa menggunakan fasilitas digital yang semakin canggih dengan baik dan benar. Pada era digital ini, kita memasuki era disrupsi. *Disruption* adalah istilah yang dipopulerkan oleh Clayton Christensen sebagai kelanjutan dari tradisi berpikir “harus berkompetisi untuk bisa menang (*for you to win, you’ve got to make some body lose*)”. Agar sukses dan mampu bersaing dalam era digital, kita dituntut untuk menciptakan pasar baru tanpa perlu bersaing dengan yang sudah ada. Dengan demikian, kita akan sukses besar tanpa harus mengalahkan siapa pun. Implikasi era digital dan disrupsi membuat satu per satu kosakata tumbang karena dianggap usang dan tidak sejalan dengan perkembangan teknologi. Jadi, agar kosakata bahasa Indonesia tetap eksis perlu ada upaya penambahan kosakata baru untuk menggantikan kata usang. Demikianlah, leksikografi merupakan ilmu yang di dalamnya mempelajari prinsip-prinsip untuk mempersiapkan aturan-aturan ejaan, tata bahasa, dan kamus sebagai panduan penulis dan penutur dalam sebuah masyarakat bahasa yang majemuk. Berdasarkan hal tersebut, dapat diketahui bahwa standarisasi peristilahan hanyalah salah satu bagian kecil dari kompetensi leksikografi.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, berikut dirumuskan masalah yang akan dianalisis dan dijawab dalam penelitian ini.

- 1) Upaya apa yang ditempuh leksikografi agar siap berperan dalam era digital yang disebut sebagai era disrupsi?
- 2) Kompetensi apa saja yang dimiliki leksikografi untuk mengintegrasikan perkembangan ilmu dan teknologi dalam penambahan kosakata baru?

1.3. Tujuan Pembahasan

Sesuai dengan rumusan masalah di atas, garis-garis besar hasil pokok yang ingin dicapai setelah permasalahan dianalisis dan dijawab dalam penelitian ini, yaitu sebagai berikut.

- 1) Mendeskripsikan upaya yang ditempuh leksikografi agar siap berperan dalam era digital yang disebut sebagai era disrupsi.
- 2) Kompetensi dan kesiapan leksikografi untuk mengintegrasikan perkembangan ilmu dan teknologi dalam penambahan kosakata baru.

1.4. Tinjauan Pustaka

Untuk menjawab setiap permasalahan, digunakan ruang lingkup kajian dan tinjauan pustaka yang meliputi teori leksikografi yang dikemukakan oleh Zgusta (1971), Svensen (2009), dan Kridalaksana (2010). Zgusta (1971: 5) mengatakan bahwa tidak dapat disangsikan lagi bahwa leksikografi merupakan suatu bidang yang kurang akrab dalam kegiatan ilmu bahasa. Di samping itu, ada beberapa alasan yang dapat dikemukakan mengapa leksikografi merupakan suatu bidang yang sangat sukar dipahami dalam ilmu bahasa. Menurut Svensen (2009:3), leksikografi adalah cabang leksikologi, dan leksikologi sama dengan metasikologafi. Leksikografi merupakan aktivitas yang terdiri dari penelitian, pengumpulan, seleksi, analisis dan penjelasan satuan leksikal (kata-kata, elemen kata, dan gabungan kata) yang dimiliki oleh satu atau lebih bahasa di dalam kamus. Lebih jauh lagi Svensen (2009: 3) mengatakan bahwa leksikografi merupakan bidang terapan linguistik, sedangkan leksikologi adalah ilmu tentang studi kosakata, stuktur, dan karakteristik lainnya.

Menurut Kridalaksana (2010) Leksikografi adalah bidang linguistik terapan yang mencakup metode dan teknik penyusunan kamus. Kridalaksana (2010: 53) berpendapat bahwa karya leksikografis yang bermutu semata-mata ditentukan oleh penelitian leksikologis yang bermutu. Lebih jauh lagi Kridalaksana (2010: 53) mengemukakan pembinaan perkamusan merupakan proses yang panjang. Setiap tahap dalam proses itu merupakan akumulasi dari penelitian dan analisis bahasa serta kegunaan praktis kamus hasil proses sebelumnya.

II. KERANGKA TEORI DAN METODE PENELITIAN

Leksikologi (Yunani: *lexiko-*, "leksikon") adalah cabang ilmu linguistik yang mempelajari kata, sifat dan makna, unsur, hubungan antarkata (semantis), kelompok kata, serta keseluruhan leksikon. Ilmu ini terkait erat dengan leksikografi yang juga mempelajari kata, terutama dalam kaitannya dengan penyusunan kamus. Secara sederhana, leksikografi disebut sebagai penerapan praktis dari leksikologi.

Leksikologi adalah ilmu mengenai leksikon yang satuannya disebut leksem. Leksikografi dapat dimaknai sebagai usaha untuk memperbaiki komunikasi bahasa dengan menciptakan subsistem-subsistem baru atau menyempurnakan sesuai dengan kebutuhan masyarakat yang memakainya. Standardisasi peristilahan dalam bidang-bidang kehidupan tertentu adalah contoh leksikografi (Kridalaksana 2010).

Leksikografi merupakan sebuah kegiatan yang dilakukan oleh pemerintah atau badan yang ditunjuk oleh pemerintah, dalam hal ini Badan Bahasa. Badan ini memiliki kewenangan antara lain, untuk menentukan variasi bahasa mana yang berhak untuk dikukuhkan sebagai bahasa nasional. Salah satu tugas lain yang juga menjadi tanggung jawab Badan Bahasa adalah mengatur dan menyempurnakan sistem ejaan mana yang dianggap paling komprehensif untuk mengantisipasi masuknya kosakata asing ke dalam khazanah bahasa Indonesia serta menentukan kosakata baru mana yang berhak memasuki khazanah bahasa nasional kita.

Metode penelitian yang digunakan ialah metode analisis deskriptif yang menganalisis bahasa yang menggunakan korpus. Korpus atau korpora sangat bermanfaat dalam dunia

kebahasaan, baik dalam pengajaran ataupun penelitian. Beberapa bidang yang mengandalkan korpus antara lain adalah leksikografi (penyusunan kamus), tata bahasa, sosiolinguistik, terjemahan, pengajaran, stilistika (ilmu gaya bahasa), dialektologi, dan linguistik historis. Dalam era yang serba digital ini kita, penelitian leksikografi semakin dimudahkan dalam melakukan segala sesuatu termasuk pemanfaatan korpus untuk pengembangan ilmu kebahasaan ini. Mungkin bagi beberapa kalangan, teknologi semacam ini masih sangat asing. Akan tetapi, sesuai dengan tujuan utama dari pengembangan ilmu pengetahuan, yaitu demi kemaslahatan umat manusia, sudah seharusnya kita memanfaatkan teknologi yang ada untuk masa depan yang lebih baik.

III. PENDAHULUAN

3.1. Kesiapan Leksikografi dalam Era Digital

Leksikografi dapat dijadikan sebagai alat untuk menghasilkan inovasi dan wajah baru bahasa Indonesia yang modern dan siap bersaing dalam era digital. Leksikografi adalah ilmu yang diterapkan dan digunakan dalam pengembangan istilah dan kosakata. Dengan leksikografi, penyerapan unsur bahasa asing dan penentuan padanannya dalam bahasa Indonesia dapat mempertajam daya ungkap penutur dan memungkinkan penutur menyatakan isi hatinya dengan tepat dan cermat.

Pada awal pemakaiannya kata-kata baru akan terasa asing. Akan tetapi, dalam jangka panjang usaha ini akan sangat menunjang pengembangan ilmu pengetahuan dan teknologi karena memberi sarana untuk meneruskan gagasan atau ilmu pengetahuan kepada mereka yang belum mengenal bahasa asing secukupnya. Usaha leksikografi di bidang keilmuan bertujuan agar setiap makna istilah, baik yang berupa kata maupun yang berupa ungkapan dapat dijabarkan dari strukturnya. Ketaatan diperlukan untuk menjadikan pemakaian yang seragam dalam hal sarana tata bahasa sehingga penafsiran akan seragam pula.

Leksikografi telah mampu dan berhasil menciptakan istilah dan kata-kata baru yang sifatnya menambah kosakata dan menambah medan makna yang dapat diungkapkan dalam Bahasa Indonesia, sehingga suatu pengalaman atau gagasan dapat diungkapkan dengan simbol kata yang tepat. Kata-kata baru tersebut banyak yang sudah diterima baik di kalangan akademik maupun masyarakat umum.

Sebagian masyarakat memiliki kekhawatiran bahwa jika bahasa Indonesia berkembang pesat dan semua literatur dan bahan bacaan yang sudah ditulis dalam bahasa Indonesia menyebabkan kemampuan pelajar dan mahasiswa berbahasa asing menjadi berkurang. Memajukan bahasa Indonesia pada era global tidak berarti mematikan bahasa asing. Agar tidak tertinggal dengan negara lain, sudah saatnya wawasan pemelajar dan pemakai bahasa Indonesia terhadap pengetahuan dan teknologi diperluas sejak dini tanpa harus menunggu fasih berbahasa asing. Cakrawala pengetahuan kita dapat dikembangkan dan luas dengan cara mentransfer perkembangan ilmu dan teknologi dengan mengalihkannya dalam kosakata dan istilah bahasa Indonesia. Akan tetapi, tidak dapat disangkal bahwa buku-buku keilmuan berbahasa Indonesia kurang memadai sehingga pemelajar (kalangan bisnis, akademik, dan ilmiah) terpaksa harus menguasai bahasa asing (khususnya bahasa Inggris).

Leksikografi dapat memberikan dampak yang signifikan terhadap penyebaran dan pemahaman ilmu pengetahuan dan teknologi. Pada hakikatnya, dalam era disrupsi, bahasa Indonesia sesungguhnya sedang bersaing ketat dengan bahasa asing. Seyogianya sikap apriori terhadap pengembangan bahasa merupakan salah satu faktor yang menghambat pengembangan bahasa itu sendiri. Bahasa Indonesia masih dipandang hanya sebagai bahasa politis atau sebagai simbol persatuan dan belum dikembangkan menjadi sarana komunikasi untuk pengungkapan informasi yang kompleks dalam bidang keilmuan.

Leksikografi memiliki potensi yang besar untuk dijadikan sarana pengembangan bahasa dalam konteks keilmuan yang canggih dan futuristik. Dengan demikian, pemelajar dan masyarakat luas dapat menikmati karya-karya sastra, ilmu pengetahuan, dan teknologi yang tinggi tanpa harus menunggu kefasihan berbahasa asing. Justru kompetensi dan keterampilan berbahasa Indonesia akan sangat membantu proses dan pemahaman dalam mempelajari bahasa asing itu sendiri.

Pembentukan istilah yang konsisten dan berkaidah akan memudahkan pengertian makna atau gagasan yang terkandung dalam simbol berupa rangkaian kata. Pembentukan istilah yang cermat ini akan sangat terasa manfaatnya dalam kajian keilmuan yang mensyaratkan kecermatan ekspresi. Pembentukan istilah semata-mata bukan melulu logika dan kaidah bahasa, tetapi dengan perasaan atau pengalaman saja. Cara seperti ini dapat saja dilakukan, tetapi hasilnya sering tidak mengena atau bahkan menyesatkan. Pengembangan pengetahuan dan bahasa sering menjadi terhambat karena orang mempertahankan apa yang sudah kaprah, tetapi secara kaidah dan makna bahasa keliru sehingga membuat penangkapan dan pemahaman suatu konsep dalam pengetahuan tertentu juga ikut keliru (walaupun tidak disadari).

3.2. Kompetensi Leksikografi untuk Memunculkan Kosakata Baru

3.2.1. Kompetensi Pekamus

Seorang pekamus (ahli perkamusan) haruslah memiliki wawasan perkamusan yang luas. Pekamus tidak saja memusatkan perhatian pada kata, tetapi juga pada masalah lebih luas. Teori leksikografi bersangkutan paut dengan semua bidang yang mempelajari sistem kata seperti semantik, leksikologi, tata bahasa, dan stilistika. Untuk menetapkan makna yang tepat, pekamus acap kali menghadapi kesulitan. Makna kata akan didukung oleh bentuk kata, tanpa bentuk maka tidak ada makna. Kesatuan bentuk dan maknalah yang akan membuat bahasa berfungsi sebagai alat penghubung antarmanusia. Kompetensi pekamus ialah kemampuan analisis secara ilmiah, tetapi hasil karyanya harus disajikan kepada masyarakat dalam bentuk praktis. Pertimbangan komersial dan ilmiah merupakan hal yang dihadapi oleh pekamus dalam menjalankan tugasnya. Berdasarkan data di lapangan, terdapat kamus-kamus yang diterbitkan hanya untuk kepentingan komersial saja. Sebaliknya, banyak pula kamus yang betul-betul ilmiah, tetapi patut disayangkan pemakaiannya hanya dimanfaatkan oleh kalangan ilmuwan yang sangat terbatas.

3.2.2. Skema Alir Pemunculan Kosakata Baru

Berikut ini akan dipaparkan dan dieksplanasikan skema alir pemunculan kosakata. Tahap pertama kita mencari apakah ada kemungkinan tersedia kata dalam bahasa Indonesia yang lazim dipakai. Kalau ada, maka kata tersebut kita ambil dan hasilnya adalah istilah baru 1. Ini belum sebuah keputusan, tetapi baru alternatif 1. Tahap kedua, kita mencari apakah ada kata dalam bahasa Indonesia yang sudah tidak lazim dipakai? Kalau ada, maka kita ambil, sehingga menghasilkan istilah baru 2 yang juga masih merupakan alternatif 2.

Tahap ketiga, kita mencari apakah ada kata dalam bahasa serumpun yang lazim dipakai, yaitu bahasa-bahasa daerah di Indonesia dan bahasa di luar Indonesia yang masuk dalam rumpun bahasa Melayu Polinesia. Dalam kenyataannya, kita mengutamakan pengambilan istilah dari bahasa-bahasa daerah kita di Indonesia. Kalau dengan tahap ini kita menemukan istilah baru, maka itu menjadi alternatif ketiga dan berupa istilah baru 3. Tahap keempat adalah mencari kata yang sudah tidak lazim dipakai di dalam bahasa serumpun atau bahasa daerah yang dapat diambil sebagai istilah. Hasilnya akan berupa istilah baru 4 yang juga masih bersifat alternatif 4.

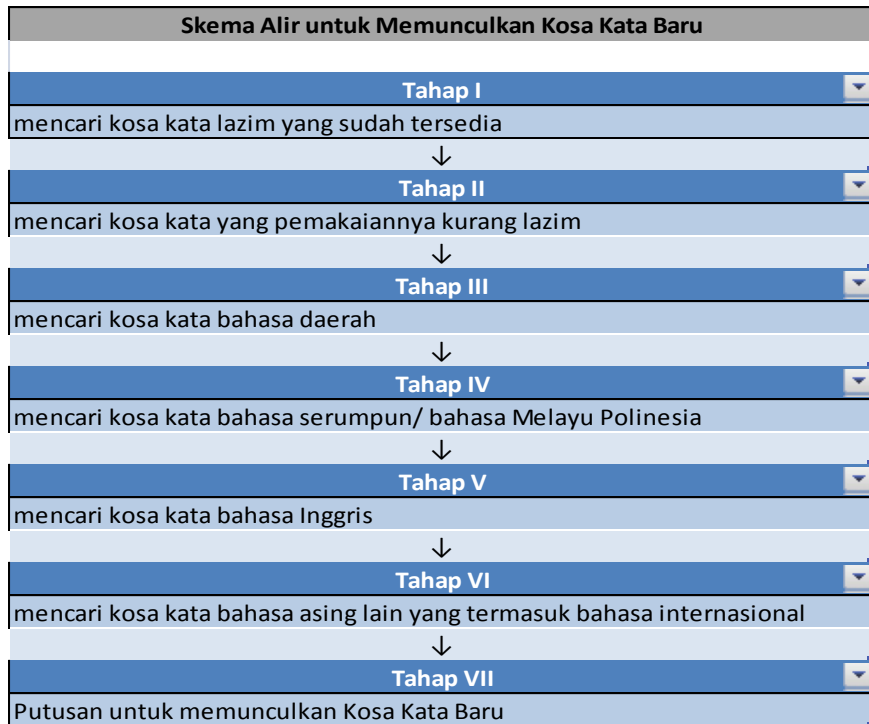
Walaupun kita sudah sampai pada tahap keempat, keputusan masih belum dapat diambil. Kita telah menghasilkan empat istilah baru yang semuanya masih bersifat alternatif. Ada syarat-syarat yang diutamakan dalam tahap pertama sampai dengan keempat. Syarat-syarat tersebut yaitu ungkapan harus singkat, makna tepat atau tidak menyimpang, dan memiliki ungkapan dengan nilai rasa (konotasi) baik serta sedap didengar.

Setelah tahap pertama sampai dengan keempat kita laksanakan, tanpa mengambil keputusan kita melangkah ke tahap lima, yaitu mengambil kata dari bahasa Inggris. Jika istilah yang akan diindonesiakan ialah kata dari bahasa Hongaria, misalnya, bukan kata Hongaria itu yang kita acu, melainkan kita melihat padanannya dalam bahasa Inggris. Apakah kata tersebut ada atau tidak. Jika dalam bahasa Inggris hasilnya ada, kata tersebut menjadi istilah baru 5 dan menjadi alternatif 5 pula. Tahap keenam dalam penelitian ini adalah pengambilan kata dari bahasa asing lain yang diakui secara internasional oleh PBB yaitu bahasa Inggris, bahasa Prancis, bahasa Spanyol, bahasa Cina, bahasa Rusia, atau sekurang-kurangnya digunakan oleh lima negara misalnya, bahasa Arab dan bahasa Jerman. Hasil dari tahap ini dilabeli sebagai istilah baru 6 dan digunakan sebagai alternatif 6. Pada tahap kelima dan keenam tersebut harus dilandasi kriteria, yaitu ungkapan asing lebih cocok, lebih singkat, memudahkan pengalihan antarbangsa, dan memudahkan kesepakatan.

Sesudah tahap pertama sampai dengan enam kita ambil, sampailah pada tahap ketujuh, enam istilah baru yang masih merupakan alternatif tersebut diputuskan dan dipilih satu yang terbaik dan dijadikan istilah baru. Dengan demikian muncul kosakata atau istilah baru.

Berdasarkan uraian di atas dapat kita lihat skema alir pemunculan atau penambahan kosakata baru tampak seperti dalam skema di bawah ini.

Bagan 1 Skema Alir Pemunculan Kosakata Baru



3.2.3. Perluasan Kosakata sebagai Pemerdayaan

Pesatnya kemajuan di segala bidang ilmu pengetahuan dapat memperkaya/memperluas kosakata. Sumber perluasan kosakata itu pada garis besarnya ada dua macam, yaitu (1) sumber dalam dan (2) sumber luar.

3.2.3.1 Sumber Dalam

Sumber dalam adalah swadaya bahasa Indonesia sendiri. Swadaya itu dapat berwujud:

- (1) pengaktifan kata-kata lama;
- (2) pembentukan baru;
- (3) penciptaan baru; dan
- (4) pengakroniman.

Kata-kata lama yang dapat diaktifkan kembali dengan arti yang (1) sama atau (2) baru, contohnya, mengelola, pengelola (manajer), pengelolaan (manajemen), pemerhati (orang yang memperhatikan pidato/ceramah/khotbah dalam siaran TV), abdi (abdi negara = pegawai), bahari (wawasan bahari = wawasan laut) pakar (para pakar = para ahli), dan pemantauan (*monitoring*), dan kemas (peti kemas = *container*). Contoh yang diaktifkan dengan makna baru ialah senjang (Minangkabau): genjang; tidak seimbang, tidak sama benar; *arti baru*: jurang pemisah (*gap*); meliputi (meliputi): menutupi, menyelubungi, melingkupi; *arti baru*: padanan kata meng-*cover*; acuan: cetakan (klise, peluru, dan sebagainya); *arti baru*: padanan kata *referensi* (buku acuan): selia, menyelia = melihat, mengawasi; *arti baru*: penyelia = *supervisor* menyelia = *mensupervisi*; sigi, menyigi: menyuluh dengan sigi (suluh

atau obor yang dibuat dari kayu atau bilah-bilah bambu); *arti baru*: padanan kata menyurvei (*to survey*)

Bentukan baru dapat dibentuk dari kata-kata yang sudah ada dengan proses *pengimbuhan*.

Bentuk dasar kata asal

masuk
keluar
unggul
kaji
temu
tatar
satu
langgan

Bentukan baru

masukan (*input*)
keluaran (*output*)
unggulan (*seeded*)
kajian (*studi*)
temuan
tataran (*level*)
satuan (*unit*)
pelanggan (*customer*)

Bentuk baru seperti contoh di atas dibentuk menurut pola pembentukan kata yang telah ada dalam bahasa Indonesia. Manfaat bentukan baru itu adalah menampung pengertian (konsep) baru.

Bentuk dasar kata turunan dan frasa

pemimpin
pendidik
pengurus
seragam
sejajar

Bentukan baru

kepemimpinan
kependidikan
kepengurusan
keseragaman
kesejajaran

Kata-kata ciptaan baru digunakan sebagai pengganti pengertian-pengertian yang belum ada.

Contoh:

sinambung; kesinambungan (*continuity*/kontinuitas)
rujukan (*reference*); rinujuk (*referent*) = yang dirujuk
nirguna (*useless*) = tanpa guna
taninsan (*nonhuman*) = bukan insan
narasumber (*resource person*)

Akronim ialah singkatan yang berupa gabungan huruf awal/gabungan suku kata/gabungan kombinasi huruf dan suku kata dari deret kata yang ditulis dan dilafalkan seperti kata yang wajar.

Contoh:

japri (jalur pribadi)
daring (dalam jaringan)
luring (luar jaringan)

Pembentukan akronim seperti contoh-contoh di atas tampak semakin meluas dan digemari orang. Hal tersebut merupakan salah satu ciri yang paling menonjol dari perkembangan bahasa Indonesia.

3.2.3.2. Sumber Luar

Sumber luar peluasan kosakata bahasa Indonesia adalah kata-kata dari (1) bahasa serumpun (bahasa-bahasa daerah yang terdapat di Indonesia/rumpun bahasa Austronesia) dan (2) bahasa asing (Sanskerta, Arab, Portugis, Belanda, Inggris, dan sebagainya). Kata-kata yang diambil dari sumber luar itu disebut kata-kata pungutan (pinjaman/serapan). Misalnya, ajek, baku, ejawantah (bahasa Jawa), anjangsana, anjangkarya, sewajarnya (bahasa Sunda), acuh, heboh, lambung (Minangkabau), kosakata dari sumber bahasa asing juga melimpah sejalan dengan perkembangan ilmu dan teknologi.

IV. SIMPULAN

Demikianlah, leksikografi merupakan ilmu yang di dalamnya mempelajari prinsip-prinsip untuk mempersiapkan aturan-aturan ejaan, tata bahasa, dan kamus sebagai panduan penulis dan penutur dalam sebuah masyarakat bahasa yang majemuk. Leksikografi memiliki potensi yang besar dalam era digital untuk dijadikan sarana pengembangan bahasa dalam konteks keilmuan yang canggih dan futuristik. Dengan demikian, pemelajar dan masyarakat luas dapat menikmati karya-karya sastra, ilmu pengetahuan, dan teknologi yang tinggi tanpa harus menunggu kefasihan berbahasa asing. Justru kompetensi dan keterampilan berbahasa Indonesia akan sangat membantu proses dan pemahaman dalam mempelajari istilah dan kosakata yang diserap dari bahasa asing.

Daftar Pustaka

- Alwi, Hasan. 2005. *Kamus Besar Bahasa Indonesia*. Jakarta: Balai Pustaka.
- Badan Bahasa. 2009. *Kamus Besar Bahasa Indonesia*. Jakarta: Balai Pustaka.
- Kridalaksana, Harimurti. (2010). Sendi-sendi ilmiah bagi pembinaan bahasa. Jakarta: Laboratorium Leksikologi dan Leksikografi.
- Moeliono, Anton M. 1985. *Pengembangan dan Pembinaan Bahasa: Ancangan Alternatif di dalam Perencanaan Bahasa*. Jakarta: Djambatan.
- , 1997. "Pembakuan Istilah dalam Pemodernan Bahasa" dalam *Peranan Bahasa Kebangsaan dalam Pengembangan Iptek*. Jakarta: Departemen Pendidikan Nasional.
- Panitia Pengembangan Bahasa Indonesia. 2004. Edisi ketiga. *Pedoman Umum Pembentukan Istilah*. Jakarta: Departemen Pendidikan Nasional.
- Poerwadarminta, W. J. S. 1952. *Kamus Umum Bahasa Indonesia*. Djakarta: Balai Pustaka.
- Puryadi, D. 2003. "Pemanfaatan Kosakata Daerah dalam *Kamus Besar Bahasa Indonesia*" dalam *Rampak Serantau* 10:143-152.
- Safioedin, A. 1977. *Kamus Bahasa Madura-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Steinhauer, H. 1994. "The Indonesian Language Situation and Linguistic: Prospects and Possibilities". *Bijdragen tot de Taal-, Land-, en Volkenkunde* 150:755-784.
- , 2003. "Perkembangan Bahasa Indonesia dari Aspek Teori Linguistik". Makalah dalam *Kongres Bahasa Indonesia VIII*, Jakarta Oktober 2003.

Sugiarto dkk. 1999. *Kamus Indonesia – Daerah (Jawa, Bali, Sunda, Madura)*. Edisi ke-2.
Jakarta: Gramedia Pustaka Utama.

Svensen, Bo. (2009). *The handbook of lexicography, the theory and practice of dictionary-making*. Cambridge: Cambridge University Press.

Zgusta, Ladislav. (1991). *Manual of lexicography*. The Hague – Paris: Mouton.

KAMUS DIGITAL BAHASA INDONESIA MASA DEPAN MENGACU PADA LIMA KAMUS DIGITAL BAHASA INGGRIS TERNAMA

Almira F. Artha
Universitas Airlangga (UA)
airafidel@gmail.com

Rosita Masfirotul Uyun
Universitas Airlangga (UA)
uyunarose92@gmail.com

Abstrak

Pada era digital ini pembuatan kamus diharapkan dapat membantu masyarakat mendapatkan informasi secara cepat dan murah. Dengan demikian, dibutuhkan kamus dengan konten baik, nirkertas, sangkil, dan mangkus yang dapat memberikan informasi memadai. Penelitian ini berfokus pada perbandingan struktur internal lima kamus digital bahasa Inggris, yaitu *Cambridge Dictionary (CD)*, *Oxford Learner's Dictionaries (OLD)*, *Longman Dictionary of Contemporary English (LDOCE)*, *Collins English Dictionary (CED)*, dan *Macmillan English Dictionary (MED)*. Dengan melihat struktur internal masing-masing berdasarkan lima lema berbahasa Inggris, yaitu *batik*, *sarong*, *bamboo*, *durian*, dan *orangutan*, skala prioritas struktur internal dapat dirumuskan. Hasil penelitian ini menunjukkan bahwa kamus digital bahasa Indonesia masa depan selayaknya memiliki tiga skala prioritas struktur internal. Skala prioritas utama terdiri atas lema kepala, kelas kata, audio, lafal, definisi, dan contoh kalimat. Skala prioritas kedua terdiri atas kata turunan dan gabungan kata, asal kata, dan gambar, serta skala prioritas ketiga yang meliputi sinonim dan tren. Penelitian ini diharapkan mampu memberikan informasi dan rekomendasi bagi pekamus, pakar teknologi informasi, pembuat aplikasi, dan peminat bahasa mengenai pengembangan kamus digital bahasa Indonesia di masa depan.

kata kunci: struktur internal kamus, kamus digital, leksikografi perbandingan

I. PENDAHULUAN

Kemajuan teknologi kini berkembang semakin pesat dan memberi dampak pada semua bidang, termasuk leksikografi. Salah satu dampak terhadap dunia leksikografi adalah pembuatan dan penggunaan kamus yang terdigitalisasi. Meskipun telah memanfaatkan teknologi, pembuatan kamus tetap melalui beberapa tahap, seperti pengumpulan data korpus, analisis dan penetapan definisi, dan penentuan tampilan kamus (Atkins dan Rundell, 2008).

Berkaitan dengan penyajian kamus digital, penelitian ini berfokus pada penyajian kamus digital bahasa Indonesia masa depan dengan sasaran pengguna awam. Di Indonesia, saat ini telah ada Kamus Besar Bahasa Indonesia (KBBI) yang dapat diakses secara daring. Melalui KBBI daring, pengguna mendapatkan informasi dari sebuah kata berdasarkan definisi, contoh kata, kata turunan, dan gabungan kata. Struktur internal kamus yang demikian tentunya memberikan keterangan yang bermanfaat kepada pengguna. Dengan menggabungkan dan membandingkan struktur internal kamus dalam *Cambridge Dictionary (CD)*, *Oxford Learner's Dictionaries (OLD)*, *Longman Dictionary of Contemporary English*

(LDOCE), Collins English Dictionary (CED), dan Macmillan English Dictionary (MED), hasil penelitian ini diharapkan dapat merumuskan struktur internal yang bisa menjadi acuan dalam pengembangan KBBI daring atau kamus digital masa depan lainnya.

Untuk merumuskan struktur internal entri yang sangkil dan mangkus dengan informasi yang memadai, penelitian ini mengambil lima kata bahasa Indonesia yang diserap ke dalam bahasa Inggris, yaitu bambu (*bamboo*), batik (*batik*), durian (*durian*), orangutan (*orangutan*), dan sarung (*sarong*). Kelima kata tersebut diambil dari CD, OLD, LDOCE, CED, dan MED.

II. KERANGKA TEORI DAN METODE PENELITIAN

2.1 Kerangka Teori

2.1.1 Struktur internal kamus

Teori leksikografi mengenai struktur internal kamus sangat bervariasi dan beragam. Tidak ada teori definit yang menyebutkan jumlah absolut struktur internal yang harus dimiliki oleh sebuah kamus. Struktur internal kamus tampaknya bervariasi berdasarkan jenis, tujuan, dan pengguna kamus. Teori pertama diajukan oleh Hartmann dan James (1998) yang menyebutkan sesungguhnya terdapat tiga struktur internal kamus, yakni kata kunci, definisi, dan kategori informasi lainnya (pelafalan, fungsi sintaksis, dan label). Struktur internal kamus ini sepatutnya mengisyaratkan fungsi kamus yang sangat normatif sebagai deskriptif yang lengkap dengan sematan fungsi sintaksis dalam sebuah entri.

Lebih kompleks lagi, Hanks (2006) menyebutkan, ada delapan struktur internal dalam kamus. Struktur pertama adalah *ortografi*. Ortografi merupakan sistem ejaan bunyi suatu bahasa dan lema kepala dalam sebuah kamus. Struktur kedua adalah *pelafalan* yang didefinisikan sebagai panduan cara pengucapan entri untuk pengguna kamus. Struktur ketiga adalah *tata bahasa* yang mengandung informasi mengenai kelas kata serta indikasi ketransitifan suatu verba. Struktur keempat adalah *definisi* sebagai penentu arti dari suatu entri. Struktur kelima adalah *contoh* berupa sematan entri dalam sebuah kalimat. Struktur keenam adalah *frasa*, sedangkan struktur ketujuh adalah *pemakaian* yang menjelaskan panduan mengenai pemakaian suatu entri secara tepat, baik secara preskriptif maupun deskriptif. Struktur terakhir adalah *etimologi* sebagai penunjuk sejarah asal suatu entri. Namun, tidak semua kamus, seperti kamus pemelajar bahasa asing, mencantumkan etimologi.

Berbeda dengan penelitian terdahulu, Atkins dan Rundell (2008) menyebutkan, sebenarnya struktur internal kamus memiliki 23 tiga komponen yang meliputi kata dasar, homograf, bentuk varian, kelas kata, tata bahasa, makna, contoh, referensi silang, dan komentar. Kompleksitas struktur internal yang diajukan oleh Atkins dan Rundell (2008) tersebut menunjukkan bahwa kelengkapan struktur internal kamus sangatlah normatif, bergantung pada kebutuhan kamus itu sendiri. Berbeda dengan yang lainnya, dalam Kamus Budaya Bahasa Inggris Australia Online, struktur internalnya juga berbeda. Struktur internal kamus tersebut terdiri atas bidang inti (kata dasar, pelafalan, dan kelas kata), definisi, contoh, daftar, riwayat kata, gambar, audio, dan video (Kwary & Miller, 2013).

Struktur entri kamus harus meliputi tiga kebutuhan, yaitu *real needs*, *ancillary needs*, dan *ultimate needs* (Kwary, 2011). *Real needs* berarti, sebuah kamus harus memenuhi kebutuhan informasi untuk menjawab masalah leksikografis dengan tepat dan benar. Dalam hal ini, *real needs* dari pengguna kamus ini adalah mengetahui definisi dari suatu kata. *Ancillary needs* adalah kebutuhan tambahan yang mungkin akan muncul pada saat konsultasi kamus seperti penggunaan kata di masyarakat. *Ultimate needs* adalah kebutuhan yang terjadi pada akhir proses pembelajaran yang meliputi konsultasi pembuatan kamus agar pengguna kamus memiliki pemahaman yang lengkap mengenai. Dengan adanya perumusan struktur internal yang merujuk pada kamus terdahulu yang telah terkenal diharapkan kamus digital bahasa Indonesia masa depan mampu memenuhi kebutuhan dan harapan masyarakat.

2.2 Metode Penelitian

Penelitian ini berfokus pada perbandingan struktur internal lima kamus digital bahasa Inggris ternama, yaitu CD, OLD, LDOCE, CED, dan MED dengan menggunakan lima kata berbahasa Inggris yang diserap dari bahasa Indonesia, yaitu *batik*, *sarong*, *bamboo*, *durian*, dan *orangutan*. Hasil dari rumusan struktur internal kamus ini menghasilkan tiga skala prioritas, yakni prioritas pertama, kedua, dan ketiga. Cara untuk menentukan struktur internal berdasarkan prioritas dilaksanakan dengan melakukan daftar periksa entri dan struktur internal yang mengikuti (Lihat Tabel 1).

Tabel 1. Tahapan Penentuan Struktur Internal Kamus Digital Masa Depan.

Struktur Internal	Cambridge					Oxford					Longman					Collins					MacMillan									
	batik	sarong	bamboo	durian	orangutan	batik	sarong	bamboo	durian	orangutan	batik	sarong	bamboo	durian	orangutan	batik	sarong	ambo	durian	orangutan	batik	sarong	bamboo	durian	orangutan					
Lema Kepala	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kelas Kata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Audio	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Lafal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Definisi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Contoh Kalimat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kata turunan dan Gabungan Kata								✓			✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Asal							✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓										

Kata																											
Gambar			✓					✓	✓				✓					✓									
Sinonim	✓	✓	✓	✓	✓															✓	✓	✓	✓	✓	✓	✓	
Tren													✓	✓	✓	✓	✓	✓									

Berdasarkan tahapan daftar periksa di atas ditemukan bahwa terdapat enam struktur internal yang digunakan oleh kelima kamus dengan merujuk pada kelima kata yang diteliti. Skala prioritas utama yang harus ada berdasarkan kelima kamus tersebut adalah lema kepala, kelas kata, audio, lafal, definisi, dan contoh kalimat. Skala prioritas kedua adalah kata turunan dan gabungan kata, asal kata, dan gambar. Skala prioritas terakhir meliputi sinonim dan tren.

III. ANALISIS PENELITIAN

3.1 Lema Kepala

Di dalam sebuah kamus, terdapat lema kepala yang merupakan kata atau frasa di posisi awal entri sebelum definisi dan penjelasan. Hartmann dan James (1998) menyebutkan bahwa lema kepala biasanya disajikan dengan tipografi yang berbeda, mulai dari bentuk tulisan tebal sampai dengan adanya keterangan dengan tulisan subskrip dan superskrip untuk menjelaskan lema kepala yang sama dengan informasi yang berbeda. Lema kepala dalam CD, OLD, LDOCE, CED, dan MED ditulis dengan huruf cetak tebal dan besar untuk membedakannya dengan definisi dan penjelasan. Biasanya ketika pengguna kamus hendak mencari definisi dari sebuah kata atau frasa, lema kepala juga akan muncul dalam kolom rekomendasi pencarian.

Sebuah kamus cetak lazimnya menyajikan satu lema kepala dengan beberapa definisi dan penjelasan. Demikian juga halnya dengan kelima kamus di atas. Hanya saja, ada beberapa di antara kamus tersebut, salah satunya CD, yang menyajikan lema kepala yang sama secara berulang-ulang dalam satu laman. Pada kata *batik*, CD menampilkan dua lema kata yang sama dengan definisi yang sedikit berbeda. Hal itu berbeda dengan CED yang menyajikan hanya satu lema kata diikuti dengan sublema yang menjelaskan kata *batik* di beberapa variasi bahasa Inggris, seperti *American English* dan *British English*.

Lema kepala sudah seharusnya ada di kamus digital bahasa Indonesia masa depan sebab ini adalah kata atau frasa pertama yang menjadi awalan di kamus dan umumnya dicari maknanya oleh pengguna. Jika melihat CD, OLD, LDOCE, CED, dan MED, lema kepala bisa disajikan berulang-ulang dalam satu laman dengan definisi dan informasi yang berbeda atau dengan sublema. Di samping itu, lema kata juga bisa disajikan dengan tampilan yang berbeda, mulai dari bentuk dan ukuran yang lebih mudah terlihat. Hal ini sebenarnya sudah terlihat di KBBI daring dan luring.

3.2 Kelas Kata

Kelas kata adalah informasi dari kata atau frasa yang menunjukkan peran gramatikal dalam kalimat (Hartmann & James, 1998). Dalam bahasa Inggris, kelas kata yang diadopsi oleh Atkins dan Rundell (2008) berisi sembilan kelas kata dasar, yakni adjektiva, adverbia, konjungsi, determinator, interjeksi, nomina, preposisi, pronomina, dan verba. Selain itu, dalam kamus digital masa depan di Indonesia diharapkan memiliki fitur tata bahasa yang lengkap dan matang. Bagian dari informasi kelas kata harus lengkap dengan fitur tata bahasa lainnya. Fitur tersebut dapat berupa subkelas kata, seperti transitif dan intransitif sebagai subkelas kata verba dan terhitung dan takterhitung sebagai subkelas kata nomina. Contoh penggunaan kelas kata dan subkelas kata dalam LDOCE dapat dilihat pada Gambar 2.

Gambar 2. Kelas Kata dan Subkelas Kata dalam LDOCE

bamboo

From Longman Dictionary of Contemporary English

Related topics: **Kelas kata**

bam·boo /ˌbæmˈbuː/ **noun** [countable, uncountable]

a tall tropical plant with hollow stems that is used for making furniture

SubKelas kata



Examples from the Corpus

bamboo

- Grazing buffalo knocked over the satellite dish, so a bamboo fence has had to be built around it.

3.3 Audio



Pada pembahasan sebelumnya telah dijelaskan bahwa sistem pelafalan *International Phonetic Alphabet* (IPA) juga perlu dimasukkan dalam kamus digital bahasa Indonesia masa depan. Hanya saja, tidak semua orang akrab dengan IPA sehingga audio diperlukan untuk menjadi penjelas cara pelafalan kata. Audio yang berisi pelafalan sebuah kata telah disediakan disamping IPA dalam CD, OLD, LDOCE, CED, dan MED.

Di samping itu, audio juga bisa dalam bentuk file audio yang berisi bunyi atau suara dari sebuah entri. Contohnya adalah kata *angklung*. Di KBBI, kata *angklung* didefinisikan sebagai ‘alat musik yang dibuat dari tabung bambu’. Dengan adanya penyediaan file audio di kamus digital, pengguna dapat mendengarkan bunyi alat musik tersebut ketika dimainkan. Dengan demikian, file audio seharusnya menjadi bagian dari struktur internal kamus digital bahasa Indonesia masa depan.

3.4 Lafal

Kelima kamus yang menjadi subjek penelitian menunjukkan bahwa pelafalan merupakan struktur internal kamus prioritas utama. Kelima kamus tersebut menggunakan IPA. IPA merupakan sistem pelafalan yang paling mendunia yang dipakai dalam CD, OLD, LDOCE, CED, dan MED (lihat Gambar 2). Sistem IPA sangat dipertimbangkan untuk diadopsi dalam kamus digital masa depan di Indonesia sebab pelafalan yang keliru atau yang tidak tepat dapat mengubah makna dan mungkin dirasakan berbeda dalam sebuah percakapan.

Gambar 2. Penggunaan IPA dalam CD dan CED

Sistem IPA dalam Cambridge Dictionary	Sistem IPA dalam Collins Dictionary
	

Pelafalan merupakan hal yang sangat penting untuk menunjukkan bahwa sebuah kata diproduksi. Salah satu fungsi pelafalan yang paling sederhana adalah untuk membedakan entri yang berhomograf, yakni memiliki ejaan yang identik, tetapi bentuk fonetis yang berbeda. Ada berbagai cara untuk menyajikan bentuk fonetik dalam kamus tertentu seperti Alphabet Fonetik Internasional, sistem penulisan ulang ortografi, dan sebagainya. Namun, sebagian besar leksikograf juga berpendapat bahwa bentuk fonetis yang terbaik yang harus disematkan dalam kamus adalah notasi IPA (Atkins & Rundell, 2008; Sterkenburg, 2003; Crystal, 2008).

Adapun alasan penyematan IPA meliputi dua hal paling penting yang telah disimpulkan. Pertama, IPA adalah sistem yang paling banyak digunakan untuk menyajikan bentuk fonetik. IPA dirancang untuk digunakan untuk mewakili suara semua bahasa manusia (Richards & Schmidt, 2010). Kedua, IPA adalah sistem paket yang lengkap. Penggunaan IPA tidak hanya diakui secara internasional, tetapi transkripsi IPA memungkinkan kamus untuk memasukkan, (1) bunyi bahasa, (2) panjang vokal, (3) stres, dan fitur khusus bahasa lainnya seperti nada bahasa (Crystal, 2008).

3.5 Definisi

Dua hal yang perlu diperhatikan dalam menampilkan definisi adalah isi dan bentuk (Atkins & Rundell, 2008, hlm. 450). Kamus seharusnya bisa menyajikan definisi yang bisa memenuhi kebutuhan penggunanya. Jika kamus dibuat dengan sasaran pengguna secara umum, definisi seharusnya disajikan dengan cara yang mudah dipahami. Contohnya, sebuah definisi yang hanya berdasarkan sinonim saja mungkin bisa dimengerti dengan mudah, tetapi di sisi lain definisi yang demikian tidak bisa memberikan isi atau informasi yang lebih memadai. Selain itu, tujuan penyajian definisi dapat tidak tercapai jika definisi disajikan

dengan bahasa yang terlalu spesifik untuk pengguna yang tidak akrab dengan gaya bahasa yang demikian.

Sebagai contoh, definisi *durian* dalam CD adalah ‘*a SE (South East) Asian bombacaceous tree, Durio zibethinus, having very large oval fruits with a hard spiny rind containing seeds surrounded by edible evil-smelling aril*’ dan ‘*the fruit of this tree, which has an offensive smell but a pleasant taste: supposedly an aphrodisiac*’. Pengertian yang demikian memang memberikan lebih banyak informasi bahwa durian bukan hanya merupakan buah, melainkan juga pohon. Walaupun demikian, jika pengguna tidak akrab dengan kata-kata dalam definisi tersebut, mereka akan sulit memahaminya.

Demikian pula bila sebuah kata dijelaskan terlalu sederhana. Contohnya adalah kata *durian* dalam MED yang didefinisikan sebagai ‘*tropical fruit with a strong, unpleasant smell but a sweet flavour*’. Jika pengguna tidak pernah melihat bentuk buah durian, mereka akan kesulitan membayangkan buah durian, apalagi bila pengertiannya hanya disampaikan berdasarkan aroma dan rasa.


Penelitian ini diharapkan mampu memberi kontribusi untuk kamus digital bahasa Indonesia masa depan dalam menyajikan definisi yang proposional dan membantu pengguna secara umum. Karena sasaran kamus digital bahasa Indonesia masa depan adalah pengguna secara luas, proses dalam penentuan definisi kata dalam kamus tetap dilakukan berdasarkan properti semantis, seperti hiponimi dan sinonimi, keterkaitan sebuah kata dengan objek di dunia nyata (meronimi), dan polisemi (Atkins & Rundell, 2008, hlm. 131). Dengan demikian, definisi yang dipaparkan menjadi tetap sangkil dan mangkus yang memberi informasi yang memadai.

3.6 Contoh Kalimat

Contoh kalimat dalam kamus tidak hanya menggambarkan data leksikografi dalam entri, tetapi secara implisit juga menyediakan informasi leksikografi baru yang belum disajikan di tempat lain (Burkhanov, 2003). Untuk membuat contoh yang baik, sebagian besar leksikograf menggunakan bantuan korpus. Namun, penggunaan korpus merupakan sebuah masalah lain di Indonesia karena korpus berbahasa Indonesia yang dapat diakses secara gratis dalam jumlah besar masih terbatas saat ini. Cara lain yang dapat dilakukan adalah membuat contoh kalimat secara manual dengan bantuan ahli bahasa atau praktisi bahasa. Adapun cara untuk menyajikan contoh yang lebih baik dan rasional, menurut Atkins dan Rundell (2008), harus ada tiga kriteria yang diadopsi, yaitu alami dan khas, informatif, dan dimengerti. Pada kata *bamboo*, semua kamus telah menyertakan contoh kalimat. Hanya saja contoh kalimat dalam CD dan CED merupakan contoh yang paling menarik dan patut dicontoh (lihat Gambar 3).

Gambar 3. Contoh kalimat dalam kata *bamboo* dalam CD dan CED

Contoh kalimat dalam kata “bamboo” dalam CD	Contoh kalimat dalam kata “bamboo” dalam CED
<i>Use bamboo canes to support tomato plants.</i>	<i>...huts with walls of bamboo. ...bamboo shoots.</i>

Examples for 'bamboo'	Example sentences containing 'bamboo'
<p>These examples are from external sources. Click on the  icon to tell us what you think.</p>	<p> These examples have been automatically selected and may contain sensitive content. Read more...</p>
<p>The seafood and rice were pressed using wooden (usually bamboo) molds. From <i>Wikipedia</i> </p>	<p>Seats spill into a charming courtyard back garden with a terrace and bamboo plants. <i>TIMES, SUNDAY TIMES (2014)</i></p> <p>We walk and walk until we enter a village of low houses and bamboo fences. <i>Hyland, Paul INDIAN BALM - TRAVELS IN THE SOUTHERN SUBCONTINENT (1994)</i></p>
<p>Kawayan is the most common local term for bamboo. From <i>Wikipedia</i> </p>	<p>Play around with bamboo poles stuck in the ground. <i>TIMES, SUNDAY TIMES (2016)</i></p>

Setelah definisi, CD dan CED secara langsung memberi contoh dalam bentuk kalimat yang ditunjukkan dalam gambar pertama. Namun, hal yang lebih menarik adalah kedua kamus ini memberikan contoh kalimat yang aktual serta berintegrasi dengan teknologi laman web dan korpus. Dalam CD dapat dilihat bahwa beberapa contoh kalimat yang mengandung kata *bamboo* dimunculkan dari sumber laman Wikipedia, sedangkan dalam CED contoh kalimat tersebut ditulis oleh banyak sumber, baik perorangan maupun majalah yang disertai dengan tahun kutipannya.

3.7 Kata Turunan dan Gabungan Kata

Kata turunan dan gabungan kata tidak terdapat di lima kamus digital Bahasa Inggris. Kamus-kamus digital tersebut hanya memiliki kata-kata terkait dan juga thesaurus. Dalam kamus Bahasa Indonesia, kata turunan dan gabungan kata tidak bisa dihilangkan sepenuhnya. Kata turunan adalah kata berimbuhan yang memiliki makna berbeda dari kata dasarnya. Contohnya kata *teman* di KBBI didefinisikan sebagai ‘kawan; sahabat; orang yang bersama-sama bekerja’. Kata tersebut berkelas kata nomina. Namun, bila sudah diberi imbuhan *me-* dan *-i*, menjadi *menemani*, kelas kata berubah menjadi verba dan memiliki arti ‘menyertai dan mengiringi’. Sementara itu, gabungan kata dapat berupa frasa yang terdiri atas lema kepala dengan lema kepala lain. Contoh gabungan kata adalah *teman nasi*. Gabungan kata tersebut dibentuk dari *teman* dan *nasi* yang masing-masing memiliki arti yang berbeda di KBBI. Namun, bila kedua kata ini digabung menjadi *teman nasi*, gabungan kata tersebut memiliki arti ‘lauk pauk atau sayur’.

3.8 Asal Kata

Asal kata merupakan struktur internal kamus yang masuk dalam prioritas kedua. Struktur internal kamus ini muncul dalam tiga kamus, yakni OLD, LDOCE, dan CED. Asal kata atau *word origin* ini merupakan struktur internal kamus yang penting untuk mengetahui etimologi dalam sebuah entri (Atkins & Rundell, 2008). Sebagai contoh, dalam CED disebutkan bahwa *bamboo* berasal dari bahasa Melayu yaitu *bambu*, sedangkan dalam OLD

kata tersebut berasal dari *mambu*. Bahkan, dalam OLD juga dijelaskan bahwa kata *bamboo* dipopulerkan pada abad ke 16 oleh Belanda sebagai *bamboes*.

Gambar 4. Word origin dalam CED dan OED

CED:

Word origin of 'bamboo'
Malay *mambu* < ? a Dravidian language

OED:

— Word Origin
late 16th cent.: from Dutch *bamboes*, based on Malay *mambu*.

3.9 Gambar

Salah satu kelebihan kamus digital adalah penyajian yang lebih menarik dengan multimedia. Multimedia yang berupa audio dan gambar dapat membantu pengguna untuk mendapat gambaran utuh dari sebuah kata. Pada kata-kata tertentu memang tidak dibutuhkan adanya gambar jika definisi dan contoh kalimat sudah jelas. Namun, ada kata-kata tertentu yang tidak dapat dijelaskan hanya dengan sinonim atau deskripsinya saja.

Kamus-kamus digital bahasa Inggris, seperti CD dan CED, telah menyediakan beberapa gambar di beberapa kata sebagai informasi tambahan di samping definisi. Contohnya adalah kata *bamboo* dalam CD yang didefinisikan sebagai '*a tall tropical grass with hard, hollow stems, or the stems of this plant*'. CD telah menyediakan gambar sebagai fitur tambahan dalam kamus digital. Meskipun begitu, gambar tersebut akan lebih membantu pengguna dalam mendapatkan gambaran nyata tanaman bambu jika di dalam gambar terdapat mode yang menunjukkan tinggi bambu secara nyata atau terukur.

3.10 Sinonim

Sinonim merupakan semua kata yang memiliki arti sama atau sebuah kata yang memiliki hubungan sangat spesial dan erat satu sama lain (Atkins & Rundell, 2008). CD dan MED mencantumkan sinonim dari lema *batik*, *sarong*, *bamboo*, *durian*, dan *orangutan*. Kelebihan dari kamus yang lengkap dengan sinonim diibaratkan seperti memenuhi suatu kebutuhan bagi mereka yang terlibat dalam tugas *encoding* dengan menawarkan pilihan leksikal untuk mengekspresikan makna tertentu. Hartmann dan James (1998) menjelaskan bahwa penggunaan sinonim dalam kamus dapat disusun dalam abjad atau tatanan tematik atau alfabetikal indeks (THESAURUS). Berikut merupakan contoh sinonim dalam bentuk thesaurus dalam kamus MED.

Gambar 5. Tesaurus atau Sinonim Kata *sarong*

Skirts and dresses - thesaurus



Explore other meanings

Using it

Related words

ballgown NOUN

a special long dress worn by a woman for dancing at a ball (=a formal social event)

caftan NOUN

another spelling of kaftan

chemise NOUN

a long loose dress or piece of underwear, worn by women

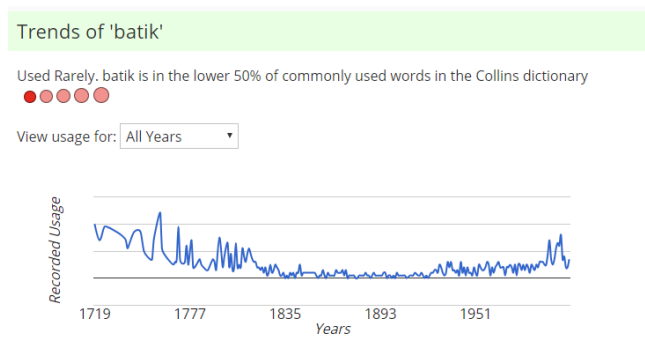
cheongsam NOUN

a tight silk dress with a high neck, short sleeves, and a slit (=long narrow space) on each side by the leg that is worn by Chinese and Indonesian women

3.11 Tren

Dari lima kamus monolingual bahasa Inggris yang diteliti, CED merupakan satu-satunya kamus yang menyematkan tren sebagai salah satu internal struktur kamusnya (lihat Gambar 6). Tren menunjukkan frekuensi penggunaan sebuah entri, misalnya dalam kata *batik*. *Batik* sangat jarang dicari dalam CED. Bahkan, pencarian *batik* dibawah 50%. Tren juga dapat menunjukkan tingkat popularitas sebuah entri secara diakronik. Gambar 6 menunjukkan bahwa kata *batik* sering digunakan pada tahun 1759 daripada tahun 2008.

Gambar 6. Tren entri kata batik dalam Collins



IV. SIMPULAN

Struktur internal kamus telah dirumuskan oleh para leksikograf dengan berbagai teori dan argumen yang berbeda-beda. Adapun perbedaan struktur internal kamus dapat dipengaruhi oleh tujuan dan pengguna kamus itu sendiri. Dalam kamus digital monolingual masa depan di Indonesia diharapkan dapat mengadopsi struktur kamus yang sudah lama ada dalam kamus monolingual berbahasa Inggris, seperti dalam CD, OLD, LDOCE, CED, dan

MED. Struktur internal kamus yang dapat diadopsi dari kelima kamus tersebut terdiri atas tiga struktur berdasarkan skala prioritasnya. Skala prioritas utama adalah lema kepala, kelas kata, audio, lafal, definisi, dan contoh kalimat. Skala prioritas kedua adalah kata turunan dan gabungan kata, asal kata, dan gambar. Skala prioritas yang terakhir meliputi sinonim dan tren. Penelitian ini diharapkan mampu memberikan informasi dan rekomendasi bagi pekamus, pakar teknologi informasi, pembuat aplikasi, dan peminat bahasa untuk mengembangkan kamus digital bahasa Indonesia masa depan berdasarkan tiga skala prioritas tersebut.

Daftar Rujukan

- Atkins, B. T., dan Rundell, M. (2008). *The oxford guide to practical lexicography*. New York: Oxford University Press.
- Burkhanov, I. (2003). Pragmatic specifications: Usage indications, labels, examples; dictionaries of style, dictionaries of collocations. In P. v. Sterkenburg, *A practical guide to lexicography* (pp. 102-113). John Benjamins Publishing Company: Amsterdam/Philadelphia.
- Crystal, D. (2008). *A dictionary of linguistics and phonetics*. Malden: Blackwell publishing.
- Hanks, P. (2006). Lexicography: Overview. In K. Brown, *Encyclopedia of language and linguistics* (pp. 113-128).
- Hartmann, R. R., dan James, G. (1998). *Dictionary of lexicography*. London and New York: Routledge.
- Kwary, D. A. (2011). Adaptive hypermedia and user-oriented data for online dictionaries: A case study on an English dictionary of finance for Indonesian students. *International Journal of Lexicography*, 30-49.
- Kwary, D. A., dan Miller, J. (2013). A model for an online Australian English cultural dictionary database. *Terminology*, 258–276.
- Sterkenburg, V. P. (2003). *A practical guide to lexicography*. Amsterdam: John Benjamins Publishing.

PENGEMBANGAN APLIKASI KAMUS DARING TIGA BAHASA
*(THE DEVELOPMENT OF THE THREE-LANGUAGES DICTIONARY
APPLICATION)*

Sri Andayani¹

Prodi Sastra Inggris, Fakultas Sastra dan Filsafat, Universitas Panca Marga Probolinggo
sriandayani@upm.ac.id

Sugeng Edy Mulyono²

Prodi Sastra Inggris, Fakultas Sastra dan Filsafat, Universitas Panca Marga Probolinggo
masedymulyono@gmail.com

Ahmad Izzuddin³

Prodi Teknik Elektro, Fakultas Teknik, Universitas Panca Marga Probolinggo
ahmad.izzuddin@upm.ac.id

Abstrak

Pesatnya perkembangan Iptek serta adanya kecenderungan masyarakat untuk menggunakan kamus elektronik menjadi pendorong munculnya berbagai aplikasi kamus elektronik, baik kamus daring maupun luring. Namun, aplikasi kamus elektronik yang telah ada rata-rata masih sebatas membandingkan dua bahasa, baik bahasa lokal dengan bahasa nasional maupun bahasa nasional dengan bahasa asing. Mestinya, aplikasi tersebut dapat dikembangkan menjadi tiga bahasa sekaligus, yaitu bahasa lokal, bahasa nasional, sekaligus bahasa asing. Hal ini dilakukan dalam penyusunan kamus daring tiga bahasa: Madura-Indonesia-Inggris. Kamus daring ini membandingkan tiga bahasa sekaligus, yaitu bahasa Madura, bahasa Indonesia, dan bahasa Inggris. Sistem pencariannya yaitu, ketika kata dicari dalam bahasa Madura, dua bahasa lain akan ikut muncul, begitu pula ketika pencarian dilakukan dari dua bahasa lain. Kamus ini nantinya juga memiliki tambahan fitur, sehingga fungsi kamus elektronik saat ini dapat menjadi lebih lengkap dan luas. Fitur tambahan berupa bentuk pelafalan sekaligus transkripsi fonetis. Jadi, lema dalam kamus daring ini mencakup ejaan, pelafalan, transkripsi fonetis, serta gambar visual. Pengguna kamus dapat memilih satu dari tiga fitur yang disediakan, atau memanfaatkan tiga fitur sekaligus dari ketiga bahasa ini., Kamus ini diharapkan dapat dimanfaatkan oleh pengguna ketiga bahasa ini, baik masyarakat awam dalam tingkat lokal, nasional, dan internasional; maupun para pegiat bahasa. Masyarakat awam dapat dengan mudah mengetahui bentuk ejaan pada setiap lema sekaligus bentuk pelafalannya. Sedangkan, bagi para pegiat bahasa dapat memanfaatkan bentuk transkripsi fonetisnya. Selain itu, kamus ini nantinya juga dapat digunakan sebagai salah satu data dalam penelitian linguistik.

Abstract

The development of science has a tendency of public to change a printed dictionary to electronic one. It becomes the encouraging factors of the new version of dictionaries, like online edition. However, most of them is just to compare in two languages, like local and national language. Supposedly, the application can be developed into three languages at once, either in the local, national, and foreign languages. It is done in making this three-language online dictionary: Madurese-Indonesian-English. This dictionary can compare those three languages at once. When the searching is done in Madurese, two other languages will be the target languages, and vice versa. In this online dictionary, there is also an application development that includes some additional features. It makes the current electronic dictionary function more complete and more widespread. The additional features are in the form of the pronunciation and phonetic transcription. So, the entry of the dictionary is included in the form of spelling, pronunciation, phonetic transcription, and visual images. The users can choose one of the three features provided, or use three features at once, from these three languages. It is hoped that this dictionary can be used by the ordinary speakers of the languages, either at the local, national, and international levels; as well as the language activists. They can easily know the spelling form of each entry, as well as the pronunciation. Meanwhile, the language activists can also use the phonetic transcription as the data in the linguistic researches.

I. PENDAHULUAN

Selama 16 tahun (sejak tahun 1991 sampai dengan 2017), Badan Pengembangan dan Pembinaan Bahasa Kementerian Pendidikan dan Kebudayaan telah mengidentifikasi dan memvalidasi 652 bahasa dari 2.452 daerah pengamatan (Republika, 27 Oktober 2017). Sementara, pada tahun 1988, Grimes menyatakan bahwa terdapat sejumlah 672 bahasa lokal dimiliki oleh Indonesia (Laksono, 2007:19). Dari kedua data itu, terdapat selisih yang cukup signifikan. Hal itu berarti bahwa selama kurun waktu 29 tahun kurang lebih terdapat 20 bahasa lokal yang punah. Kepunahan bahasa-bahasa tersebut dapat terjadi seiring dengan perkembangan waktu karena ditinggalkan oleh penuturnya. Oleh karena itu, upaya-upaya tertentu harus dilakukan agar kepunahan bahasa dapat dihindari atau setidaknya dapat diminimalisasi. Oleh sebab itu, perlu upaya pemertahanan atau pelestarian terhadap bahasa-bahasa yang masih memiliki penutur.

Upaya pemertahanan dan pelestarian bahasa-bahasa lokal di Indonesia perlu dilakukan secara serius dan berkelanjutan. Upaya tersebut dapat dilakukan melalui pendokumentasian bahasa lokal dan penyusunan kamus sebagai upaya lanjutan agar bahasa-bahasa lokal tidak punah. Kamus dapat mendorong keberlanjutan dan kelestarian bahasa di lingkungan masyarakat. Oleh sebab itu, penyusunan kamus bahasa menjadi hal penting yang harus dilakukan.

Di Indonesia, sebagian besar kamus diproduksi dalam bentuk cetak. Hal itu karena masyarakat Indonesia cenderung menggunakan kamus cetak. Seiring perkembangan ilmu pengetahuan dan teknologi, kamus elektronik mulai digunakan dan bahkan lebih diminati. Bentuk kamus cetak yang tebal dan tidak fleksibel merupakan alasan masyarakat lebih

memilih kamus elektronik daripada cetak, terlebih telepon pintar, laptop, dan berbagai gawai lainnya sudah akrab di lingkungan keseharian masyarakat.

Maraknya penggunaan barang-barang elektronik menjadi faktor utama populernya kamus elektronik di lingkungan masyarakat. Kamus elektronik memiliki kelebihan dibanding dengan kamus cetak, terutama dari segi akses internet. Fasilitas internet saat ini sudah menjangkau daerah-daerah terpencil mampu mendukung penyebaran kamus elektronik. Hal itu berdampak pada penurunan jumlah pengguna kamus cetak.

Penelitian ini bertujuan untuk mengembangkan kamus daring agar akses menjadi lebih mudah, fleksibel, dan murah. Kamus bahasa daerah biasanya disusun dalam format dua bahasa yaitu, bahasa sumber dan padanannya. Misalnya, Kamus Lengkap Bahasa Madura-Indonesia (Adrian Pawitra, 2008), Kamus Dwibahasa Indonesia-Madura (Balai Bahasa Provinsi Jawa Timur, 2013), Kamus Suroboyoan-Indonesia (M. Jupri, 2008). Kamus-kamus itu umumnya hanya mengakomodasi kebutuhan penutur bahasa Indonesia terhadap bahasa lokal.

Selain itu, kamus ini dibuat untuk mempermudah pengguna mengetahui bentuk bahasa sumber dan padanannya. Hal itu dilatarbelakangi karena pengguna kamus elektronik tidak hanya dari penutur bahasa lokal, namun juga dari penutur bahasa Indonesia bahkan bahasa asing. Bentuk kamus elektronik yang dipilih pada penelitian ini adalah kamus daring bahasa Madura, bahasa Indonesia, dan bahasa Inggris.

Bahasa Madura dipilih menjadi salah satu bahasa sumber dari kamus elektronik karena memiliki jumlah penutur terbanyak keempat di Indonesia, (Laksono, 2009:20; Davies, 2010:1). Bahasa Madura dituturkan oleh sekitar 13,7 juta orang (Laksono, 2009:20). Bahasa ini dituturkan oleh penduduk di kepulauan Madura, beberapa kota di daerah tapal kuda di Jawa Timur, Kalimantan, Sumatera, serta penutur yang tersebar di beberapa kota besar di Jawa Tengah dan Jawa Barat (Davies, 2010:1-2). Karakteristik penduduk pulau Madura yang gemar merantau membuat penutur bahasa Madura tersebar hingga menjangkau manca negara seperti Malaysia, Singapura, Australia, dan Arab Saudi. Hal ini membuat bahasa Madura ini menjadi populer.

Kepopuleran bahasa Madura ini seharusnya didokumentasikan dengan baik. Hal itu ditujukan agar bahasa Madura terpelihara dengan baik, dengan harapan dapat membantu para penutur, terutama generasi muda untuk lebih mengenal kosakata bahasa Madura, baik secara tertulis maupun lisan. Secara formal, bahasa Madura hanya dikenal penduduk Madura di lingkungan sekolah, sedangkan di luar Pulau Madura, penutur hanya mengenal bahasa Madura sebagai bahasa lisan.

Bahasa Madura yang akan diinventarisasi dalam kamus ini akan diberikan padanan dalam bahasa Indonesia dan bahasa Inggris. Pemberian padanan bahasa Indonesia bertujuan untuk mempermudah pengguna kamus mencari kosakata yang diinginkan. Kemudian, padanan bahasa Inggris digunakan untuk mempermudah penutur bahasa Madura untuk mencari kosakata bahasa Inggris, begitu pula sebaliknya. Jumlah kosakata yang akan diinventaris pada penelitian ini sejumlah 1.000 kosakata bahasa Madura pada tingkatan yang paling populer, yakni tingkat tutur *enjâ'-iyâ* (Balai Bahasa Provinsi Jawa Timur, 2014:3). Selain itu, fitur lain yang akan ditambahkan antara lain, lafal, transkripsi fonetis ketiga bahasa, serta gambar visualnya. Transkripsi fonetis, digunakan untuk mempermudah

pelafalan agar lebih tepat, sebab salah satu keunikan bahasa Madura ada dalam pelafalannya.

II. KERANGKA TEORI DAN METODE PENELITIAN

Cabang ilmu bahasa yang mempelajari metode dan teknik penyusunan kamus adalah leksikografi (Kridalaksana, 2010: 53). Aspek-aspek kamus mencakup segi semantis, morfologis, fonetis, fonemis, hingga tata bahasa. Selain itu, hal-hal yang harus diperhatikan ketika menyusun kamus, yaitu: 1) jumlah dan jenis kata, 2) metode pengumpulan kosakata dari berbagai sumber dan penyusunannya, 3) cara pemberian definisi pada setiap kata, 4) cara membedakan variasi makna kata, 5) cara mengindikasikan bentuk tuturnya (Hartmann, 2001:14). Kamus daring tiga bahasa terdiri atas 1.000 kosakata lengkap dengan padanannya. Kosakatanya meliputi nomina (n), verba (v), adjektiva (a), adverbial (adv), dan pronomina (pron). Kode singkatan mengikuti KBBI V (2016). Kosakata dikumpulkan dari hasil wawancara dan observasi pada penelitian terdahulu (Andayani, Sri & Sutrisno, Adi, 2017), ditambah dengan kosakata hasil wawancara dan observasi lanjutan. Identifikasi tuturan dilakukan dengan menyertakan pelafalan oleh penutur asli, lengkap dengan transkripsi fonetis, agar pelafalan dapat diketahui dengan lebih jelas.

Transkripsi fonetis pada setiap kosakata dalam kamus tiga bahasa ini menggunakan simbol IPA (International Phonetics Association). Simbol IPA yang digunakan pada transkripsi fonetis pada pelafalan ketiga bahasa ini menggunakan simbol fonetis bahasa Inggris-Amerika (Fromkin, et. al., 2014: 198-215). Alasan penggunaan simbol-simbol fonetis ini adalah bahasa Inggris-Amerika memiliki lambang fonetis yang lebih banyak daripada simbol bunyi bahasa Madura maupun bahasa Indonesia. Hasilnya diharapkan dapat mengakomodasi bunyi-bunyi dalam bahasa Madura dan bahasa Indonesia. Kosakata Bahasa Madura yang tidak terakomodasi, lambang fonetisnya menggunakan simbol pada Tata Bahasa: Bahasa Madura (Balai Bahasa Provinsi Jawa Timur, 2014:24-42), serta beberapa lambang fonetis yang lazim digunakan pada bahasa-bahasa lainnya.

Bentuk ejaan bahasa Inggris mengikuti bentuk ejaan yang digunakan pada Kamus Inggris-Indonesia karangan Jhon M. Echols dan Hassan Shadily (2017). Ejaan bahasa Indonesia mengikuti KBBI V (2016), sedangkan ejaan bahasa Madura sesuai dengan Kamus Dwibahasa Indonesia-Madura (Balai Bahasa Provinsi Jawa Timur, 2013).

Adapun aplikasi kamus daring tiga bahasa ini dibangun menggunakan *web interface* dengan menggunakan bahasa pemrograman *PHP* dan *CSS Bootstrap* agar menjadi *responsive*. Bahasa pemrograman *PHP* digunakan untuk menerjemahkan perintah yang diberikan programmer menjadi bahasa komputer agar dapat digunakan untuk mengolah data kamus tiga bahasa ini. Sedangkan, *CSS Bootstrap* berfungsi untuk membuat tampilan kamus menjadi sesuai ketika di buka pada layar laptop maupun layar telepon pintar.

Penelitian pada penyusunan Kamus Tiga Bahasa (Madura-Indonesia-Inggris) ini merupakan penelitian deskriptif kualitatif. Penelitian ini berusaha mendeskripsikan bentuk bahasa yang digunakan oleh penutur bahasa Madura, bahasa Indonesia, dan bahasa Inggris. Kosakata bahasa Madura yang menjadi lema pada kamus ini adalah kosakata bahasa Madura dialek Sumenep. Dialek Sumenep dipilih dengan alasan sebagai tindak lanjut dari penelitian yang dilakukan sebelumnya.

Data penelitian ini dikumpulkan dengan metode wawancara dan observasi. Wawancara dilakukan pada penutur asli bahasa Madura. Terdapat sejumlah 1.000 kosakata bahasa Indonesia yang menjadi daftar tanya yang ditanyakan kepada narasumber bahasa Madura secara langsung. Sementara, data bahasa Indonesia dan bahasa Inggris dikumpulkan dengan cara observasi pada KBBI V (2016) dan Kamus Inggris-Indonesia karangan Jhon M. Echols dan Hassan Shadily (2017).

Pelafalan kosakata pada ketiga bahasa dilakukan dengan merekam tuturan para penutur asli. Para penutur berasal dari Madura berdialek Sumenep. Pelafalan kosakata bahasa Indonesia dilakukan pada penutur asli bahasa Indonesia. Sedangkan, pelafalan bahasa Inggris dilakukan oleh penutur asli bahasa Inggris-Amerika.

Penyusunan lema kamus ini menggunakan pangkalan data Excel 97-2003. Jadi 1.000 kosakata disusun dalam tabel Excel yang terdiri atas tujuh kolom, yaitu kosakata bahasa Madura (md), lambang fonetis kosakata bahasa Madura (fnmd), kosakata bahasa Indonesia (ina), lambang fonetis kosakata bahasa Indonesia (fnina), kosakata bahasa Inggris (en), lambang fonetis kosakata bahasa Inggris (fnen), dan kelas kata (kl). Kemudian, pangkalan data excel tersebut dimasukkan ke dalam pangkalan data SQL (Structured Query Language) dari aplikasi Kamus Tiga Bahasa (Madura-Indonesia-Inggris). Selain itu, fitur pelafalan dan gambar visual juga disediakan pada menu tersendiri dalam aplikasi kamus ini. Selanjutnya, aplikasi kamus ini siap untuk didaringkan.

III. ANALISIS PENELITIAN

Kamus Daring Tiga Bahasa ini menyajikan daftar kosakata dalam tiga bahasa yaitu bahasa Madura, bahasa Indonesia, dan bahasa Inggris. Kamus ini menyajikan kosakata tiga bahasa dalam bentuk ejaan, transkripsi fonetis, pelafalan, serta bentuk visualnya. Data kamus ini disusun dalam bentuk pangkalan data Excel sebelum dimasukkan dalam basis data *SQL*. Data kamus daring tiga bahasa dalam bentuk Excel digambarkan pada tabel 1 di bawah ini.

Tabel 1. Data Excel Kamus Tiga Bahasa

md	fnmd	ina	fnina	en	fnen	kl
ennol	[ənnɔl]	nol	[nɔl]	zero	[ziro]	num
səttong	[settɔŋ]	satu	[satu]	one	[wʌn]	num
Duwá'	[duweʔ]	dua	[duwa]	two	[tʰu]	num
tello'	[təllɔʔ]	tiga	[tiga]	three	[θri]	num
empa'	[əmpʌʔ]	empat	[əmpʌt]	four	[fɔ(r)]	num
ləma'	[ləmʌʔ]	lima	[lima]	five	[faɪv]	num
ennem	[ənnəm]	enam	[ənʌm]	six	[sɪks]	num
pətto'	[pəttɔʔ]	tujuh	[tuʤuh]	seven	[seɪvən]	num
bəllu'	[belluʔ]	delapan	[dəlapʌn]	eight	[eɪθ]	num
sanga'	[sʌŋʌʔ]	sembilan	[səmbilan]	nine	[naɪn]	num

Kolom satu adalah bentuk ejaan tulis kosakata bahasa Madura yang diadopsi dari Kamus Dwibahasa Indonesia-Inggris (Balai Bahasa Provinsi Jawa Timur, 2013). Ada

beberapa simbol alfabetis yang digunakan pada ejaan bahasa Madura ini, yakni terdiri atas 31 simbol huruf konsonan dan 11 simbol huruf vokal. Lambang fonetis konsonan yang digunakan antara lain: <p, t, T, c, k, q, ‘, b, d, D, j, g, bh, dh, Dh, jh, gh, f, s, sy, kh, h, z, m, n, ny, ng, r, l, w, y>. Sedangkan lambang fonetis vokal pada ejaan bahasa yang digunakan adalah <a, â, i, u, è, e. o. ay. ây, uy, oy>

Kolom kedua pada basis data Excel berupa transkripsi fonetis kosakata bahasa Madura. Masing-masing huruf yang menjadi bentuk ejaan bahasa Madura dilambangkan bunyinya sesuai dengan cara pelafalan kosakata bahasa Madura yang dilakukan oleh penutur asli bahasa Madura. Simbol huruf dan simbol bunyi pada bahasa Madura cenderung bersifat stabil. Dengan kata lain, satu simbol huruf dilambangkan juga dengan satu simbol bunyi, begitu sebaliknya. Adapun simbol bunyi konsonan dan bunyi vokal yang sering digunakan dalam kosakata bahasa Madura antara lain tercantum pada tabel 2 di bawah ini.

Tabel 2. Padanan simbol huruf dan simbol bunyi bahasa Madura

No	Huruf Konsonan	Bunyi Konsonan	Fitur Fonetis	Huruf Vokal	bunyi vokal	Fitur Fonetis
1	p	[p]	konsonan hambat tak bersuara bilabial	a	[a]	vokal pusat bawah
2	t	[t]	konsonan hambat tak bersuara alveolar	â	[e]	vokal pusat tengah-bawah
3	T	[ʈ]	konsonan hambat tak bersuara retrofleks	i	[i]	vokal depan atas
4	c	[ç]	konsonan hambat tak bersuara palatal	u	[u]	vokal belakang atas
5	k	[k]	konsonan hambat tak bersuara velar	è	[e]	vokal depan tengah bawah
6	q	[q]	konsonan hambat tak bersuara uvular	e	[ə]	vokal pusat tengah
7	‘	[ʔ]	konsonan hambat tak bersuara glotal	o	[o]	vokal belakang tengah bawah
8	b	[b]	konsonan hambat bersuara bilabial	av	[aj]	diftong
9	d	[d]	konsonan hambat bersuara alveolar	ây	[ej]	diftong
10	D	[dʂ]	konsonan hambat bersuara retrofleks	uy	[uj]	diftong
11	j	[dʃ]	konsonan hambat bersuara palatal	oy	[oj]	diftong
12	g	[g]	konsonan hambat bersuara velar			
13	bh	[bʰ]	konsonan hambat bersuara bilabial beraspira			
14	dh	[dʰ]	konsonan hambat bersuara alveolar beraspira			
15	Dh	[dʂʰ]	konsonan hambat bersuara retrofleks beraspira			
16	jh	[dʃʰ]	konsonan hambat bersuara palatal beraspira			
17	gh	[gʰ]	konsonan hambat bersuara velar beraspira			
18	f	[f]	konsonan frikatif tak bersuara labiodental			
19	s	[s]	konsonan frikatif tak bersuara			
20	sv	[ʃ]	konsonan frikatif tak bersuara palatal			
21	kh	[x]	konsonan frikatif tak bersuara uvular			
22	h	[h]	konsonan frikatif glotal			
23	z	[z]	konsonan frikatif bersuara alveolar			
24	m	[m]	konsonan nasal bilabial			
25	n	[n]	konsonan nasal alveolar			
26	ny	[ɲ]	konsonan nasal palatal			
27	ng	[ŋ]	konsonan nasal velar			
28	r	[r]	konsonan getar alveolar			
29	l	[l]	konsonan lateral alveolar			
30	w	[w]	semivokal bilabial			
31	y	[j]	konsonan semivokal palatal			

- Tata Bahasa Bahasa Madura (2014: 27 dan 37-38)

Kolom ketiga berisi daftar kosakata bahasa Indonesia. Ejaan yang digunakan pada kosakata bahasa Indonesia adalah mengikuti KBBI edisi V (2016). Bahasa Indonesia menggunakan 26 simbol huruf latin baik berupa konsonan dan vokal, yakni <a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z>.

Sedangkan, kolom keempat berisi transkripsi fonetis untuk kata-kata bahasa Indonesia. Simbol bunyi yang secara signifikan sering digunakan pada bahasa Indonesia tidaklah sebanyak simbol bunyi bahasa Madura. Pelafalan kosakata dalam bahasa Indonesia cenderung lebih sederhana dibanding bahasa Madura. Namun, sejauh ini belum ada standar baku pelafalan bahasa Indonesia. Pengaruh dialek bahasa lokal masih sangat kuat mewarnai kosakata bahasa Indonesia. Oleh karena itu, pelafalan oleh penutur bahasa Indonesia dialek Jawa Timuran menjadi data pelafalan pada kamus ini. Simbol bunyi konsonan yang sering digunakan oleh penutur bahasa Indonesia dialek Jawa Timuran adalah [b, ʃ, d, f, g, h, dʒ, k, l, m, n, p, r, s, t, w, j], sedangkan bunyi vokal yang sering digunakan adalah [ʌ, i, u, e, ε, ə, o, ɔ].

Kolom kelima berisi daftar kosakata bahasa Inggris, sebagai padanan dari kosakata bahasa Madura dan bahasa Indonesia. Bahasa Inggris memiliki sejumlah 26 simbol huruf, sama seperti pada bahasa Indonesia, perbedaan terletak pada pelafalannya, yaitu <a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z>.

Kolom keenam berisi transkripsi fonetis pelafalan dalam bahasa Inggris. Pelafalan direkam sesuai dengan pelafalan asli bahasa Inggris-Amerika. Pada bahasa Inggris, ejaan dan pelafalan seringkali tidak memiliki padanan yang konsisten, seperti halnya pada bahasa Madura dan bahasa Indonesia. Oleh karena itu, tidak dapat dibuat daftar padanan simbol huruf konsonan dan vokal sesuai dengan simbol bunyi konsonan dan vokal seperti pada pada tabel 2. Pada tabel 3 di bawah ini, dideskripsikan bunyi-bunyi yang secara signifikan sering digunakan pada pelafalan bahasa Inggris beserta fitur fonetisnya. Sesuai dengan pelafalan yang didapatkan, 30 simbol bunyi konsonan, dan 15 vokal yang secara signifikan sering digunakan terdapat pada tabel 3.

Tabel 3. Simbol Bunyi Konsonan dan Vokal Bahasa Inggris

No	Simbol Bunyi Konsonan	Fitur Fonetis	Simbol bunyi Vokal	Fitur Fonetis
1	[p]	konsonan hambat tak bersuara bilabial	[i]	vokal depan tinggi-atas
2	[p ^h]	konsonan hambat tak bersuara bilabial beraspira	[ɪ]	vokal depan tinggi-bawah
3	[b]	konsonan hambat bersuara bilabial	[e]	vokal depan tengah-atas
4	[m]	konsonan nasal bilabial	[ɛ]	vokal depan tengah-bawah
5	[t]	konsonan hambat tak bersuara alveolar	[æ]	vokal depan rendah
6	[t ^h]	konsonan konsonan hambat tak bersuara alveolar beraspira	[ə]	vokal sentral tengah-bawah
7	[d]	konsonan konsonan hambat bersuara alveolar	[u]	vokal belakang tinggi-atas
8	[n]	konsonan nasal alveolar	[ʊ]	vokal belakang tinggi-bawah
9	[k]	konsonan hambat tak bersuara velar	[o]	vokal belakang tengah
10	[k ^h]	konsonan hambat tak bersuara velar beraspira	[ɔ]	vokal belakang rendah
11	[g]	konsonan hambat bersuara velar	[ɑ]	vokal sentral rendah-bawah
12	[ŋ]	konsonan nasal velar	[ʌ]	vokal sentral rendah-atas
13	[f]	konsonan frikatif tak bersuara labiodental	[aj]	diftong
14	[v]	konsonan frikatif bersuara labiodental	[aw]	diftong
15	[s]	konsonan frikatif tak bersuara alveolar	[ɔj]	diftong
16	[z]	konsonan frikatif bersuara alveolar		
17	[θ]	konsonan frikatif tak bersuara interdental		
18	[ð]	konsonan frikatif bersuara interdental		
19	[ʃ]	konsonan frikatif tak bersuara palatal		
20	[ʒ]	konsonan frikatif bersuara palatal		
21	[tʃ]	konsonan afrikatif tak bersuara palatal		
22	[tʃ ^h]	konsonan afrikatif tak bersuara palatal beraspira		
23	[dʒ]	konsonan afrikatif bersuara palatal		
24	[l]	konsonan lateral liquid		
25	[r]	konsonan sentral liquid		
26	[j]	konsonan semivokal tak bersuara palatal		
27	[w]	konsonan semivokal bersuara bilabial		
28	[ʌ]	konsonan semivokal tak bersuara bilabial		
29	[h]	konsonan frikatif tak bersuara glotal		
30	[ʔ]	konsonan hambat glotal		

- Phonetics, An Introduction to Language (Fromkin, et. al., 204 & 2017)

Adanya fitur transkripsi fonetis pada kamus daring tiga bahasa ini dimaksudkan agar pelafalan kosakata pada ketiga bahasa ini juga memiliki bentuk representasi bunyi bahasanya yang diwujudkan dalam simbol-simbol bunyi, baik bunyi konsonan maupun bunyi vokal. Kosakata dituliskan dalam bentuk transkripsi alfabetis, sebagai bentuk penyimbolan ejaan tulisnya dan transkripsi fonetis, sebagai bentuk penyimbolan pelafalannya dimaksudkan juga agar data pada kamus ini juga dapat dimanfaatkan sebagai data penelitian linguistik. Jadi kamus ini tidak hanya berguna bagi masyarakat awam, namun juga bagi para pegiat bahasa.

Kolom ketujuh berisi identifikasi kelas kata yang menjadi lema. Kelas kata yang digunakan pada kamus meliputi nomina (n), verba (v), adjektiva (a), adverbial (adv), dan pronomina (pron). Jika proses tabulasi pangkalan data Excel 3.000 kosakata pada ketiga bahasa telah lengkap, selanjutnya proses memasukkan pangkalan data Excel tersebut ke dalam aplikasi kamus daring tiga bahasa.

Selanjutnya memasukkan data berupa rekaman pelafalan setiap kosakata ketiga bahasa pada fitur pelafalan. Setiap kosakata dibuat menjadi satu berkas rekaman. Jadi pada setiap bahasa terdapat 1.000 berkas rekaman, dan setiap berkas diberi nama sesuai dengan ejaan kosakatanya. Berkas rekaman ini kemudian dimasukkan ke dalam menu khusus yang tersedia dalam aplikasi kamus.

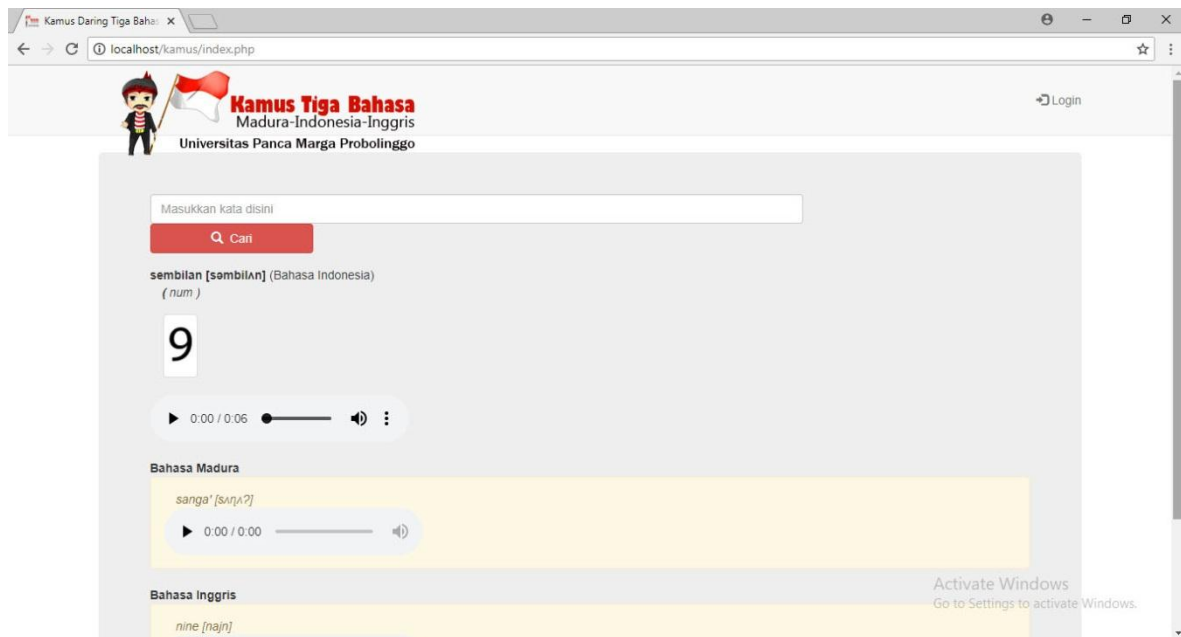
+ Options								
id	md	fnmd	ina	fnina	en	fnen	kl	vi
0	ennol	[ənnɔl]	nol	[nɔl]	zero	[ziro]	num	NULL
1	sèttong	[sɛttɔŋ]	satu	[sʌtu]	one	[wʌn]	num	NULL
2	Duwá'	[duwɛʔ]	dua	[duwʌ]	two	[tʰu]	num	NULL
3	tello'	[tɛllɔʔ]	tiga	[tigʌ]	three	[θri]	num	NULL
4	empa'	[ɛmpʌʔ]	empat	[ɛmpʌt]	four	[fɔ(r)]	num	NULL
5	lèma'	[lɛmʌʔ]	lima	[limʌ]	five	[fajv]	num	NULL
6	ennem	[ənnɛm]	enam	[ɛnʌm]	six	[sɪks]	num	NULL
7	pètto'	[pɛttɔʔ]	tujuh	[tudʒuh]	seven	[sɛvɛn]	num	NULL
8	bállu'	[belluʔ]	delapan	[dɛlʌpʌn]	eight	[eɪθ]	num	NULL
9	sanga'	[sʌŋʌʔ]	sembilan	[sɛmbilʌn]	nine	[najn]	num	NULL

Gambar 1. Tampilan Data dalam Basis Data SQL

Fitur visual pada aplikasi kamus daring ini berupa gambar-gambar visual yang juga dibuat dalam berkas gambar. Setiap gambar visual dibuat dalam satu berkas. Jadi, pada ketiga bahasa ini terdapat 3000 berkas gambar. Berkas-berkas gambar tersebut juga dimasukkan ke dalam menu yang sudah tersedia dalam aplikasi kamus daring tiga bahasa ini.

Setelah proses memasukkan pangkalan data Excel, fitur pelafalan, dan fitur visual pada aplikasi kamus selesai dilakukan, tahap selanjutnya adalah mengunggah aplikasi tersebut secara daring. Setelah berhasil, kamus daring tiga bahasa (Madura-Indonesia-Inggris) dapat diakses oleh para pengguna dalam laman Universitas Panca Marga dengan

alamat *kamus3bahasa.upm.ac.id*. Adapun tampilan Kamus Daring Tiga Bahasa: Madura-Indonesia-Inggris adalah sebagai berikut.



Gambar 2. Tampilan Kamus pada Laptop



Gambar 3. Tampilan Kamus pada Telepon Pintar

IV. SIMPULAN

Penyusunan kamus daring ini merupakan salah satu upaya nyata pendokumentasian bahasa lokal guna mempertahankan dan melestarikan bahasa Madura. Kamus daring tiga bahasa ini merupakan pengembangan dari aplikasi kamus daring yang telah ada sebelumnya. Tiga bahasa itu yakni bahasa Madura, bahasa Indonesia, dan bahasa Inggris. Fitur dalam aplikasi kamus daring ini meliputi bentuk ejaan dalam ketiga bahasa, bentuk pelafalan, transkripsi fonetis, kelas kata, serta bentuk visualnya. Aplikasi kamus daring ini dibangun dengan menggunakan *web interface* dengan menggunakan bahasa pemrograman *PHP* dan *CSS Bootstrap*. Sasaran pengguna kamus ini adalah masyarakat awam penutur bahasa Madura, bahasa Indonesia, dan bahasa Inggris, baik di tingkat lokal, nasional, maupun internasional. Dalam ranah akademik, kamus ini dapat berguna dalam penelitian dan pengembangan ilmu linguistik.

V. CATATAN AKHIR

Penyusunan Kamus Daring Tiga Bahasa: Madura-Indonesia-Inggris ini merupakan kegiatan dalam program Hibah Penelitian Dosen Pemula Kemristekdikti 2018. Kegiatan ini berlangsung selama 8 bulan dari bulan April sampai dengan November 2018. Jadi, kamus ini masih dalam proses penyusunan. Aplikasi kamus daring tiga bahasa ini masih terus dilakukan perbaikan dan penyempurnaan.

Daftar Pustaka

- Andayani, Sri. & Sutrisno, Adi. 2017. PDP Bahasa Madura di Pulau Jawa, Madura, dan Bawean (Kajian Dialektologi). Laporan Penelitian. PDP Kemristekdikti.
- Djupri, M. 2008. *Kamus Suroboyoan-Indonesia*. Surabaya: Henk Publika.
- Echols, John M. & Shadily, Hassan. 2017. *Kamus Inggris-Indonesia*. Jakarta: PT. Gramedia.
- 2017. *Kamus Indonesia-Inggris*. Jakarta: PT. Gramedia.
- Fromkin, Victoria, et. al. 2014. *An Introduction to Language*. Tenth Edition. Wardsworth: Cengage learning.
- Hartmann, R.R.K. 2001. *Teaching and Researching Lexicography*. England: Pearson Education Limited.
- Kementerian Pendidikan dan Kebudayaan, Badan Pengembangan dan Pembinaan Bahasa. 2016. *Kamus Besar Bahasa Indonesia*. Edisi V. Jakarta: Badan Bahasa.
- Badan Pengembangan dan Pembinaan Bahasa. 2013. *Kamus Dwibahasa Indonesia-Madura*. Sidoarjo: Balai Bahasa Provinsi Jawa Timur.
- Badan Pengembangan dan Pembinaan Bahasa. 2014. *Tata Bahasa Bahasa Madura*. Edisi Revisi. Sidoarjo: Balai Bahasa Provinsi Jawa Timur.
- Kridalaksana, Harimurti. 2010. *Kamus Linguistik*. Jakarta: PT. Gramedia Pustaka Utama.
- Laksono, Kisyani & Savitri, Agusniar Dian. 2009. *Dialektologi*. Surabaya: Unesa University Press.
- Republika, Jakarta: 27 Oktober 2017

**PEMBENTUKAN KAMUS BAHASA JAWA BERDASARKAN
TINGKAT BAHASA UNTUK ANAK SEKOLAH (SD, SMP, SMA)
BERBASIS KORPUS
(THE PREPARATION OF CORPUS-BASED JAVANESE DICTIONARY
FOR STUDENTS [ES, JHS, SHS])**

Afwin Sulistiawati
Universitas Negeri Surabaya
afwinsw@gmail.com

Abstrak

Kamus memiliki peran penting dalam membantu seseorang memahami kosakata, apalagi kosakata dalam bahasa Jawa yang pada kenyataannya memiliki tiga tingkatan yang bahkan sebagian besar katanya berbeda. Jika seseorang akan menggunakan bahasa Jawa, seseorang itu harus memahami tingkat bahasa Jawa sehingga ia dapat menggunakan bahasa Jawa dengan baik dan benar. Tingkat bahasa Jawa itu akan dianggap penting jika bahasa Jawa berada pada ranah pendidikan, maka pebelajar harus memiliki pegangan (baca: kamus bahasa Jawa khusus) untuk mengetahui tingkat bahasa Jawa. Hal tersebut dikarenakan belum banyak kamus bahasa Jawa yang menyuguhkan tingkat bahasanya. Tujuan pembentukan kamus ini adalah untuk membantu anak sekolah agar lebih mudah memahami tingkat bahasa Jawa dalam penggunaannya. Selain itu juga untuk memudahkan proses pembelajaran bahasa Jawa.

Tahap-tahap yang digunakan dalam pembuatan kamus ini adalah (1) mengubah file buku-buku pelajaran Bahasa Jawa siswa SD, SMP, dan SMA ke dalam bentuk file txt, (2) memasukkan file txt tersebut ke dalam *Ant Cont 3.4* untuk mendapatkan kosakata yang berfrekuensi tinggi (kata yang sering muncul), (3) menyeleksi kosakata berfrekuensi tinggi sebanyak 500 kata untuk dijadikan lema dalam kamus, (4) menentukan makna, tingkat bahasa, dan beberapa fitur kamus lainnya, (5) menyusun kamus menggunakan *software Lexique Pro 3.6* dalam rangka membentuk kamus bahasa Jawa berdasarkan tingkat bahasa untuk anak sekolah. Hasil yang didapat dari tulisan ini adalah formula kamus bahasa Jawa berdasarkan tingkat bahasa untuk anak sekolah (SD, SMP, SMA) sebagai penunjang proses pembelajaran anak sekolah. Pembuatan kamus khusus ini dianggap penting karena pebelajar bahasa Jawa pasti perlu mengetahui tingkat bahasa Jawa, tidak hanya dibutuhkan dalam proses reseptif bahasa (mamahami, menyimak, dan membaca) saja, tetapi juga dalam proses produksi bahasa (berbicara dan menulis) yang sesuai dengan kaidah bahasa Jawa.

Kata kunci: kamus, tingkat bahasa, anak sekolah, korpus

Abstract

Dictionary has an important role for helping someone to understand the vocabulary, Javanese vocabulary has three levels and most of the words are different. If someone will be using Javanese language, they should understand the level of Javanese language so that they can use the Javanese language very well and correctly. Javanese language levels will be considered important if the Javanese language is in realm of education, then the learners must have a role model (a special Javanese dictionary) to know the level of Javanese language. Due to not many dictionaries of Javanese language that presenting the language level. The purpose of forming this dictionary is to help students to more easily understand the level of Javanese language in its use. In addition, to facilitate the process of learning Javanese language.

The steps that been used in making this dictionary are (1) change the Java language textbook files of elementary, junior and senior high school students into txt file, (2) insert the txt file into Ant Cont 3.4 to get the vocabulary (3) selects a high-frequency vocabulary of 500 words to be used in the dictionary, (4) determines meanings, language levels, and some other dictionary features; (5) constructs a dictionary using Lexique Pro 3.6 software in order to establish a Javanese language dictionary based on language level for students. The results obtained from this paper is a formula of Javanese language dictionary based on the level of language for students (elementary, junior, and senior high school) as a supporter of students' learning process. The making of this special dictionary is considered important because the Javanese learners will need to know the level of Javanese, not only in the receptive process of language (understanding, listening, and reading), but also in the process of producing language (speaking and writing) in accordance with the rules of the Javanese language.

Keywords: *dictionary, language level, students, corpus*

I. PENDAHULUAN

Kamus menjadi hal yang penting jika dikaitkan dengan bahasa. Suatu permasalahan bahasa dapat diselesaikan dengan menggunakan kamus sebagai sumber pemecahannya. Oleh sebab itu kamus yang baik selalu dibutuhkan oleh pengguna bahasa, baha apapun dan dimanapun, baik bahasa daerah maupun bahasa negara. Di Indonesia, bahasa Jawa merupakan salah satu bahasa tradisional namun penuturnya tergolong banyak. Hal tersebut terbukti dalam sebuah artikel Badan Bahasa yang ditulis oleh Sugiyono dengan judul “Pelindungan Bahasa Daerah dalam Kerangka Kebijakan Nasional Kebahasaan”, disebutkan bahwa di antara bahasa di Indonesia, terdapat tiga bahasa daerah yang memiliki jumlah penutur lebih dari 10 juta jiwa yakni bahasa Jawa (84,3 juta jiwa), bahasa Sunda (34 juta jiwa), dan bahasa Madura (13,6 juta jiwa).

Dibandingkan dengan bahasa-bahasa daerah yang ada di Indonesia, jumlah penutur bahasa Jawa termasuk paling banyak. Menurut data sensus 2000, penutur bahasa Jawa di Indonesia adalah sebanyak 84 juta jiwa lebih (84.300.000) (AMM, 2017). Selain jumlah penutur yang banyak, bahasa Jawa juga dijadikan sebagai mata pelajaran (muatan lokal) yang

ada di jenjang Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), dan Sekolah Menengah Atas (SMA) di Jawa Timur dan Jawa Tengah, bahkan bahasa Jawa termasuk muatan lokal yang ada di dalam kurikulum 2013. Dari hal-hal yang telah disebutkan di muka, dapat diprediksi bahwa pengguna kamus bahasa Jawa pun berada dalam jumlah yang tinggi. Namun kenyataannya, kamus bahasa Jawa yang berada di pasaran kebanyakan hanya menyuguhkan definisi kata, arti kata, atau pengalihan bahasa (misalnya dari kata berbahasa Jawa diartikan ke dalam bahasa Indonesia atau dari kata berbahasa Indonesia diartikan ke dalam bahasa Jawa).

Sebagian aplikasi kamus yang dapat dibeli atau diunduh di Play-Store telah menyuguhkan tingkat bahasa Jawa namun hanya kata-katanya saja yang disuguhkan, aplikasi kamus tersebut tidak memberikan definisi atau contoh penggunaan kata sehingga dianggap belum cukup membantu para pengguna kamus. Bahasa Jawa terdiri atas tiga tingkat bahasa, yakni bahasa Jawa Ngoko yang biasa digunakan dalam ragam santai dan cenderung kurang sopan, Krama Madya yang cenderung setengah sopan, dan Krama Inggil yang bisa dikatakan sangat sopan. Tiga tingkat bahasa ini merupakan budaya Jawa yang mengandung nilai karakter bangsa.

Pembelajaran bahasa Jawa dianggap penting karena di dalam pembelajaran bahasa Jawa terdapat penanaman karakter bangsa yang kuat, yakni dengan bahasa, manusia dapat menempatkan dirinya sesuai dengan unggah-ungguh atau nilai moral yang berlaku di masyarakat. Misalnya ketika seorang siswa berkomunikasi dengan sesama teman sebaya menggunakan bahasa Jawa Ngoko, dalam berkomunikasi dengan teman yang lebih tua menggunakan bahasa Jawa Krama Madya, dan dalam berkomunikasi dengan orang tua atau dengan gurunya menggunakan bahasa Jawa Krama Inggil. Dengan demikian, generasi penerus ini dapat tetap menjadi pelaku budaya atau bahkan pewaris budaya sesuai dengan pendapat Sapir (dalam Blount, 1974) menyatakan bahwa kandungan tiap budaya terwujud di dalam bahasanya.

Oleh sebab hal di atas, para siswa membutuhkan kamus bahasa Jawa yang dapat menjadi referensi tingkat bahasa Jawa sesuai dengan kebutuhannya. Tidak hanya itu, selain tingkat bahasa, definisi arti dan contoh kalimat pun dibutuhkan demi terwujudnya pemahaman siswa terhadap bahasa Jawa. Dengan demikian, kamus bahasa Jawa berdasarkan tingkat bahasa Jawa untuk anak sekolah perlu dibentuk dengan tujuan agar dapat dijadikan sebagai penunjang keberhasilan pembelajaran bahasa Jawa bagi para siswa. Karena itulah rumusan masalah dalam tulisan ini adalah bagaimana pembentukan kamus bahasa Jawa berdasarkan tingkat bahasa Jawa untuk siswa (SD, SMP, dan SMA).

Penelitian terkait yang relevan dengan tulisan ini adalah penelitian berjudul “Pembentukan Kamus Pewayangan Bahasa Jawa–Indonesia Berbasis Korpus” oleh Nurfarida dan Suryawati pada tahun 2016. Hasil yang didapat dari penelitian ini adalah kamus saku (*pocket dictionary*) yang berisi kosakata-kosakata bahasa yang digunakan dalam pentas pewayangan (Ika Nurfarida, 2016). Persamaan dengan tulisan ini adalah sama-sama membahas tentang kamus bahasa Jawa khusus, yakni penelitian tersebut membentuk kamus pewayangan berbahasa Jawa. Namun dalam penelitian tersebut tidak ada penjelasan tentang tingkat bahasa Jawa dalam kamusnya sehingga tulisan ini dianggap dapat melengkapi kekurangan yang ada pada penelitian terdahulu yang relevan.

II. KERANGKA TEORI DAN METODE PENELITIAN

Kamus merupakan wadah yang berisi daftar kata dengan diikuti dengan arti, definisi, dan berbagai fitur kamus yang lain. Wadah tersebut bisa berupa buku atau suatu *software* (*online dictionary*). Menurut Kridalaksana (1985) kamus adalah buku referensi yang memuat daftar kata atau gabungan kata dengan keterangan mengenai berbagai segi maknanya dan penggunaannya dalam bahasa; biasanya disusun menurut abjad (dalam tradisi Yunani-Romawi menurut urutan abjad Yunani-Romawi), kemudian menurut abjad bahasa bersangkutan; dalam tradisi Arab menurut urutan jumlah konsonan.

Jika berbicara tentang teori perkamusan, maka kebanyakan orang langsung teringat pada istilah leksikografi. Menurut Kridalaksana (1985) leksikografi merupakan satu cabang linguistik terapan dan garapannya mencakup teknik dan metodologi penyusunan kamus. Terdapat pendapat lain yakni menurut Malkiel (dalam Sariyan, 2014: 363) yang mengemukakan bahwa leksikografi, dalam makna yang sempit, merupakan subbagian linguistik terapan yang berkaitan dengan teknik mengumpulkan data leksikal dan mempersembahkannya dalam bentuk kamus. Dari dua pendapat tersebut, masuk akal jika leksikografi diartikan sebagai cabang ilmu linguistik yang mempelajari teknik penyusunan kamus.

Sehubungan dengan pembuatan kamus, berdasarkan bahasa sasaran kamus dapat dibedakan menjadi kamus ekabahasa (*monolingual dictionary*), kamus dwibahasa (*bilingual dictionary*), dan kamus aneka bahasa (*multilingual dictionary*) (Chaer, 2007). Kamus ekabahasa adalah kamus yang bahasa sumbernya sama dengan bahasa sasarannya. Contohnya adalah Kamus Besar Bahasa Indonesia (KBBI), yakni lema yang menggunakan bahasa Indonesia dijelaskan arti dan berbagai fiturnya menggunakan bahasa Indonesia juga. Kamus dwibahasa berarti kamus yang terdiri atas dua bahasa, misalnya kamus bahasa Inggris-bahasa Indonesia atau sebaliknya. Kamus aneka bahasa adalah kamus yang terdiri atas lebih dari dua bahasa, misalnya Kamus Indonesia-Arab-Inggris.

Hanks (2006) menyebutkan ada delapan fitur atau unsur yang umumnya ada di dalam entri kamus, yakni lema kepala, pengucapan, gramatika, definisi, contoh, frasa, penggunaan, dan etimologi. Kedelapan fitur tersebut dapat diatur oleh pembuat kamus dengan menggunakan *software* tertentu. Dalam tulisan ini, pembentukan kamus menggunakan *software* Lexique Pro. *Lexique Pro is an interactive lexicon viewer, with hyperlinks between entries, category views, dictionary reversal, search, and export tools. It can be configured to display your Toolbox/Shoebox database in a user-friendly format so that you can distribute it to others* (SIL, 2015). Lexique Pro merupakan penampil leksikon interaktif, dengan sistem *hyperlink* antara entri, tampilan kategori, pembalikan kamus, pencarian, dan alat-alat ekspor dalam bentuk kamus cetak ataupun kamus online. Hal ini dapat dikonfigurasi untuk menampilkan *database* dalam format yang sederhana dan memudahkan sehingga pengguna *Lexique Pro* dapat mendistribusikannya kepada orang lain.

Metode yang digunakan dalam tulisan ini adalah dengan mengumpulkan kosakata yang didapat dari buku-buku Pelajaran Bahasa Jawa siswa SD, SMP, dan SMA. Buku yang dimaksudkan adalah buku teks Pelajaran Bahasa Jawa kelas 4—6 Sekolah Dasar (SD), buku teks Pelajaran Bahasa Jawa kelas 7—9 Sekolah Menengah Pertama (SMP), dan buku teks

Pelajaran Bahasa Jawa kelas 10—12 Sekolah Menengah Atas (SMA). Buku-buku tersebut berupa buku elektronik yang kemudian *file* buku-buku Pelajaran Bahasa Jawa itu diubah ke dalam bentuk *file* txt. *File* txt tersebut kemudian dimasukkan ke dalam *Ant Cont 3.4* untuk mendapatkan kosakata yang berfrekuensi tinggi (kata yang sering muncul) sebab kosakata itulah yang dianggap penting dan dibutuhkan oleh siswa.

Dari kosakata berfrekuensi tinggi yang dapat dilihat di *Ant Cont 3.4* diambil kosakata sebanyak 500 kata untuk dijadikan lema dalam kamus. Namun pengambilan 500 kata teratas itu pun tidak asal ambil, ada seleksi khusus berupa penentuan bahwa kata-kata yang muncul adalah kata-kata yang memiliki makna secara jelas. Selanjutnya, dengan menggunakan *software* atau alat Leksikografi berupa *Lexique Pro 3.6*, fitur-fitur kamus dapat dibentuk sesuai kebutuhan pembuat kamus. Dalam tulisan ini, fitur kamus yang dipilih adalah lema, gloss, leksem, definisi, dan contoh kalimat. Secara umum, *Lexique Pro* merupakan *software* atau alat untuk membuat kamus dengan berbagai pilihan bahasa bahkan bahasa daerah. Dalam tulisan ini, bahasa yang digunakan adalah bahasa Indonesia sebagai bahasa sumber dan bahasa Jawa sebagai bahasa untuk pendefinisian atau pengartian. Metode terakhir yang perlu dilakukan adalah memasukkan semua data sesuai dengan fitur kamus yang sudah ditentukan di dalam *software Lexique Pro 3.6* dalam rangka membentuk kamus bahasa Jawa berdasarkan tingkat bahasa untuk anak sekolah yang kemudian diakhiri dengan proses ekspor data dari dalam *Lexique Pro 3.6* ke dalam bentuk kamus cetak atau kamus *online*.

III. ANALISIS PENELITIAN

Dalam proses pembelajaran Bahasa Jawa, siswa selalu membutuhkan kosakata bahasa Jawa yang sesuai dengan tingkat bahasa Jawanya namun belum banyak kamus yang menyuguhkan lema dengan turunan kosakata sesuai dengan tingkat bahasa Jawa. Biasanya siswa mencari tingkat bahasa Jawa itu di *Pepak Bahasa Jawa*, itupun kosakatanya masih terbatas, maka perlulah dibentuk kamus bahasa Jawa berdasarkan tingkat bahasa untuk anak sekolah. Pembentukan kamus dalam penulisan ini adalah kamus berbasis korpus sehingga sumber data yang digunakan adalah buku Pelajaran Bahasa Jawa yang biasa digunakan siswa dalam proses pembelajaran. Data utama berupa data korpus yang didapat dari *Ant Conc 3.4*. Dari sinilah penentuan 500 lema ditentukan, yakni kata yang berfrekuensi tinggi yang sering muncul di buku teks. Adapun munculnya lema dengan frekuensi kemunculan terbanyak dapat dilihat pada gambar berikut ini.

Rank	Freq	Word
1	5165	ing
2	4244	lan
3	3393	kang
4	3267	sing
5	3040	teks
6	2833	basa
7	1930	bisa
8	1910	iki
9	1886	a
10	1707	ora
11	1594	kanthi
12	1446	ana
13	1428	b
14	1378	apa
15	1300	tembung
16	1265	c
17	1186	d
18	1144	yen
19	1125	karo
20	1120	jawa
21	1097	utawa
22	1093	saka
23	1084	tembang
24	1083	yaiku
25	1080	dadi

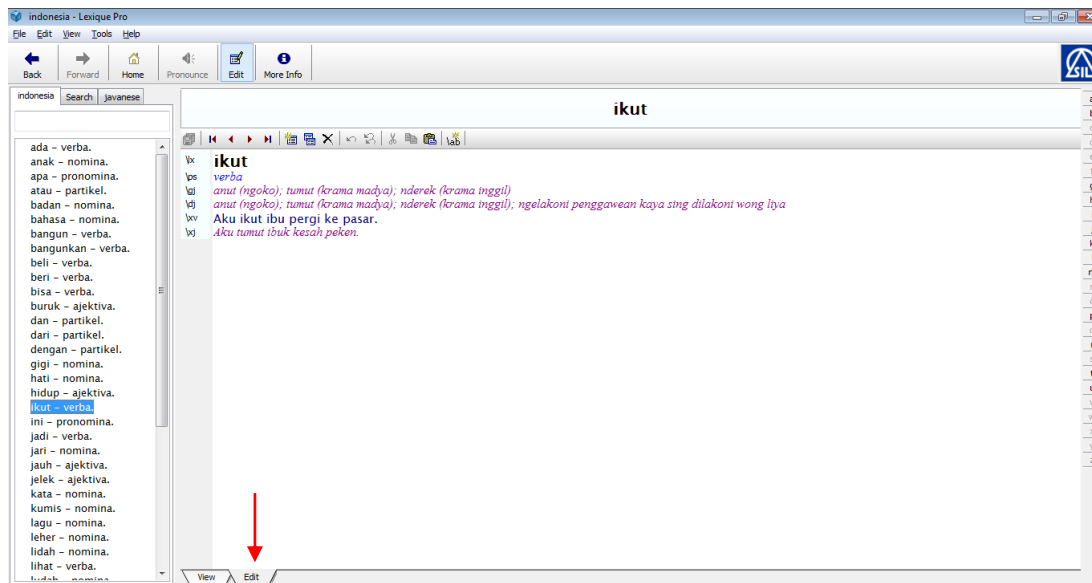
Gambar 3.1: Tampilan Kata Berfrekuensi Tinggi pada *Ant Conc 3.4*

Kosakata yang terlihat itu merupakan sampel dari keseluruhan kata yang ada dari sembilan data yang telah dimasukkan, yakni secara keseluruhan terdapat 254.131 kata yang sebenarnya terdiri atas 20.491 kosakata. Dari keseluruhan kosakata yang ada, diambil 500 kosakata yang nantinya akan dijadikan lema dalam kamus. Pengambilan kosakata teratas tersebut tidak asal ambil, namun ada penentuan terlebih dahulu untuk memastikan kata tersebut memiliki arti yang jelas. Contoh terlihat pada gambar di muka, yakni kata-kata yang ada dalam lingkaran merah dianggap kata-kata yang tidak memiliki arti yang jelas dan tidak perlu dimasukkan menjadi lema dalam kamus. 500 kosakata tersebut kemudian ditentukan tingkat bahasanya, yakni akan berubah tampilan data menjadi seperti ini.

Nmr	Bahasa Indonesia	bahasa Jawa		
		Ngoko	Krama Madya	Krama Inggil
1	Anak	Anak	Yoga	Putra
2	Bisa	Bisa	Saged	Saged
3	Gigi	Untu	Waos	Wojo
4	Mau	Gelem	Purun	Kersa
5	Pergi	Lunga	Kesah	Tindak
6	Uang	Duwit	Yatra	Artha
...

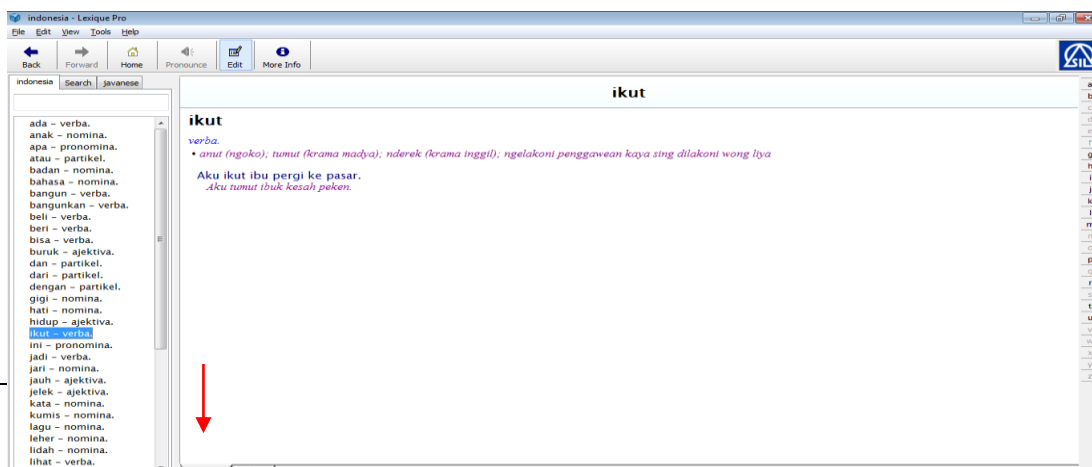
Tabel 3.1: Tampilan Kata Berfrekuensi Tinggi pada *Ant Conc 3.4*

Tabel data tersebut masih dianggap sebagai data mentah sebab data belum dimasukkan ke dalam *software* pembuat kamus. Berdasarkan beberapa data mentah di muka, dapat diketahui bahwa tidak semua kata memiliki kosakata yang berbeda pada setiap tingkat bahasa Jawanya. Hal tersebut wajar terjadi karena tidak semua kata memiliki kesantunan yang berdeda, akan tetapi ada beberapa kata yang memiliki tingkat kesantunan yang sama dengan tingkat bahasa Jawa yang lain. Data mentah tersebut kemudian dimasukkan ke dalam *software Lexique Pro* dengan memasukkan beberaps fitur yang sudah ditentukan. Hal itu dapat dilihat pada gambar berikut ini.

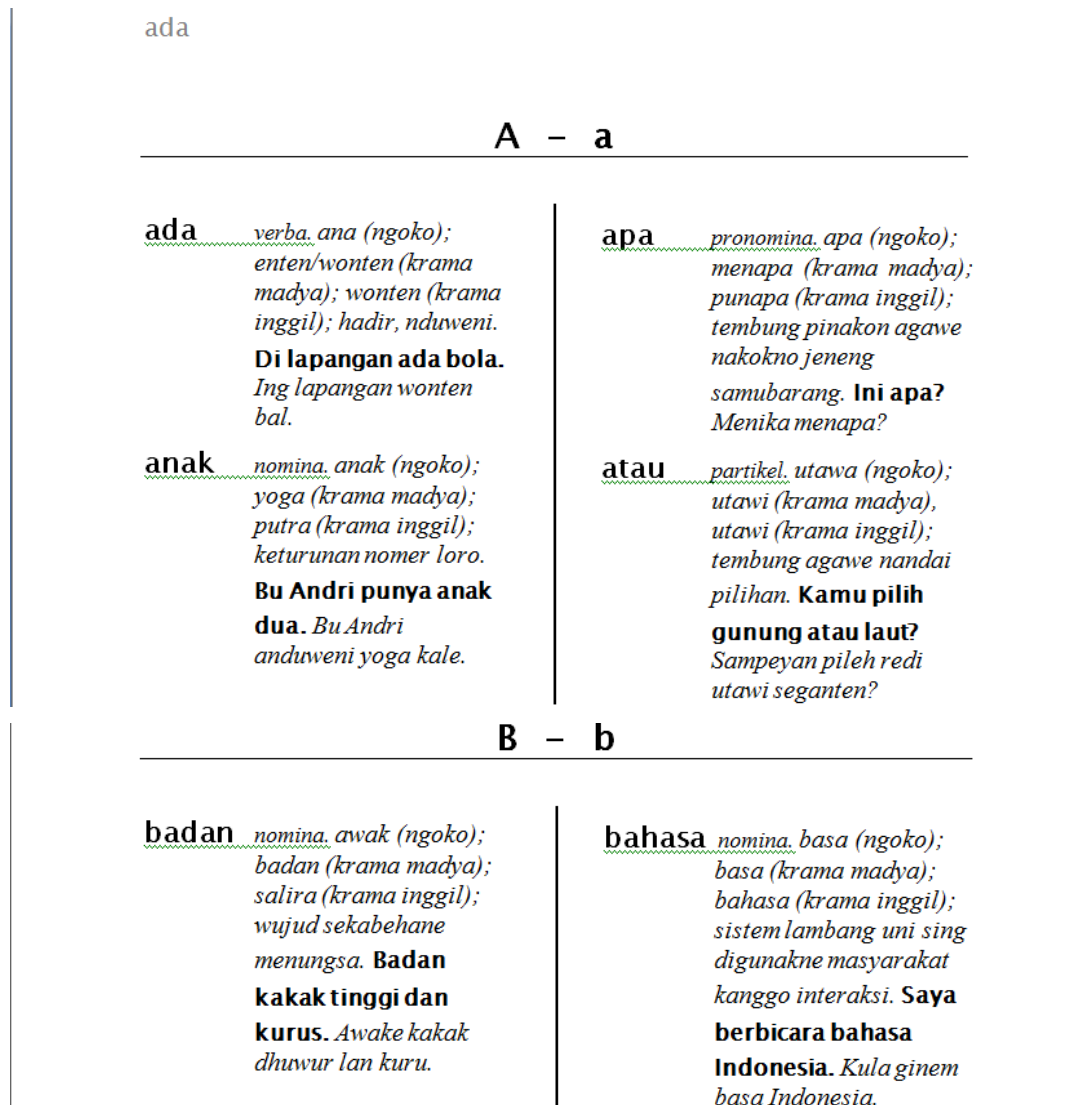


Gambar 3.2: Tampilan edit atau pemasukan data per lema pada *software Lexique Pro*

Data-data dimasukkan ke dalam *Lexique Pro* dengan menggunakan enam fitur kamus atau istilah yang digunakan dalam *Lexique Pro* adalah *marker*. *Marker* yang digunakan dalam pembentukan kamus ini adalah *lx* (*lexeme*), *ps* (*part of speech*), *gj* (*gloss in Javanese*), *dj* (*deffinition in Javanese*), *xv* (*example in Indonesian* atau contoh bahasa sumber), dan *xj* (*example in Javanese* atau contoh bahasa sasaran). Perlu diketahui contoh kalimat yang ada dalam kamus ini adalah contoh kalimat dalam tingkat bahasa Jawa Krama Madya. Setiap data dimasukkan ke dalam *Lexique Pro* dengan cara yang sama namun isi setiap fiturnya berbeda. Gambar di atas adalah wujud dari pengeditan atau pemasukan data ke dalam *Lexique Pro*. Namun jika diklik dalam pilihan *view*, maka tampilannya akan berubah menjadi seperti gambar di bawah ini.



Setelah pengisian semua data, data yang ada dalam *Lexique Pro* kemudian diekspor ke dalam bentuk kamus cetak atau kamus *online* yang tampilannya akan terlihat seperti gambar di bawah ini.



Gambar 3.4: Tampilan kamus setelah diekspor ke dalam bentuk kamus cetak (1)

Gambar di atas adalah tampilan dari satu halaman kamus setelah proses ekspor ke dalam bentuk kamus cetak ukuran A5. Pada *software Lexique Pro*, terdapat pilihan kamus berwarna atau monokrom, namun pada pembuatan kamus ini telah dipilih kamus dengan warna monokrom. Di halaman itu terdapat beberapa lema yang memiliki huruf awal A dan B, sedangkan gambar berikut adalah tampilan halaman selanjutnya.

<p>bangun</p> <p>bangun <i>verba. tangi (ngoko); tangi (krama madya); wungu (krama inggil); ngadhek utawa lungga soko turu. Adik menangis ketika bangun tidur. Adik nangis wektu tangi turu.</i></p> <p>bangunkan <i>verba. gugah (ngoko); gigah (krama madya); wungu (krama inggil); nggarai wong liyo ngadhek utawa lungga soko turu. Aku membangunkan ibu karena ada tamu. Aku gugah ibu amerga ana dayo.</i></p> <p>beli <i>verba. tuku (ngoko); tumbas (krama madya); mundhut (krama inggil); olehi apa-apa amerga ditukar nganggo duwit. Aku beli baju baru ketika lebaran. Aku tuku klambi anyar wektu riyaya.</i></p>	<p>beri <i>verba. awèh (ngoko); sukani (krama madya); paringi (krama inggil); nyerahake apa-apa kanggo wong liya. Ada teman yang suka memberi. Ana kanca sing seneng awèh-awèh.</i></p> <p>bisa <i>verba. bisa (ngoko); saged (krama madya); saged (krama inggil); kuasa agawe ngelakoni pekerjaan. Dono bisa memasak nasi. Dono bisa masak sego.</i></p> <p>buruk <i>ajektiva. ala (ngoko); awon (krama madya); awon (krama inggil); ora kepenak didhelok. Kelakuan buruk tidak boeh dilakukan. Kelakuan ala ora oleh dilakoni.</i></p>
--	---

Gambar 3.5: Tampilan kamus setelah diekspor ke dalam bentuk kamus cetak (2)

Gambar di muka merupakan wujud kamus bahasa Jawa berdasarkan tingkat bahasa Jawa untuk anak sekolah (SD, SMP, SMA). Kamus ini tidak hanya berisi padanan kata yang sesuai dengan tingkat bahasa Jawa saja, tetapi juga dilengkapi dengan kelas kata, definisi makna, contoh kalimat dalam bahasa Indonesia, dan contoh kalimat dalam bahasa Jawa. Hal tersebut dimaksudkan untuk menunjang pembelajaran siswa agar tujuan pembelajaran dapat tercapai dengan baik.

IV. SIMPULAN

Pembentukan kamus bahasa Jawa berdasarkan tingkat bahasa Jawa untuk anak sekolah yang berbasis korpus dapat membantu para siswa, pengguna kamus, dan pengguna bahasa Jawa untuk menemukan tingkat bahasa Jawa secara mudah, urut, dan rinci. Kamus tersebut juga merupakan suatu sarana pembelajaran untuk membantu mencapai tujuan pembelajaran siswa. Selain itu, pembentukan kamus ini dapat dikategorikan sebagai proses pelestarian bahasa daerah di Indonesia yang di dalamnya terkandung karakter dan budaya kesantunan bahasa yang dimiliki masyarakat Indonesia, khususnya masyarakat Jawa.

Daftar Pustaka

- AMM. (2017, Oktober 30). *SINDONEWS.com*. Retrieved Juli 1, 2018, from SINDONEWS.com: <https://nasional.sindonews.com/read/1252853/15/ini-dia-10-bahasa-daerah-dengan-penutur-terbanyak-di-indonesia-1509344321/>
- Blount, B. G. (1974). *Language Culture and Society*. Cambridge: Winthrop.
- Chaer, A. (2007). *Leksikologi & Leksikografi Indonesia*. Jakarta: Rineka Cipta.
- Hanks, P. (2006). Lexicography: Overview. Dalam *Encyclopedia of Language and Linguistics*. Elsevier .
- Ika Nurfarida, N. S. (2016). Pembentukan Kamus Pewayangan Bahasa Jawa-Indonesia Berbasis Korpus. *Seminar Leksikografi Indonesia* (p. 165). Jakarta: Badan Bahasa Kemdikbud.
- Kridalaksana, H. (1985). *Tata bahasa deskriptif bahasa Indonesia: Sintaksis*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Sariyan, A. (2014). *Teras Pendidikan Bahasa Melayu*. Selangor: PTS Malaysia.
- SIL. (2015, Oktober 5). *SIL Language Technology*. Retrieved Juli 10, 2016, from SIL Language Technology: http://www-01.sil.org/computing/catalog/show_software.asp?id=92

**PERANCANGAN KAMUS SAKU REGISTER HAJI DAN UMRAH
BERBASIS KORPUS MENGGUNAKAN APLIKASI
LEXIQUE PRO**

*(DESIGNING POCKET DICTIONARY REGISTER HAJJ AND UMRAH
BASED CORPUS USING LEXIQUE PRO APPLICATION)*

Muhammad Ridwan, S.S., M.A.

Azmil Fikri Abdil Haq

Universitas Sebelas Maret Surakarta

azmilfikri@student.uns.ac.id

ABSTRAK

Penelitian ini bertujuan merancang sebuah kamus saku yang berisikan kosakata register haji dan umrah dengan metode korpus data. Data diambil dari buku “Tuntunan Manasik Haji dan Umrah” milik Kemenag Indonesia dan buku “*Dalīl Al-Hājj Wal-Mu‘Tamir*” karya Tholal bin Ahmad Al-Aqil yang diterbitkan oleh Kementerian Urusan Agama dari Kerajaan Saudi Arabia.

Metode penyediaan data yang digunakan adalah metode simak dengan teknik dasar teknik sadap. Metode analisis data yang digunakan adalah metode padan translasional dengan teknik hubung banding menyamakan hal pokok dan teknik lanjutan teknik pilah unsur penentu. Daya pilah yang digunakan ialah daya pilah translasional.

Hasil penelitian ini menunjukkan tentang pengembangan korpus kosakata register haji dan umrah berserta pengaplikasiannya dalam bentuk perancangan kamus saku dwibahasa (Arab-Indonesia). Pengembangan korpus kosakata register haji dan umrah memiliki beberapa tahapan. Pertama, pengolahan objek data dengan beberapa proses yaitu: pengalihan teks, pengoperasian *AntConc*, penghitungan jumlah dan frekuensi kosakata. Kedua, penyepadanan kosakata berdasarkan Kamus Besar Bahasa Indonesia. Ketiga, analisis pola serapan kamus saku register haji dan umrah. Keempat, klasifikasi waktu beserta frekuensi kosakata register haji dan umrah dalam *Leipzig Corpora*. Perancangan kamus saku register haji dan umrah menggunakan aplikasi *Lexique Pro* yang ditampilkan dalam versi media cetak.

Kata kunci: kamus saku, register, korpus, haji umrah, dan *Lexique Pro*, *AntConc*, *Liepzig Corpora*.

ABSTRACT

This study aims to design a pocket dictionary containing the vocabulary register hajj and umrah with the method of data corpus. The data is taken from Dalīl Al-Hājj Wal-Mu‘Tamir book by Tholal bin Ahmad Al-Aqil published by the Ministry of Religious Affairs of

the Kingdom of Saudi Arabia and the book of the hajj and Umrah guide of the Ministry of Religious Affairs of Indonesia.

The method of providing data used is the method refer to the basic techniques of tapping techniques. The method of data analysis used is the method of translational padan with the apparatus of appeal to equate the principal and advanced techniques of the technique of the decisive element. The power used is translational.

The results of this study show about the development of the vocabulary system of Hajj and Umrah register and its application in the form of bilingual pocket dictionary (Arab-Indonesia). The development of the vocabulary of the Hajj and Umrah registers has several stages. First processing of data objects with several processes namely: text transfer, AntConc operation, counting the number and frequency of vocabulary. Second, matching vocabulary based on Big Indonesian Dictionary. Third, the pattern analysis of the pocket dictionary register Hajj and Umrah. Fourth, the classification of time along with the frequency of the vocabulary of Haj and Umrah register in Leipzig Corpora. The design of pocket dictionaries Haj and Umrah register using Lexique Pro application that can be displayed in the form of electronic media supported by the application or in the form of print media.

Keywords: *pocket dictionary, register, corpus, Hajj Umrah, and Lexique Pro*

I. PENDAHULUAN

Bahasa merupakan sistem lambang bunyi yang dipergunakan oleh para anggota suatu masyarakat untuk bekerja sama, berinteraksi, dan mengidentifikasikan diri (Kridalaksana, 2008:24). Ilmu linguistik memiliki beberapa cabang kajian interdisipliner, diantaranya adalah sosiolinguistik. Termasuk didalam ilmu sosiolinguistik adalah ilmu tentang register yang merupakan bagian dari variasi bahasa. Chaer (2004: 68) mengemukakan bahwasanya variasi bahasa berkenaan dengan penggunaannya, pemakaiannya atau fungsinya disebut fungsiolek atau register.

Haji adalah rukun islam yang kelima setelah syahadat, shalat, zakat, dan puasa (Rouf, 2012: 122). Sedangkan, Umrah adalah salah satu kegiatan ibadah dalam agama Islam yang hampir mirip dengan ibadah haji dan dilakukan dengan menjalankan beberapa ritual ibadah di kota Mekah (Lukmanulhakim, 2015: 7). Setiap tahun jumlah jamaah haji Indonesia selalu bertambah. Tercatat pada tahun 2017 dalam situs resmi milik kemenag (kemenag.go.id) bahwasanya jumlah kuota haji Indonesia mencapai 221.000 jamaah. Begitu pula pada ibadah umrah yang tercatat mencapai 800.000 jamaah yang berangkat pada tahun 2016 (hidayatullah.com). Kegiatan tersebut membutuhkan pemahaman leksikon yang berkaitan dengannya.

Bahasa sebagai alat untuk menyampaikan informasi memiliki peran yang sangat penting bagi manusia. Komunikasi yang dilakukan dengan bahasa pun memiliki ragam yang berbeda-beda. Keragaman tersebut ditentukan dengan kepentingan dan tujuan pemakaian dari suatu bahasa atau disebut register bahasa. Salah satu kepentingan yang mempengaruhi keragaman bahasa adalah pada kegiatan haji dan umrah.

Perkembangan ilmu bahasa (linguistik) yang sangat pesat berdampak positif bagi perkembangan cabang kajian ilmu bahasa lainnya. Salah satunya adalah pada ilmu perkamusan atau leksikologi dan leksikografi. Perkembangan tersebut ditandai dengan banyaknya muncul berbagai jenis kamus.

Sebagai bukti dari wujud perkembangan ilmu linguistik adalah adanya sistem CADC telah banyak digunakan dalam dunia perkamusan di dunia, salah satu diantaranya adalah aplikasi *Lexique Pro*. Selain CADC, terdapat pula ilmu korpus yang dapat membantu dalam pembuatan kamus khususnya pada pengumpulan, penentuan kosakata yang menjadi lema, pembelajaran dan penguasaan suatu bahasa.

Dari ulasan latarbelakang ditemukan beberapa rumusan masalah yang diuraikan dalam penelitian ini yaitu: cara mengembangkan korpus kosakata register haji dan umrah dan cara merancang kamus saku register haji dan umrah dengan menggunakan aplikasi *Lexique Pro*. Manfaat pada penelitian ini adalah untuk mengetahui cara mengembangkan korpus kosakata register haji dan umrah dan cara merancang kamus tersebut dengan menggunakan aplikasi *Lexique Pro*.

II. KERANGKA TEORI DAN METODE PENELITIAN

Penelitian ini menggunakan teori sebagai berikut:

1. Register

Variasi bahasa terjadi akibat dari penutur bahasa dan juga kegiatan interaksi sosial didalam masyarakat yang bersifat heterogen. Sebab-akibat tersebut menjadikan variasi bahasa dapat diklasifikasikan berdasarkan adanya keragaman sosial dan fungsi kegiatan di dalam masyarakat sosial (Chaer dan Agustina, 2004: 61-62).

2. Korpus Bahasa

Menurut Budiwiyanto dalam artikelnya yang berjudul “Korpus dalam Penyusunan Kamus” (2014) mengatakan bahwa korpus adalah kumpulan teks yang bersifat alami yaitu yang diproduksi dan digunakan secara wajar dan tidak dibuat-buat, baik dalam bentuk lisan maupun tulisan yang disusun secara sistematis. Menurut Enery (2012: i) linguistik korpus adalah studi tentang analisis bahasa dalam skala yang besar dengan bantuan komputer yang berasal dari kumpulan ucapan tertulis atau teks tertulis.

3. Kamus

Kamus adalah “sebuah karya yang berfungsi sebagai referensi pada umumnya berupa senarai kata yang disusun secara alfabetis dan disertakan informasi mengenai ejaan, pelafalan, kelas kata, makna kata, kadang kala sejarah kata, dan contoh pemakaian kata dalam kalimat” (Kushartanti 2007: 223). Menurut Chaer kamus dikelompokkan menjadi 3 (tiga): berdasarkan bahasa sasaran, berdasarkan ukurannya, dan berdasarkan isinya. Berdasarkan ukurannya kamus terbagi menjadi dua: kamus besar dan kamus terbatas (Chaer 2007: 198).

4. Penyerapan bahasa Arab ke dalam bahasa Indonesia

Teori yang digunakan dalam proses penyerapan bahasa Arab ke dalam bahasa Indonesia menggunakan teori milik Erwina Burhanuddin (1993). Beliau berpendapat bahwasanya penyerapan bahasa Arab kedalam bahasa Indonesia memiliki beberapa cara yaitu: Penyerapan penuh yang berarti terjadi pemertahanan bahasa Arab yang diserap kedalam bahasa Indonesia baik pada fonem seperti fonem /ب/ yang diserap menjadi /b/, maupun pada katanya seperti kata ادب diserap menjadi /adab/. Kedua, Penyesuaian Fonem yaitu sebuah proses penyerapan bahasa Arab dengan disesuaikannya fonem-fonem Arab. Seperti fonem /ص/ dan /ث/ di serap sama menjadi /s/. Ketiga, penyesuaian Lafal yaitu proses penyerapan dengan disesuaikannya lafal-lafal Arab dikarenakan terdapat perbedaan antara pola suku kata bahasa Indonesia dengan pola suku kata bahasa Arab.

5. Perancangan Kamus

Praktik perancangan kamus menggunakan aplikasi *Lexique Pro*. Kelebihan aplikasi tersebut adalah pada penggunaannya yang mudah dan fitur yang lengkap. Sebelum dimasukkan dalam aplikasi *Lexique Pro* diperlukan beberapa proses tahapan seperti berikut:



III. ANALISIS PENELITIAN

3.1. Pengembangan Korpus Kosakata Register Haji Dan Umrah

Proses pengembangan korpus register haji dan umrah diambil dari 2 (dua) buku yang menjadi data primer. Yaitu, buku “Tuntunan Manasik Haji dan Umrah” yang diterbitkan oleh Kemenag RI dan buku “*Dalil Al-Hajj Wal Mu'tamir*” karya Tholal bin Ahmad Al-Aqil yang diterbitkan oleh Kementerian Urusan Agama dari Kerajaan Saudi Arabia.

a.) Proses Korpus

1.) Buku “Tuntunan Manasik Haji dan Umrah”

Proses korpus buku “Tuntunan Haji dan Umrah” mencakup beberapa hal yaitu: proses pengalihan teks, proses pengoperasian *AntConc*, proses penghitungan jumlah kosakata dan frekuensinya melalui *AntConc*.

Proses pengalihan teks diperlukan agar teks dapat diuraikan dan diproses oleh aplikasi *AntConc*. Pengalihan teks memiliki 2 (dua) proses yaitu:

Pengalihan teks yang berformat *PDF* menjadi teks yang berformat *Docx*

Buku “Tuntunan Manasik Haji dan Umrah” Kemenag Indonesia diunduh dalam format *PDF* di situs resmi Kemenag Indonesia kemenag.go.id. Kemudian *file* dikonversi menjadi format *Docx* atau *ms.word* menggunakan aplikasi *UniPDF*. Berikut adalah gambar hasil *file PDF* yang telah dikonversi menjadi format *Docx*:

Pengalihan Teks Berformat *Docx* Menjadi Teks Berformat *TXT*

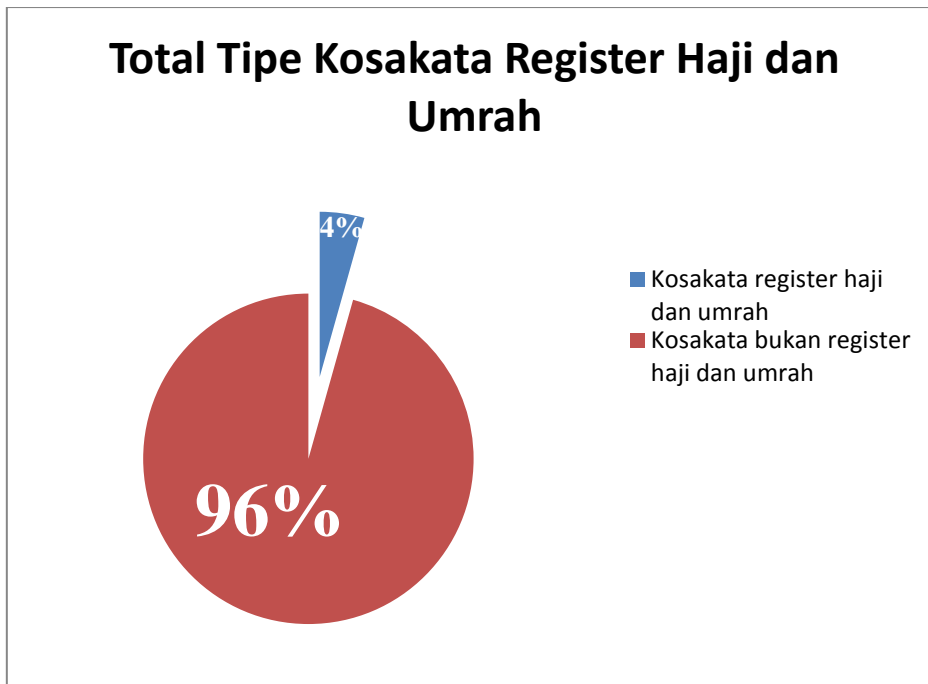
Pengalihan teks menggunakan microsoft word, dengan cara menyimpan *file* buku Tuntunan Haji dan Umroh Kemenag RI dengan *Plain Text (*TXT)* yang terdapat pada menu “*Save as type*”. Kemudian klik “*save*”, maka *file* akan tersimpan dalam format *TXT* yang hanya dapat dibuka dengan aplikasi *Notepad*. Seperti gambar berikut:

- a) Proses Pengoperasian Teks dengan aplikasi *AntConc*
Proses ini dilakukan untuk mendapatkan hasil korpus yang diinginkan. Pilih menu “*word list*” pada aplikasi *AntConc* kemudian pilih “*file*” kemudian klik “*start*”. Maka akan muncul hasil seperti gambar berikut:
- b) Setelah teks dioperasikan pada *AntConc*, proses selanjutnya adalah penghitungan jumlah kosakata dan frekuensinya. Hasil dapat dilihat pada kolom *word list* berupa jumlah tipe kata (*word types*), jumlah keseluruhan kata (*word tokens*) dan penjelasan secara rinci frekuensi kemunculan setiap kosakata.

Tipe Kata	Jumlah Kata
4056	30560

Tabel 1: Hasil dari *AntConc*.

Hasil kata yang di scan oleh peranti lunak *AntConc* sebanyak 4056 tipe kata, sedangkan jumlah keseluruhan kata yang berhasil di scan oleh peranti lunak tersebut sebanyak 30560. Dari 4056 tipe kata tersebut terdapat 169 kosakata yang termasuk kosakata register haji dan umrah atau sekitar 4%.



Gambar 1: Diagram kosakata register haji dan umrah buku “Tuntunan Manasik Haji dan Umrah” yang diterbitkan Kemenag RI.

Tabel berikut menyajikan kosakata register haji dan umrah dari buku “Tuntunan Manasik Haji dan Umrah” yang diterbitkan Kemenag RI. Akan tetapi untuk menghemat halaman hanya ditampilkan 3 kosakata saja

No	Kosakata dalam Bahasa Indonesia	Kosakata dalam Bahasa Arab
1	Air Zamzam	ماء زمزم
2	Armina	عرفة مزدلفة و منى
3	Asrama Haji	سكن الحجاج

Tabel 2: Daftar kosakata register haji dan umrah yang terdapat dalam buku “Tuntunan Manasik Haji dan Umrah”.

- c) Pengecekan frekuensi setiap kosakata register haji dan umrah berdasarkan dari hasil korpus melalui peranti lunak *AntConc* dengan pengetikan secara manual satu persatu pada menu *concordance* yang terdapat pada peranti lunak *AntConc* seperti gambar berikut:

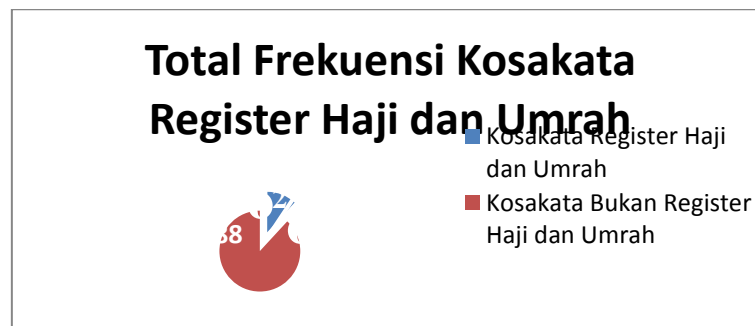
Gambar di atas merupakan proses pengecekan frekuensi kemunculan kosakata. Data yang menjadi contoh adalah kata /thawaf/. Kolom *Concordance Hit* menunjukkan terdapat 210 jumlah kosakata /thawaf/ yang muncul. Sedangkan kolom *KWIC* sebagai bukti kemunculan kosakata tersebut beserta kalimat tempat dimana terdapat kata /thawaf/.

Penyajian hasil dari pengecekan frekuensi kemunculan kosakata register haji dan umrah buku “Tuntunan Manasik Haji dan Umrah” adalah sebagai berikut. Namun sama seperti tabel sebelumnya untuk menhemat halaman hanya ditampilkan 3 kosakata saja.

No	Kosakata dalam Bahasa Indonesia	Kosakata dalam Bahasa Arab	Frekuensi
1	Air Zamzam	ماء زمزم	4
2	Armina	عرفة مزدلفة و منى	8
3	Asrama Haji	سكن الحجاج	23

Tabel ٧: Hasil pengecekan frekuensi kemunculan kosakata register haji dan umrah buku “Tuntunan Manasik Haji dan Umrah” yang diterbitkan Kementerian Agama RI.

Jika dikalkulasi total dari seluruh frekuensi kosakata terdapat 3577 kosakata register haji dan umrah dari 30560 kosakata dalam buku “Tuntunan Manasik Haji dan Umrah” yang diterbitkan Kemenag RI. Setara dengan 12% kosakata untuk Kamus Saku Register Haji Dan Umrah yang terdapat dalam buku tersebut.



Gambar ٧: Diagram total frekuensi kemunculan kosakata register haji dan umrah buku” Tuntunan Manasik Haji dan Umrah”.

2.) Buku “*Dalīl Al-Hajj Wal Mu’tamir*”.

Buku “*Dalīl Al-Hajj Wal Mu’tamir*” milik Kementerian Urusan Agama, Waqaf, Da’wah dan Penyuluhan Kerajaan Saudi Arabia menggunakan bahasa Arab dalam versi *file* yang berformat *PDF*. *File PDF* tersebut pun berbeda, dikarenakan tulisan dalam *file PDF* tersebut tidak tercetak dengan jelas. Sehingga *file* tidak dapat di olah dan diekstrak ke dalam bentuk lain. Sedangkan, aplikasi *AntConc* membutuhkan *file* yang berformat *.TXT*. oleh karena itu pembinaan data korpus yang dilakukan pada buku tersebut menggunakan cara manual. Pemilihan kosakata yang merupakan register haji dan umrah dan penghitungan jumlah frekuensi kemunculan kosakata dilakukan secara manual.

a.) Pemilihan Kosakata Register Haji dan Umrah

Kosakata dicari dengan cara manual, yaitu dengan membaca, memilah dan menandai disetiap halamannya. Terdapat 79 kosakata register haji dan umrah dari 80 halaman yang terdapat dalam buku tersebut. Tabel berikut menyajikan kosakata register haji dan umrah buku “*Dalīl Al-Hajj Wal Mu’tamir*” milik Kementerian Urusan Ke-Islaman, Waqaf, Da’wah dan Penyuluhan Kerajaan Saudi Arabia. Akan ditampilkan 5 contoh saja.

No	Kosakata	Transliterasi
1	عرفة	‘Arafah
2	ثياب الإحرام	Tsiyābul-Ichrām
3	بقيع	Baqī’

Tabel ٣: Kosakata register haji dan umrah buku “*Dalīl Al-Hajj Wal Mu’tamir*” milik Kementerian Urusan Ke-Islaman, Waqaf.

b) Penghitungan Jumlah Frekuensi Kemunculan Kosakata

Penghitungan dilakukan secara manual dengan menghitung kemunculan kosakata di setiap halamannya. Tabel berikut menyajikan hasil dari pengecekan frekuensi kemunculan kosakata register haji dan umrah buku “*Dalīl Al-Hajj Wal Mu’tamir*” milik Kementerian Urusan Ke-Islaman, Waqaf, Da’wah dan Penyuluhan Kerajaan Saudi Arabia. Akan tetapi sama seperti tabel sebelumnya untuk menhemat halaman hanya ditampilkan 3 contoh saja.

No	Kosakata	Transliterasi	Frekuensi
1	عرفة	‘Arafah	23
2	ثياب الإحرام	Tsiyābul-Ichrām	5
3	بقيع	Baqī’	1
	Jumlah Total Frekuensi		34

Tabel ٤: Tabel jumlah frekuensi kosakata register haji dan umrah dalam buku “*Dalīl Al-Hajj Wal Mu’tamir*”

Tabel diatas menunjukkan total kemunculan atau frekuensi kosakata register haji dan umrah adalah 994 kali dalam buku “*Dalīl Al-Hajj Wal Mu’tamir*” milik Kementerian Urusan Ke-Islaman, Waqaf, Da’wah dan Penyuluhan Kerajaan Saudi Arabia.

3.) Gabungan Kosakata Dan Padanannya

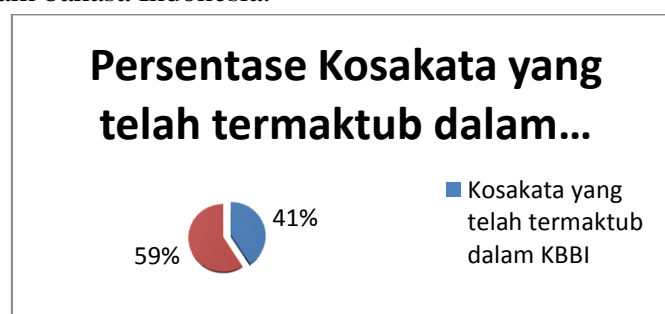
Kosakata yang diambil untuk menjadi lema dalam kamus saku register haji dan umrah merupakan kosakata pilihan dari 2 (dua) buku panduan haji milik kemenag RI

dan Kementerian Urusan Ke-Islaman, Waqaf, Da'wah dan Penyuluhan Kerajaan Saudi Arabia yang digabungkan. Penggabungan dilakukan untuk mengeliminasi kosakata yang sama pada 2 (dua) data diatas. Kemudian kosakata dipadankan dengan bahasa Indonesia untuk mengetahui kata-kata yang telah diserap menjadi bahasa resmi bahasa Indonesia. Oleh karena itu, data yang ada dipadankan dengan kosakata yang terdapat dalam Kamus Besar Bahasa Indonesia (KBBI) untuk mengetahui seberapa banyak kosakata haji dan umrah yang telah dibakukan menjadi bahasa resmi. Dari gabungan kosakata tersebut didapatkan hasil 187 kosakata register haji dan umrah. Untuk menghemat tabel hanya menyajikan 3 dari kosakata register haji hasil penggabungan dari 2 (dua) buku:

	Kosakata	Kosakata dalam Bahasa Arab	Transliterasi	Padanan Kosakata Dalam KBBI
1	Air Zamzam	ماء زمزم	<i>Mā'u Zamzam</i>	Air Zamzam
2	Arafah	عرفة	<i>'Arafah</i>	-
3	Armina	عرفة و منى	<i>'Arafah Wa Minā</i>	-

Tabel ٦: Tabel kosakata register haji dan umrah beserta padanannya dalam KBBI.

Hasil tersebut juga menunjukkan bahwasanya terdapat 76 kosakata yang telah termaktub dalam KBBI dan 111 kosakata yang belum termaktub dalam KBBI. Dilihat dari total tersebut maka, terdapat 40,641711% atau setara dengan 41% kosakata yang telah dibakukan dalam bahasa Indonesia.



Gambar ٤: Diagram persentase kosakata yang telah termaktub dalam KBBI.

Jika diperhatikan lagi sebenarnya masih banyak kosakata tidak lagi di adaptasi, diterjemahkan ataupun dikresasi oleh masyarakat, melainkan kosakata-kosakata tersebut telah diadopsi oleh masyarakat yang artinya kosakata tersebut diambil bentuk dan maknanya secara keseluruhan. Seperti, kata mabit, tahalul awal, tahalul tsani dan lain sebagainya.

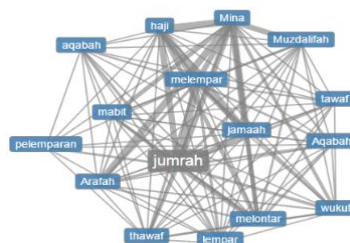
4.) Analisis Pola Serapan Lema Kamus saku register haji dan umrah

Analisis pola serapan lema kamus saku register haji dan umrah dilakukan untuk mengetahui sejauh mana kata dari bahasa Arab yang diserap kedalam bahasa Indonesia. Dalam hal ini, ditemukan beberapa kosakata dalam Kamus Saku Register Haji Dan Umrah yang telah diserap menjadi bahasa Indonesia dan telah dibakukan dan dibukukan dalam Kamus Besar Bahasa Indonesia (KBBI) khususnya pada edisi ke V. Adapun data dan uraian penjelasan berdasarkan teori yang disampaikan oleh Burhanudin Erwina,dkk. Dari data tersebut ditemukan 6 (enam) pola penyerapan kosakata, dengan rincian 6 (enam) kosakata dengan pola penyerapan sebagian, 8 (delapan) kosakata dengan pola penyerapan penuh, 14 (empat belas) kosakata dengan pola penyesuaian Lafal, 9 (sembilan) kosakata dengan pola penyerapan penyesuaian fonem, 39 (tiga puluh sembilan) kosakata dengan pola penyerapan penyesuaian fonem+penyesuaian lafal, dan 1 (satu) kosakata dengan pola penyerapan penyesuaian fonem+penyerapan penuh.

5.) Klasifikasi dan Frekuensi Kosakata Register Haji dan Umrah dalam Leipzig Corpora

Klasifikasi kosakata diperlukan untuk mempermudah masyarakat dalam penggunaan kamus. Kemudian frekuensi dalam Leipzig Corpora berfungsi untuk mengetahui bentuk kolokasi dan keterkaitan kosakata register haji dan umrah dengan kata haji dan umrah. Klasifikasi dibagi menjadi 3 (tiga) tema besar berdasarkan waktu pelaksanaannya. Yaitu, pra haji umrah, haji umrah dan pasca haji umrah. Dewasa ini *Leipzig Corpora* dapat dijadikan sebagai tolak ukur untuk kelayakan sebuah kosakata dalam ranah internasional.

Untuk dapat melihat keterkaitan dan kolokasi sebuah kata dengan kata lainnya *Leipzig Corpora* menyediakan grafik yang terletak halaman paling bawah. Terdapat 3 (tiga) jenis kolokasi yang ditemukan pada kosakata register haji dan umrah dalam *Leipzig Corpora* yaitu, kosakata yang memiliki kolokasi dan hubungan dengan haji dan umrah, kosakata yang memiliki kolokasi dan hubungan dengan haji dan umrah secara tidak langsung dan kosakata yang tidak memiliki kolokasi dan hubungan dengan haji dan umrah. Seperti contoh berikut:



Gambar 5: Kolokasi dari kata /Jumrah/ dalam *Leipzig Corpora*

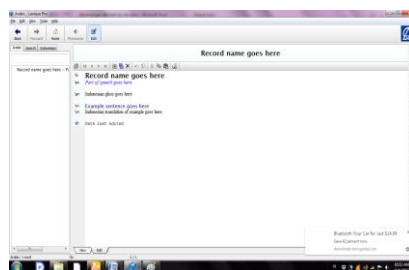
Pada gambar 5 menunjukkan bahwasanya kata /jumrah/ berkolokasi dengan kata /haji/. Hal tersebut menunjukkan bahwasanya kata /jumrah/ memiliki keterkaitan atau hubungan dengan kata /haji/.

Perancangan Kamus Saku Register Haji dan Umrah dengan Menggunakan Aplikasi *Lexique Pro*

Aplikasi *Lexique Pro* digunakan sebagai alat untuk membuat kamus saku register haji dan umrah. Penggunaan aplikasi *Lexique Pro* dalam pembuatan kamus saku register haji dan umrah dikarenakan aplikasi ini memiliki cara kerja yang mudah dan dapat membantu dalam pembuatan kamus dengan lema yang lengkap.

Software ini merupakan alat khusus untuk membuat kamus dengan berbagai macam pilihan bahasa. Bahasa sumber pada penelitian ini adalah bahasa Arab yang akan dialihbahasakan kedalam bahasa Indonesia. Langkah berikutnya adalah dengan memilih menu *file* dan klik “*create New Lexicon*” untuk membuat proyek kamus baru. Kemudian klik “*next*”, kemudian muncul tampilan untuk memasukan nama kamus yang akan dibuat. Penelitian ini menggunakan nama “Kamus saku register haji dan umrah” sebagai nama dari kamus yang akan dibuat. Kemudian pada bar ‘*Language Code*’ diisi dengan “*arabic*”. Untuk mendapatkan kata “*arabic*” dilakukan dengan memilih tombol “*Find Code*” kemudian bisa mencarinya secara manual atau langsung diketik pada menu “*search*”.

Kemudian akan disuguhkan dengan berbagai pilihan bahasa. Pilih menu bahasa Arab dan bahasa Indonesia. Kemudian klik tombol “*next*” hingga muncul tampilan pemberitahuan bahwasanya *Lexique Pro* melambangkan definisi bahasa Indonesia dengan lambang “n” kemudian klik ‘*next*’. Kemudian klik “*next*” lalu akan muncul tampilan layar “*Field Markes*” lalu klik “*next*” lalu muncul tampilan layar “*Filename*” untuk memberikan nama pada proyek kamus yang dirancang lalu klik “*next*”. Kemudian akan muncul perintah untuk memilih membuat shortcut pada desktop lalu dicentang untuk mempermudah dalam pencarian proyek yang telah dibuat. Kemudian klik “*next*” lalu klik tombol “*create*”. Kemudian akan muncul calon proyek kamus yang telah siap di entrikan kosakata. Seperti gambar dibawah ini yang menunjukkan bahwasanya basis data sudah siap dientrikan lema kamus.



Gambar 1: Tampilan awal perancangan

Lexique Pro menyediakan berbagai macam pilihan kode untuk memaksimalkan proyek dalam membuat suatu kamus. Kamus saku register haji dan umrah memilih beberapa kode yang masing-masing memiliki fungsi tersendiri, dengan uraian sebagai berikut: *Lexicon* (lx) untuk diisi leksikon haji dan umrah. *Gloss* (gr) merupakan sebuah kode untuk mengisi terjemah leksikon haji dan umrah atau leksikon dalam bahasa Indonesia. *Phonetic form* (ph) untuk diisi transliterasi leksikon haji dan umrah, *Part of Speech* (ps) merupakan sebuah kode untuk mengisi kelas kata. Akan tetapi, fungsi (ps) difungsikan untuk mengisi jenis klasifikasi

kata berdasarkan 3 (tiga) waktu yaitu: pra haji dan umrah, haji dan umrah dan pasca haji dan umrah. *Encyclopedic Information* (ev) untuk diisi dengan definisi leksikon haji dan umrah. *Picture* (pc) merupakan sebuah kode untuk mengisi gambar leksikon haji dan umrah. Berbeda dengan kode lainnya, penulisan atau pembuatan pada kode *picture* membutuhkan perhatian lebih. Penulisan harus disesuaikan dengan nama folder tersimpannya gambar.

Lexique Pro juga memberikan menu ‘lihat’ dan ‘edit’ yang berfungsi untuk menunjukkan tingkat keberhasilan proyek yang telah dibuat pada pilihan tombol ‘lihat’. Kemudian jika terjadi kesalahan dapat menekan tombol ‘edit’ untuk mengedit kembali proyek. Langkah terakhir perancangan kamus saku register haji dan umrah adalah mengeksport proyek kamus yang termuat pada *Lexique Pro* menjadi file dokumen yang dapat di cetak. Dengan menggunakan menu “*Export as Document*”. Pada menu tersebut juga dapat memilih tipe kamus cetak yang berupa leksikon dan keterangannya saja atau ditambah dengan penampilan gambar pada kamus. Pengaturan ukuran, posisi gambar serta jenis paragraf juga tersedia pada menu tersebut untuk disesuaikan dengan kamus yang diinginkan.

IV. SIMPULAN

Hasil penelitian ini menunjukkan tentang pengembangan korpus kosakata register haji dan umrah dengan menggunakan objek buku “Tuntunan Manasik Haji dan Umrah” yang diterbitkan oleh KEMENAG RI dan buku “*Dalīl Al-Hājj Wal-Mu’tamir*” karya Tholal bin Ahmad Al-Aqil yang diterbitkan oleh Kementerian Urusan Agama dari Kerajaan Saudi Arabia, beserta pengaplikasiannya dalam bentuk kamus dwibahasa (Arab-Indonesia).

Pengembangan korpus kosakata register haji dan umrah memiliki beberapa tahapan. Yaitu, pengolahan objek yang terdiri dari dua buku yaitu buku “Tuntunan Manasik Haji dan Umrah” yang diterbitkan oleh KEMENAG RI dan buku “*Dalīl Al-Hājj Wal-Mu’tamir*” karya Tholal bin Ahmad Al-Aqil yang diterbitkan oleh Kementerian Urusan Agama dari Kerajaan Saudi Arabia. Pengolahan tersebut terdapat beberapa proses yaitu: pengalihan teks, pengoperasian *AntConc*, penghitungan jumlah dan frekuensi kosakata. Penyepadanan koskaata berdasarkan Kamus Besar Bahasa Indonesia. Analisis pola serapan kamus saku register haji dan umrah, yang kemudian didapatkan klasifikasi berdasarkan pola serapan tersebut, dan klasifikasi waktu beserta frekuensi kemunculan kosakata register haji dan umrah dalam *Leipzig Corpora*.

Pengaplikasian kamus dwibahasa (Arab-Indonesia) atau Perancangan kamus saku register haji dan umrah menggunakan aplikasi *Lexique Pro* yang dapat ditampilkan berupa media elektronik yang didukung dengan adanya aplikasi tersebut maupun berupa media cetak.

Ucapan Terima Kasih

Puji syukur penulis panjatkan kepada Allah *Subha>nahu wa-Ta’ala>* atas berkah, rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan penulisan skripsi dengan judul “Perancangan Kamus Saku Register Haji dan Umrah Berbasis Korpus Menggunakan Aplikasi *Lexique Pro*”.

Terimakasih kepada program studi sastra arab yang telah memberikan bantuan, motivasi, dukungan, semangat dan bimbingan dari banyak pihak yang dirasa sangat membantu kelancaran dalam menyelesaikan penelitian ini

Daftar Pustaka

- Burhanuddin Erwina, Abdul Gaffar Ruskhan, dan Crismanto. 1993. *Penelitian Kosakata Bahasa Arab dalam Bahasa Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Chaer, Abdul dan Leonie Agustina. 2004. *Sosiolinguistik Perkenalan Awal*. Jakarta: Rineka Cipta.
- Chaer, Abdul. 2007. *Leksikologi dan Leksikografi Indonesia*. Jakarta: Rineka Cipta.
- Kridalaksana, Harimurti. 2008. *Kamus Linguistik*. Jakarta: Gramedia
- Kushartanti, Dkk. 2007. *Pesona Bahasa: Langkah Awal Memahami Linguistik*. Jakarta: PT SUN Printing.
- Kwari, Deny Arnos dkk. "Pengembangan Kamus Elektronik Akuntansi Keuangan Berbasis Korpus". Laporan Penelitian A2. Fakultas Sastra Jurusan Sastra Inggris, Universitas Airlangga.
http://www.kwary.net/about/Kwary_Laporan%20A2%202007_kamus.pdf
- Lukmanulhakim, Muhammad. 2015. *Cara Cepat Tepat dan Hemat Pergi Umroh*. Bandung: SBI SAHABAT.
- Mcenery, Tony dan Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory And Practice*. New York: Cambridge University Press.
- Nurafida, Ika dan Nita Suryawati. 2016. *Pembentukan Kamus Pewayangan Bahasa Jawa-Indonesia Berbasis Korpus*. Dalam Prosiding Seminar Leksikografi Indonesia. 2016. *Tantangan Leksikografis Bahasa-Bahasa Daerah di Indonesia*. Hlmn. 165-180. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.
http://www.academia.edu/30191714/PROSIDING_SEMINAR_LEKSIKOGRAFI_INDONESIA_2016
- Puspita, Dewi. 2016. *Pemanfaatan Korpus dalam Analisis Makna Kata Bersinonim mau, ingin, hendak, dan akan*. Dalam Prosiding Seminar Leksikografi Indonesia. 2016. *Tantangan Leksikografis Bahasa-Bahasa Daerah di Indonesia*. Hlmn. 31-40. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.
http://www.academia.edu/30191714/PROSIDING_SEMINAR_LEKSIKOGRAFI_INDONESIA_2016
- Rizqi, Fardianto Rakhman. 2008. "Penyerapan Istilah Asing Register Kedokteran Pada Rubrik Kesehatan Surat Kabar Republika Edisi Januari-Maret 2008". Skripsi. Fakultas Keguruan dan Ilmu Pendidikan Pendidikan Bahasa Sastra Indonesia, Universitas Muhamadiyah Surakarta.
- Rouf, Muh. 2012. *Buku panduan Pintar Anak Muslim*. Jakarta: Anak Kita.
- Shodiq, Muhammad Anwar. 2015. *Kosakata Bahasa Arab Laras Teknologi Smartphone Studi Kasus dalam Situs AITNEWS.com (Analisa Morfo-Semantik)*. Skripsi. Fakultas Sastra dan Seni Rupa Sastra Arab, Universitas Sebelas Maret Surakarta.

Winanto, Yugo. 2014. "Register Sepakbola Acara Lensa Olahraga". Skripsi. Fakultas Keguruan dan Ilmu Pendidikan Pendidikan Bahasa Sastra Indonesia, Universitas Muhamadiyah Surakarta.

Daftar Laman

Budiwiyanto, Adi. tt. Korpus dalam Penyusunan Kamus. [http:// badanbahasa.kemdikbud.go.id/lamanbahasa/artikel/1580](http://badanbahasa.kemdikbud.go.id/lamanbahasa/artikel/1580). Diakses pada 5 april 2017 09:44 WIB.

Gomuslim. 2017. <http://www.gomuslim.co.id/read/news/2016/09/12/1496/badan-statistik-saudi-jumlah-keseluruhan-jemaah-haji-2016-capai-1-862-909-orang.html>. Diakses pada 10 November 2017 13.30 WIB.

PENYUSUNAN KAMUS SINONIM BAHASA JAWA: PENENTUAN SINONIM BERBASIS KORPUS DALAM DESKRIPSI LEKSIKOGRAFIS BAHASA JAWA

Atin Fitriana, Totok Suhardijanto, Dien Rovita
Universitas Indonesia

Abstrak

Korpus bahasa dapat dimanfaatkan dalam berbagai penelitian, tidak terkecuali dalam penyusunan kamus. Pada makalah ini, akan disajikan berbagai aspek atau fungsi dalam korpus bahasa Jawa yang akan dimanfaatkan dalam penyusunan Kamus Sinonim Bahasa Jawa (KSBJ). Adapun korpus yang akan dimanfaatkan dalam penyusunan kamus adalah korpus bahasa Jawa yang berjumlah lebih dari 3 juta kata. Korpus ini merupakan korpus yang disusun oleh Program Studi Sastra Jawa, Universitas Indonesia sejak 2016. Korpus bahasa Jawa tersebut terdiri atas berbagai dialek bahasa Jawa, seperti dialek Yogya-Solo, Banyumas, dan Jawa Timur. Selain itu, korpus tersebut juga terdiri atas berbagai jenis teks, mulai dari teks sastra, buku pelajaran, karya tulis ilmiah, hingga undangan berbahasa Jawa dari tahun 1940-an hingga 2010-an. Dalam penyusunan kamus sinonim dan antonim berdasarkan korpus, penelitian ini menggunakan fungsi kolokasi dan konkordansi. Kedua fungsi tersebut dimanfaatkan dalam penentuan dan pengklasifikasian berbagai kata di dalam korpus bahasa Jawa yang dianggap sebagai bentuk sinonim.

Dalam makalah ini, lima kata emosi *seneng-bungah* dan *sedhik-nelangsa* digunakan sebagai contoh dasar pengembangan klaster sinonimi dan antonimi dalam upaya penyusunan KSBJ. Kemudian, dari kelima kata tersebut masing-masing, dipilih sepuluh satuan multikata kata (multiword units) yang paling umum berdasarkan frekuensinya. Berdasarkan hasil yang diperoleh, ditentukan berapa jauh sinonim berjarak satu sama lain ketika dibandingkan dengan kata lain yang merupakan bagian dari sinonim sebagian alih-alih sinonim absolut. Lepas dari konteks yang memungkinkan kedua sinonim saling menggantikan, dapat dikonfirmasi bahwa hanya satu yang signifikan kemunculannya pada beberapa satuan multikata. Pendekatan berbasis korpus menekankan cara yang berbeda dalam mendefinisikan sinonim absolut dalam kamus—terlepas dari penggambaran fitur semantis dasar pada kata, definisi harus mempertimbangkan representasi kontekstual dan perspektif kolokasional dari kata-kata yang bersinonim. Hasil korpus dapat meningkatkan hubungan kesinoniman antarkata dalam deskripsi leksikografis.

Kata kunci : korpus, kamus, bahasa Jawa, sinonim, kolokasi, konkordansi

I. PENDAHULUAN

Bahasa Jawa merupakan salah satu bahasa daerah di Indonesia dengan jumlah penutur terbanyak. Ethnologue.com mencatat ada lebih dari 84 juta penutur bahasa Jawa yang tersebar di seluruh wilayah di Indonesia. Adapun wilayah dengan jumlah penutur terbanyak ada di pulau Jawa, yaitu di wilayah Jawa Tengah, Yogyakarta, dan Jawa Timur. Berdasarkan dialektanya, bahasa Jawa terdiri atas beberapa dialek, seperti dialek Yogya-Solo, Banyumas, dan Jawa Timur. Selain memiliki beberapa dialek, bahasa Jawa juga memiliki variasi bahasa berdasarkan tingkat tutur. Rahyono (2002:14) menjelaskan bahwa bahasa Jawa terdiri atas beberapa tingkat tutur, yaitu *ngoko*, *madya*, dan *krama*. Perbedaan ketika tingkat tutur tersebut bergantung pada konteks situasi penggunaan bahasa Jawa, seperti memperhatikan kedudukan partisipan tutur dan situasi formal dan nonformal.

Dengan banyaknya variasi bahasa Jawa tersebut (dialek dan tingkat tutur), penyusunan korpus bahasa Jawa sangat penting dilakukan sebagai upaya dalam pendokumentasian bahasa. Selain sebagai upaya pendokumentasian, korpus dapat dimanfaatkan untuk berbagai penelitian, seperti penelitian di bidang linguistik, leksikografi, dan pengajaran bahasa. Sinclair (2005:16) dalam Cheng (2012:30) menjelaskan bahwa korpus berisi koleksi dari berbagai teks bahasa yang disimpan dalam bentuk digital, diseleksi berdasarkan beberapa kriteria untuk dihadirkan kembali, dan variasi bahasa dijadikan sebagai sumber data untuk penelitian linguistik. Sinclair (1991:171) dalam Cheng (2012:30) juga menambahkan, suatu korpus tidak hanya terdiri atas berbagai koleksi teks saja melainkan berisi variasi bahasa yang dipilih berdasarkan karakteristik tertentu. Dengan demikian, suatu korpus didisain dan disusun berdasarkan prinsip tertentu (Cheng, 2012:30).

Melihat pentingnya penyusunan korpus, Program Studi Sastra Jawa, Universitas Indonesia sejak tahun 2016 telah melakukan penyusunan korpus bahasa Jawa sebagai upaya revitalisasi bahasa daerah. Korpus bahasa Jawa tersebut tersusun atas berbagai dialek di wilayah Jawa Tengah, Yogyakarta, dan Jawa Timur. Selain itu, korpus bahasa Jawa tersebut juga terdiri atas berbagai jenis teks, seperti teks mulai dari teks sastra, buku pelajaran, karya tulis ilmiah, hingga undangan berbahasa Jawa dari tahun 1940-an hingga 2010-an. Hingga saat ini, korpus bahasa Jawa sudah berjumlah lebih dari 3 juta kata. Selain melakukan penyusunan korpus bahasa Jawa, Program Studi Sastra Jawa juga melakukan pemanfaatan korpus tersebut, seperti pemanfaatan dalam penyusunan buku ajar dan penyusunan kamus bahasa Jawa. Dalam penyusunan kamus bahasa Jawa, salah satu kamus bahasa Jawa yang akan disusun yaitu kamus sinonim bahasa Jawa.

Penelitian mengenai kamus bahasa Jawa telah dilakukan oleh beberapa orang, seperti oleh Ika Nurfaridan dan Nita Suryawati (2015) dalam penelitiannya yang berjudul *Pembentukan Kamus Pewayangan Bahasa Jawa-Indonesia Berbasis Korpus* dan penelitian oleh Afwin Sulistiawati (2015) dalam penelitiannya yang berjudul *Redesain Kamus Bahasa Jawa Berdasarkan Tingkatan Bahasa Jawa (Bahasa Jawa Ngoko, Krama Madra, Krama Inggil)*. Dalam penelitian yang dilakukan oleh Ika dan Nita (2015), data yang digunakan merupakan transkripsi dari pertunjukkan wayang semalam suntuk dan diolah menggunakan piranti lunak *AntCont*. 500 kosakata dengan frekuensi kemunculan tertinggi didefinisikan dan disajikan dalam *Lexique Pro* 3.6. Sementara itu, dalam penelitian yang dilakukan oleh Afwin (2015), data yang digunakan yaitu kosakata yang memiliki perubahan berdasarkan tingkatan

bahasa Jawa. Kosakata tersebut kemudian dimasukkan ke dalam kamus. Adapun kosakata tersebut dianalisis berdasarkan tingkat kesopanan atau tingkatan bahasa Jawa.

Selain penelitian mengenai kamus bahasa Jawa, penelitian mengenai kesinoniman bahasa Jawa juga telah dilakukan oleh beberapa orang, seperti Siti Marlina (2014) dan Erika Rahmatika (2013). Siti (2014) dalam skripsinya yang berjudul *Nomina Konkret Bahasa Jawa dalam Kamus Baoesastra Djawa Karya W.J.S. Poerwadarminta* membahas kesinoniman nomina konkret yang terdiri atas kesinoniman bentuk dan kesinoniman makna. Adapun keabsahan data dilakukan melalui pengatan dan pengecekan oleh pertimbangan ahli, dan analisis dilakukan dengan metode agih dan metode padan intralingual. Sementara itu, Erika (2013) dalam skripsinya yang berjudul *Sinonim Nomina dan Adjektiva Dialek Banyumasan* membahas mengenai wujud dan faktor penyebab sinonim yang terdapat dalam dialek Banyumasan. Data yang digunakan dalam penelitiannya, yaitu tuturan masyarakat desa Sirau. Data tersebut dianalisis menggunakan metode padan.

Berdasarkan beberapa penelitian di atas, terlihat bahwa penelitian mengenai kamus bahasa Jawa berdasarkan korpus dan penelitian mengenai sinonim bahasa Jawa dilakukan dengan tujuan dan sumber data yang berbeda. Dengan banyaknya variasi (dialek, jenis teks, dan tahun) dan banyaknya jumlah kata yang tersimpan dalam korpus bahasa Jawa (Prodi Sastra Jawa UI), penelitian mengenai pemanfaatan korpus dalam penyusunan kamus, khususnya kamus sinonim dapat dilakukan. Apalagi dengan jumlah kamus sinonim bahasa Jawa yang masih sangat terbatas. Hingga saat ini, penyusunan kamus sinonim bahasa Jawa dengan memanfaatkan korpus bahasa Jawa belum dilakukan. Oleh karena itu, rumusan masalah pada penelitian ini, yaitu bagaimana aspek atau fungsi dalam korpus bahasa Jawa dimanfaatkan dalam penyusunan kamus sinonim bahasa Jawa?. Adapun tujuan penelitian ini yaitu menyajikan aspek atau fungsi dalam korpus bahasa Jawa yang dimanfaatkan dalam penyusunan kamus sinonim bahasa Jawa.

II. KERANGKA TEORI

Berdasarkan tujuan yang sudah dipaparkan pada bagian Pendahuluan, penelitian ini akan membahas aspek dan fungsi dalam korpus bahasa Jawa dalam penyusunan kamus sinonim bahasa Jawa. Berdasarkan hal itu, pada bagian ini akan dipaparkan beberapa konsep yang berkaitan dengan jenis kamus, khususnya mengenai kamus sinonim serta pembahasan mengenai sinonimi, khususnya sinonimi di dalam kamus.

2.1 Kamus Sinonim

Dalam beberapa referensi, disebutkan berbagai jenis kamus, yang dibedakan berdasarkan berbagai kriteria. Atkins dan Rundell (2008: 24—25) memaparkan beberapa jenis kamus berdasarkan jumlah bahasanya, cakupan isinya, ukurannya, medianya, pengorganisasiannya, penggunaannya, kemahiran penggunaannya, serta berdasarkan tujuan penggunaan kamus.

Sementara itu, Svensen (2009: 21—36) membahas pengelompokan kamus berdasarkan tipologi kamus umum dan tipologi kamus khusus. Svensen membagi kamus berdasarkan tipologi umum menjadi beberapa kelompok, yaitu jenis kamus berdasarkan karakteristik

umum, berdasarkan karakteristik kuantitatif, sesuai dengan karakteristik struktural, berdasarkan pengorganisasian, fungsi, dan penggunaannya, serta berdasarkan media penyimpanan dan distribusinya. Pengelompokan kamus berdasarkan tipologi kamus khusus dibedakan lagi menjadi kategori kamus khusus yang bersifat sintagmatis, kamus khusus yang bersifat paradigmatis, serta kamus terbatas.

Selanjutnya, Svensen (2009: 30—32) memaparkan bahwa kamus khusus yang bersifat sintagmatis adalah kamus yang melihat keterkaitan antara unsur-unsur leksikal dan unsur-unsur lain serta bagaimana pengombinasian unsur-unsur tersebut. Yang termasuk di dalam pengelompokan ini adalah kamus valensi, kamus kolokasi, dan kamus idiom. Sementara itu, kamus khusus yang bersifat paradigmatis adalah kamus yang memperhatikan keterkaitan antara unsur-unsur leksikal dan unsur-unsur lain yang dapat diganti. Penggantian unsur-unsur tersebut dapat dibedakan berdasarkan konten dan ekspresinya.

Berdasarkan kontennya, kamus khusus yang bersifat paradigmatis dibedakan lagi atas kamus sinonim, kamus antonim, kamus thesaurus, dan kamus bergambar. Svensen menyatakan bahwa kamus sinonim dapat berupa kamus bilingual yang ditujukan untuk pengguna yang bukan merupakan penutur asli bahasa yang didefinisikan dalam kamus. Beberapa contoh kamus sinonim adalah *Longman Synonym Dictionary*, *Webster's New Explorer Dictionary of Synonyms & Antonyms*, dan *Dictionary of German Synonyms* (Svensen, 2009: 31).

Chaer (2007: 202—203) mengelompokkan kamus sinonim ke dalam jenis kamus berdasarkan isinya. Chaer menyatakan bahwa kamus sinonim adalah kamus yang penjelasan makna lemanya hanya berupa sinonim dari kata-kata tersebut, baik dalam bentuk sebuah kata maupun dalam bentuk gabungan kata. Penjelasan dalam bentuk definisi atau keterangan yang panjang tidak ada. Kalau sebuah lema memiliki sinonim lebih dari satu, sinonim itu dideretkan dari yang maknanya paling banyak kesamannya sampai pada yang paling sedikit kesamannya. Contoh kamus sinonim dalam bahasa Indonesia adalah *Kamus Sinonim Bahasa Indonesia* karangan Harimurti Kridalaksana.

Sinonim dalam bahasa Jawa dikenal dengan nama *dasanama*, *dasa* berarti sepuluh dan *nama* berarti nama. Dengan demikian *dasanama* berarti sepuluh nama. Dalam bahasa Jawa, terdapat beberapa kumpulan *dasanama* yang ditulis dalam sebuah buku. Jika dilihat sepintas mungkin kumpulan *dasanama* tersebut tidak mirip dengan kamus sinonim karena tidak diurutkan secara alfabetis. Dalam laman <https://www.sastra.org>, terdapat 5 entri *dasanama*, yaitu *Dasanama Kawi Padmasusastra* (1897), *Dasanama Mawi Jarwo* (1904), *Dasanama Anonim* (1932), *Dasanama Dwijo* (-), dan *Dasanamajarwa Dirjaatmaja* (1913). Hingga saat ini belum ada kamus sinonim bahasa Jawa. *Dasanama Kawi Padmasusastra* (1897) ditulis menggunakan aksara Jawa dan berisi daftar kata bahasa Jawa dengan sinonim dalam bahasa Kawi. Selain itu penulisan diurutkan berdasarkan urutan huruf Jawa, yaitu *ha* sampai *tha*. *Dasanama Mawi Jarwo* (1904) berisi daftar padanan kata (sinonim) dalam bahasa Jawa yang dipakai dalam tembang bahasa Jawa. Daftar kata tersebut ditulis menggunakan aksara Jawa. *Dasanama Dwijo* (-) memiliki 431 kata yang diurutkan secara alfabetis, mulai dari kata *akeh* sampai *windu*. Buku tersebut tidak diketahui tahun terbitnya. Sementara itu, *Dasanamaarja* (1913) berisi daftar padanan kata (sinonim) yang ditulis menggunakan aksara Jawa dan berisi daftar kata yang digunakan dalam tembang.

2.2 Sinonim dalam Bahasa Jawa

Sinonim adalah kata-kata yang kemiripan maknanya lebih menonjol dibandingkan perbedaan maknanya (Cruse, 2004: 154). Cruse memaparkan tiga tingkat sinonimi, yaitu *absolute synonymy*, *propositional synonymy*, dan *near-synonymy*, yang dapat diuraikan sebagai berikut (Cruse, 2004: 154—157).

- (1) *Absolute synonymy* atau sinonimi mutlak adalah relasi antara kata-kata yang memiliki kesamaan makna dalam semua konteks. Misalnya, kata *brave* dan *courageous*, *calm* dan *placid*, *big* dan *large*, *almost* dan *nearly*, serta *die* dan *kick the bucket*.
- (2) *Propositional synonymy* atau sinonimi proposisional adalah relasi antara dua kata yang dapat saling menggantikan berdasarkan kondisi-kondisi tertentu. Perbedaan makna pada sinonimi proposisional perlu memperhatikan beberapa hal yang berkaitan dengan makna nonproposisional, yaitu perbedaan makna perasaan, perbedaan berdasarkan tingkatan ragam bahasa, dan perbedaan berdasarkan presuposisi dalam wacana. Misalnya, kata *violin* dan *fiddle*.
- (3) *Near-synonymy* atau sinonimi dekat adalah relasi yang dekat secara semantik tetapi tidak dapat dikatakan sebagai sinonim, misalnya kata *animal* dan *plant*, *animal* dan *bird*, serta *dog* dan *cat*. Secara semantik, makna kedua kata dekat. Akan tetapi, kedua kata tersebut tidak bisa menjadi sinonim.

Saaed (2016: 61) memberikan batasan sinonimi, yaitu kata-kata yang memiliki perbedaan fonologis, yang memiliki makna yang sama atau serupa. Selanjutnya, Saaed menjelaskan bahwa kata-kata bersinonim dapat dilihat antara lain pada contoh: *couch* dan *sofa*, *boy* dan *lad*, *lawyer* dan *attorney*, *toilet* dan *lavatory*, serta *large* dan *big*. Kata-kata bersinonim sering memperlihatkan perbedaan distribusi pemakaian berdasarkan beberapa parameter, misalnya kata-kata bersinonim tersebut berasal dari dua dialek yang berbeda dan menyebabkan penutur bahasa menjadi familiar dengan kedua dialek tersebut. Kata-kata bersinonim dapat memperlihatkan ragam bahasa yang berbeda berdasarkan situasi pemakaian yang berbeda, misalnya, kata *wife* atau *spouse* lebih formal dibandingkan dengan *old lady* atau *missus*. Kata-kata bersinonim juga dapat memperlihatkan sikap positif atau negative penutur, misalnya kata *ingenuous* dibandingkan kata *naïve* dan *gullible*. Selain itu, beberapa kata bersinonim dapat dilihat berdasarkan kolokasi. Misalnya, kata *lad* dan *boy* pada kalimat (1) *She called out to the young lad* dan pada kalimat (2) *She called out to the young boy*, memperlihatkan konotasi yang berbeda dengan kalimat (3) *He always was a bit of a lad* dan kalimat (4) *He always was a bit of a boy* (Saaed, 2016: 62—63).

Sinonimi merupakan salah satu ciri kebahasaan yang bersifat universal. Dalam berbagai bahasa dapat ditemukan kata-kata bersinonim, tidak terkecuali dalam bahasa Jawa. Suwadji, dkk., 1992: 6—9) menyatakan bahwa kesinoniman dalam bahasa Jawa antara lain didorong oleh beberapa hal, yaitu tingkat tutur, kolokasi, nilai rasa, ragam bahasa, dan kata serapan, sebagai berikut.

- (1) Berdasarkan tingkat tutur, pemakaian kata-kata yang bersinonim menerapkan kaidah tingkat tutur yang telah disepakati, dengan mempertimbangkan situasi atau lingkungan pemakaiannya. Misalnya kata *pejah* ‘meninggal’ lebih sesuai dipakai dibandingkan kata *séda* ‘wafat’.

- (2) Berdasarkan kolokasi atau asosiasi hubungan makna kata yang satu dengan kata yang lain yang keduanya memiliki hubungan ciri yang relative tetap, dapat dilihat contoh sinonimi pada kata *ayu* ‘cantik’ dan *bagus* ‘tampan’ yang memiliki makna referensial yang sama, tidak dapat digunakan dalam konteks pemakaian *jaka ayu* ‘jejak cantik’ atau *prawan bagus* ‘perawan cantik’.
- (3) Berdasarkan nilai rasa dapat dikaitkan dengan perasaan pemakai bahasa yang diwujudkan dengan perasaan halus, kasar, indah, atau anggun. Misalnya, kata *tindak* ‘pergi’ dan *sare* ‘tidur’ digunakan untuk orang lain, sedangkan untuk diri sendiri digunakan kata *lunga* ‘pergi’ atau *kesah* ‘pergi’ dan *turu* ‘tidur’ atau *tilem* ‘tidur’.
- (4) Berdasarkan ragam bahasa, yaitu ragam bahasa formal dan nonformal. Misalnya dapat dilihat pada kata-kata formal *wonten* ‘ada’, *saka* ‘dari’, *besuk* ‘nanti’, *kadospundi* ‘bagaimana’, serta *punika* ‘ini/itu’ yang dapat dibandingkan dengan kata-kata nonformalnya, yaitu *onten* ‘ada’, *ka* ‘dari’, *suk* ‘nanti’, *dospundi* ‘bagaimana’, serta kata *nika* ‘ini/itu’.
- (5) Berdasarkan kata serapan dari bahasa lain, terutama dari bahasa Indonesia, muncul pasangan sinonimi dalam bahasa Jawa. Misalnya pada kata *bebana* ‘hadiah’ dan *hadiah*, kata *lelahan* ‘cuma-cuma’ dan *gratis*, kata *piagem* ‘piagam’ dan *piagam*, kata *béker* ‘piala’ dan *piala*, serta kata *pacangan* ‘tunangan’ dengan kata *pacar*.

Uraian mengenai teori dan konsep-konsep dasar dalam kajian leksikografi dan semantik di atas akan menjadi landasan di dalam melakukan analisis terhadap data berupa kata-kata bersinonim dalam bahasa Jawa serta penerapannya di dalam kamus sinonim.

III. METODE

Dalam penelitian ini digunakan rancangan metode gabungan untuk menelusuri dan mengidentifikasi relasi makna sinonimi antarkosakata dalam bahasa Jawa. Menurut Creswell (2014, hlm. 4) penelitian dengan metode gabungan merupakan pendekatan yang memanfaatkan data kuantitatif dan data kualitatif. Untuk menjawab permasalahan dalam penelitian, bagaimana aspek atau fungsi dalam korpus bahasa Jawa dimanfaatkan dalam penyusunan kamus sinonim bahasa Jawa, digunakan metode korpus untuk menggali informasi yang ada pada koleksi korpus untuk memperoleh bentuk-bentuk yang bersinonim.

Analisis sinonimi dengan pendekatan korpus telah dimulai dua dasawarsa lalu dengan berbagai metode kuantitatif (Church et al. 1994; Lin et al. 2003). Meskipun demikian, pada umumnya, pendekatan kuantitatif tersebut dilengkapi dengan pengamatan terhadap perbedaan stilistika di antara kedua bentuk ketika muncul dalam konteks yang serupa (De Cesare 2003). Menurut De Cesare, apabila dua bentuk serupa muncul dalam konteks yang berhubungan, tidak hanya dapat diungkapkan ranah penggunaan yang paling sering, tetapi juga dapat diketahui penggunaan yang paling jarang untuk kedua bentuk tersebut. Hal ini merupakan informasi mengenai perbedaan stilistik di antara dua bentuk yang bersinonim.

Dalam penelitian ini digunakan alat (*tool*) pengolah korpus Sketch Engine untuk analisis data (). Sketch Engine merupakan alat lengkap untuk mengeksplorasi bagaimana bahasa bekerja. Algoritmanya menganalisis teks autentik yang berjumlah miliaran kata untuk

mengenali secara cepat apa yang khas dan mana yang tidak, atau mana yang tidak biasa dan menonjol dalam sebuah korpus.

IV. PEMBAHASAN

Pada bagian ini dijelaskan mengenai fungsi-fungsi dalam korpus linguistik yang digunakan dalam penyusunan kamus sinonim bahasa Jawa. Pada piranti lunak *SketchEngine*, penelitian ini memanfaatkan fungsi kolokasi dan konkordansi. Pada *SketchEngine*, kolokasi digunakan untuk mengetahui lingkungan dari kosakata yang disasar. Sementara itu, konkordasi digunakan untuk mengetahui konteks penggunaan kata tersebut di dalam korpus. Pada penelitian ini, lima kosakata emosi yang digunakan sebagai contoh dasar pengembangan klaster sinonimi dalam penyusunan kamus sinonim bahasa Jawa yaitu, *bungah-seneng*, *sedhik-nelangsa*, *getun-kuciwa*, *gandrung-kasmaran*, dan *tega-tegel*.

4.1. *Bungah-Seneng* ‘senang’

Berdasarkan hasil analisis kolokasi menggunakan program *SketchEngine*, terlihat bahwa penggunaan kata *bungah* dan *seneng* di dalam korpus bahasa Jawa banyak memiliki persamaan. Meskipun demikian, terdapat beberapa hal yang menunjukkan adanya perbedaan kolokasi dari kedua kata tersebut. Perhatikan tabel di bawah ini.

Tabel 1

	<i>Bungah</i>	<i>Seneng</i>
Tempat	<i>Ati, rasa</i>	<i>Ati, rasa</i>
Antonim	<i>Meri, susah</i>	<i>Susah, sedhik, prihatin, wedi</i>
Sinonim	<i>Seneng, suka</i>	<i>Seneng, bungah</i>
Verba (panca indra)	<i>Katon, krungu</i>	<i>Katon, krungu, weruh, mesem, nyawang, ngguyu,</i>
Verba (tindakan)	<i>Ndadekake</i>	<i>nulung, nonton, maca, mangan, dolanan, sinau, nulis, bal-balan, plesir, srawung</i>
Kata sifat lain	-	<i>Pinter, becik, tentrem, rukun, sregep, eklas, lali</i>

Di dalam korpus, penggunaan kata *seneng* lebih banyak dibandingkan dengan kata *bungah*. Oleh karena itu, kata yang berkolokasi dengan kata *seneng* lebih banyak daripada kata *bungah*. Berdasarkan tabel 1 di atas, kata *bungah* dan *seneng* sama-sama berkolokasi dengan kata *ati* dan *rasa*. Akan tetapi, pada baris berikutnya, terlihat bahwa terdapat beberapa perbedaan kata yang berkolokasi. Pada bagian *antonim*, kata *seneng* berkolokasi dengan kata *susah*, *sedhik*, *prihatin*, dan *wedi*. Sementara itu, kata *bungah* hanya berkolokasi dengan kata *meri* dan *susah*. Pada bagian lain, yaitu pada bagian verba, terlihat bahwa kata *seneng* diekspresikan dengan verba *mesem* ‘senyum’ dan *ngguyu* ‘tertawa’, sedangkan kata *bungah*

tidak diekspresikan dengan kedua verba tersebut. Selain itu, kata *seneng* juga berhubungan dengan tindakan. Berdasarkan baris konkordansi, kata *seneng* tidak hanya berhubungan dengan perasaan tetapi juga berhubungan dengan tindakan. Sementara itu, kata *bungah* hanya berhubungan dengan perasaan.

4.2. *Sedhih-Nelangsa* ‘sedih’

Berdasarkan fungsi kolokasi, kata *sedhih* dan *nelangsa* memiliki beberapa perbedaan. Perhatikan tabel 2 di bawah ini.

Tabel 2

	<i>sedhih</i>	<i>nelangsa</i>
Tempat	<i>Rasa, ati</i>	<i>Rasa, ati</i>
Sinonim	<i>Prihatin, sungkawa, susah</i>	<i>Kuciwa, prihatin, susah</i>
Antonim	<i>Seneng</i>	<i>legawa</i>
Verba tindakan	<i>nangis</i>	-
Nomina	<i>Crita, asmara</i>	<i>Tembang, urip</i>

Pada tabel 2 di atas dapat dilihat bahwa persamaan keduanya terletak pada tempat, yaitu di rasa dan di hati. Sementara itu pada kelompok lain, terdapat perbedaan. Pada bagian sinonim, terdapat persamaan, namun juga terdapat perbedaan. Pada bagian tersebut, *sedhih* berkolokasi dengan kata *sungkawa*, sedangkan *nelangsa* berkolokasi dengan kata *kuciwa*. Pada bagian verba tindakan, *sedhih* diekspresikan dengan tindakan nangis ‘menangis’, sedangkan kata *nelangsa* tidak diekspresikan dengan tindakan menangis. Selain itu, perbedaan lain yang cukup terlihat yaitu pada bagian nomina. *Sedhih* berkolokasi dengan *asmara* atau percintaan, sedangkan *nelangsa* berkolokasi dengan urip ‘hidup’. Berdasarkan konkordansi, terlihat bahwa kata *nelangsa* digunakan dalam konteks untuk menjelaskan tentang keadaan hidup dan nasib yang kurang baik. Sementara kata *sedhih* digunakan dalam konteks yang berhubungan dengan kekecewaan dan perasaan cinta, serta diekspresikan dengan tindakan menangis.

4.3. *Getun-Kuciwa* ‘kecewa’

Di dalam korpus bahasa Jawa, kata-kata yang berkolokasi dengan kata *getun* dan *kuciwa* tidak banyak. Meskipun demikian, terlihat beberapa perbedaan di antara keduanya. Perhatikan tabel 3 di bawah ini.

Tabel 3

		<i>a</i>
tt		<i>sa</i>
m	<i>kuwatir</i>	<i>nelangsa, prihatin</i>
tindakan		

Berdasarkan tabel di atas, terlihat beberapa perbedaan kata-kata yang berkolokasi dengan kata *getun* dan *kuciwa*. Kata *getun* berkolokasi dengan kata *gela* ‘kecewa’ dan *kuwatir*. Sementara itu, kata *kuciwa* berkolokasi dengan kata *gela*, *nelangsa* dan *prihatin*.

Pada tabel 2, kata *nelangsa* juga berkolokasi dengan kata *kuciwa*. Pada bagian verba tindakan, kata *getun* berkolokasi dengan kata *nesu* ‘marah’, sedangkan kata *kuciwa* tidak berkolokasi dengan kata *nesu*. Berdasarkan baris konkordansi, kata *kuciwa* digunakan dalam konteks kekecewaan yang berhubungan dengan *nelangsa* dan *prihatin*. Sementara kata *getun* digunakan dalam konteks kekecewaan yang dirasakan setelah tindakan yang dilakukan selesai.

4.4. *Gandrung-Kasmaran* ‘Tergila-gila’

Sama seperti kata *getun* dan *kuciwa*, kata yang berkolokasi dengan kata *gandrung-kasmaran* juga tidak banyak. Terdapat perbedaan di antara kedua kata tersebut. Perhatikan tabel 4 di bawah ini.

Tabel 4

	<i>Gandrung</i>	<i>Kasmaran</i>
Sinonim	<i>kasmaran</i>	<i>Gandrung, tresna</i>
Nomina	<i>Nom-noman, wong</i>	<i>Wong, macapat</i>

Berdasarkan tabel di atas, kata *kasmaran* berkolokasi dengan kata *gandrung* dan *tresna* ‘cinta’, sedangkan kata *gandrung* tidak berkolokasi dengan kata *tresna*. Kata *tresna* ‘cinta’ lebih sering berhubungan dengan kata *kasmaran* dibandingkan dengan kata *gandrung*. Pada bagian nomina, keduanya sama-sama berkolokasi dengan kata *wong*. Akan tetapi juga terdapat perbedaan pada bagian nomina. Kata *gandrung* berkolokasi dengan kata *nom-noman* ‘pemuda’, sedangkan kata *kasmaran* berkolokasi dengan kata *macapat*. Kata *macapat* berhubungan dengan *kasmaran* karena salah satu tembang *macapat* ada yang berisi mengenai *kasmaran*, yaitu tembung *asmarandhana*. Sementara kata *gandrung* berkolokasi dengan pemuda karena pemuda lebih dekat hubungannya dengan rasa tergil-gila. Adapun berdasarkan konkordansi, kata *gandrung* digunakan pada konteks tergil-gila pada suatu hal, bisa berupa cinta di antara orang yang muda ataupun kepada suatu benda. Sementara kata *kasmaran* digunakan pada konteks cinta kepada seseorang dan juga konteks yang digunakan dalam tembang *macapat*.

4.5. *Tega-Tegel* ‘Tega’

Kata yang berkolokasi dengan kata *tega* dan *tegel* tidak banyak ditemukan di dalam korpus bahasa Jawa. Oleh karena itu, untuk melihat perbedaan di antara kedua kata tersebut tidak terlalu signifikan. Perhatikan tabel 5 di bawah ini.

Tabel 5

	<i>Tega</i>	<i>Tegel</i>
Sinonim	-	<i>tega</i>
nomina	<i>Patine, larane</i>	<i>rasa</i>
Verba tindakan	-	<i>nangis</i>

Pada tabel di atas terlihat bahwa kata *tega* dan kata *tegel* memiliki kata berkolokasi yang berbeda. Pada bagian sinonim kata *tega* tidak berkolokasi dengan kata *tegel*, sedangkan kata *tegel* berkolokasi dengan kata *tega*. Begitu juga pada bagian selanjutnya, pada bagian verba tindakan kata *tega* tidak berkolokasi dengan verba tertentu. Hal ini dimungkinkan karena keterbatasan penggunaan kedua kata di dalam korpus bahasa Jawa. Jika dilihat berdasarkan baris konkordansi, kata *tegel* digunakan dalam dua konteks yang berbeda yaitu konteks yang digunakan untuk menyatakan *tega* dan konteks yang menyatakan bagian dari bahan pembuatan lantai. Hal tersebut karena kata *tegel* yang berarti *tega* mempunyai relasi hononimi dengan kata *tegel* yang berarti bahan pembuat lantai. Untuk mengetahui perbedaan keduanya, dibutuhkan konteks pemakaian kata tersebut.

V. KESIMPULAN

Berdasarkan analisis yang telah dilakukan terhadap kelima pasang kata yang bermakna emosi dalam bahasa Jawa, penggunaan fungsi kolokasi dan konkordansi membantu dalam membandingkan dua kata yang bersinonim. Melalui fungsi kolokasi, diketahui kata-kata apa saja yang berada di lingkungan sebuah kata tertentu dan membedakan antara satu kata dengan kata yang lain. Fungsi konkordansi membantu untuk mengetahui konteks penggunaan kata. Hasil analisis pada fungsi konkordansi dapat mendukung hasil dari analisis kolokasi yang kurang lengkap. Akan tetapi, berdasarkan hasil analisis, terlihat bahwa masih terdapat beberapa kekurangan dalam membedakan kosakata yang bersinonim. Hal tersebut disebabkan tidak banyaknya penggunaan kosakata sinonim tersebut di dalam korpus bahasa Jawa. Oleh karena itu, beberapa hasil analisis menunjukkan tidak signifikannya persamaan dan perbedaan antara dua kata yang bersinonim.

Daftar Referensi

- Adam Kilgarriff, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, Vít Suchomel. The Sketch Engine: ten years on. *Lexicography*, 1: 7-36, 2014.
- Atkins, B. T. & Rundell, Michael. (2008). *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- Chaer, Abdul. (2007). *Leksikologi dan Leksikografi Indonesia*. Jakarta: Rineka Cipta.
- Cheng, Winnie. (2012). *Exploring Corpus Linguistics: Language in Action*. London: Routledge
- Cruse, Alan. (2004). *Meaning in Language: An Introduction to Semantics and Pragmatics*. Second Edition. New York: Oxford University Press.
- Church, Kenneth Ward, William Gale, Patrick Hanks, Donald Hindle, and Rosamund Moon. 1994. Lexical substitutability. In B. T. S. Atkins and A. Zampolli, editors, *Computational Approaches to the Lexicon*. Oxford University Press, pages 153–177.
- De Cesare, Anna Maria. (2003). A Corpus-based approach to the concept of synonymy: A case study of Italian *anche* and *pure*. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, Number 2/2003, pp. 273-293.

- Lin, Dekang, Shaojun Zhao, Lijuan Qin, and Ming Zhou. 2003. Identifying synonyms among distributionally similar words. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1492–1493, Acapulco, Mexico.
- Marlina, Siti. (2014). *Kesinoniman Nomina Konkret Bahasa Jawa dalam Kamus Baoesastra Djawa Karya W.J.S. Poerwadarminta*. Skripsi. Yogyakarta: Universitas Negeri Yogyakarta
- Nurfarida, Ika dan Nita Suryawati. (2016). Pembentukan Kamus Pewayangan Bahasa Jawa-Indonesia Berbasis Korpus. *Tantangan Leksikografis Bahasa-Bahasa Daerah di Indonesia*. Prosiding Seminar Leksikografi Indonesia. Hlm 165-180. Jakarta: Badan Pengembangan dan Pembinaan Bahasa, Kemendikbud.
- Rahmawati, Erika. (2013). *Sinonim Nomina dan Adjektiva Dialek Banyumasan*. Skripsi. Semarang: Universitas Negeri Semarang.
- Rahyono, F.X. (2002). *Ekspresi Deiktis Bahasa Jawa*. Depok: FIB UI.
- Saeed, John I. (2016). *Semantics*. Forth Edition. West Sussex: Blackwell Publishing.
- Sulistiawati, Afwin. (2017). Redesain Kamus Bahasa Jawa Berdasarkan Tingkatan Bahasa Jawa (Bahasa Jawa Ngoko, Krama Madya, Krama Inggil). *Tantangan Leksikografis Bahasa-Bahasa Daerah di Indonesia*. Prosiding Seminar Leksikografi Indonesia. Hlm 181-192. Jakarta: Badan Pengembangan dan Pembinaan Bahasa, Kemendikbud.
- Suwadji dkk. (1992). *Sistem Kesinoniman dalam Bahasa Jawa*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Svensen, Bo. (2009). *A Handbook of Lexicography*. New York: Cambridge University Press.

**PENGEMBANGAN KAMUS PENGUCAPAN BERBASIS GAWAI
UNTUK PEMELAJAR BAHASA INDONESIA DI TIONGKOK**
*(THE DEVELOPMENT OF A PRONUNCIATION DICTIONARY MOBILE-
APP FOR LEARNERS OF INDONESIAN
LANGUAGE IN CHINA)*

Hery Yanto The

Universitas Bahasa Asing Yuexiu Zhejiang, herythe@outlook.com

Latifah

Dharma Achariya Sekolah Tinggi Agama Buddha Kertarajasa, efi.latifah@gmail.com

Batari Oja Andini

Universitas Bahasa Asing Yuexiu Zhejiang, batarioja@outlook.com

Abstrak

Kamus elektronik dan aplikasi-aplikasi belajar berbasis gawai sangat diperlukan untuk mendukung proses belajar mengajar bahasa Indonesia. Bagi mahasiswa pemelajar bahasa Indonesia di Yuexiu Tiongkok, belum tersedianya kamus elektronik dan aplikasi-aplikasi belajar untuk mendukung proses pembelajaran pengucapan berusaha diatasi oleh pengajar dengan merancang dan membuat sendiri media belajar. Media belajar tersebut berbentuk modul dan audio yang didistribusikan secara elektronik. Kedua media tersebut akan lebih mudah digunakan oleh mahasiswa jika bisa disatukan dalam bentuk aplikasi kamus elektronik berbasis gawai. Analisis kebutuhan dan rancangan prototipe yang akan digunakan untuk mengembangkan Kamus Pengucapan disajikan pada tulisan ini. Analisis kebutuhan secara khusus diprioritaskan pada mahasiswa pemelajar bahasa Indonesia di Yuexiu. Prototipe disajikan dalam bentuk contoh-contoh sketsa tampilan kamus yang diharapkan dapat dikembangkan pada aplikasi gawai. Hasil analisis menunjukkan, mahasiswa yang telah terbiasa menggunakan gawai dan berbagai aplikasi tentu saja tidak akan mengalami kesulitan untuk mempelajari cara penggunaan Kamus Pengucapan yang akan dikembangkan. Kamus Pengucapan yang akan dibuat perlu memiliki struktur dan elemen-elemen interaktif yang memungkinkan mahasiswa pembelajar pemula Bahasa Indonesia untuk memahami arti kata. Elemen-elemen tersebut adalah kata, kelas kata, fitur suara, gambar, dan terjemahan kata dalam bahasa Inggris dan bahasa Mandarin.

Kata kunci: BIPA, kamus elektronik, pengucapan

Abstract

The electronic dictionary and mobile-apps are two essential learning media for language learners, especially for Chinese' students who learn Indonesian Language Pronunciation. Recently, in China mobile-apps market, it is uneasy for these learners to get the proper apps for learning Indonesian Language. Teachers of the language, Bahasa Indonesia, then made an effort to design their own learning modules and audio files. These learning modules and audio files have been distributed electronically. Students will certainly feel ease of using these learning resources if they are turned into an electronic dictionary that is immediately accessible to their mobile devices. Researchers, in this paper, conduct the need analysis and create the mobile-device app prototypes of the Indonesian Language Pronunciation Dictionary. The need analysis is conducted specifically in the Indonesian Language Department of Yuexiu, and the prototypes are showed as simple sketches of the prospective dictionary design. From the need analysis, it appears that students have been using various mobile-device learning apps extensively. It is not going to be difficult for them to adapt and use the pronunciation dictionary when it is available. The dictionary should be made almost similar to other apps that they have been using. This new app has to be interactive and enables these students, the novice learners of Bahasa Indonesia, to quickly understand the meaning of words. The interactive features to be included in this app are words, word classes, the voice feature, images and the translation of words in English and Chinese-Mandarin.

Keywords: *BIPA, pronunciation, electronic dictionary*

I. PENDAHULUAN

Penyelenggaraan program pendidikan Bahasa Indonesia di Tiongkok telah berlangsung dalam kurun waktu yang cukup panjang (Xiaoqiang 2014). Universitas Peking dengan program khusus di bidang Bahasa Indonesia sudah ada sejak 1950 dan pada tahun 1970 Universitas Bahasa Asing Guangdong juga membuka program yang sama (Sudaryanto 2014). Pada skala nasional pemelajar Bahasa Indonesia di Tiongkok terus mengalami peningkatan, meskipun jumlahnya masih jauh dibandingkan dengan peminat bahasa-bahasa Asia lainnya (misalnya Bahasa Jepang dan Bahasa Korea), Sebelum kuliah di jurusan Bahasa Indonesia, mahasiswa tersebut sama sekali belum mengenal bahasa Indonesia dan tidak memiliki keterampilan berbahasa Indonesia. Mahasiswa pemelajar Bahasa Indonesia di Tiongkok mulai mempelajari materi-materi Bahasa Indonesia dari dasar ketika mereka memulai kuliah pada tahun pertama. Hampir sama dengan program Pendidikan Bahasa Indonesia pada universitas-universitas lainnya di Tiongkok, mahasiswa di *Zhejiang Yuexiu University of Foreign Languages* (selanjutnya disebut Yuexiu) tahun pertama dibekali dengan mata kuliah Pengucapan, Tata Bahasa Indonesia, Pengenalan Budaya Indonesia, Berbicara, Mendengarkan Intensif, dan Membaca Intensif. Mahasiswa diberi kuliah oleh dosen yang berasal Tiongkok dan juga dosen yang berasal dari Indonesia.

Pemberian kuliah oleh dosen yang berasal dari Tiongkok dilakukan dalam dua bahasa, yaitu bahasa Mandarin dan bahasa Indonesia. Penyampaian kuliah oleh dosen yang

berasal dari Indonesia dilakukan dalam dua atau tiga bahasa, tergantung pada kemampuan berbahasa dosen yang bersangkutan. Jika dosen hanya menguasai bahasa Indonesia dan bahasa Inggris, maka perkuliahan dilakukan dengan menggunakan kedua bahasa tersebut, namun jika dosen menguasai bahasa Mandarin, maka perkuliahan dapat dilakukan dengan memadukan tiga bahasa, yaitu bahasa Indonesia, bahasa Inggris, dan bahasa Mandarin. Selain memperoleh materi kuliah dari dosen, mahasiswa juga melakukan studi secara mandiri dengan memanfaatkan berbagai media pembelajaran yang tersedia di kampus, seperti buku berbentuk cetak dan elektronik, rekaman audio dan video, laman web, dan aplikasi-aplikasi gawai. Fitur dan aplikasi pada gawai cerdas sangat membantu pemelajar untuk memperoleh kemudahan dan kepraktisan dalam belajar (Gangaiamaran & Pasupathi 2017; Godwin-Jones 2017; Nushi & Eqbali 2018). Pembelajar bahasa asing lain di Tiongkok (seperti bahasa Italia, bahasa Jerman, bahasa Korea, dan bahasa Jepang) sudah cukup lama disokong berbagai aplikasi belajar berbasis gawai. Namun, pasar aplikasi gawai cerdas di Tiongkok belum banyak menyediakan aplikasi-aplikasi untuk belajar bahasa Indonesia. Aplikasi-aplikasi yang ada pun seringkali tidak sesuai dengan kebutuhan pemelajar.

Tingkat penguasaan bahasa yang masih belum memadai seringkali membuat mahasiswa kesulitan menggunakan kamus dua bahasa ataupun kamus ekabahasa (Hamilton 2012). Sebagai alternatif, mahasiswa kemudian mengandalkan aplikasi daring, seperti *YouDao*, *Baidu* Terjemahan, dan *Google* Terjemahan. Aplikasi-aplikasi ini cukup membantu mereka, namun sangat tidak memadai dalam memberikan pengetahuan mengenai konteks kebudayaan Indonesia yang berhubungan dengan kata-kata tertentu (Vidhayasai, Keyurauong & Bunsoun 2015). Terjemahan singkat yang diberikan tidak jarang memberikan pemahaman yang kurang tepat. Kata “becak” misalnya diterjemahkan oleh *Google* Terjemahan sebagai “*sān lún chē*” yang oleh mahasiswa dapat dipahami sebagai kendaraan beroda tiga yang digerakkan dengan menggunakan motor. Terjemahan tersebut tentu saja berbeda dengan pengertian becak yang dimaksud dalam bahasa Indonesia karena dalam KBBI (2016) becak diartikan sebagai “kendaraan umum seperti sepeda tidak bermotor beroda tiga, bertutup (tutupnya dapat dibuka), satu sadel di belakang, tempat duduk untuk penumpang di depan, dijalankan dengan tenaga manusia (pengemudinya duduk di belakang).” Bagi mahasiswa pembelajar pemula, penggunaan gambar untuk mendampingi terjemahan yang diberikan tentu saja akan lebih membantu mereka memahami kata yang diterjemahkan (Stacey 2007). Gambar yang diberikan juga memperjelas makna kata yang diterjemahkan sesuai dengan konteks budaya dari bahasa target belajar (Krcelic & Matijevic 2015). Inilah salah satu kekurangan dari aplikasi-aplikasi pembelajaran Bahasa Indonesia di Tiongkok yang masih perlu dilengkapi dengan pengembangan yang direncanakan pada penelitian ini.

Mahasiswa pemelajar pemula bahasa Indonesia di Tiongkok juga sangat membutuhkan fitur suara pada aplikasi belajar untuk membantu mereka berlatih sendiri untuk mengulang-ucap kata-kata yang dipelajari. *YouDao*, *Baidu* Terjemahan, dan *Google* Terjemahan memang memiliki fitur suara, namun menurut sebagian besar mahasiswa, fitur suara pada kata-kata tertentu masih terdengar kurang alamiah. Kadang mereka kesulitan untuk menentukan pengucapan yang benar dengan mendengarkan fitur suara pada ketiga aplikasi tersebut. Mahasiswa di Yuexiu bahkan lebih memilih untuk merekam ucapan yang dilakukan oleh guru menggunakan fitur rekaman suara pada gawai cerdasnya sehingga mereka dapat mendengarkannya kembali ketika berlatih sendiri.

Belum tersedianya kamus elektronik dan aplikasi-aplikasi belajar yang dapat mendukung perkuliahan mahasiswa pemelajar bahasa Indonesia di Tiongkok berusaha diatasi oleh pengajar dengan merancang dan membuat sendiri media belajar untuk memenuhi kebutuhan mahasiswa. Di Yuexiu misalnya, pengajar membuat modul latihan “Pengucapan” yang disertai dengan gambar. Modul belajar tersebut selama ini telah dibagikan kepada mahasiswa untuk digunakan secara elektronik dalam bentuk PDF. Rekaman suara juga dibuat menggunakan aplikasi Audacity dan dibagikan kepada mahasiswa sebagai pendamping modul yang disediakan. Modul dan rekaman suara saat ini masih merupakan moda-moda pembelajaran yang tersedia secara terpisah. Pengajar merasa, mahasiswa akan lebih mudah untuk mengakses dan menggunakan media belajar tersebut jika kedua moda tersebut dapat disatukan dalam aplikasi berbasis gawai. Aplikasi berbasis gawai yang akan dikembangkan tidak untuk menggantikan kedua media yang telah dibuat, namun dapat menjadi moda pilihan yang dapat digunakan secara bersama-sama dengan modul dan rekaman suara.

Aplikasi berbasis gawai yang pengajar rencanakan berupa “Kamus Pengucapan” berbasis gawai dengan fitur multimodal. Fitur multimodal tersebut terdiri dari kata atau frase, simbol fonetik, suara, gambar, dan terjemahan kata atau frase ke dalam Bahasa Inggris dan Bahasa Mandarin. Pengembangan kamus ini mengacu pada modul Pengucapan yang telah dibuat oleh pengajar. Saat ini pengembangan kamus masih dalam tahap perencanaan. Tujuan dari tulisan ini adalah menyajikan analisis kebutuhan dan prototipe Kamus Pengucapan tersebut. Analisis kebutuhan secara khusus diprioritaskan pada mahasiswa pemelajar Bahasa Indonesia di Yuexiu. Prototipe akan disajikan dalam bentuk contoh-contoh sketsa tampilan kamus yang selanjutnya pada saat implementasi akan dikembangkan pada aplikasi gawai. Peneliti juga akan menjelaskan secara singkat perangkat lunak yang akan dipilih untuk mengembangkan aplikasi gawai tersebut.

II. KERANGKA TEORI DAN METODE PENELITIAN

Pemelajar bahasa kedua memerlukan pengetahuan kosa kata yang memadai supaya dapat berhasil menguasai keterampilan berbahasa secara komprehensif (Alqahtani 2015). Pengetahuan pemelajar bahasa kedua terhadap kosa kata dapat diukur dari kemampuan mengetahui arti kata, mengucapkan kata tersebut dengan jelas, dan menggunakan kata tersebut di dalam komunikasi tertulis maupun lisan (Alharbi 2015). Pembelajaran kosa kata menjadi proses berkelanjutan di dalam perkembangan kemampuan berbahasa pemelajar bahasa kedua (Laufer 1998). Kamus merupakan media yang digunakan oleh pemelajar dewasa bahasa kedua untuk mendukung pembelajaran kosa kata (Yunhao 2011). Perkembangan teknologi telah memungkinkan pemelajar memilih berbagai bentuk kamus untuk mendukung proses belajarnya (Kruyt 1995). Dari aspek kebahasaan, kamus yang tersedia bisa berupa kamus ekabahasa atau kamus dua bahasa (El-Sayed & Siddiek 2013). Dari aspek akses, kamus bisa berbentuk cetakan, CD-ROM, perangkat lunak yang dapat dipasang pada komputer, laman web, atau aplikasi pada gawai cerdas (Tono 2000). Dari aspek ketersediaan informasi, kamus bisa berupa kamus definisi kata, kamus bergambar, kamus ungkapan, kamus peribahasa, dan kamus khusus di bidang-bidang keilmuan Tarp (2008).

Pada penelitian ini, kamus yang menjadi pusat perhatian adalah kamus berbentuk elektronik, multibahasa, dapat diakses melalui gawai, dan menyediakan informasi khusus untuk belajar arti dan cara pengucapan kata. Sasaran pengguna kamus ini adalah pembelajaran bahasa Indonesia berusia dewasa yang menguasai bahasa jati, Mandarin dan bahasa kedua, bahasa Inggris. Bahasa Indonesia merupakan bahasa kedua selain bahasa Inggris yang dipelajari oleh pemelajar ini sebagai spesialisasi keahliannya di perguruan tinggi. Kamus yang akan dikembangkan di sini merupakan kamus elektronik berbasis gawai yang diharapkan dapat digunakan pada aplikasi android maupun iphone. Pengembangan kamus elektronik ini menggunakan pendekatan tematik yang didasarkan pada satuan bahasa, mulai dari fonem, kata, frase, dan kalimat. Bahan yang akan digunakan untuk pemilihan kata, frase, dan kalimat adalah modul pengucapan yang telah dikembangkan dan digunakan di Yuexiu. Peneliti merasa perlu untuk melakukan pengembangan kamus ini karena pemelajar saat ini menghadapi kesulitan untuk memperoleh aplikasi yang sesuai dengan kebutuhannya untuk melakukan pembelajaran ulang materi-materi pengucapan yang telah diperoleh di kelas.

Banyak usaha telah dilakukan oleh pengembang aplikasi untuk membuat kamus elektronik berbasis gawai (Tarp 2008). Tentu saja pengembangan akan dilakukan terhadap aplikasi yang telah teridentifikasi memiliki banyak pengguna. Oleh sebab itu, akan sangat sulit bagi pemelajar bahasa Indonesia di Tiongkok untuk mengharapakan pengembang aplikasi gawai memenuhi kebutuhan mereka karena jumlah mereka yang masih jauh di bawah target pasar pengguna aplikasi gawai. Salah satu upaya yang dapat dilakukan untuk mengatasi persoalan tersebut adalah peneliti yang sekaligus pengajar mata kuliah mengembangkan secara berkelanjutan media belajar yang diperlukan oleh mahasiswa, termasuk aplikasi kamus berbasis gawai.

Pada tahap implementasi, penelitian ini termasuk jenis penelitian pengembangan. Untuk tulisan ini, penelitian dapat diklasifikasikan sebagai studi kasus untuk memaparkan analisis kebutuhan dan perancangan prototipe. Pada analisis kebutuhan akan dipaparkan mengenai karakteristik pemelajar dan pelaksanaan proses belajar mengajar pengucapan. Pada paparan mengenai perancangan prototipe akan dijelaskan ide-ide mengenai karakteristik aplikasi, fitur-fitur interaktif pada aplikasi, dan pilihan perangkat lunak untuk membuat aplikasi. Data untuk penelitian ini terutama berupa dokumen, saran mahasiswa, dan sketsa tampilan kamus. Dokumen yang akan digunakan terdiri dari modul bahan ajar pengucapan dan administrasi pembelajaran. Saran mahasiswa adalah pernyataan secara lisan yang disampaikan mahasiswa kepada pengajar pada saat perkuliahan ataupun di luar perkuliahan. Sketsa tampilan kamus dibuat dengan menggunakan fitur gambar pada WPS Writer berdasarkan hasil diskusi antar-peneliti. Hasil penelitian akan ditampilkan dalam bentuk deskripsi eksplanatif yang disertai dengan sketsa gambar. Tampilan sketsa gambar berfungsi untuk memperjelas deskripsi eksplanatif. Sketsa gambar juga akan disertai dengan anotasi untuk mengidentifikasi kluster-kluster informasi yang penting guna menjelaskan karakter aplikasi dan fitur-fitur interaktif yang direncanakan.

III. ANALISIS PENELITIAN

3.1 Karakteristik Pemelajar

Pengguna utama aplikasi yang akan dikembangkan adalah mahasiswa tahun pertama di jurusan Bahasa Indonesia di Yuexiu. Meskipun demikian, aplikasi juga dapat digunakan oleh mahasiswa lain di Tiongkok yang mempelajari Bahasa Indonesia. Mahasiswa yang menjadi pengguna aplikasi berusia sekitar 17-19 tahun pada saat mulai belajar di Yuexiu. Mereka berasal dari provinsi yang berbeda-beda di Tiongkok, seperti Zhejiang, Yunan, Hunan, dan Guandong. Selain berbicara dalam bahasa Mandarin, sebagian besar mahasiswa ini juga berbicara dalam berbagai dialek lokal sesuai dengan daerah asalnya. Sebagian besar mahasiswa memiliki kemampuan menyimak dan berbicara dalam bahasa Inggris yang sangat baik, sehingga mereka dapat mengikuti perkuliahan di Yuexiu dengan menggunakan pengantar bahasa Inggris. Meskipun tidak semua mahasiswa di jurusan ini memilih bahasa Indonesia sebagai pilihan utama, secara umum mereka memiliki minat terhadap bahasa dan kebudayaan Indonesia.

Teknologi seperti komputer dan telepon cerdas telah menjadi gawai yang digunakan dalam frekuensi sangat tinggi oleh mahasiswa dalam kehidupan sehari-hari, baik di dalam maupun di luar kampus. Untuk melakukan komunikasi sosial secara daring, mereka menggunakan *QQ messenger* dan WeChat. Dalam proses komunikasi mengenai kegiatan belajar mengajar kedua media sosial tersebut juga dijadikan sebagai media penghubung utama antara mahasiswa dengan dosen. Media sosial ini digunakan untuk mengirim pesan antar-pribadi atau pesan untuk kelompok. Mahasiswa juga menggunakan WeChat, Alipay, JD, dan Taobao sebagai aplikasi untuk berbelanja dan melakukan pembayaran. Transaksi elektronik tersebut digunakan juga oleh mahasiswa untuk memperoleh buku dan berbagai peralatan belajar. Untuk mendukung pembelajaran bahasa, *Youdao*, *Baidu Terjemahan*, dan *Google Terjemahan* digunakan secara bersamaan untuk mencari arti kata. Penggunaan laman internet asing, seperti Google Pencarian, Instagram, Facebook, dan YouTube tidak dapat dilakukan tanpa VPN. Sebagian kecil dari mahasiswa memasang VPN pada gawai cerdasnya sehingga dapat mengakses laman-laman tersebut untuk mencari bahan-bahan pendukung belajar bahasa Indonesia.

3.2 Pelaksanaan Proses Belajar Mengajar Pengucapan

Mata kuliah Pengucapan di Yuexiu diberikan kepada mahasiswa pada semester pertama dan diampu oleh dosen yang berasal dari Indonesia. Perkuliahan dilakukan dalam durasi 90 menit per minggu, setara dengan dua satuan kredit semester. Materi perkuliahan disusun sendiri oleh dosen pengampu mata kuliah dengan mengacu pada satuan bahasa, mulai dari fonem sampai dengan kalimat sederhana. Materi tersebut disajikan dalam modul berformat PDF yang didistribusikan secara elektronik melalui *QQ messenger*. Selain modul, pengajar juga menyiapkan powerpoint presentasi dan rekaman audio. Kegiatan kuliah dilaksanakan di dalam ruang kelas yang dilengkapi dengan papan tulis, komputer, layar proyeksi, dan pengeras suara. Aktivitas selama kegiatan kuliah mencakup praktik ulang-ucap berkelompok, berpasangan, dan individu. Selain itu, pengajar juga merancang aktivitas

permainan dan memberikan penjelasan makna kata-kata khusus yang memiliki konteks khusus terkait dengan budaya Indonesia, seperti becak, petai, dan tapai. Di luar jam kuliah, mahasiswa semester satu juga dibimbing oleh pengajar untuk berlatih ulang-ucap. Kegiatan tersebut dilaksanakan setiap pagi dari pukul 07.00 - 07.50 waktu Beijing. Tujuan dari kegiatan latihan ini adalah untuk membiasakan mahasiswa mengucapkan kata-kata dalam bahasa Indonesia sehingga pengujaran yang dilakukan menjadi lebih jelas bagi pendengarnya.

3.3 Struktur Kamus Berbasis Gawai

Kamus berbasis gawai yang akan dikembangkan untuk memenuhi kebutuhan pengguna disusun dengan menggunakan pendekatan tematik. Tema yang dipilih adalah satuan bahasa. Lema yang akan disertakan di dalam kamus ini diambil dari modul belajar pengucapan yang digunakan oleh mahasiswa Yuexiu. Terdapat 10 tema di dalam modul tersebut, yaitu pengenalan abjad, vokal, konsonan, suku kata, kata dasar, kata ulang, rima, frase, dan kalimat sederhana. Untuk setiap tema, mahasiswa diberikan pengetahuan mengenai kata-kata. Setiap kata dijelaskan artinya menggunakan terjemahan ke dalam bahasa Inggris atau bahasa Mandarin disertai dengan gambar. Pada kamus yang akan dikembangkan, tema ini digunakan untuk mengorganisasi menu yang ditampilkan pada laman gawai.

Table 1. Kata-kata Berawalan Huruf Vokal

Aa	Ee	Ee	Ii	Oo	Uu
Air	Elang	Enak	Ikan	Obat	Ubi
Aku	Enam	Eka	Ini	Odol	Udang
Anak	Emas	Ekor	Itik	Omelet	Ular
Apel	Emak		Itu	Oval	Ungu
Api	Erat				
Arak	Es				

Jumlah dan jenis kata yang dapat ditemukan pada setiap tema tentu saja akan terbatas pada lema yang digunakan pada modul sebagai latihan bagi siswa. Kata-kata yang dipilih untuk latihan pengucapan vokal, misalnya, adalah kata-kata yang diawali dengan huruf vokal (lihat tabel 1). Pemilihan kata-kata untuk latihan disesuaikan dengan kemampuan mahasiswa sebagai pemelajar tingkat pemula, yaitu diusahakan hanya berupa kata dasar dan lebih banyak mencakup kata-kata nomina sehingga gambar yang disertakan pada kata dapat membantu memberikan kejelasan makna bagi pemelajar.

3.4 Sketsa Tampilan dan Fitur Interaktif Kamus

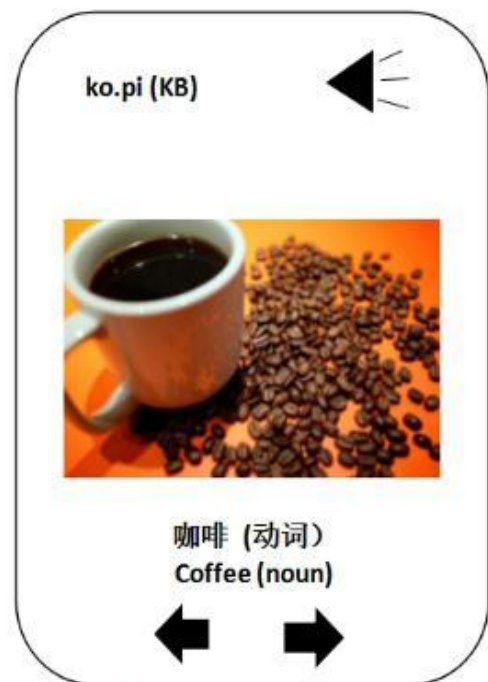
Sketsa tampilan dan fitur interaktif kamus dirancang berdasarkan pengamatan dan uji coba terhadap berbagai aplikasi kamus dan aplikasi belajar bahasa berbasis gawai, seperti Pleco, Anki, Memerise, dan Duolingo. Sketsa yang ditampilkan bukan merupakan produk akhir melainkan rancangan awal yang pada tahap implementasi dapat mengalami perubahan atau modifikasi sesuai dengan kemampuan desain pada perangkat lunak yang digunakan untuk mengembangkan kamus ini. Secara struktural, aplikasi ini akan memiliki tiga laman utama, yaitu laman beranda, laman menu, dan laman tampilan.



Figur 1. Laman Beranda

Pada laman tampilan hasil rinci, pengguna akan disajikan tampilan kata, jenis kata, arti kata, gambar, fitur sentuh untuk mendengarkan pengucapan kata, dan fitur sentuh untuk menuju pada laman kata sebelumnya atau sesudahnya dalam tema kategori kata tersebut. Arti kata akan ditampilkan sesuai dengan pengaturan pilihan yang ditentukan oleh pengguna. Tampilan arti dapat terdiri atas terjemahan dalam bahasa Inggris, bahasa Mandarin, atau keduanya. Fitur sentuh untuk mendengarkan suara juga akan menampilkan pengucapan sesuai dengan pengaturan pilihan yang ditentukan pengguna. Pengguna dapat memilih pengucapan oleh laki-laki atau pengucapan oleh perempuan.

Pada laman beranda, pengguna dapat memilih masuk ke laman menu atau memasukkan kata-kata yang akan dicari pada kolom pencarian. Ada tiga metode pemasukan kata, yaitu menggunakan papan ketik pada gawai, menuliskan kata dengan menggunakan pena elektronik atau jari, dan melakukan pemindaian terhadap tulisan pada artikel/wacana. Jika kata yang dicari ada pada direktori kamus, maka kata dan artinya akan ditampilkan pada tampilan hasil di bagian bawah ketiga fitur cara pemasukan kata. Tampilan hasil tersebut dapat disentuh dan hasilnya akan membawa pengguna kamus masuk ke laman tampilan hasil rinci.



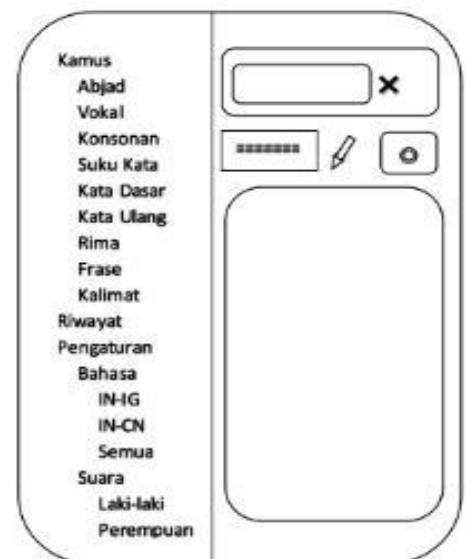
Figur 2. Laman Detail

Menu riwayat akan menampilkan daftar kata-kata yang telah dimasukkan dalam pencarian selama tiga hari terakhir. Kata-kata tersebut dikelompokkan sesuai dengan hari dan waktu pencarian. Riwayat pencarian dapat dibersihkan dengan cara menyentuh dan menggeser kata pada setiap daftar ke kiri. Menyentuh kata pada daftar riwayat dengan ketukan dua kali akan membawa pengguna ke laman tampilan rinci kata tersebut. Untuk menu pengaturan terdapat dua jenis pengaturan yang dapat dilakukan oleh pengguna, yaitu pengaturan pilihan bahasa terjemahan dan pengaturan pilihan suara. Ada tiga pilihan bahasa terjemahan, yaitu bahasa Inggris, bahasa Mandarin, atau bahasa Inggris dan bahasa Mandarin.



Figur 3. Laman Riwayat

Laman menu menampilkan tiga struktur utama, yaitu kamus, riwayat pencarian, dan pengaturan. Untuk masuk ke laman menu, pengguna dapat menyentuh simbol berupa tiga garis pada posisi kiri atas di laman pencarian. Sentuhan pada simbol ini akan menampilkan menu pada pinggir laman pencarian. Setiap kategori menu memiliki sub-menu. Untuk menu kamus akan terdapat tampilan lema-lema utama kamus. Jika lema tersebut disentuh, maka pengguna akan diarahkan menuju kata pertama pada kategori/tema.



Figur 4. Struktur Menu

3.5 Perangkat Lunak untuk Membuat Aplikasi

Peneliti berencana menggunakan perangkat lunak *Dictionary App Builder* untuk mengembangkan aplikasi yang telah dirancang. Perangkat lunak ini tersedia secara bebas bayar pada laman *SIL International*, sebuah lembaga yang telah mendedikasikan penelitian dan pengembangan bahasa dan kebudayaan. Selain didasarkan pada pertimbangan ketersediaan aplikasi secara bebas bayar, peneliti juga telah mempelajari kemungkinan perangkat lunak ini untuk merancang dan mengintegrasikan fitur interaktif yang diperlukan pembelajar pada kamus berbasis gawai, yaitu gambar dan suara. Perangkat lunak ini harus dipasang pada komputer dan perancangan aplikasi juga harus dilakukan pada komputer. Aplikasi kemudian dapat didistribusikan kepada pengguna setelah dirancang dalam bentuk paket aplikasi dan selanjutnya dipasang pada gawai. Perangkat lunak dapat digunakan untuk merancang aplikasi yang dapat digunakan pada gawai dengan sistem operasi android maupun

iOS. Pada tahap ini, peneliti masih melakukan eksplorasi terhadap perangkat lunak dan mempelajari fungsi-fungsinya.

IV. SIMPULAN

Pengucapan merupakan aspek berbahasa yang krusial dalam pembelajaran bahasa asing. Dalam pembelajaran bahasa Indonesia di Universitas Bahasa Asing Yuexiu Zhejiang, kesulitan ini dirasakan mahasiswa meskipun porsi waktu belajar Pengucapan di kelas dan luar kelas dengan bimbingan mentor cukup memadai. Untuk mengatasi kesulitan itu, mahasiswa berinisiatif membuat rekaman sendiri agar bisa berlatih secara mandiri. Tujuan penyusunan kamus tiga bahasa, bahasa Indonesia, bahasa Inggris, dan bahasa Mandarin, berbasis gawai ini adalah membantu proses belajar mandiri mahasiswa tersebut. Karena targetnya adalah pemelajar pemula, kamus multimodal interaktif ini disusun secara sederhana dan tematik berdasarkan tingkatan satuan bahasa. Oleh karena itu, kamus ini memerlukan pengembangan terus-menerus sesuai dengan perkembangan tingkat kemahiran berbahasa pemelajar bahasa.

Daftar Pustaka

- Alharbi, A.M. 2015. Building Vocabulary for Language Learning: Approach for ESL Learners to Study New Vocabulary. *Journal of International Students*, 5(4), 501-511.
- Alqahtani, M. 2015. The Importance of Vocabulary in Language Learning and How to be Taught. *International Journal of Teaching and Education*, III(3), 21-34.
- El-Sayed, A-N A-A A, & Siddiek, A.G. 2013. Monolingual and Bilingual Dictionaries as Effective Tools of the Management of English Language Education. *Theory and Practice in Language Studies*, 3(10), 1744-1755.
- Gangaiamaran, R. & Pasupathi, M. 2017. Review on Use of Mobile Apps for Language Learning. *Journal of Applied Engineering Research*, 12(21), 11242-11251.
- Godwin-Jones, R. 2017. Smartphones and Language Learning. *Language Learning and Technology*, 21(2), 3-17.
- Hamilton, H. 2012. The Efficacy of Dictionary Use while Reading for Learning New Words. *American Annals of the Deaf*, 157(4), 1-46.
- Krcelic, P. & Matijevic, A.S. 2015. *A Picture and a Thousand Words: Visual Tool in ELL*. The International Language Conference on The Importance of Learning Professional Foreign Languages for Communication Between Cultures, 18-19 September, 110-114.
- Kruyt, J.G. 1995. Technologies in Computerized Lexicography. *Lexikos* 5, 117-137.
- Laufer, B. 1998. The Development of Passive and Active Vocabulary in Second Language: Same or Difference? *Applied Linguistics*, 19(2), 255-271.
- Nushi, M. & Eqbali, M.H. 2018. 50 Languages: A Mobile Language Learning Application. *Teaching English with Technology*, 18(1), 93-104.
- Stacey, K. 2007. *A Teaching Resource for Using the Picture Dictionary for New Learners of English*. Centre for Refugee Education. Auckland University of Technology.

- Sudaryanto. 2014. Studi Tentang Perkembangan Bahasa Indonesia di Tiongkok dan Dampak Kesarjanaan Tiongkok Bagi Pengajar Bahasa Indonesia. *Kajian Linguistik dan Sastra*, 26(1), 61-70.
- Tarp, S. 2008. *Lexicography in the Borderland Between Knowledge and Non-knowledge General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer Verlag.
- Tono, Y. 2000. On the Effects of Different Types of Electronic Dictionary Interfaces on L2 Learners' Reference Behaviour in Productive/Receptive Tasks. *Proceeding of EURALEX*, 855-861.
- Vidhayasai, T., Keyurauong, S., & Bunsoun, T. 2015. *Investigating the Use of Google Translate in "Terms and Conditions" in an Airline's Official Website: Errors and Implications*. PASAA 49 (Jan-Jun 2015), 137-169.
- Yunhao. 2011. *The Use of Vocabulary Learning Strategies By Good and Poor Language Learners: A Case Study of Chinese Non-English Major Sophomores*, Thesis: Kristianstad University School of Teacher Education.
- Xiaoqiang, Y. 2014. Eksistensi Bahasa Indonesia dalam Kemitraan Strategis RI-RRT: Persepektif Tiongkok. Dalam Rina Ratih Sri Suryani dkk (ed.). *Prosiding Seminar Internasional PIBSI XXXVI Membangun Citra Indonesia di Mata Internasional Melalui Bahasa dan Sastra Indonesia*. Yogyakarta: Prodi PBSI FKIP UAD.

**KORPUS DATA BAHASA INDONESIA DALAM TEKS
INDO-BELANDA**
(INDONESIAN CORPUS DATA IN INDO-DUTCH TEXT)

Namira Choirani Fajri

Magister Kajian Sastra dan Budaya Universitas Airlangga
namirachoirani22@gmail.com

Abstrak

Makalah yang berjudul “Korpus Data Bahasa Indonesia dalam Teks Indo-Belanda” ini bertujuan untuk (1) memaparkan korpus berbahasa Indonesia dalam teks Indo-Belanda; (2) memaparkan latar belakang penggunaan korpus berbahasa Indonesia pada teks Indo-Belanda. Objek yang digunakan dalam esai ini adalah teks Indo-Belanda, yaitu teks yang ditulis dan diterbitkan oleh kaum Indo-Belanda. Teks tersebut berasal dari majalah dan teks humor karya kaum Indo-Belanda yang beredar di Indonesia. Metode pengumpulan data yang digunakan adalah metode simak. Melalui metode ini data bahasa pada objek akan peneliti simak. Selanjutnya, peneliti menggunakan metode analisis data padan sub-jenis translasional. Pendataan korpus berbahasa Indonesia pada teks Indo-Belanda dapat memperluas informasi yang disajikan pada kamus bahasa Indonesia. Selain itu, pendataan korpus tersebut juga dapat memetakan pandangan kaum Indo-Belanda terhadap Indonesia dari bahasa yang digunakan dalam teks. Proses pendataan ini juga dapat digunakan sebagai media agar data bahasa yang merupakan bagian sejarah dari perkembangan bahasa di Indonesia ini tidak hilang tertelan zaman.

kata kunci: korpus, bahasa Indonesia, teks Indo-Belanda

Abstract

The paper entitled "Indonesian Data Corps in the Indo-Dutch Texts" is intended to (1) expose the Indonesian corpuspression in the Indo-Dutch text; (2) background exposure using Korea on Indo-Dutch Indonesia. The object used in this essay is the Indo-Dutch text, the text written and published by the Indo-Dutch. The text is from the content and humor texts of Indo-Dutch works circulating in Indonesia. The data method used is the method refer. Through this method the data on the object will researchers refer. Furthermore, the researcher uses data analysis method in the sub-type of translational. Indonesian corpus data collection on Indo-Dutch texts can update the information spoken in the Indonesian dictionary. In addition, the data collection of the corpus can also map the number of Indo-Dutch to Indonesia from the language in the text. This data collection process can also be used as a medium for data language that become part of language development in Indonesia is not lost swallowed era.

keywords: corpus, Indonesian, Indo-Dutch text

I. PENDAHULUAN

Pada abad XVI, orang-orang Barat yang sampai ke Indonesia menghadapi kenyataan bahwa bahasa Melayu merupakan bahasa resmi pergaulan. Sebagai upaya untuk mempermudah komunikasi, orang-orang tersebut akhirnya belajar bahasa yang ada di suatu daerah agar dapat berinteraksi dengan penduduk asli. Saat orang-orang Barat berupaya masuk ke sekolah-sekolah yang ada di Indonesia, usaha mereka terhalang oleh bahasa. Puncak dari kegagalan ini menghasilkan sebuah keputusan dari pemerintah kolonial yang menyatakan bahwa pengajaran di sekolah-sekolah bumiputera menggunakan bahasa daerah atau bahasa Melayu (Keraf, 1984: 22-23). Hal ini memberikan pengaruh kepada orang-orang Barat yang berada di Indonesia karena secara tidak langsung mereka harus akrab dengan bahasa daerah atau bahasa Melayu pasar yang kemudian menjadi cikal bakal bahasa Indonesia.

Persebaran bahasa Indonesia pada orang-orang Barat semakin cepat dikarenakan perkawinan campuran. Saat pria Belanda dan wanita Indonesia menikah, maka anak mereka berdarah Indo-Belanda. Bahasa yang digunakan oleh kaum Indo-Belanda beragam, mulai dari bahasa Belanda, bahasa Melayu pasar, bahasa daerah, dan bahasa pergaulan mereka yang disebut bahasa Petjoek. Bahasa Petjoek merupakan bahasa percampuran antara Belanda dan Melayu pasar. Bahasa Petjoek dianggap sebagai bahasa yang kasar sehingga bahasa ini tidak boleh digunakan di rumah (Eddy Samson, wawancara, 26 Oktober 2016). Keempat bahasa ini memang tak dapat dipisahkan begitu saja dari kaum Indo-Belanda karena bahasa Belanda tak dapat memenuhi kebutuhan mereka untuk berkomunikasi, sedangkan bahasa Indonesia, bahasa daerah dan bahasa Petjoek dianggap rendah.

Sejak akhir abad ke-18 hingga awal abad ke-20 bahasa Melayu pasar yang berbau dengan bahasa Belanda banyak digunakan oleh kaum Indo-Belanda. Bahasa ini berkembang di berbagai daerah dan digunakan oleh sebagian masyarakat pendukung kebudayaan Indis sehingga menimbulkan bahasa pidgin atau bahasa campuran yang pada umumnya digunakan oleh orang-orang keturunan Belanda dengan ibu Jawa (Soekiman, 2000: 45).

Teks Indo-Belanda telah ada sejak permulaan usaha penjajahan bangsa Belanda di Indonesia yaitu awal abad ke-17 sampai beberapa tahun sesudah pengakuan kedaulatan NKRI tahun 1949. Pada teks Indo-Belanda atau teks yang ditulis serta ditujukan untuk kaum Indo-Belanda, ketiga bahasa ini, yaitu bahasa Belanda, bahasa Indonesia, dan bahasa Petjoek muncul. Bahasa pergaulan yang digunakan dalam teks Indo-Belanda dapat terbilang unik. Sistem bahasa yang mereka gunakan diambil dari berbagai bahasa sehingga menghasilkan bahasa campuran. Kaum Indo-Belanda memang lebih dominan menggunakan kata-kata berbahasa Belanda di dalam tulisannya, namun beberapa menggunakan unsur bahasa Indonesia. Hal ini menimbulkan pertanyaan tentang penyebab munculnya unsur-unsur bahasa Indonesia beserta latar belakang kemunculannya di tengah bahasa tingkat atas pada saat itu yaitu bahasa Belanda.

Pada abad XIX – XX, bahasa Indonesia berkembang bersama kaum Indo-Belanda, yaitu kaum berdarah campuran Indonesia dan Belanda. Berbagai teks Indo-Belanda yang menggunakan campuran bahasa Indonesia dan Belanda pun mulai bermunculan pada masa itu. Teks tersebut hingga saat ini masih beredar dan diproduksi oleh kaum Indo-Belanda.

Seiring berkurangnya kuantitas kaum Indo-Belanda di Indonesia, korpus berbahasa Indonesia dalam teks Indo-Belanda pun terabaikan. Hal ini sangat disayangkan mengingat ragam diksi berbahasa Indonesia dalam teks Indo-Belanda merupakan bagian dari sejarah bahasa di Indonesia. Dampak yang terjadi akibat terabaikannya korpus berbahasa Indonesia dalam teks Indo-Belanda adalah berkurangnya pemahaman kesejarahan terhadap data bahasa di dalam kamus bahasa Indonesia. Diperhatikannya korpus data bahasa Indonesia diharapkan mampu menambah wawasan pembaca kamus agar memahami konteks kebahasaan dari sudut pandang sejarah.

Pembahasan mengenai teks Indo-Belanda dan bahasa Indonesia yang terdapat pada teks Indo-Belanda masih sangat minim. Supriadi (2011) dalam skripsi yang berjudul “Bahasa Petjoek sebagai Pembebasan Ekspresi Kelompok Indo Kecil pada Empat Belas Sketsa Pierkerans Van Een Straatslijper I dan II karya Tjalie Robinson”, menganalisis tentang penggunaan bahasa Petjoek sebagai pembebasan ekspresi kelompok Indo kecil pada empat belas Sketsa *Piekerans Van Een Straatslijper I dan II* karya Tjalie Robinson. Skripsi ini membahas penggunaan dan permainan bahasa Petjoek sebagai pembebasan ekspresi mereka untuk membangun dan mempertahankan identitas. Skripsi ini mengangkat bahan kajian bahasa Petjoek yang diambil dari naskah sketsa *Piekerans Van Een Straatslijper* karya Tjalie Robinson, bukan diksi bahasa Indonesia yang terselip pada bahasa Petjoek. Bahasa Petjoek di sini hanya dianalisis fungsinya sebagai media pembangunan dan pemertahanan identitas, tidak secara langsung menganalisis kemunculan diksi bahasa Indonesia di dalamnya.

Fatimah (2017) dalam skripsi yang berjudul “Kaum Indo sebagai Pengguna Bahasa Perjoek di Surabaya Tahun 1870-1942” memaparkan tentang kaum Indo-Belanda sebagai pengguna bahasa Petjoek. Skripsi tersebut membahas akulturasi budaya yang merupakan hasil perkawinan yang terjadi antara orang Eropa dan Jawa. Melalui skripsi ini, penulis dapat mengetahui sejarah bahasa yang digunakan oleh kaum Indo-Belanda sehingga dapat menambah bahan mengenai latar belakang munculnya korpus data bahasa Indonesia dalam teks Indo-Belanda.

II. KERANGKA TEORI DAN METODE PENELITIAN

2.1. Kerangka Teori

2.1.1 Korpus Data

Korpus data adalah data yang digunakan sebagai sumber penelitian. Korpus data dapat diperoleh dari berbagai sumber lisan maupun tulis. Minimnya jumlah komunitas Indo-Belanda saat ini menyebabkan pelacakan bahasa Indonesia dalam komunikasi mereka sangat sulit dilakukan, oleh karena itu pada penelitian ini penulis menggunakan sumber tertulis sebagai sumber korpus data. Adapaun sumber tertulis dapat diperoleh dari berbagai media, seperti media massa, jurnal, dan karya sastra. Fokus penelitian ini adalah teks Indo-Belanda yang ditulis oleh kaum Indo-Belanda.

Menurut pedoman penyusunan kamus, dalam upaya pembinaan penggunaan bahasa Indonesia serta sebagai upaya penyatuan konsep dan pemahaman maka korpus data baiknya mengemban amanat dan wawasan yang luas dan mendalam mengenai bahasa Indonesia.

2.2 Metode Penelitian

Penelitian ini menggunakan pendekatan kualitatif, metodologi kualitatif merupakan prosedur yang menghasilkan data deskriptif. Istilah deskripsi ini mengisyaratkan bahwa penelitian yang dilakukan semata-mata hanya berdasarkan fakta yang ada atau fenomena yang memang dihasilkan atau dicatat berupa pemberian yang biasa dikatakan sebagai potret atau paparan seperti adanya.

Metode pengumpulan data yang digunakan adalah metode simak. Metode simak adalah metode yang digunakan dalam penelitian bahasa dengan cara menyimak penggunaan bahasa pada objek yang akan diteliti. Kemudian, dilakukan teknik sadap, yaitu pelaksanaan metode simak dengan menyadap penggunaan bahasa seseorang atau beberapa orang.

Metode yang digunakan dalam penelitian ini adalah metode padan sub-jenis translasional. Teknik dasar yang digunakan dalam metode ini adalah teknik pilah unsur penentu dengan daya pilah sebagai pembeda sifat dan watak aneka *language*. Selanjutnya, dilakukan teknik lanjutan yang menggunakan teknik hubung banding memperbedakan (Sudaryanto, 2015: 25-35). Penelitian ini menggunakan unsur bahasa Belanda pembeda sifat dengan unsur bahasa Indonesia.

Hasil penelitian ini adalah adanya unsur-unsur berbahasa Indonesia dalam teks Indo-Belanda yang kemudian diteliti latar belakang penggunaannya serta fungsinya. Metode penyajian data ada dua macam, yaitu bersifat formal dan informal (Sudaryanto, 2015: 240). Pemaparan hasil analisis data dalam penelitian ini menggunakan metode informal. Menurut Sudaryanto (2015: 241) metode penyajian informal adalah perumusan dengan kata-kata biasa, namun ada pengalihan dengan bahasa lain. Dengan demikian, sajian hasil analisis data dalam penelitian ini tidak memanfaatkan berbagai lambang, tanda, dan singkatan, seperti yang biasa digunakan dalam metode penyajian formal karena penelitian ini tidak memerlukan notasi formal.

III. Pembahasan

Pada makalah ini, penulis akan memaparkan korpus bahasa Indonesia dalam teks Indo-Belanda yang berjudul *Je-lâh-je-kripoet*. *Je-lâh-je-kripoet* adalah kumpulan humor dari Hindia-Belanda yang memuat tentang cerita terbaik dan lelucon dari tujuh jilid Tongtong. Kumpulan humor *Je-lâh-je-kripoet* merupakan antologi dari majalah Tongtong, sebuah majalah Indis yang telah ada sejak tahun 1956. Terbit pada tahun 1980, *Je-lâh-je-kripoet* ditulis oleh Tjalie Robinson, seorang penulis Indo-Belanda. Kumpulan humor ini memuat tentang humor yang sering didengar di pasar malam.

Kumpulan humor *Je-lâh-je-kripoet* menceritakan keseharian kaum Indo-Belanda serta interaksinya dengan pribumi, totok, dan kaum Indo-Belanda. Latar tempat dari cerita yang dimuat dalam *Je-lâh-je-kripoet* berada di Indonesia. Teks tersebut banyak berkisah tentang kesalahpahaman bahasa dan budaya kaum Indo-Belanda dengan pribumi. Kesalahpahaman tersebut kemudian menimbulkan efek humor pada teks *Je-lâh-je-kripoet*.

3.1 Korpus Bahasa Indonesia dalam Teks Indo-Belanda *Je-lâh-je-kripoet*

Berikut merupakan data bahasa Indonesia yang terdapat pada teks *Je-lâh-je-kripoet*.

No	Ejaan dalam Teks	Ejaan Bahasa Indonesia	Jumlah Kemunculan	Kelas Kata
1	ada	ada	1	v
2	adoe/ adoech	aduh	7	p
3	air	air	3	n
4	ajam	ayam	4	n
5	ajo	ayo	3	p
6	ajo	ayo	1	p
7	akar	akar	1	n
8	akoe	aku	4	n
9	amblas	ambblas	1	v
10	antik/ antiek	antik	2	a
11	apa	apa	12	pron
12	appel	apel	2	n
13	asia	asia	2	n
14	astagah	astaga	1	p
15	atro	atau	1	p
16	awas	awas	2	v
17	babi	babi	2	n
18	baboe	baba	3	n
19	bagaimana	bagaimana	1	pron
20	baik	baik	1	a
21	bajem	bayem	1	n
22	bakoel	bakul	1	n
23	bangoen	bangun	2	v
24	banjak	banyak	2	a
25	barang	barang	3	n
26	baroe	baru	1	a
27	batok	batok	2	n
28	bawah	bawah	1	n
29	bawang	bawang	3	n
30	belandja	belanja	1	v
31	beli	beli	1	v
32	berani	berani	1	a
33	berapa	berapa	1	pron
34	berdjalan	berjalan	1	v
35	beres	beres	1	a
36	besar	besar	2	a
37	betoel	betul	6	a
38	bikin	bikin	1	v
39	bikinan	bikinan	1	n
40	bilang	bilang	4	v

No	Ejaan dalam Teks	Ejaan Bahasa Indonesia	Jumlah Kemunculan	Kelas Kata
41	bisa	bisa	1	v
42	bloestroe	blustru	1	n
43	boeah	buah	1	n
44	boeang	buang	4	v
45	boeat	buat	1	v
46	boekan	bukan	1	adv
47	boeloe	bulu	1	n
48	boemboe	bumbu	1	n
49	boeng	bung	1	n
50	bohong	bohong	4	a
51	boleh	boleh	2	adv
52	boloe	bulu	1	n
53	cari	cari	1	v
54	dapoer	dapur	2	n
55	dari	dari	1	p
56	datang	datang	8	v
57	depot	depot	2	n
58	di	di	2	p
59	dia	dia	2	n
60	dia	dia	2	n
61	diam	diam	1	v
62	diam	diam	1	v
63	djuga	juga	2	v
64	djago	jago	1	n
65	djahat	jahat	1	a
66	djaman	zaman	2	n
67	djamboe air	jambu air	1	n
68	djangan	jangnan	6	adv
69	djoega	juga	1	n
70	djongkok	jongkok	1	v
71	djongos	jongos	10	n
72	dodol	dodol	2	n
73	doea/ doewa	dua	4	num
74	doekoen	dukun	3	n
75	doeren	durian	2	n
76	doeri	duri	1	n
77	doewit	duit	1	n
78	dokter	dokter	3	n
79	dong	dong	1	p
80	dongeng	dongeng	1	n

No	Ejaan dalam Teks	Ejaan Bahasa Indonesia	Jumlah Kemunculan	Kelas Kata
81	enak	enak	1	A
82	gadoh-gadoh	gado-gado	1	n
83	gajoeng	gayung	1	n
84	gantoeng	gantung	1	v
85	gemoek	gemuk	3	a
86	gendoet	gendut	5	a
87	goea	gua	2	n
88	goela batu	gula batu	2	n
89	goeling	guling	5	n
90	goenoeng	gunung	2	n
91	goreng	goreng	2	n
92	habis	habis	4	v
93	haloes	halus	1	a
94	hari	hari	2	n
95	hati	hati	1	n
96	hebat	hebat	1	a
97	ija	iya	1	p
98	ikan	ikan	1	n
99	ini	ini	3	pron
100	itoe	itu	9	pron
101	jang	yang	1	p
102	kabaja	kebaya	1	n
103	kaja	kaya	2	p
104	kajoepoetih	kayu putih	1	n
105	kaki	kaki	1	n
106	kalau	kalau	2	p
107	kalong	kalong	3	n
108	kamar/kamer	kamar	3	n
109	kami	kami	1	n
110	kamoe	kamu	1	n
111	kampong/ kampoeng	kampung	3	n
112	kan	kan	1	adv
113	kankoeng	kangkung	1	n
114	kantong	kantong	1	n
115	kapan	kapan	2	pron
116	kapok	kapuk	1	n
117	kasih	kasih	1	v
118	katholiek	katolik	1	n
119	katjong	kacong	7	n

No	Ejaan dalam Teks	Ejaan Bahasa Indonesia	Jumlah Kemunculan	Kelas Kata
177	nanti	nanti	1	n
178	nasi goreng	nasi goreng	1	n
179	nganggoer	nganggur	1	v
180	ngobrol	ngobrol	2	v
181	ngrokok	ngerokok	1	v
182	-nja	-nya	1	adv
183	nja/ njah	nya	2	n
184	njai	nyai	3	n
185	njamoe	nyamuk	1	n
186	njo	nyo	4	n
187	njonja	nyonya	12	n
188	obat	obat	1	n
189	oebie	ubi	1	n
190	oeroet	urut	1	v
191	oleh-oleh	oleh-oleh	1	n
192	oma	oma	1	n
193	orang	orang	2	n
194	paman	paman	1	n
195	panggil	panggil	1	v
196	pasar	pasar	1	n
197	pedot	pedot	1	v
198	pegel	pegal	1	a
199	pendek	pendek	1	a
200	penghoeloe	penghulu	1	n
201	perintah	perintah	1	n
202	petis	petis	1	n
203	pidjit	pijit	2	v
204	pigi	pergi	2	v
205	piring	piring	1	n
206	poetih	putih	1	n
207	pohon waringin	pohon beringin	1	n
208	rami	rami	1	n
209	redjeki	rezeki	1	n
210	robek	robek	1	a
211	roedjak	rujak	2	n
212	roemah	rumah	1	n
213	roepanja	rupanya	1	adv
214	roewet	ruwet	2	a
215	sabar	sabar	1	a

120	katok	katok	1	n
121	kawat	kawat	2	n
122	kebon	kebun	4	n
123	kedjeblos	kejeblos	1	v
124	kejantol	kejantol	1	v
125	kembang poekoel teroes	kembang pukul tujuh	1	n
126	keris	keris	1	n
127	kerok	kerok	1	v
128	ketak	ketak	1	v
129	ketebroek	ketubruk	1	v
130	ketjap	kecap	1	n
131	ketjil	kecil	2	a
132	ketjoeboeng	kecubung	1	n
133	ketumbar	ketumbar	1	n
134	kira-kira	kira-kira	1	adv
135	klamboe	kelambu	1	n
136	klentang	klentang	1	n
137	koerang atjar	kurang ajar	1	adv
138	kokki	koki	2	n
139	kopjes toebroek	kopi tubruk	1	n
140	kripiiek	keripik	1	n
141	kripoet	keriput	1	n
142	kroepoek	kerupuk	2	n
143	kwatir	khawatir	1	a
144	kwee/ koewe	kue	3	n
145	lagi	lagi	1	adv
146	lakoe	laku	1	a
147	lari	lari	1	v
148	loe	lo	3	n
149	loempiah	lumpia	1	n
150	loepa	lupa	2	v
151	loepoet	luput	2	v
152	loerah	lurah	1	n
153	loewak	luwak	1	n
154	loh	lo	2	p
155	mabok	mabuk	1	a
156	makan	makan	3	v

216	sadja	saja	4	adv
217	saja	saya	1	n
218	salah	salah	1	a
219	sama	sama	2	a
220	sampe	sampai	1	v
221	sapoe lidi	sapu lidi	1	n
222	saroendeng	serundeng	1	n
223	sate	satai	4	n
224	satoe	satu	2	num
225	sekarang	sekarang	2	n
226	semoea	semua	2	num
227	serabi	serabi	1	n
228	setengah	setengah	3	num
229	si	si	2	p
230	singkong	singkon	1	n
231	sinjo	sinyo	1	n
232	sioong	siung	1	n
233	sirih	sirih	1	n
234	soedah	sudah	5	adv
235	soeka	suka	1	a
236	soekoen goreng	sukun goreng	1	n
237	soekoer	sukur	2	n
238	soeroh	suruh	1	n
239	soesah	susah	1	a
240	sore	sore	1	n
241	sosis	sosis	1	n
242	soto	soto	1	n
243	tak	tak	1	adv
244	tambak	tambak	1	n
245	tamoe	tamu	1	n
246	tanah	tanah	1	n
247	tanja	tanya	1	v
248	taoe	tau	2	v
249	tapi/ tetapi	tetapi	2	p
250	taroeh	taruh	2	n
251	taroek	taruh	1	v
252	tawon	tawon	1	n

157	makanan	makanan	1	n
158	malam	malam	1	n
159	maloe	malu	1	a
160	mampoes	mampus	1	v
161	mantri	mantri	2	n
162	maoe	mau	1	adv
163	marah	marah	1	v
164	masa	masa	1	n
165	masah	masa	1	adv
166	masak	masak	2	a
167	mati	mati	1	v
168	melotot	melotot	2	v
169	mie	mie	1	n
170	minjak	minyak	1	n
171	minta	minta	5	v
172	moeloet	mulut	2	n
173	moesti	mesti	2	adv
174	montor	motor (mobil)	2	n
175	nah	nah	1	p
176	nama	nama	1	n
253	telor	telur	2	n
254	telor-asin	telur asin	1	n
255	tempe	tempe	2	n
256	tempieling	tempiling	1	v
257	terima kasih	terima kasih	1	n
258	tida	tidak	4	adv
259	tikar	tikar	1	n
260	tikoos	tikus	3	n
261	tjabe	cabai	1	n
262	tjatjing	cacing	1	n
263	tjeleng	celeng	5	n
264	tjemploeng	cemplung	1	v
265	tjilaka	celaka	3	a
266	tjitjak	cicak	1	n
267	tjoba	coba	4	v
268	tjobek	cobek	1	n
269	tjoklat	coklat	1	n
270	toean/toewan	tuan	16	n
224	satoe	satu	2	num

Tabel 1. Korpus Bahasa Indonesia dalam teks Indo-Belanda *Je-lâh-je-kripoet*

Melalui paparan beberapa data bahasa Indonesia dalam teks Indo-Belanda *Je-lâh-je-kripoet*, dapat terlihat bahwa kelas kata yang mendominasi adalah nomina. Salah satu data bahasa Indonesia yang banyak terdapat dalam teks Indo-Belanda *Je-lâh-je-kripoet* adalah *djongos*. Apabila meninjau penggunaan kata *djongos* pada teks tersebut, sekalipun kata tersebut berusaha menunjukkan perbedaan derajat, namun tidak digunakan untuk merendahkan. Kata *djongos* atau yang dalam Ejaan Bahasa Indonesia ditulis dengan *jongos* artinya pembantu atau pelayan rumah tangga yang berjenis kelamin laki-laki (KBBI Daring, <https://kbbi.kemdikbud.go.id/entri/jongos>, 2017, diakses pada 13 Maret 2017). *Jongos* merupakan sebutan kaum Belanda pada pelayan laki-laki mereka. Pelayan ini sebenarnya adalah kuli di wilayah perkebunan di Jawa Tengah dan Jawa Timur. Mereka dibayar rendah oleh tuan tanah Belanda untuk dilatih mengadakan perjamuan makan bergaya kolonial di rumah mereka (Rahman, 2016: 61)

Pada masa Hindia-Belanda, profesi ini identik dengan profesi seorang pribumi maka layak jika unsur bahasa Indonesia *jongos* digunakan agar lebih mudah diucapkan lidah pribumi. Kata *jongos* menurut Onghokham (1996: 25), secara etimologi kemungkinan berasal dari bahasa Belanda yaitu *jongen* yang artinya anak laki-laki. Asal kata tersebut dapat menunjukkan bahwa kata *jongos* bukan merupakan variasi kata kasar.

3.2 Latar Belakang Kemunculan Korpus Bahasa Indonesia dalam Teks Indo-Belanda

Bahasa Indonesia yang muncul pada teks Indo-Belanda didominasi kata serapan dari bahasa Jawa. Kemunculan kata serapan berbahasa Jawa ini disebabkan latar tempat peristiwa yang diceritakan dalam teks berada di Jawa. Selain itu, interaksi orang totok, berdarah campuran, dan pribumi juga menjadi salah satu faktor dominasi serapan bahasa Jawa. Tidak hanya itu, diksi bahasa Jawa memang beragam dan sangat detail, sehingga penutur kesulitan menemui padanan yang tepat dalam bahasa lain.

Kosa kata berbahasa Indonesia dalam teks Indo-Belanda *Je-lâh-je-kripoet* kebanyakan digunakan untuk mengutarakan emosi dikarenakan orang Belanda memang memiliki kebiasaan ekspresif dalam mengutarakan perasaannya. Selain itu, teks-teks di dalam kumpulan humor *Je-lâh-je-kripoet* merupakan teks yang hampir keseluruhan berisi interaksi antara penutur dan mitra tutur, oleh karena itu satuan bahasa yang dimunculkan memang kerap kali ingin menyampaikan emosi dari penutur kepada mitra tuturnya. Melalui bahasa, mitra tutur diharapkan mampu merasakan ungkapan perasaan atau emosi penutur. Walaupun antara penutur dan mitra tutur tidak memiliki konvensi bahasa yang sama, namun lewat fungsi bahasa Indonesia yang terdapat di dalam teks diharapkan mitra tutur mampu memahami keinginan penutur sekalipun mitra tutur belum tentu memahami keseluruhan maksud penutur. Bahasa Indonesia dalam kumpulan humor Indo-Belanda *Je-lâh-je-kripoet* diharapkan mampu berfungsi sebagai penunjuk perasaan penutur agar emosi dapat tersampaikan pada mitra tutur.

Berikut merupakan contoh pemaparan latar belakang penggunaan bahasa Indonesia dalam teks Indo-Belanda.

Data :

Totok *mevrouwtje knikte tevreden tót ze opeens die haan zag!!* (hlm. 12)

Terjemahan kalimat di atas adalah *nyonya totok mengangguk puas sampai tiba-tiba melihat ayam itu!!* Kalimat tersebut terdapat pada teks berjudul “De NJONJA-KIP” yang artinya “Nyonya Ayam”. Pada Kamus Besar Bahasa Indonesia kata *totok* artinya adalah asli; sejati atau bukan peranakan (KBBI Daring, <https://kbbi.kemdikbud.go.id/entri/totok>, 2017, diakses pada 13 Maret 2017). Teks berjudul “De NJONJA-KIP” ini menceritakan tentang seorang nyonya yang meminta kokinya membuatkan hidangan ayam. Pada data di atas, sebutan *totok* ditujukan pada seorang nyonya yang berdarah asli, bukan seorang campuran Indo-Eropa.

Data di atas menggunakan kata *totok* sekalipun bahasa Belanda sebenarnya memiliki padanan katanya yaitu *volbloed*, namun dalam kumpulan humor Indo-Belanda *Je-lâh-je-kripoet* tidak menggunakan diksi itu mengingat kalimat ini digunakan dalam teks yang dekat dengan kebudayaan Indis. Bagi yang berdarah pribumi, lebih mudah menyebut kebalikan dari mereka, yaitu orang berdarah Eropa, dengan sebutan yang akrab di lidah mereka. Kata *volbloed* lebih sulit diucapkan dan diingat bagi mereka yang bukan penutur bahasa Belanda. Pada masa itu kata *totok* memang tak asing dan selalu digunakan untuk menyebut mereka yang berdarah asli dan tidak campuran.

Diksi *totok* merupakan istilah bahasa Indonesia yang diserap dari bahasa Jawa yang artinya asli; tulen, dan tidak peranakan (Kamus Bahasa Jawa, 2010: 190). Kata ini kerap kali

digunakan untuk mendeskripsikan para pendatang yang lahir di luar negeri atau berdarah murni. Pada masa Hindia Belanda, kata *totok* merujuk pada orang Belanda atau Eropa, namun sebenarnya kata ini juga dapat digunakan untuk orang Tionghoa asli yang datang ke Indonesia.

Kata *totok* pada saat ini sudah jarang digunakan. Hal ini dikarenakan minimnya perbedaan ras yang ditemui, tidak seperti masa kolonial dahulu, sehingga penjelasan tentang yang *totok*, yang campuran atau Indo-Eropa atau peranakan, serta pribumi perlu ditekankan untuk mengetahui penempatan mereka di mata sosial. Seperti yang diketahui, status dan perlakuan antara *totok*, campuran, dan pribumi tidak dapat dicampuradukkan (Niwandhono, 2011: 27), oleh karena itu pembeda ini menjadi sering diucapkan.

Penggunaan unsur bahasa Indonesia yaitu *totok* yang diikuti dengan kata *mevrouwtje* menunjukkan identitas kebudayaan Indis yang sangat jelas. Penggabungan kedua kata yang berbeda bahasa ini menjadi satu frasa memunculkan pesan humor karena kedua kata ini memiliki ketimpangan yang kentara. Penyebutan *mevrouwtje* tanpa diawali dengan kata *totok* sebenarnya telah mampu menunjukkan identitas sang nyonya yang asli Belanda, namun untuk memperdalam kesan humor dan menunjukkan sudut pandang pribumi dengan kepolosannya dalam kalimat tersebut dibubuhi dengan kata *totok*. Faktor superioritas tampak pada teks berjudul “De NJONJA-KIP” sehingga hal ini menjadi salah satu hal yang menunjang humor di dalamnya.

Berikut merupakan beberapa penggunaan bahasa Indonesia dalam teks Indo-Belanda *Je-lâh-je-kripoet* beserta latar belakang penggunaannya.

No	Diksi/frasa	Kalimat	Terjemahan	Latar Belakang Penggunaan
1	kripoet	<i>Je-lâh-je-kripoet</i>	Anda tertawa Anda keriput	Mengharapkan pembaca tertawa hingga menampakkan garis keriput. Pada budaya Indonesia, tertawa terbahak-bahak tidak sopan, sehingga kaum Indo Belanda yang kemudian membaca judul tersebut diharapkan dapat membebaskan diri dari ketegangan di antara perbedaan yang pada kedua budaya.
2	paman	<i>Vele indischgasten herinneren zich vooral van het platteland dat wanneer een paman of b.v. onze baboe</i>	banyak orang Hindia ingat kampung asalnya, ketika paman atau babu kami	Sebutan untuk pribumi yang belum dikenal penulis secara mendalam
3	baboe	<i>onze baboe plotseling in de lach schoot</i>	babu kami tiba-tiba tertawa	Orang Belanda terbiasa membedakan melalui

				gender
4	totok	<i>Totok mevrouwtje knikte tevreden tót ze opeens die haan zag!!</i>	Nyonya totok mengangguk sampai tiba-tiba melihat ayam itu!!	Menunjukkan perbedaan status
5	doekoen	<i>De vriend bracht hem bij een doekoen, aan wie hij zijn probleem voorlegde.</i>	temannya membawanya ke seorang dukun , yang bisa menyelesaikan masalahnya	Berdasarkan konteksnya yaitu guna-guna, penggunaan diksi dukun menunjukkan spesifikasi klenik khas Indonesia
6	djongos	<i>...djongos Marto voor het diner de tafel met het moostie damast gedekt...</i>	djongos Marto menutupi meja dengan kain sutera terbaik	Orang Belanda terbiasa membedakan melalui gender. Menurut salah satu sumber berasal dari kata berbahasa Belanda <i>jongen</i> yang berarti anak laki-laki
7	toewan	<i>Verdomde lummel, toewan</i>	orang udik sialan, tuan	Menunjukkan perbedaan status
8	kalong	<i>Nah, drie kalong tuurlijk</i>	nah, tentu itu tiga kalong	Varietas hewan yang belum tentu sama antara Indonesia dan Belanda
9	tjeleng	<i>Neen, deze keer eens niet op tjelèng, maar op meliwis</i>	tidak, kali ini bukan berburu tjeleng tetapi meliwis	Varietas hewan yang belum tentu sama antara Indonesia dan Belanda
10	katok	<i>katok en badjoe tjina</i>	katok dan baju cina	<i>Katok</i> yang dimaksudkan pada teks ini bukan merupakan pakaian khas Eropa. Pada masa itu, mereka menggunakan celana panjang berkain motif batik dan berbaju potongan Tionghoa (baju takwa) untuk kaum laki-laki
11	Goena-goena	Goena-goena	Guna-guna	Tidak terdapat padanan yang tepat dalam bahasa Belanda
12	gendoet	<i>Gendoet is ook een Indo, maar voor deze toch geen nadere beschrijving nodig; zijn bijnaam zegt al alles.</i>	gendoet juga seorang Indo, tetapi tidak perlu disebutkan penjelasan secara rinci, julukan yang diberikan sudah	<i>Gendut</i> merupakan kata yang spesifik karena keadaan gemuk pada perut saja. Kata ini digunakan untuk mengolok-olok.

			menunjukkan keadaan tubuhnya	
13	njai	<i>“H’m! De mannen van vroeger die hádden njai’s; de mannen van nu zijn njai’s...”</i>	“H’m! Laki-laki pada masa lalu ini masih memiliki njai ; Laki-laki pada masa sekarang tidak ada lagi njai .”	<i>Nyai</i> pada masa kolonial identik dengan gundik
14	pidjit	<i>In de medische wetenschappen krijgt de studie van het pidjit en oeroet een vooraanstaande plaats.</i>	Ilmu kedokteran akan mempelajari pidjit dan oeroet sebagai sesuatu yang populer	Tidak terdapat padanan yang tepat dalam bahasa Belanda
15	Oeroet	<i>In de medische wetenschappen krijgt de studie van het pidjit en oeroet een vooraanstaande plaats.</i>	Ilmu kedokteran akan mempelajari pidjit dan oeroet sebagai sesuatu yang populer	Tidak terdapat padanan yang tepat dalam bahasa Belanda
16	koerang atjar	<i>“Wat heb je met mijn rotte kiezen te maken!... Koerang atjar!”</i>	apa yang telah Anda lakukan dengan gigi busuk saya! Kurang ajar!	Ditujukan pada pribumi
19	mampoes	<i>... als achttien jaar jij prik jou vinger en dan mampoes.</i>	ketika kamu berumur delapan belas tahun, jarimu akan ditusuk dengan jarum dan kemudian mampoes	Tidak memiliki padanan yang tepat dalam bahasa Belanda
22	roewet	Als assimilatie, roewet-ah!	jika asimilasi, ruwet ah!	Menunjukkan unsur budaya Jawa
31	kerok	Vroeger: Kerok met een <i>benggol</i> Nou: met een <i>stuiver</i> ...	dulu: kerok menggunakan koin <i>benggol</i> , sekarang: dengan koin sen	Tidak terdapat padanan yang tepat dalam bahasa Belanda

Tabel 2. Latar Belakang Penggunaan Bahasa Indonesia dalam Teks Indo-Belanda

IV. Simpulan

Penggalan data bahasa Indonesia dalam teks Indo-Belanda dapat melahirkan pengetahuan yang lebih luas. Korpus data bahasa Indonesia di atas dapat digunakan sebagai pijakan awal peneliti untuk menggali lebih dalam mengenai penggunaan bahasa Indonesia dalam teks Indo-Belanda. Dilakukannya penggalan korpus data bahasa Indonesia dalam teks Indo-Belanda dapat menjadi salah satu alternatif peneliti bahasa Indonesia untuk menelaah konteks pada masa kolonial dari sudut pandang kaum Indis. Manfaat yang kemudian dapat diperoleh adalah peneliti dapat mengetahui konteks kebahasaan pada suatu masa, sehingga tidak menyimpulkan dari satu sudut pandang saja. Selanjutnya peneliti dapat mengetahui

perubahan kebahasaan yang terjadi dari masa ke masa. Dari korpus bahasa Indonesia yang telah dipaparkan di atas, dapat dilihat nukilan dari latar belakang penggunaan bahasa Indonesia pada masa itu. Penggunaan bahasa Indonesia yang pada masa kini dianggap kasar belum tentu memiliki kesejarahan demikian. Proses pendataan ini juga dapat digunakan sebagai media agar data bahasa yang merupakan bagian sejarah dari perkembangan bahasa di Indonesia ini tidak hilang tertelan zaman.

Daftar Pustaka

- Badan Pengembangan dan Pembinaan Bahasa. 2017. KBBI Daring. Daring.<https://kbbi.kemdikbud.go.id/> diakses pada 11 Januari-20 April 2017.
- Darminto, dkk. 2010. *Bausastra Jawa Kamus Jawa Besar*. Surakarta: Kharisma.
- Faisal, Sanapiah. 1990. *Penelitian Kualitatif: Dasar-dasar dan Aplikasi*. Malang: Yayasan Asih Asah Asuh Malang.
- Fatimah, Annisa Nurani. 2017. "Kaum Indo sebagai Pengguna Bahasa Petjoek di Surabaya Tahun 1970-1942". Skripsi. Fakultas Ilmu Budaya Universitas Airlangga.
- Keraf, Gorys. 1984. *Tata Bahasa Indonesia*. Flores: Nusa Indah.
- Niwandhono, Pradipto. 2011. *Yang Ter(di)lupakan*. Yogyakarta: Djaman Baroe.
- Sastrowardoyo, Subagio. 1990. *Sastra Hindia Belanda dan Kita*. Jakarta: Balai Pustaka.
- Soekiman, Djoko. 2000. *Kebudayaan Indis: Dan Gaya Hidup Masyarakat Pendukungnya di Jawa (Abad XVIII-Medio Abad XX)*. Yogyakarta: Bentang Pustaka
- Supriadi, Bakti. 2011. "Penggunaan Bahasa Petjoek sebagai Pembebasan Ekspresi Kelompok Indo Kecil pada Empat Belas Sketsa Piekerans Van Een Straatslijper I dan II karya Tjarlie Robinson" <http://budaya-indonesia.org/a/e/m/8957/e/> diakses pada 6 Juni 2016.

PEMANFAATAN KORPUS DALAM PENYUSUNAN KAMUS KONSTRUKSI PREPOSISIONAL

Totok Suhardijanto
Universitas Indonesia

Miel Slager
Utrecht University

Abstrak

Makalah ini menyajikan upaya kolaboratif dalam menyusun kamus yang mencerminkan intuisi penutur jati bahasa Indonesia terkait dengan konstruksi seperti V + preposisi + obyek (*berpengaruh terhadap; tergantung pada*), Adj + preposisi + obyek (*iri dengan; cinta akan*), atau N + preposisi + obyek (*sikap terhadap; ingatan atas*). Intuisi tersebut terkait dengan penggunaan bahasa yang memunculkan kombinasi tertentu (ragam formal atau nonformal, dsb), keterbatasan makna (hanya digunakan dalam teks bertopik tertentu), dan berbagai keterbatasan sintaktis, baik di kiri atau kanan—misalnya, subyek harus bernyawa/tak bernyawa, adjektiva hanya dapat digunakan untuk seseorang atau sesuatu dengan karakteristik tertentu, obyek harus bernyawa/tak bernyawa, atau dapat berupa klausa, baik interogatif maupun tidak.

Dalam penyusunan kamus, digunakan korpus sebagai dasar penentuan makna maupun tipe konstruksi berdasarkan informasi frekuensi maupun konkordansi. Korpus berjumlah 18,3 juta kata dan disusun dari berbagai jenis teks tulis mencakup artikel surat kabar dan majalah, karya imajinatif, tulisan blog, surat, karya ilmiah, dan buku teks. Pemanfaatan korpus, terutama terkait dengan metode linguistik korpus—misalnya konkordansi, kolokasi, dan koligasi, berfungsi sangat efektif dalam tahap pengidentifikasian konstruksi, penentuan makna, dan penjelasan konteks pemakaian sebuah konstruksi preposisional. Tiap kata dikelompokkan berdasarkan akar (*root*) yang sama. Jadi, kita akan menemukan verba *berpengaruh* di dalam entri yang sama dengan verba *mempengaruhi* dan nomina *pengaruh*, serta adjektiva *peka* dikelompokkan bersama dengan nomina *kepekaan*. Kamus akan berisi 6500 hingga 7000 entri. Saat ini kamus sedang dalam tahap pengerjaan akhir.

Kata kunci: korpus, kolokasi, koligasi, struktur preposisional, bahasa Indonesia

I. PENDAHULUAN

Menurut Atkins dan Rundell (2008: 2), kamus sering kali dianggap sebagai rekaman otoritatif tentang bagaimana seorang penutur bahasa “seharusnya” menggunakan bahasa. Dalam

banyak kasus, kamus lebih sering dianggap sebagai panduan tentang pemakaian yang “benar”. Tentu saja, bagi seorang leksikograf, hal ini merupakan “tuduhan” yang mengganggu pekerjaan menyusun kamus. Dalam sudut leksikograf, kamus merupakan deskripsi bahasa yang digunakan oleh anggota masyarakat bahasa. Jadi jelas bahwa bagi leksikograf, kamus merupakan dokumen yang menggambarkan bagaimana bahasa digunakan.

Makalah ini menyajikan upaya kami menyusun kamus yang menggambarkan intuisi penutur jati bahasa Indonesia berkenaan dengan konstruksi preposisional. Penting dijelaskan apa yang dimaksud dengan konstruksi preposisional dalam tulisan. Konstruksi preposisional merupakan entitas yang serupa dengan “bergantung pada FN (Frasa Nomina)”, “berpikir tentang FN”, “pendapatku tentang FN” dan sebagainya. Sebuah kamus yang memuat konstruksi preposisional bertujuan untuk merepresentasikan pengetahuan penutur jati tentang konstruksi tersebut dalam bahasa mereka.

Menurut Hagege (2010) preposisi—termasuk ke dalamnya posposisi dalam istilah payung adposisi—dapat dibatasi sebagai alat gramatikal yang memarkahi hubungan di antara dua bagian kalimat: salah satunya merupakan unsur yang dikendalikan oleh adposisi. Comrie (1991: 91) menyebutkan bahwa preposisi merupakan anggota kelas tertutup yang (i) muncul sebelum atau setelah sebuah pelengkap membentuk FN, nomina, pronomina, atau klausa yang berfungsi sebagai frasa nomina dan (ii) membentuk sebuah struktur tunggal dengan pelengkap untuk mengungkapkan relasi gramatikal dan semantis dengan unsur lain di dalam klausa.

Sejalan dengan Hagege, Alwi et al. (1998: 288) mengusulkan bahwa preposisi atau kata depan dalam bahasa Indonesia digunakan untuk menunjukkan relasi semantis di antara konstituen depan dan konstituen belakang. Dari sudut pandang sintaktis, preposisi hadir di depan adjektiva, adverbial, atau nomina. Kemudian, kelompok kata ini membentuk frasa preposisional. Meskipun demikian, menurut Sneddon et al (2010: 191), frasa preposisional terdiri atas preposisi dan FN. Preposisi menghubungkan FN yang mengikuti komponen lain dalam sebuah kalimat. Pada tataran klausa, preposisi menunjukkan hubungan frasa nomina yang mengikuti dan predikatnya. Berdasarkan sudut pandang ini, dalam makalah ini, kami mendefinisikan konstruksi preposisional sebagai struktur yang di dalamnya sebuah preposisi muncul bersama dengan konstituennya. Konstruksi ini mungkin berupa verba/frasa verbal, nomina/frasa nomina, atau adjektiva/frasa adjektiva.

Dalam bahasa Indonesia, dari sudut pandang bentuk preposisional, terdapat dua jenis preposisi: preposisi dasar atau sederhana dan preposisi gabungan. Menurut Neddon (2010), di antaranya yang paling penting adalah preposisi yang mengindikasikan posisi dan arah, yakni preposisi lokatif: *di*, *ke*, dan *dari*. Kombinasi ketiga preposisi ini dengan tiga pronomina lokatif *sini*, *sana*, dan *situ* menghasilkan sembilan cara membedakan lokasi dan arah (lihat Moeljadi 2010, Suhardi & Suyiata 2013, Ramlan 1987, Hakim 1995).

Dalam kaitannya dengan penggunaan pendekatan korpus dalam penyusunan kamus konstruksi preposisional ini, terdapat konsep yang disebut dengan koligasi. Kolokasi—sebuah konsep yang mirip—lebih banyak dikenal daripada koligasi. Pada dasarnya, koligasi adalah hubungan di antara satuan leksikal dan struktur gramatikal yang mencakup kecenderungan sebuah satuan leksikal muncul bersama dengan kosakata lain. Jika kolokasi menyangkut bentuk leksikal, koligasi berfokus pada relasi antara bentuk leksikal dan pola gramatikal.

Menurut Lehecka (2015), istilah koligasi telah digunakan dengan banyak pengertian berbeda, bahkan lebih banyak variasinya daripada kolokasi. Salah satunya adalah Firth (1968: 181) yang menggunakan istilah ini untuk merujuk pada daya tarik-menarik sintagmatis di antara kategori gramatikal, misalnya kelas kata atau fungsi sintaktis. Meskipun demikian, penggunaan koligasi yang lebih lazim sekarang ditujukan pada daya tarik-menarik di antara satuan leksikal dan kategori gramatikal (Sinclair 1998: 15). Lebih lanjut, koligasi tidak saja diterapkan pada kata tunggal, misalnya *menduga*, melainkan juga pada frasa yang merupakan konstruksi dengan kata lebih dari satu, misalnya *tanggung jawab*.

Kajian koligasi dan kolokasi merupakan bagian dari pendekatan terhadap bahasa yang memandang informasi sintaktis terkandung di dalam unsur kosakata pada leksikon mental. Dengan pendekatan korpus, khususnya yang berfokus pada upaya menghasilkan data pemakaian lengkap dengan informasi frekuensinya. Dengan demikian, pola mendasar pada preposisi tentu saja lebih mudah diidentifikasi. Korpus memungkinkan kita mengenali cara koligasi yang berbeda membentuk makna yang berbeda pula pada kata yang nyaris bersinonim. Selain itu, dengan pendekatan korpus, pola koligasi yang teridentifikasi sangat penting untuk mengonsepskan relasi di antara leksikon dan sintaksis.

II. METODE PENYUSUNAN KAMUS

Dalam penyusunan kamus konstruksi preposisional ini, digunakan dua macam korpus. Korpus pertama adalah korpus yang disusun oleh tim kamus ini dengan jumlah 18 juta kata. Korpus ini terdiri atas koleksi teks fiksi dan nonfiksi dengan rasio 30:70. Teks nonfiksi terdiri atas artikel jurnalistik, karya ilmiah, buku pelajaran, buku pengetahuan umum, dan sebagainya. Sementara itu, korpus kedua adalah IndonesiaWac yang tersedia di Sketch Engine (Jakubíček, et al., 2010; Kilgarriff et al. 2014). Korpus ini terdiri atas 109 juta kata yang disusun dari berbagai teks yang pada umumnya tersedia di internet. Korpus IndonesiaWac hanya digunakan sebagai pembandingan atau sebagai alternatif apabila pada korpus utama tidak dijumpai contoh pemakaian.

Pada bagian sebelumnya disebutkan bahwa tujuan kamus kami adalah untuk merepresentasikan pengetahuan penutur jati tentang struktur preposisional pada bahasa mereka. Perhatikan contoh berikut ini.

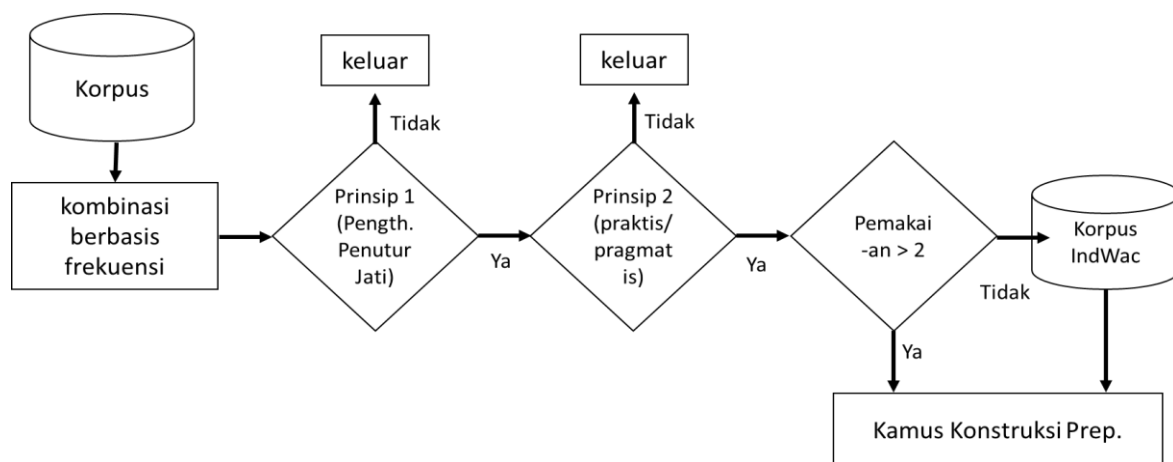
- 1) Kami makan nasi kapau di restoran itu.

Jelas bahwa kemungkinan untuk mengombinasikan *makan* dan *di* + frasa nomina (FN) tidak terkait dengan pengetahuan leksikal penutur asli. Untuk memastikan apakah hal tersebut sebuah ungkapan bahasa Indonesia yang valid, penutur perlu tahu bahwa kegiatan seperti *makan nasi kapau* membutuhkan **tempat**. Hal yang sama juga berlaku untuk **waktu** (*Kami makan nasi kapau sepanjang acara itu*), **peserta** (*Kami makan nasi kapau dengan teman-teman*), **instrumen** (*Kami makan nasi kapau dengan tangan*), **tujuan** (*Kami makan nasi kapau untuk menggantal perut*), **cara** (*Kami makan nasi uduk dengan nikmat*), dan sebagainya. Semua kombinasi tersebut tidak tersedia dalam kamus internal kita, leksikon kita. Semua kombinasi tersebut tidak terkait dengan bahasa, tetapi dengan dunia di luar bahasa.

Sebaliknya, jika kita adalah penutur asli bahasa Indonesia, sudah menjadi pengetahuan leksikal kita bahwa kata *berbakti* berkombinasi dengan preposisi *pada* atau *kepada* (mungkin juga dengan preposisi yang lain). Tidak cukup dengan itu, kita pun tahu arti kombinasi tersebut. Jika sebuah kata memiliki preposisi yang pasti, makna dari frasa preposisional sangat bergantung padanya. Selain itu, seorang penutur asli bahasa Indonesia juga tahu bagaimana struktur ungkapan yang mengikuti setelah kata *berbakti*, misalnya kita dapat mengatakan *berbakti pada bangsa dan negara, kepada orang tua*, dan sebagainya. Namun, kita tidak bisa mengatakan, misalnya, *berbakti pada membahagiakan kedua orang tua kita*. Informasi semacam ini tentu saja penting untuk dimuat di dalam kamus dan kamus umum seperti KBBI tidak memuatnya.

Proses penentuan entri kamus dilakukan berdasarkan korpus dengan menggunakan analisis koligasi sebagaimana yang dilakukan oleh Tognini-Bonelli (2001: 96–100). Dengan analisis korpus, Tognini-Bonelli menunjukkan bahwa bentuk infleksional verba bahasa Italia SAPERE ‘mengetahui’ muncul dalam konstruksi gramatikal yang berbeda. Menariknya, dalam bahasa Italia terdapat dua alternatif bentuk infinitif verba *saper* dan *sapere*. *Saper* secara signifikan lebih sering digunakan dalam konteks dengan infinitif lain, sementara *sapere* digunakan lebih banyak dalam klausa subordinatif.

Jadi, dalam penyusunan kamus ini, penentuan konstruksi preposisional yang patut menjadi entri kamus mengikuti prinsip berikut ini. Prinsip pertama yang paling mendasar adalah jika kombinasi merupakan bagian dari pengetahuan penutur jati bahasa Indonesia, kombinasi tersebut patut menjadi entri kamus. Misalnya, pada *minum jus di teras, di teras* bukan merupakan pengetahuan penutur jati bahasa Indonesia. Sementara itu, pada *minum jus dengan gelas* terdapat informasi yang menjadi bagian dari pengetahuan penutur jati bahasa Indonesia. Hal itu terjadi karena pada bahasa Inggris yang digunakan adalah *drink juice from a glass* atau *out of a glass*, sedangkan pada bahasa Perancis digunakan *boire du jus dans un verre*. Kemudian, prinsip yang kedua adalah pertimbangan praktis dan pragmatis, misalnya apakah orang ingin mengetahui hal tersebut pada sebuah kamus? Apakah hal tersebut membantu orang asing yang sedang belajar bahasa Indonesia? Jika berbicara soal penerjemahan, apakah ada kemungkinan timbul ketidaksepakatan tentang konstruksi preposisional tertentu? Keseluruhan prosedur penentuan entri kamus dapat dilihat pada Gambar 1 berikut ini.



Gambar 1 Garis Besar Prosedur Penyusunan Kamus

III. KAMUS KONSTRUKSI PREPOSISIONAL

Pada bagian sebelumnya telah dijelaskan pentingnya menyediakan informasi penggunaan preposisional pada sebuah kamus karena konstruksi preposisional merupakan permasalahan yang terkait dengan bahasa tertentu. Hal inilah yang menjadi landasan penyusunan kamus konstruksi yang memuat kombinasi kata dengan proposisinya yang tetap. Terdapat fakta yang nyata bahwa perilaku preposisi bahasa Indonesia tidak terlalu banyak perbedaannya dengan perilaku preposisi bahasa Indogermania seperti bahasa Inggris.

Hal yang perlu dibahas selanjutnya pada bagian ini adalah bagaimana bentuk kamus konstruksi preposisional ini? Dalam proyek ini, kami bermaksud untuk mengikuti tradisi leksikografi Indonesia, yaitu entri kamus tidak didasarkan pada kata melainkan pada akar kata. Dalam satu entri yang sama pada kamus, akan dijumpai verba, nomina, atau adjektiva sepanjang semua kategori kata tersebut berkombinasi dengan preposisi yang sama.

Selain itu, entri pada kamus tidak akan didefinisikan sebagaimana kamus biasanya. Namun, hampir seluruh informasi akan muncul dalam bentuk contoh. Jika sebuah kombinasi dapat diikuti baik oleh sebuah nomina atau klausa, hal tersebut tidak akan disampaikan secara langsung melainkan melalui contoh. Juga ketika fleksibilitas penggunaan sebuah kombinasi menurut gaya bahasanya, hal tersebut tidak akan disebutkan melainkan hanya melalui contoh. Informasi lain yang akan disertakan antara lain adalah jumlah frekuensi penggunaan dan catatan tambahan pada kasus-kasus tertentu.

Jadi, dari segi format, paling tidak ada tiga hal yang perlu dijelaskan di sini. Pertama, tiap entri diawali oleh akar kata; kamus kami berbasis akar sebagaimana kamus bahasa Indonesia yang lain. Pada tiap contoh, tiap posisi kata yang berhubungan dengan akar dicetak miring. Dalam kamus, cetak miring hanya digunakan untuk tujuan ini sehingga informasi lain yang biasa ditulis dengan huruf kursif diganti dengan lambang « ».

Kedua, preposisi yang berkombinasi dengan akar kata ini muncul dalam urutan alfabetis. Tidak dibedakan antara preposisi dan kata lain yang berperilaku sebagai sebuah preposisi, misalnya kata *menuju*. Jika relevan, dalam kamus juga dimasukkan penggunaan non-preposisional sebuah kata yang menggunakan preposisi.

Kemudian, yang ketiga, tiap kombinasi sebuah akar dan preposisinya akan ditampilkan dalam sebuah contoh atau lebih, bergantung dari banyak tidaknya variasi pemakaian kombinasi tersebut. Semua contoh yang digunakan diambil dari teks riil dalam korpus yang kami susun, namun tidak disebutkan sumbernya. Tiap contoh dimulai pada garis baru dan didahului oleh tanda butir (*bullet operator*).

Keempat, tiap entri dapat saja dilengkapi dengan komentar, yang mungkin saja bersifat lokal (memberikan informasi tentang satu contoh pemakaian preposisi) atau bersifat umum (memberikan informasi tentang entri secara menyeluruh). Selain itu, pada kamus juga digunakan rujuk silang untuk entri tertentu, misalnya:

pede
→ percaya (diri)

Pada gambar berikut ini terdapat contoh tata letak kamus. Pada Gambar 2, semua entri terkait dengan bentuk dasar *dalih*, *dalil*, *dakwa*, *damba*, dan *dampak*.

dalih

dengan

· Awalnya Mahmudin masih bisa *berdalih* dengan alasan ini dan itu.

untuk

· Dengan *dalih* untuk mengamankan pasokan pangan dalam negeri pemerintah melakukan impor. Membanjirnya produk luar negeri.

tanpa preposisi

· Maka, begitu pemuda itu diterima ayah ibunya duduk di ruangan tamu, ia lalu mengundurkan diri dengan *dalih* membantu pelayanan di belakang.

dalil

untuk

· Mengapa banyak sekali orang yang mengambil ayat ini sebagai *dalil* untuk berpoligami?

damai

antara

· Dengan demikian, ke depan, ada *damai* di antara sesama anak bangsa.

· Hakim bersifat aktif dalam mengupayakan adanya *perdamaian* antara pihak-pihak yang berperkar.

dengan

· Dan untuk kesekian kalinya Belanda mengajak *damai* dengan Sultan.

· Rakyat Merdeka dituduh melakukan

· Jika setiap warga di negeri ini memiliki *dambaan* untuk memiliki lingkungan hidup yang sehat dan nyaman, saya kira akan terwujud.

· Sejak kecil ia *mendambakan* untuk menjadi musisi.

tanpa preposisi

· Dia *mendambakan* seorang wanita yang tidak terlalu serius dan bisa menghiburnya dengan segala tingkahnya manjanya.

· Umumnya manusia tidak menyukai kegelisahan dan *mendambakan* kebahagiaan.

· Kaum muda yang selalu *mendambakan* terjadinya perubahan atas kondisi sosial yang ada.

dampak

atas

· Bagaimana *dampak* atas penerapan kebijakan tax holiday di Indonesia dilihat dari perspektif makro?

· Pemikiran seperti ini tidak hanya mempengaruhi kita, tetapi *berdampak* atas cara berpikir semua manusia yang terdidik di seluruh dunia.

bagi

· Saya ingin mengetahui *dampak* penjajahan bangsa asing bagi kehidupan bangsa Indonesia.

buat

· Jadi, kondisi ayah dan ibu sedikit nggak-nya akan memberi *dampak* psikologis buat

Gambar 2 Tampilan dan Tata Letak Kamus

Sebagaimana dijelaskan sebelum dan tampak pada Gambar 2, semua informasi dalam sebuah entri akan diberikan dalam bentuk contoh-contoh. Jadi, jika dinyatakan bahwa A berkombinasi dengan B, akan langsung dijumpai sebuah kalimat yang diambil dari teks yang digunakan secara nyata. Dengan kata lain, pengguna tidak akan dikonfrontasikan dengan apa yang dipikirkan oleh penyusun kamus atau leksikograf, tetapi dengan apa yang ditulis oleh orang biasa, misalnya bukan ahli bahasa, dan pengguna dapat mengonfrontasikan intuisi mereka dengan apa yang dibaca pada contoh kamus. Dalam kamus, juga sering ditampilkan lebih dari satu contoh dengan tujuan untuk memperlihatkan bahwa kombinasi tertentu muncul dengan berbagai tingkatan ragam (*style*) misalnya formal, tidak terlalu formal, sangat formal, dan sebagainya atau bahwa kombinasi dapat digunakan untuk benda dan makhluk hidup sekaligus. Atau bisa juga frasa nomina di sebelah kanan preposisi dapat digantikan oleh klausa atau kalimat. Apa yang tidak dapat diungkapkan melalui contoh membutuhkan cara penyampaian yang lain. Untuk itu, akan ditambahkan label atau lambang untuk mengatakan bahwa sesuatu hanya digunakan dalam bidang tertentu, misalnya olahraga, teknologi komputer, dan sebagainya, atau bahwa sebuah kombinasi tertentu lebih sedikit dari segi frekuensi daripada kombinasi lain. Obyek preposisional sering bersaing dengan obyek langsung, seperti dalam *belajar (tentang) matematika*. Hal ini juga akan dinyatakan dalam kamus, tentu saja dengan contoh. Namun, jika dibutuhkan akan diperjelas dengan tambahan komentar sebagaimana yang tampak pada contoh dalam Gambar 3.

pada

- Sebenarnya Dian tak terlalu *paham* pada kalimat itu.
- Maksud saya, saya *paham* pada apa yang dilakukan Arjuna.
- Saya puas mendengar komentar mereka. Komentar-komentar ini mencerminkan tingkat *pemahaman* mereka pada materi pelatihan.
- Hal ini dilakukan untuk menghindari *salah paham* pada produk yang dijual.
- * Kombinasi *pemahaman kepada* + nama seseorang pada umumnya ditemukan setelah kata kerja seperti *memberikan*.

sebagai

- Tingginya kasus kekerasan terhadap perempuan dan anak sebenarnya sudah sangat lama berlangsung, namun sebagian besar masyarakat belum *memahaminya* sebagai bentuk pelanggaran hak Asasi Manusia (HAM).

Gambar 3 Contoh Penempatan Komentar pada Kamus untuk entri *Paham* dan kombinasi *paham* + *pada/kepada*

Kamus konstruksi preposisi ini berupaya untuk deskriptif, tidak preskriptif, tetapi kadang-kadang ditambahkan informasi bahwa pemakaian tertentu dihindari oleh sebagian penutur, yang menyiratkan bahwa semuanya bergantung pada pengguna, apakah akan menggunakannya atau tidak. Pemberian informasi tersebut tidak muncul terlalu sering meskipun seperti pada banyak bahasa lainnya, terdapat banyak perbedaan di antara data bahasa yang tersedia dan opini tentang penggunaan bahasa.

Hingga saat penulisan artikel ini, penyusunan entri kamus nyaris selesai. Telah terkumpul kurang lebih 7500 entri kamus. Meskipun tentu saja masih banyak kombinasi yang mungkin terlepas dari pengamatan kami terbatasnya korpus yang kami susun dan gunakan—korpus kamus berjumlah 18 juta kata dan korpus IndonesianWac sebesar 109 juta kata, kombinasi yang kami masukkan sebagai entri kamus cukup representatif sebagai kombinasi koligasional yang signifikan dalam bahasa Indonesia.

Sasaran pengguna kamus konstruksi ini adalah mereka yang membutuhkan informasi pelengkap dari kamus umum yang kita miliki. Mereka bisa merupakan orang asing yang sedang belajar tentang bahasa Indonesia, atau dapat juga mereka yang ingin membandingkan intuisi mereka dengan contoh relevan yang diambil dari teks riil.

IV. KESIMPULAN

Telah ditampilkan dalam tulisan ini bahwa sebuah kamus tidak sekadar daftar kosakata yang disusun dengan definisi logis dan relasi semantisnya. Namun, sebuah kamus dapat mencerminkan intuisi penutur jati sebuah bahasa terkait dengan struktur tertentu. Untuk contoh ini, dalam upaya ini, entri kamus konstruksi ini tidak dilengkapi dengan sebuah definisi tentang apa makna kata yang dimaksud, sebagaimana yang biasa diberikan oleh sebuah kamus. Oleh sebab itu, kita lebih sering perlu untuk merujuk pada makna kata: dalam sebuah entri mungkin perlu disampaikan kepada pengguna bahwa kombinasi tertentu hanya

muncul dengan makna itu atau itu. Tentu saja, hal tersebut disampaikan dalam bahasa yang sederhana, mudah dipahami oleh orang kebanyakan.

Selain itu, telah disampaikan pula bahwa dalam penyusunan kamus ini, digunakan pendekatan korpus dalam dua tahap. Pertama adalah tahap penentuan entri. Pada tahap ini, digunakan analisis koligasi untuk menentukan kombinasi kata leksikal dan konstruksi gramatikalnya. Setelah itu, pada tahap pengambilan contoh pemakaian, korpus kembali digunakan. Korpus sangat reliabel untuk mendapatkan contoh-contoh pemakaian yang relevan karena korpus disusun dengan menggunakan teks riil yang terpilih dan disusun secara sistematis. Selain itu, metode korpus—khususnya konkordansi—juga digunakan untuk menampilkan informasi mengenai perbedaan pemakaian menurut ragam bahasa atau jenis konstituen penyusunnya.

Daftar Acuan

- Atkins, B.T.S. & Rundell, M. (2008). *The Oxford guide for practical lexicography*. Oxford: Oxford University Press.
- Bergenholtz, H., & Gouws, R. H. (2013). What is lexicography? *Lexikos* 22, 31-42.
- Collins COBUILD Advanced dictionary (2009). Boston MA: Heinle Cengage Learning
- Comrie, B. (1989). *Language universals and linguistic typology*. 2nd edition. Chicago: University of Chicago.
- Effendi, S. & Aritonang, B. (1993). *Preposisi dan frase berpreposisi*. Departemen Pendidikan dan Kebudayaan (Indonesia). Pusat Pembinaan dan Pengembangan Bahasa.
- Hagege, C. (2010). *Adpositions*. Oxford: Oxford University Press.
- Hakim, L. Tasai, A. (1995). Preposisi bermakna ganda. *Pembinaan Bahasa Indonesia* (15 nov.1995), 1995.
- Jakubiček, M., Kilgarriff, A., McCarthy, D., Rychlý, P., Fast Syntactic Searching in Very Large Corpora for Many Languages. *PACLIC*: 741-47, 2010.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M. Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. The Sketch Engine: ten years on. *Lexicography*, 1: 7-36, 2014.
- Moeljadi, D. (2010). Frase preposisi bahasa Indonesia : analisis X-bar. *Kajian Sastra* Vol. 34, No. 1 (Jan. 2010), p. 1-12
- Quinn, G. (2001). *The learner's dictionary of today's Indonesian*. Crows Nest NSW: Allen & Unwin.
- Ramlan, M. (1987) *Kata depan atau preposisi dalam bahasa Indonesia*. Yogyakarta: CV Karyono.
- Slager, M. dan Suhardijanto, T. (2016). A Dictionary of prepositional constructions in Indonesian. *Proceeding of Asialex Internasional Conference*, Juni 1-3 2016, Manila The Philippines.
- Sneddon, J.N., Adelaar, A., Djenar, D.N. & Ewing. M.C. (2010). *Indonesian reference grammar*. 2nd edition. Crows Nest NSW: Allen & Unwin.
- Suhardi & Suyiata, P. (2013). *Pelengkap verba ± preposisi dalam bahasa Indonesia*. Yogyakarta: UNY Press.
- Umiyati, M., Artawa, K., Yadnya, I.B.P. & Suparwa, N. (2014). The valency change strategy of adjectives in Indonesian. *e-Journal of Linguistics*, 01 August 2014, Vol.8(1).

**KERAGAMAN KOSAKATA DALAM KORPUS SASTRA:
SEBUAH KONTRIBUSI UNTUK PENGEMBANGAN LEKSIKOGRAFI
INDONESIA**
*(LEXICAL DIVERSITY IN LITERATURE CORPUS:
A CONTRIBUTION FOR DEVELOPMENT OF INDONESIAN LEXICOGRAPHY)*

Neneng Nurjanah
PBSI UIN Syarif Hidayatullah Jakarta
neneng.nurjanah@uinjkt.ac.id

Abstrak

Makalah ini menyelidik keragaman kosakata dalam korpus sastra, yaitu korpus Nh. Dini. Korpus ini terdiri atas enam novel karya Nh. Dini, yaitu *Hati yang Damai*, *Pada Sebuah Kapal*, *Namaku Hiroko*, *La Barka*, *Keberangkatan*, dan *Jalan Bandungan*, dengan jumlah token sebanyak 278,585. Penyelidikan ini berfokus pada *hapax legomenon* dalam korpus Nh. Dini karena berdasarkan kajian pustaka *hapax legomenon* menunjukkan sidik jari pengarang atau personal idiosinkrasi berkaitan erat dengan pembentukan kata-kata baru, penggunaan bahasa daerah atau bahasa asing yang dimanfaatkan pengarang untuk membangun narasi. Dengan memanfaatkan metode linguistik korpus, penelitian ini menghasilkan beberapa temuan, yaitu terdapat empat kata dasar dan 88 kata turunan yang belum terekam dalam KBBI Edisi Kelima versi daring. Temuan lain adalah pengkajian terhadap *hapax legomenon* memungkinkan pekamus untuk mencari kata-kata baru; makna baru; penggunaan yang menarik dari kata-kata yang sudah terekam dalam kamus; menjadi tinjauan bagi pekamus untuk menyerap kata dari bahasa daerah dan bahasa asing; menjadi sumber data dalam mengumpulkan informasi leksikografis.

Kata Kunci: korpus Nh. Dini; idiosinkrasi; *hapax legomenon*

Abstract

*This paper discusses lexical diversity in the literary corpus that is Nh. Dini Corpus. This corpus consists of six Nh. Dini's novels, there are *Hati yang Damai*, *Pada Sebuah Kapal*, *Namaku Hiroko*, *La Barka*, *Keberangkatan*, and *Jalan Bandungan* and amounted to 278,585 tokens. This investigation focuses on the *hapax legomenon* in the Nh. Dini's Corpus because based on literature review, *hapax legomenon* indicates the author's fingerprint or personal idiosyncrasy, and closely related neologism, the usage of regional or foreign languages which the author uses in that novels. Employed the linguistic corpus method, this study yielded several result, there are four base and 88 derivative words that have not been recorded in the KBBI*

Fifth Edition online version. Another finding is study of the hapax legomenon allowing lexicographer to search for new words; new meaning; interesting use of words that already recorded in the dictionary; becomes a review for lexicographer to loan words from regional and foreign languages; become the source of data in collecting lexicographical information.

Keyword: *Nh. Dini Corpus; idiosyncrasy; hapax legomenon*

I. PENDAHULUAN

Sebelum teknologi komputasi bahasa berkembang pesat, penyusunan kamus dilakukan secara sederhana. Seorang pekamus atau editor secara manual mengumpulkan kata-kata, penggunaan kata-kata tersebut dalam bentuk kalimat, serta rujukannya. Salah seorang ahli bahasa kenamaan Denmark, Prof. Otto Jenpersen mengumpulkan ratusan ribu kertas slip dalam kotak sepatu yang kemudian diletakkan di sebuah villa di luar daerah Kopenhagen. Di tiap secarik kertas slip, ahli gramatika Inggris itu menuliskan kalimat-kalimat menarik dari berbagai karya sastra Inggris (Lindquist, 2009:2). Bahkan, editor OED (Oxford English Dictionary) memiliki ribuan koresponden yang bertugas untuk mengumpulkan dan mengirimkan kertas-kertas slip yang berisi kata-kata baru yang menarik, serta penggunaan kata-kata yang lama (Lindquist, 2009:51). Upaya-upaya tersebut dilakukan dengan telaten untuk menghasilkan informasi leksikografi yang lengkap sehingga bermanfaat dalam penyusunan leksikografi.

Kini, berkat kemajuan teknologi pengolahan bahasa, penyusunan leksikografi dilakukan dengan memanfaatkan komputasi bahasa. Lindquist mencontohkan kamus Cobuild yang mengkompilasikan data bahasa dalam bentuk korpora (kumpulan teks elektronik dalam jumlah besar) untuk tujuan leksikografis (2009:52). Berdasarkan informasi yang tercantum di situs www.collins.co.uk, Cobuild Dictionary disusun berdasarkan Cobuild Corpus yang berjumlah lebih dari 4.5 miliar token. Korpus ini terdiri atas ragam bahasa tulis yang diambil dari majalah dan buku yang diterbitkan di seluruh dunia, bahasa lisan dari radio, TV dan percakapan sehari-hari. Data tersebut diperbaharui setiap bulan, sehingga memudahkan editor untuk mengidentifikasi kata dan makna baru dari kata-kata yang pertama kali ditemukan.ⁱ Tidak hanya Cobuild, penerbit besar kamus di Inggris, pada umumnya telah memiliki korpora (Lindquist, 2009:52).

Terlepas dari itu, upaya memanfaatkan korpus untuk tujuan leksikografis sudah dilakukan di Indonesia, salah satunya penelitian leksikografi yang dilakukan Dora Amalia untuk disertasinya pada tahun (2014). Dengan memanfaatkan korpus buku teks Bipa, penelitian ini mencoba merumuskan bentuk definisi untuk verba dalam kamus pemelajar bahasa Indonesia dengan memanfaatkan korpus Bipa. Meski demikian, berdasarkan penelusuran pustaka, penelitian leksikografi yang memanfaatkan korpus sastra belum ditemukan. Sampai sejauh ini penelitian yang memanfaatkan korpus sastra ditujukan untuk mengetahui representasi isu yang terkandung dalam korpus sastra (Nurjanah, 2015).

Untuk itu, makalah ini mencoba mengkaji korpus sastra untuk melihat keragaman kosakata yang terkandung di dalamnya. Hal ini berangkat dari pandangan bahwa setiap pengarang memiliki gaya khas. Dalam hal ini beberapa peneliti menyebutnya sebagai idiosinkrasi personal atau sidik jari pengarang. Personal idiosinkrasi ini dapat terlihat dari kata-kata yang muncul sekali dalam sebuah korpus. Kata-kata ini kemudian dikenal sebagai *hapax legomenon*. Dengan mengkaji *hapax legomenon* diharapkan terdapat keragaman kosakata yang menjadi ciri khas penulis dan bermanfaat untuk pengembangan leksikografi Indonesia.

Penyelidikan korpus sastra untuk tujuan leksikografis dimungkinkan terlebih pada karya sastra yang berkualitas, seperti novel-novel Nh. Dini. Sebagai seorang novelis, Nh. Dini dikenal publik dan karya-karyanya menyita perhatian kritikus sastra. Tinneke Helwig (2003:203) menilai Nh. Dini sebagai perempuan pertama yang memperoleh pengakuan umum dan penghargaan atas karya-karyanya. Sederet penghargaan pernah diterimanya, di antaranya meraih juara pertama penulisan cerpen dalam bahasa Prancis se-Indonesia yang diadakan surat kabar *Le Monde* pada tahun 1988; mendapat penghargaan Bhakti Upapradana dari pemerintah daerah Jawa Tengah; meraih penghargaan Sea Write Award Literary di Bangkok, Thailand pada tahun 2003; dan menerima anugerah Ahmad Bakrie di bidang sastra pada tahun 2011.

Dengan memanfaatkan korpus Nh. Dini, penelitian ini mencoba menganalisis keragaman kosakata yang menjadi idiosinkrasi personal pengarang. Penelitian ini berangkat dari pandangan bahwa seorang pengarang akan memanfaatkan kosakata yang dimilikinya dan memungkinkan untuk membentuk kata-kata baru sebagai pengejawantahan atas *licentia poetica* yang dimilikinya. Di satu pihak, keragaman kosakata tersebut dimanfaatkan pengarang untuk membangun narasi dan memperkuat daya ungkap. Di pihak lain, keragaman kosakata diharapkan memberikan kontribusi pada pengembangan leksikografi Indonesia.

II. KERANGKA TEORETIS DAN METODE PENELITIAN

Dalam penyelidikan korpus linguistik untuk tujuan leksikografi, pada umumnya dimulai dengan menyelidik daftar frekuensi kata (*frequency list*). McEnery dan Hardie (2012:243) mendefinisikan daftar frekuensi kata sebagai daftar semua satuan dari tiap tipe dalam sebuah korpus bersama dengan hitungan berapa banyak tiap satuan itu muncul dalam korpus. Cheng menambahkan bahwa fitur ini merupakan fitur dasar yang dimiliki setiap peranti lunak pengolah korpus. Dengan memanfaatkan peranti lunak, daftar frekuensi kata dapat dengan mudah digunakan dan hasilnya selalu menarik (2012:62).

Terkait daftar frekuensi kata, Cheng (2012:66) menjelaskan bahwa dalam korpus general seperti BNC (British National Corpus), dan dua korpus khusus, yaitu HKFSC (Hong Kong Finance Service Corpus) dan HKEC (Hong Kong Engineering Corpus), kata-kata yang memiliki frekuensi paling besar adalah kata-kata fungsi atau kata-kata gramatikal yang memiliki fungsi penting dalam bahasa Inggris, seperti *the*,

of, and, to, in, dan a. Pendapat ini senada dengan temuan Ahmad (dalam Cheng, 2012:67) bahwa 50 kata berfrekuensi paling besar dalam korpus BNC adalah kata-kata gramatikal. Namun, pada HKFSC dan HKEC, kata leksikal mulai muncul pada 20 kata berfrekuensi paling besar. Dari tiga korpus ini terlihat bahwa kemunculan kata-kata leksikal lebih dimungkinkan pada korpus-korpus khusus yang memiliki karakteristik teks yang jelas dibandingkan korpus umum yang mencakup semua penggunaan dalam satu bahasa.

Jika pada kata-kata yang berfrekuensi tinggi, kata leksikal jarang ditemui, sebaliknya pada kata-kata yang berfrekuensi rendah, terutama pada kata-kata yang muncul satu kali dalam satu korpus. Coultard dan Johnson (dalam Cheng, 2012:69) jumlah tipe berfrekuensi satu kali berjumlah hampir 50% dari jumlah total tipe dalam korpus. Kata-kata yang memiliki frekuensi satu dinamai *hapax legomenon* dan hampir selalu yang muncul adalah kata-kata leksikal.

Terminologi *hapax legomenon* terambil dari bahasa Latin yang secara bahasa berarti kata yang muncul sekali.ⁱⁱ Waltke dan O' Connor menyebutkan bahwa *hapax legomenon* adalah kata, bentukan, atau kombinasi kata yang ditemukan dalam isi tulisan dari sebuah karya sastra (1990:691). Sementara itu, menurut Bauer (2004:51) *hapax legomenon* merupakan kata yang muncul satu kali dalam sebuah teks atau korpus. Yang dimaksud sebagai korpus dalam hal ini bisa berupa arsip teks dan bisa berupa dalam keseluruhan penggunaan bahasa seperti Latin Klasik. Crystal menyebut bahwa *hapax legomenon* adalah studi khusus yang memberikan perhatian tertentu atas kata-kata tersebut yang muncul satu kali dalam teks, dalam karya-karya seorang pengarang atau dalam satu bahasa (2003:67). Terma ini dipakai di pelbagai bidang, di antaranya filologi klasik, studi Bible, namun akhir-akhir ini digunakan dalam studi korpus linguistik.

Beberapa penelitian yang mengkhususkan pada *hapax legomenon* memperlihatkan bahwa terma itu mampu merefleksikan latar belakang, pengalaman, dan menyampaikan makna yang tersirat (Oakes,1998:201). Dalam kajian Coultard dan Johnson tahun 2007 (dalam Cheng, 2012:69) memanfaatkan, salah satunya, *hapax legomenon*, untuk mengidentifikasi idiolek penulis. Dalam kajian sastra idiolek penulis dikenal juga dengan gaya bahasa penulis. Middleton Murry menyebutkan ...*style means that personal idiosyncrasy of expression by which we recognize a writer*. Dari pandangan Murry, dapat diketahui bahwa gaya bahasa berarti ekspresi idiosinkrasi personal (ekspresi pribadi yang unik) yang kita kenali dari seorang pengarang. Murry pun melanjutkan pandangannya bahwa idiosinkrasi personal dapat ditunjukkan dengan penggunaan kata-kata dan frasa yang tidak umum dalam sebuah karya (Prasad, 2004:203). Jika dikaitkan pandangan sebelumnya, kata-kata yang tidak umum muncul pada kata-kata berfrekuensi rendah.

Secara morfologis, *hapax legomenon* berkaitan erat juga dengan neologismeⁱⁱⁱ (Domazakis, 2010:11). Pandangan ini senada dengan pandangan Baayen dan Renouf (Domazakis, 2010:11) yang mengatakan bahwa dalam kata-kata berfrekuensi rendah, kemunculan neologisme dapat dikatakan besar. Jika kedua pandangan itu dikaitkan dengan penelitian ini maka terdapat kemungkinan besar kata-kata baru yang akan muncul dalam *hapax legomenon* dari sebuah korpus. Terlebih korpus sastra yang

memungkinkan pengarang—dengan memanfaatkan *licentia poetica*—membentuk kata-kata baru, atau menggunakan kata yang sudah ada dengan makna baru.

Dari uraian di atas dapat disimpulkan beberapa hal. Pertama, untuk melihat keragaman kosakata, penyelidikan atas daftar kata bisa berfokus pada *hapax legomenon* karena kata-kata yang berfrekuensi tinggi cenderung diisi oleh kata fungsi atau kata gramatikal. Kedua, salah satu manfaat mengkaji *hapax legomenon* ialah dapat teridentifikasinya gaya bahasa penulis, idiolek penulis, atau idiosinkarasi penulis. Ketiga, *hapax legomenon* terkait erat dengan neologisme, hal ini memungkinkan keragaman kosakata yang muncul bisa berupa kata-kata baru yang belum terekam dalam kamus.

Selain fitur daftar frekuensi kata, salah satu fitur yang digunakan dalam penyusunan leksikografi adalah konkordansi. McEnery dan Hardie mendefinisikan konkordansi sebagai tampilan setiap contoh kata tertentu atau istilah tertentu dalam sebuah korpus, yang disertai hit atau sejumlah konteks kata atau istilah tertentu yang mendahului dan mengikutinya (2012:241). Konkordansi umumnya digunakan untuk melihat penggunaan kata dalam kalimat dan melihat kecenderungan makna sebuah kata dari pelbagai contoh penggunaan.

Editor kamus dapat memanfaatkan fitur konkordansi untuk memperoleh makna secara khusus dan mengambil susunan tiap-tiap kata seraya memilah mana yang pantas untuk ditampilkan dalam kamus. Banyak kamus mengutip kalimat-kalimat dalam korpora secara verbatim. Ada pula yang mereka-reka untuk mendapatkan contoh baru (Lindquist, 2009:52). Hal tersebut dimungkinkan karena fitur konkordansi menampilkan tipikal kata-kata yang berada sebelumnya (sebelah kiri noda) dan kata-kata yang berada setelahnya (sebelah kanan noda).

Menurut Baker (dalam Litosseliti, 2010) penelitian yang memanfaatkan linguistik korpus umumnya menggunakan pendekatan campuran (*mix method*) atau kualitatif dan kuantitatif. Hal ini disebabkan penyelidikan terhadap korpus linguistik memanfaatkan perhitungan statistik, salah satunya frekuensi (McEnery dan Hardie, 2012:49). Sementara itu, Lindquist (2009:25) menyarankan penyelidikan ini tidak boleh berhenti pada perhitungan statistik, tapi dilanjutkan dengan interpretasi dari peneliti yang dilakukan berdasarkan pada data kuantitatif.

Sumber data yang digunakan dalam penelitian ini adalah korpus Nh. Dini. Korpus ini merupakan kumpulan teks elektronik dari enam novel karya Nh. Dini, yaitu *Hati yang Damai* (1961), *Pada Sebuah Kapal* (1973), *Namaku Hiroko* (1974), *La Barka* (1975), *Keberangkatan* (1975), dan *Jalan Bandungan* (1989). Jumlah token dalam korpus ini adalah 278,585 dengan jumlah tipe 13.506.

Peranti lunak yang digunakan dalam penelitian ini adalah AntConc versi 3.4.4w 2014. Peranti lunak dirancang oleh Laurence Anthony, seorang profesor dari Waseda University. Peranti lunak ini bisa diperoleh secara cuma-cuma dengan mengunduh di situs <http://www.laurenceanthony.net>. Adapun fitur yang dimanfaatkan dalam penelitian ini adalah *wordlist* dan *concordance*.

Sementara itu, data yang digunakan dalam penelitian ini adalah daftar frekuensi yang dikategorikan sebagai *hapax legomenon*. Untuk mendapatkan *hapax legomenon* korpus Nh. Dini diolah melalui beberapa tahap. Pertama, korpus Nh. Dini

diolah menggunakan AntConc untuk memperoleh daftar frekuensi kata. Kedua, dari daftar frekuensi kata dipilihkan daftar kata yang memiliki frekuensi satu (*hapax legomenon*).

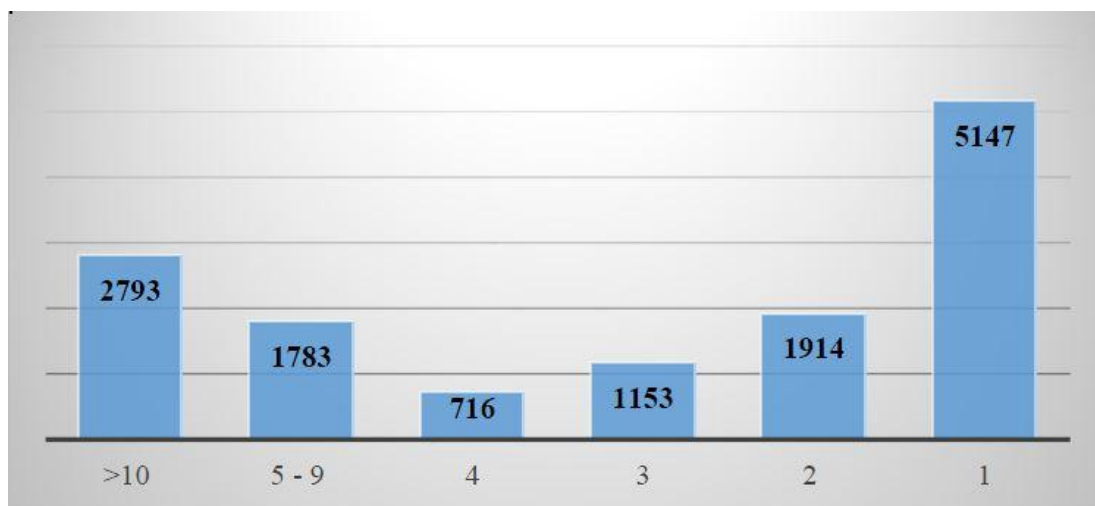
Hapax legomenon tersebut kemudian diamati untuk disaring mana kata yang sudah terekam dalam KBBI Edisi Kelima versi daring. Kemudian, kata-kata yang belum terekam diklasifikasi dan diinterpretasi.

III. ANALISIS PENELITIAN DAN PEMBAHASAN

3.1 Analisis Penelitian

Dari hasil pengolahan data, sebagian daftar kata yang memiliki frekuensi satu atau *hapax legomenon* memiliki jumlah yang paling besar dibandingkan dengan daftar kata yang memiliki jumlah frekuensi yang lain. Data selengkapnya akan disajikan dalam diagram batang berikut.

Diagram 1. Jumlah Tipe dari Pelbagai Frekuensi dalam Korpus Nh. Dini



Dari diagram di atas, dapat diperoleh hasil sebagai berikut: pertama kata yang muncul satu kali atau *hapax legomenon* berjumlah 5147 atau 38,11 persen dari total tipe dalam korpus Nh. Dini, yaitu 13.506; kedua, kata yang muncul 2 kali berjumlah 1914; ketiga, kata yang muncul 3 kali berjumlah 1153; keempat, kata yang muncul 4 kali berjumlah 716; kelima, kata yang muncul 5 sampai 9 kali berjumlah 1783; terakhir, kata yang muncul di atas 10 kali berjumlah 2793.

Data yang disajikan di atas menunjukkan bahwa *hapax legomenon* merupakan kelompok tipe terbesar dalam korpus Nh. Dini. Hal ini senada dengan pandangan Greenspahn bahwa jumlah *hapax legomenon* menjadikannya dinobatkan sebagai kelompok terbesar dalam pelbagai korpus. Jumlah ini menurut Baayen dan Renouf (dalam Domazakis, 2010:12) menyesuaikan dengan ukuran korpus yang diselidiki. Namun *hapax legomenon* dalam korpus Nh. Dini ini lebih rendah dari jumlah yang diprediksi Coultrad dan Johnson. Mereka mengatakan bahwa pada umumnya jumlah

tipe yang berfrekuensi satu berjumlah lebih dari 50 persen (dalam Lindquist, 2008:69).

Daftar kata *hapax legomenon* dalam korpus Nh. Dini, sebagian besar merupakan kata leksikal yang berkelas kata adjektiva, nomina, verba, dan numeralia. Hanya ditemukan satu konjungsi, yaitu *adapun* dan lima interjeksi, seperti *yaa*, *yaaaa*, *yaaaaa*, *yaaaah*, dan *aaaah*. Satu kata fatis, yaitu *oya*. Komposisi kelas kata ini menunjukkan bahwa kata leksikal dalam *hapax legomenon* sangat beragam. Ini memungkinkan untuk dilakukan penyelidikan yang bertujuan mencari kata baru yang belum terekam dalam kamus.

Berdasarkan hasil analisis terhadap *hapax legomenon* korpus Nh. Dini, terdapat empat kata dasar yang belum terekam dalam KBBI Edisi Kelima, di antaranya: *gercik*, *kekep*, *kerecap*, dan *pudeng*. Dari hasil analisis menggunakan teknik konkordansi didapat makna dari kata-kata tersebut. Kata *gercik* dalam korpus Nh. Dini memiliki makna tiruan bunyi air. Kata *kekep* dalam korpus Nh. Dini bermakna kekang atau tindakan membatasi seseorang. Sementara itu, kata *kerecap* dalam korpus Nh. Dini bermakna tiruan bunyi mulut yang sedang mengunyah makanan. Terakhir, *pudeng* dalam korpus Nh. Dini adalah jenis makanan penutup yang disajikan dengan buah.

Selain itu, temuan yang dihasilkan dalam penelitian ini adalah terdapat 88 kata turunan yang belum terekam dalam KBBI. Informasi lebih lengkap disajikan dalam tabel berikut ini.

Tabel 1. Daftar kata Turunan yang Belum Terekam dalam KBBI V Versi Daring

No	Imbuhan	Bentuk Kata Turunan	Jumlah
1	ber-, ber-kan, ber-an, ber-R	berbata, berbetis, berbungakan, bercampuh, bercatatan, berdaunkan, berengkuhan, bergenting, bergerisik, berhimpitan, berindustri, berjulangan, berkasut, berkekasihkan, berkenal baik, berkepingan, berkepulan, berkepulan, berkerudungkan, berketinggian, berkimono, berkondangan, berlahiran, berlangit, berlantaikan, berlimpahkan, berliter, berpanasan, berpantulan, berpendengaran, berpensiun, berperanan, berpinggirkan, berpulangan, berselakan, berselingkan, bersemburatkan, bersulamkan, bersumpalan, bertabah, bertangkaikan, bertanyakan, bertarikan, bertiduran, bertimpahan, bertitikan, bertolakan, dan bertong-tong	48
2	me-, me-i,	melendotkan, memedihi, memenuh, mencucuki,	11

	dan me-kan	mencukit, menelak, mengacai, mengharapi, menjeritkan, menyayup, dan menyebabkan.	
3	ter-	terbongkok, terikal, terisikan, terlapis, terloncat, termasak, tersambut, tersedan, tersilir, tersimpang, dan tersulir	11
4	se- dan se-an	secabik, seluasan, selubungan, selungkupan, semeja, sesehat, dan sesen	7
5	ke- dan ke-an	kebayian, kebecekan, keciap, kerandatan, kerompong, kesendatan	6
6	-an	mentahan, solekan, dan tindasan	3
7	pe-an	penjajaan dan penteroran	2
Total			88

Dari tabel di atas dapat diketahui kata turunan dengan 14 jenis imbuhan yang belum terekam dalam KBBI Edisi Kelima. Dari temuan ini ada beberapa hal yang bisa digaribawahi. Pertama, terlihat bahwa pengarang memanfaatkan afikasi secara maksimal untuk membentuk kata-kata yang mendukung narasi dalam karya sastra. Dalam ini pengarang memanfaatkan betul hak licentia poetika yang dimilikinya, yakni hak yang memungkinkan pengarang untuk menggunakan bahasa sesuai kepentingan nilai estetika. Kedua, temuan ini menguatkan pandangan Baayen dan Renouf (1996) bahwa dalam senarai kata-kata yang berfrekuensi rendah terdapat jumlah neologisme (pembentukan kata-kata baru) yang tinggi. Dalam hal ini, kata-kata baru yang muncul adalah hasil afikasi dari afiks ber-kan, ber-an, dan ber-R. Boleh jadi produktivitas pembentukan kata dengan tipikal imbuhan seperti yang tertera di atas menjadi salah satu idiosinkrasi personal atau sidik jari pengarang, yaitu Nh. Dini.

Selain memiliki keragaman bentuk kata leksikal, terdapat pula keragaman penggunaan kode dalam *hapax legomenon* korpus Nh. Dini. Sebagian besar menggunakan kode Indonesia. Sisanya menggunakan kode Jawa, kode Jepang, kode Inggris, kode Belanda, dan kode Prancis. Tipe yang menggunakan kode Jawa, yaitu *ajar, ajer, slitutan, lamis, malati, rasanan, sawal, tresno, jalaran, soko, kulino, dan setiti*. Kode bahasa Inggris, misalnya *appeal, depression, sex, show, shock, nervouse, private, photo, slow, story, VIP, dan walls*. Kode bahasa Prancis misalnya *rendez, sacra, va, le, dan vous*. Terakhir, kode bahasa Jepang, seperti *ofuro, momeji, soro, sorou, dan warasu*. Kode Belanda, yaitu *rijsttafe*.

Dengan memanfaatkan pandangan Oakes (1998:2011) penyelidikan lanjut terhadap tipe-tipe tersebut bisa mengungkapkan tentang latar belakang tokoh, pola pikir tokoh, kondisi kejiwaan tokoh, dan kondisi budaya antarnegara. Namun, terkait dengan penyusunan kamus, penggunaan kode-kode bahasa lain memungkinkan dilakukannya penyelidikan etimologi atau kemungkinan penyerapan kata-kata baru dan bahasa asing. Salah satu contohnya, kata *depression* yang dipakai dalam novel *La Barka* tahun 1975. Ini bisa menjadi informasi leksikografi yang menarik, bahwa pada tahun tersebut kata *depresi* bisa jadi belum dikenal atau belum diserap dalam bahasa Indonesia. Dalam kasus lain, penggunaan kode-kode dari bahasa asing bisa menjadi tinjauan bagi pekamus atau editor kamus untuk menentukan mana kata-kata yang patut diserap dalam bahasa Indonesia.

Selain itu, penggunaan bahasa Jawa memungkinkan pekamus menjadikan kata tersebut dimasukkan ke dalam kamus. Salah satu contohnya adalah kata *sawal*. Kata *sawal* diduga berasal dari bahasa Jawa. Selain digunakan dalam *Jalan Bandungan* pada tahun 1989. Kata *sawal* pun digunakan dalam buku *Desa* karya Sutardjo Hadikusumo (1984:224) ...*mengeluarkan perkakas tidur berupa bantal, galar, tikar, mentjutji sawal bantal dan Kasur*. Pun berdasarkan penelusuran di situs jual beli Bukalapak, terdapat judul iklan *Sarung Bantal Atau Sawal*.^{iv} Penelusuran ini bisa dilakukan dan dijadikan tinjauan bagi pekamus atau editor untuk menyerap kata-kata dalam bahasa daerah untuk memperkaya kosakata bahasa Indonesia.

Temuan terakhir, *hapax legomenon* korpus Nh. Dini memungkinkan pekamus mendapatkan penggunaan baru atas kata-kata yang sudah terekam dalam KBBI. Misalnya untuk kata *timpahan*^v (dalam KBBI Edisi Kelima tercatat *timpaan*). Dalam korpus Nh. Dini, terdapat penggunaan kata *timpahan* yang memiliki makna berbeda dengan makna yang direkam dalam KBBI Edisi Kelima versi daring. Dalam kamus tersebut secara leksikal *timpaan* berarti (1) tekanan (tindihan) berat; beban yang memberati (menyusahkan atau sebagainya); gencetan; (2) (barang) curian. Sementara itu, dalam korpus Nh. Dini, *timpahan* semakna dengan gerak udara bertiup, tiupan; terpaan. Hal ini tercermin dalam contoh kalimat di bawah ini:

*Dalam deru mesin dan **timpahan** angin yang amat keras, karena sia-sia menangkap kode-kode yang sampai pada alat radio yang macet itu.*

Dari contoh di atas, penyelidikan *hapax legomenon* bermanfaat untuk memetakan kembali makna yang terkandung dalam sebuah kata dan menjadi sumber data atas contoh kalimat yang akan ditampilkan dalam kamus.

3.3. Pembahasan

Dari temuan-temuan yang dipaparkan di atas, ada beberapa hal yang bisa digarisbawahi terkait dengan pengkajian *hapax legomenon* dalam korpus Nh. Dini. Pertama, mengenai jumlah *hapax legomenon* dalam korpus. Menurut Coultrad dan Johnson, umumnya jumlah tipe yang berfrekuensi satu berjumlah lebih dari 50

persen (dalam Lindquist, 2008:69). Berdasarkan hasil analisis, angka ambang batas 50 persen tidak berlaku secara absolut. Hal ini sesuai dengan pandangan Baayen dan Renouf (dalam Domazakis, 2010:11-12) yang mengatakan bahwa jumlah *hapax legomenon* sejalan dengan ukuran korpus. Dengan demikian, untuk sementara, jumlah *hapax legomenon* dalam korpus kecil bisa berada di bawah 50 persen dari jumlah total tipe dalam sebuah korpus. Sebaliknya, jumlah *hapax legomenon* bisa berada di atas 50 persen dalam korpus umum (*general corpus*) sebagaimana hasil penelitian yang dilakukan oleh beberapa peneliti.

Kedua, seperti yang dilakukan oleh editor Cobuild Dictionary, pemanfaatan korpus memungkinkan editor atau pekamus untuk mendapatkan kata baru baik berupa lema dasar atau lema turunan. Berdasarkan hasil analisis, penemuan terhadap lema turunan memiliki kemungkinan yang besar dibandingkan lema dasar. Hal ini menunjukkan kreativitas pengarang dalam membentuk lema-lema baru dari lema dasar yang sudah terekam kamus. Pun hal ini menunjukkan afiksasi menjadi proses morfologis yang sangat potensial untuk mengembangkan kosakata baru.

Ketiga, pengkajian *hapax legomenon* memungkinkan pekamus atau editor memperoleh informasi leksikografi mengenai penggunaan kata, seperti kapan kata itu digunakan, oleh siapa, dan dalam karya apa. Informasi berharga ini menjadi penting untuk disimpan dalam pangkalan data dan memudahkan pekamus atau peneliti bahasa lain yang akan mengkaji etimologi kata atau bahkan penyusunan kamus etimologi.

Keempat, penelusuran *hapax legomenon* dalam korpus sastra memungkinkan pekamus atau editor mendapatkan informasi tentang penggunaan baru atas lema yang sudah terekam dalam kamus. Hal ini akan berkontribusi pada dua hal, (a) memetakan kembali makna sebuah lema; memperkaya polisemi (b) menjadikan korpus sebagai sumber untuk contoh kalimat. Contoh itu bisa diambil secara verbatim atau editor bisa mereka-reka contoh berdasarkan kecenderungan penggunaan sebuah lema dalam korpus.

Terakhir, pengkajian *hapax legomenon* memungkinkan pekamus dan editor untuk mempertimbangkan kata-kata asing dan kata daerah untuk diserap dalam bahasa Indonesia. Berdasarkan hasil analisis, sejumlah kata-kata dalam bahasa asing dalam korpus Nh. Dini digunakan untuk memperlihatkan gejala psikologis, seperti *depression* dan *shock*. Dua kata ini tercantum dalam novel *La Barka* (1975) dan *Jalan Bandungan* (1989). Boleh jadi saat itu, dua kata tersebut belum diserap ke dalam bahasa Indonesia. Dengan demikian, penyelidikan seperti ini bisa menjadi tinjauan, bagi pekamus yang menemukan kata asing, melihat penggunaannya dalam bahasa Indonesia, dan mempertimbangkan untuk diserap ke dalam bahasa Indonesia.

IV. SIMPULAN

Pengkajian korpus sastra yang terfokus pada kata yang muncul satu kali atau *hapax legomenon* memiliki kontribusi terhadap pengembangan leksikografi Indonesia. Adapun bentuk kontribusinya adalah memungkinkan pekamus untuk (1) mendapatkan kata-kata baru, (2) mendapatkan penggunaan baru atas kata-kata yang

sudah terekam dalam kamus, (3) memberikan informasi leksikografis terkait penggunaan kata, dan (4) menjadi tinjauan bagi pekamus untuk menyerap kosakata baru baik dari bahasa daerah maupun bahasa asing.

Penelitian sederhana ini merupakan penelitian rintisan yang memiliki kemungkinan luas untuk dikembangkan. Objek penelitian bisa diperluas tidak hanya pada *hapax legomenon* tapi juga pada *hapax dislegomenon* (kata yang muncul dua kali). Pun memperluas objek tidak hanya pada novel, melainkan pada karya-karya Nh. Dini yang lain, yaitu catatan kenangan.

Daftar Pustaka

- Kartohadikoesoemo, S. 1984. *Desa*. Jakarta: Balai Pustaka.
- Waltke, B. K., dan C. O'Connor. 1990. *An Introduction to Biblical Hebrew Syntax*. Indiana: Eisenbrauns.
- Oakes, Michael P., 1998. *Statistics for Corpus Linguistics Edinburgh Textbooks in Empirical Linguistics*. Edinburgh:Edinburgh University Press
- Crystal, D. 2003. *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Hellwig, T. 2003. *In the Shadow of Change: Citra Perempuan dalam Sastra Indonesia*. Depok Desantara.
- Bauer, L. 2004. *A Glossary of Morphology*. Edinburgh: Edinburgh University Press.
- Prasad, A. N. 2004. *Arundhati Roy's The God of Small Things: A Critical Appraisal*. New Delhi: Sarup & Sons.
- Lindquist, Hans. 2009. *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.
- Domazakis, Nikolaos. 2010. *Septuagintal hapax legomena and neologisms in 2 Maccabees, 4-7*. Thesis: Lund University.
- Litosseliti, L. ed. 2010. *Corpus Methods in Linguistics*. London: Continuum.
- Cheng, Winnie. 2012. *Exploring Corpus Linguistics: Language in Action*. London: Routledge.
- McEnery, T. dan Hardie, Andrew. 2012. *Corpus Linguistics: Method, Theory, and Practice*. New York: Cambridge University Press.
- Dora Amalia. 2014. *Formulasi Pendefinisian dan Model Pengertian Verba dalam Kamus Pemelajar Bahasa Indonesia*. Disertasi: Universitas Indonesia.
- Nurjanah, Neneng. 2015. *Representasi Isu Perempuan dalam Novel-Novel Nh. Dini: Sebuah Analisis Kata Kunci*. Tesis: Universitas Indonesia.

Situs

<http://www.laurenceanthony.net>

<https://kbbi.kemdikbud.go.id/>

<http://bukalapak.com>

<http://Collins.co.uk/>

ⁱ <http://Collins.co.uk/pages/elt-cobuild-reference-the-collins-corpus>

ⁱⁱ Definisi ini diambil dari situs www.etymonline.com

ⁱⁱⁱ Menurut KBBI Edisi Kelima versi daring, neologisme adalah kata bentukan baru atau makna baru untuk kata lama yang dipakai dalam bahasa, yang memberi ciri pribadi atau demi pengembangan kosakata.

^{iv} Penggunaan kata sawal dalam situs jual beli Bukalapak diakses dari tautan <https://www.bukalapak.com/p/fashion-pria/celana-299/celana-pendek-2599/7xjrn7-jual-sarung-bantal-atau-sawal>

^v Kata timpahan digunakan dalam karya *Hati yang Damai* (1961). Sekarang, kata *timpahan* disesuaikan dengan EYD menjadi *timpaan*.

“SI KAMPER” SEBAGAI UPAYA PENGENALAN KOSAKATA BARU BAHASA INDONESIA KEPADA SISWA SEKOLAH MENENGAH PERTAMA

(*"SI KAMPER" AS AN INDONESIAN VOCABULARIES ACQUISITION TOOL
FOR MIDDLE SCHOOL'S STUDENTS*)

Ilmatius Sa'diyah, S.Pd., M.Hum.

Zamzam Syifa Boarding School

ilmatussadiyah@gmail.com

Abstrak

Kosakata bersifat dinamis mengikuti perkembangan ilmu pengetahuan dan teknologi. Sifat itu menyebabkan munculnya kosakata baru bahasa Indonesia yang diserap dari bahasa asing/daerah untuk menyesuaikan dengan kosakata baru yang ada. Untuk menyesuaikan dengan kata *download*, badan bahasa membuat padanan kata dalam bahasa Indonesia, yaitu *unduh*. Begitu pula dengan kata baru *outbound*, padanan kata dalam bahasa Indonesia sudah ada, yaitu *mancakrida*. Namun, padanan kata dalam bahasa Indonesia itu tidak digunakan oleh penutur jati bahasa Indonesia, terutama para remaja. Peneliti pun melakukan studi awal terhadap 15 siswa sekolah menengah pertama yang sedang mengambil kelas tambahan di bimbingan belajar. Studi awal menunjukkan bahwa siswa tidak mengetahui kata *mancakrida* dan sebagian mengetahui kata *unduh*. Saat siswa diminta untuk menggunakan kata *mancakrida* dan *unduh* dalam kehidupan sehari-hari, seluruh siswa tidak berkenan menggunakan karena menganggap kata *mancakrida* dan *unduh* aneh saat digunakan dan belum terbiasa. Kondisi itu disebabkan oleh lemahnya peran KBBI, baik secara daring maupun luring dalam menunjukkan keberadaan kosakata baru dalam bahasa Indonesia. Oleh karena itu, penelitian ini dilakukan untuk mengembangkan *Si Kamper* (Kamus Permainan) untuk mendukung penggunaan kosakata baru bahasa Indonesia secara meluas dan sejak dini. *Si Kamper* adalah kamus mini berisi kosakata baru dan kosakata baku bahasa Indonesia beserta artinya dan padanan dalam bahasa Inggris. Kosakata itu dipadukan dengan permainan berupa soal yang digunakan untuk melanjutkan permainan. Apabila tidak dapat menjawab soal terkait kosakata, permainan tidak dapat dilanjutkan. Penelitian pengembangan ini bertujuan mendeskripsikan proses pengembangan, kualitas pengembangan, dan keefektifan pengembangan. Penelitian dilakukan dengan menggunakan metode pengembangan 3P Thiagarajan (pendefinisian, perancangan, dan pengembangan). Pengembangan dilakukan menggunakan media *adobe flash player*. Revisi dilakukan terhadap penambahan jumlah kosakata, penampilan *Si Kamper*, dan perbaikan arti kata. Berdasarkan penilaian validator (guru sejawat), *Si Kamper* menunjukkan kualitas sangat bagus (89%), sedangkan menurut siswa, *Si Kamper* sangat menarik digunakan karena dapat menambah wawasan tentang kosakata baru dan baku. Setelah diuji coba secara terbatas dan luas, nilai siswa menunjukkan kenaikan sehingga *Si Kamper* menunjukkan keefektifan. Oleh karena itu, *Si Kamper* dapat

digunakan sebagai media pembelajaran oleh guru bahasa Indonesia di sekolah secara meluas. Guru dapat menggunakan *Si Kamper* saat pelajaran menulis teks, baik teks laporan hasil observasi maupun teks deskripsi sehingga siswa mengenali kosaka baru dan secara perlahan dapat menggunakannya dalam berkomunikasi sehari-hari. Siswa tidak lagi menggunakan kata *online*, tetapi menggunakan kata *daring*.

Kata Kunci: *Si Kamper* (Kamus Permainan), penelitian pengembangan, media pembelajaran, kamus mini, kosakata baru bahasa Indonesia.

Abstract

Vocabulary is dynamic following the development of science and technology. This characteristic leads to the emergence of a new Indonesian vocabulary absorbed from a foreign / regional language to conform to the existing vocabulary. To adjust to the download word, the language body makes the equivalent of the word in Indonesian, which is unduh. Similarly with the new word outbound, the equivalent of the word in the Indonesian language already exists, namely mancakrida. However, the equivalent of the word in Indonesian is not used by Indonesian teak speakers, especially teenagers. Researchers also conducted a preliminary study of 15 junior high school students who are taking additional classes in tutoring. Initial studies show that students do not know the word mancakrida and some know the word download. When students are asked to use the word mancakrida and unduh in everyday life, all students are unwilling to use because they think the word mancakrida and unduh strange when used and not used. The condition is caused by the weak role of KBBI, both online and offline in indicating the existence of new vocabulary in the Indonesian language. Therefore, this research was conducted to develop the Kamper (Dictionary Game) to support the widespread use of Indonesian vocabulary and from an early age. The Kamper is a mini dictionary containing new vocabulary and standard Indonesian vocabulary with its meanings and equivalents in English. The vocabulary is combined with a game of questions used to continue the game. If you can not answer vocabulary related questions, the game can not continue. This development research aims to describe the process of development, quality of development, and development effectiveness. The research was conducted using Thiagarajan 3P development method (definition, design, and development). The development is done using adobe flash player media. Revisions were made to the addition of vocabulary numbers, the appearance of Kamper, and the improvement of the meaning of words. Based on the validator's assessment (peer teacher), Kamper showed excellent quality (89%), whereas according to the student, Kamper is very interesting to use because it can add insight about new and standard vocabulary. Having been tested on a limited and broad scale, student grades show an increase so that Kamper shows effectiveness. Therefore, Kamper can be used as a medium of instruction by Indonesian language teachers in schools widely. Teachers can use Kamper during text writing lessons, both observational and text-only text texts so students recognize new vocabulary and can gradually use it in daily communication. Students no longer use the word online, but use the word online.

Keywords: *The Kamper (Dictionary Game), development research, instructional media, mini dictionary, new Indonesian vocabulary.*

I. PENDAHULUAN

Bahasa Indonesia bukan merupakan bahasa yang sudah terbentuk sejak awal, melainkan bahasa yang terbentuk setelah bahasa lain telah berkembang pesat. Fakta ini menyebabkan bahasa Indonesia harus menyesuaikan diri dengan menyerap banyak kosakata dari bahasa yang sudah ada. Bahkan, seiring dengan perkembangan zaman dan teknologi, bahasa Indonesia mengalami penambahan jumlah kosakata yang pesat. Melalui kemajuan teknologi, ada beberapa kata yang dimunculkan untuk menyesuaikan dengan keberadaan kosakata baru.

Pada kemunculan kata *website* dalam bidang teknologi, pusat bahasa berusaha mencari padanan kata tersebut lalu terbentuklah istilah *laman* yang memiliki makna sama dengan kata *website*. Namun, kata *laman* belum mendapat respons positif dari masyarakat luas terutama remaja. Kata *website* masih digunakan oleh mereka dalam kehidupan sehari-hari. Hanya penggiat bahasa, pusat bahasa dan ahli linguistik, yang menggunakan istilah itu. Waktu penggunaannya pun terbatas pada kegiatan formal, seperti seminar. Apabila dalam percakapan nonformal, kata *website* bisa jadi masih digunakan karena lawan bicara mereka bukan berasal dari kalangan akademik.

Remaja Indonesia, dalam hal ini siswa sekolah menengah pertama, menyatakan bahwa kata *laman* tidak pernah dikenalnya. Mereka mengetahui bahwa kata yang digunakan untuk menyebut halaman internet adalah kata *website*. Berdasarkan studi awal melalui kuesioner dan wawancara singkat, siswa SMP yang menjadi subjek dalam penelitian ini merasa aneh dengan sejumlah padanan kosakata yang diberikan, seperti *mi dadak*, *manckarida*, dan *daring* sebagai pengganti dari kata *mie instant*, *outbound*, dan *online*. Oleh karena itu, mereka enggan menggunakannya dalam kehidupan sehari-hari, apalagi teman-teman sejawatnya juga tidak menggunakannya.

Dalam hal ini, kamus bahasa Indonesia seyogianya menjadi jembatan untuk mengenalkan kosakata baru bahasa Indonesia yang sudah dibentuk. Kamus bahasa Indonesia yang sudah ada belum dapat memberikan peran itu karena jumlah kosakata di dalamnya yang sudah terlalu banyak. Kosakata baru yang sudah dibentuk menjadi tenggelam di antara kosakata baru yang sudah ada. Untuk itu, keberadaan kamus mini sangat diperlukan. Kamus mini itu hanya diisi dengan kosakata baru bahasa Indonesia agar pengenalan dan pembiasaan kosakata baru menjadi lebih mendalam.

Kamus mini yang akan diisi dengan kosakata baru bahasa Indonesia dibuat dalam bentuk permainan. Hal itu menyesuaikan dengan dampak positif dari keberadaan permainan. Permainan yang dikemas dengan menarik dapat merangsang daya pikir, meningkatkan konsentrasi siswa, dan dapat memecahkan masalah (Rahman & Tresnawati, 2016). Kamus permainan ini pun dapat digunakan sebagai media pembelajaran di kelas.

Berdasarkan uraian masalah tersebut, penelitian ini dilakukan untuk mendeskripsikan proses pengembangan kamus permainan, mendeskripsikan kualitas kamus permainan, dan mendeskripsikan keefektifan penggunaan kamus permainan. Penelitian ini diharapkan dapat

memberikan kebaruan berupa kontribusi terhadap pembelajaran kosakata bahasa Indonesia di sekolah dan sebagai upaya mengenalkan kosakata baru bahasa Indonesia sejak dini.

II. KERANGKA TEORI

Dalam hal ini, terdapat beberapa kerangka teori yang digunakan dalam penelitian ini. Berikut ini adalah penjabaran tiap teori yang digunakan.

2.1. Permainan Edukasi

Pembelajaran kosakata memang tidak dapat dilepaskan begitu saja dari pembelajaran bahasa secara umum. Melalui kosakata, keterampilan berbahasa lainnya dapat dikuasai oleh siswa. Hal itu menunjukkan bahwa kosakata berperan penting dalam pembelajaran bahasa. Untuk itulah penguasaan kosakata dalam pembelajaran bahasa Indonesia perlu ditingkatkan, terutama pada penguasaan kosakata baru bahasa Indonesia.

Pembelajaran kosakata dapat ditingkatkan penguasaannya dengan menggunakan permainan. Permainan dianggap dapat membantu siswa untuk menggunakan kosakata secara komunikatif (Derakhsan & Khatir, 2015). Selain itu, kosakata menjadi dapat dikuasai dengan baik oleh siswa dalam kondisi yang santai, penuh motivasi, dan menyenangkan (Ashraf, dkk, 2014).

Keberadaan permainan dapat mengurangi perasaan menegangkan dari siswa saat belajar kosakata. Pada pembelajaran kosakata baru yang tidak begitu digunakan oleh siswa, permainan dapat memunculkan perasaan nyaman sehingga kosakata baru dapat diterima dengan baik oleh siswa. Permainan pun dapat mengarahkan pembelajaran sesuai dengan tujuan pembelajaran yang diharapkan oleh pengajar.

2.2 Kamus sebagai Media Pembelajaran

Pembelajaran bahasa merupakan pembelajaran yang bersifat kompleks sehingga mengharuskan pengajar bahasa terampil menggunakan material pembelajaran di kelas, terutama penggunaan teknologi baru (Asadi dan Berimani, 2015). Satu di antara teknologi baru adalah *smartphone* yang memiliki banyak aplikasi. Aplikasi tersebut dapat diintegrasikan dalam praktik pembelajaran bahasa di kelas sehingga dapat meningkatkan kemampuan berbahasa siswa dan meningkatkan daya ingat siswa terhadap makna, lafal, dan struktur kata (Kalyuga, Mantai, dan Marrone, 2013; Barcena, dkk, 2015). Oleh karena itu, guru perlu kreatif mengintegrasikan teknologi, *smartphone*, ke dalam material pembelajaran bahasa kedua (Nalliveettil dan Alenazi, 2016). Teknologi memiliki dampak terhadap penggunaan media pembelajaran di kelas, yaitu memfasilitasi dan membantu guru dalam pengajaran (Tavakoli dan Esmae'li, 2013).

Dalam penggunaannya di kelas, media pembelajaran memberikan beberapa manfaat, yaitu meningkatkan motivasi siswa, meningkatkan kemampuan guru dalam mengajar, membantu guru menjelaskan materi yang bersifat abstrak dengan lebih baik, meningkatkan pemahaman siswa, membuat pembelajaran lebih bermakna, memberikan materi yang kuat/terpercaya, menyediakan kemudahan akses terhadap sumber belajar bahasa, memberikan

konteks yang autentik dan nyata, mengurangi rasa stres siswa saat di kelas, meningkatkan ketertarikan siswa karena media menggunakan dialek dan gaya yang berbeda, memfasilitasi siswa terhadap bentuk bahasa yang berbeda, meningkatkan pemerolehan kata dan kalimat, memotivasi siswa untuk mencapai target bahasa, dan memadatkan informasi serta memudahkan dalam memahami materi (Arsyad dalam Maksudin, 2006; Oroujlou, 2012; Sakat, dkk, 2012; Swiatek (2013); Tavakoli dan Esmae'li, 2013; Rao, 2014; Amelia dan Budiman, 2015; Asemota, 2015; Klentien dan Kamnungwet, 2015).

Media pembelajaran dibedakan menjadi beberapa jenis. Perbedaan tersebut berdasarkan tipe dan kegunaannya. Jenis media itu adalah media berbasis manusia (guru), media berbasis komputer (berbantuan komputer), media fotografi (gambar, slide, dan slide), media audio (radio), media visual (grafik), media audio-visual (televisi), media serba aneka (patung), media cetak (buku), simulasi dan *games*, multimedia (video animasi), internet, *smartphone* dengan dukungan sistem *android*, dan media CD-ROM atau *flash disk* (Mahsun, 2010; Putra dalam Mahsun, 2010; Naz dan Akbar, 2012; Gilakjani, 2012; Amelia dan Budiman, 2015).

III. METODE PENELITIAN

Metodologi yang digunakan dalam penelitian ini merupakan penjabaran dari tahapan-tahapan penelitian dan pengembangan. Metodologi penelitian tersebut diuraikan sebagai berikut.

3.1. Subjek Penelitian

Penelitian ini dilakukan terhadap lima belas siswa sekolah menengah pertama yang sedang belajar di kelas tambahan. Subjek itu ditentukan melalui prosedur pengambilan sampel secara acak.

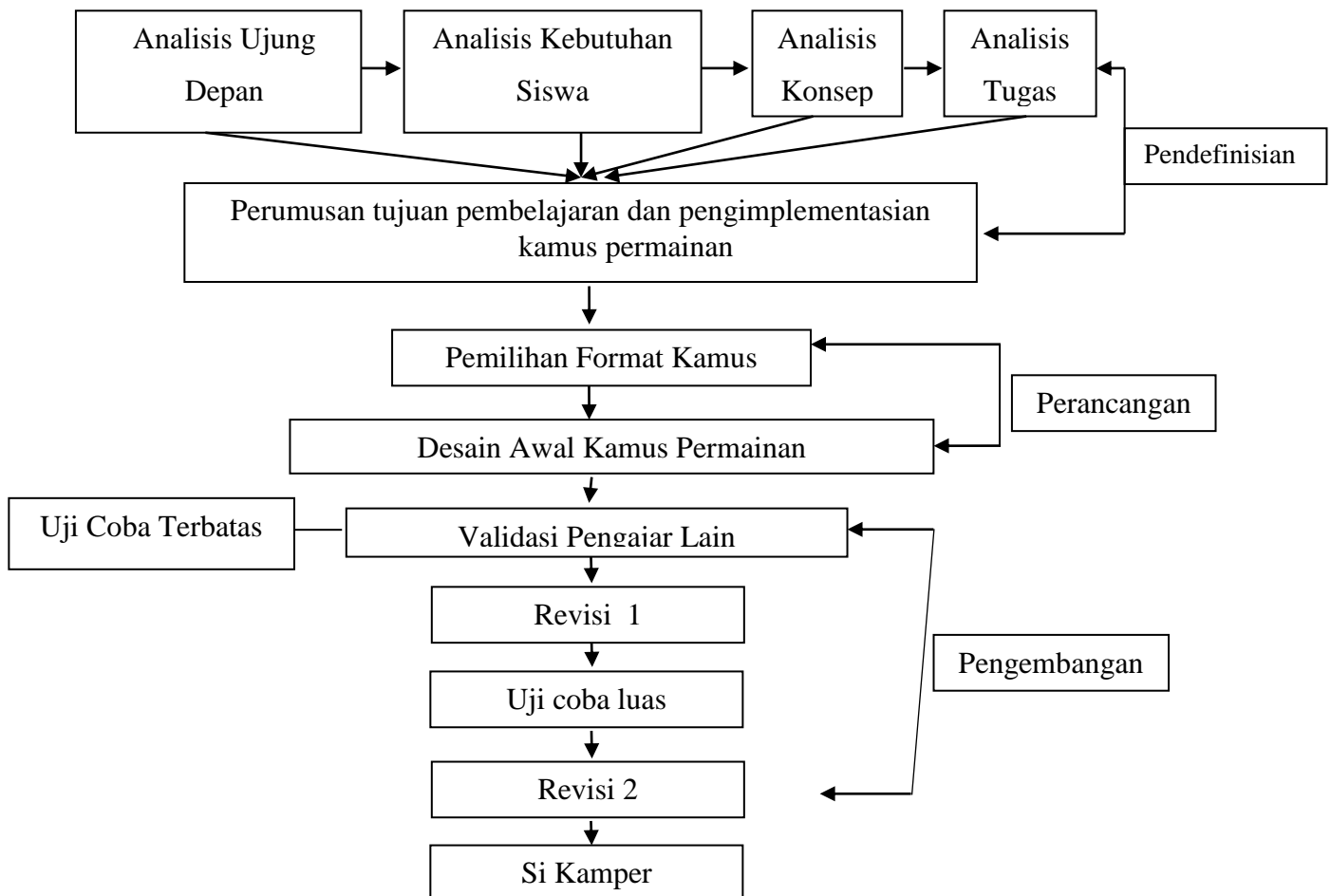
3.2. Jenis Penelitian

Penelitian ini merupakan penelitian dan pengembangan (*research and development*) karena diorientasikan untuk menghasilkan dan menguji keefektifan serta kualitas produk berupa media pembelajaran. Media pembelajaran yang dihasilkan dapat digunakan untuk menunjang pembelajaran bahasa Indonesia.

3.3. Rancangan Penelitian

Model pengembangan Thiagarajan, 4-D, digunakan dalam penelitian ini. Model pengembangan 4-D meliputi *define*, *design*, *develop*, dan *disseminate* yang diadaptasi menjadi 4-P, yaitu pendefinisian, perancangan, pengembangan, dan penyebaran. Model ini sesuai dengan tujuan penelitian, yaitu mengembangkan media pembelajaran berbasis aplikasi dan audio. Penelitian ini hanya dilakukan sampai pada tahap pengembangan karena penelitian ditujukan untuk penelitian tahap ke-1. Selanjutnya, penelitian dilanjutkan menjadi penelitian tahap ke-2 pada tahap penyebaran. Pada tahap penyebaran, media pembelajaran dapat

diujicoba keefektifannya pada konteks belajar dan subjek belajar yang berbeda. Berikut ini adalah bagan yang menggambarkan prosedur penelitian.



**Bagan Model Pengembangan Perangkat 3-P Thiagarajan
(Trianto, 2007, h.66)**

Tahap pendefinisian dilakukan untuk menetapkan dan mendefinisikan syarat-syarat pembelajaran. Penentuan dan penetapan syarat-syarat pembelajaran diawali dengan analisis tujuan dalam batasan materi pelajaran yang dikembangkan media pembelajarannya. Terdapat lima langkah pokok dalam tahap ini, yaitu analisis ujung depan, analisis tugas, analisis konsep, dan analisis tujuan pembelajaran yang dilakukan berurutan, tetapi dalam satu *framework* (Trianto, 2007, h.65).

Analisis ujung depan dalam pengembangan media pembelajaran adalah pemberlakuan sistem kurikulum. Sejalan dengan hal itu, kurikulum yang digunakan saat ini mengacu pada kurikulum di sekolah. Setelah mengidentifikasi kurikulum, analisis kebutuhan siswa dilakukan untuk mengetahui kebutuhan siswa dalam mempelajari bahasa Indonesia melalui media pembelajaran. Kebutuhan siswa yang telah didapatkan digunakan untuk menganalisis tugas sehingga dapat menentukan materi pembelajaran (Trianto, 2007, h. 67). Analisis tugas dilakukan untuk memerinci soal dalam media permainan. Selanjutnya, analisis konsep bertujuan mengidentifikasi fakta, konsep, prinsip, dan aturan yang dibutuhkan dalam

pengajaran kosakata. Analisis ini selanjutnya menjadi dasar bagi penyusunan media pembelajaran. Penjabaran tersebut terdapat pada bagan berikut.

Tahap berikutnya adalah pengimplementasian dalam merancang format media pembelajaran berupa kamus permainan. Format yang telah dibuat dijadikan dasar dalam pembuatan desain awal. Kedua tahap ini berada dalam tahap perancangan. Selanjutnya, media pembelajaran berupa desain awal diujicobakan secara terbatas dan divalidasi kepada pengajar lain. Validasi media pembelajaran meliputi kelayakan isi, kelayakan penyajian, kelayakan bahasa, dan kelayakan kegrafikaan. Setelah itu, desain awal media pembelajaran akan direvisi (revisi 1). Hasil revisi 1 diujicobakan secara luas dalam kelas besar dan jumlah siswa lebih banyak. Uji coba dilakukan untuk mengetahui kepraktisan dan keefektifan media pembelajaran dalam pembelajaran sebelum media pembelajaran tersebut digunakan secara umum. Uji coba media pembelajaran bertujuan mengetahui kemampuan dan kemudahan siswa dalam memahami pokok pembelajaran dengan menggunakan media pembelajaran, mengetahui efisiensi waktu belajar dengan menggunakan media pembelajaran, dan mengetahui keefektifan media pembelajaran dalam membantu siswa mempelajari dan menguasai materi pembelajaran. Dari uji coba luas, media pembelajaran Kamper direvisi kembali sehingga didapatkan media pembelajaran yang berkualitas dan efektif.

3.4. Pengumpulan Data

Teknik pengumpulan data yang digunakan pada penelitian ini dibedakan berdasarkan masalah yang akan diteliti. Terdapat dua teknik pengumpulan data, yaitu angket dan tes. Angket yang digunakan bersifat tertutup dan terbuka yang dibedakan menjadi dua, yaitu angket kualitas media pembelajaran dan angket respons siswa. Tes yang digunakan dalam penelitian ini adalah tes formatif. Tes formatif dilakukan setelah melangsungkan kegiatan pembelajaran.

3.5. Penganalisisan Data

1) Teknik penganalisisan proses pengembangan

Proses pengembangan dianalisis dengan metode deskriptif kualitatif. Dokumen-dokumen yang ada digunakan untuk menguatkan deskripsi tentang proses pengembangan tersebut.

2) Teknik penganalisisan lembar angket

Data hasil validasi dari pengajar dan angket siswa direkapitulasi, dianalisis secara deskriptif kuantitatif, kemudian dianalisis dengan menggunakan rumus.

$$P = \frac{\text{jumlah skor seluruh validator}}{\text{Jumlah skor maksimal}} \times 100 \%$$

Jumlah skor maksimal

P = Penilaian

(Arikunto, 2010, h. 185)

Berdasarkan hasil analisis lembar angket, dilakukan penarikan simpulan bahwa media pembelajaran yang dikembangkan layak untuk digunakan bila telah memenuhi standar pendeskripsian sesuai dengan kualifikasi penilaian dengan skala likert (lihat tabel 1).

Tabel 1 Kualifikasi Penilaian

Tingkat Pencapaian	Kualifikasi
81–100 %	Sangat layak
61–80 %	Layak
41–60 %	Cukup layak
21–40 %	Kurang layak
0–20 %	Sangat kurang layak

3) Teknik penganalisisan lembar tes

Hasil belajar siswa didapatkan dari lembar penilaian tes siswa yang akan dianalisis secara deskripsi kuantitatif dengan rumus sebagai berikut.

$$\bar{X} = \frac{\sum X}{N}$$

(Nurgiyantoro, 2011, h.219)

Keterangan:

\bar{X} = Rata-rata nilai siswa

$\sum x$ = Jumlah nilai seluruh siswa

N = Jumlah siswa

Selanjutnya, rata-rata nilai siswa dan nilai siswa sebagai individu disesuaikan dengan kriteria atau patokan dalam pencapaian kelulusan menurut Nurgiyantoro (2011, h.253). Berikut ini adalah kriteria pencapaian siswa.

Tabel 2 Kriteria Pencapaian Siswa

Interval Persentase Tingkat Penguasaan	Nilai Ubahan Skala Sepuluh	Keterangan
96–100	10	Sempurna
86–94	9	Baik Sekali
76–85	8	Baik
66–75	7	Cukup
56–65	6	Sedang
46–55	5	Hampir Sedang
36–45	4	Kurang
26–35	3	Kurang Sekali
16–25	2	Buruk
1–15	1	Buruk Sekali

IV. ANALISIS PENELITIAN

Berdasarkan hasil penelitian yang telah dilakukan, diuraikan tiga hal, yaitu proses, kualitas, dan keefektifan pengembangan. Berikut ini merupakan uraian tiap-tiap subbabnya.

4.1 Proses Pengembangan Si Kamper

Sebelum media pembelajaran Si Kamper dikembangkan, dilakukan kegiatan analisis kebutuhan dan perspektif siswa terhadap penggunaan kamus dalam pembelajaran. Analisis itu dilakukan melalui angket terbuka (lihat tabel 3).

Tabel 3. Representasi Analisis Kebutuhan dan Perspektif Siswa

No	Pertanyaan	Ya	Tidak
1	Apakah pernah belajar bahasa Indonesia sebelumnya?	20%	80%
2	Dari empat keterampilan, apakah keterampilan menulis termasuk paling susah dipelajari?	93%	7%
3	Apakah jenis huruf bahasa Indonesia susah dieja dan didengarkan?	67%	33%
4	Apakah Anda pernah belajar bahasa Indonesia dari aplikasi yang ada di <i>smartphone</i> ?	100%	0
5	Apakah Anda masih membutuhkan kamus saat belajar bahasa Indonesia di kelas?	100%	0

Selanjutnya, media pembelajaran Si Kamper divalidasi oleh ahli dan pengajar dari tiga aspek, yaitu aspek isi, penyajian, dan kegrafikaan. Selain divalidasi, media pembelajaran Si Kamper diimplementasikan di kelas secara terbatas dan luas (15 siswa). Dari hasil validasi, penilaian, dan implementasi, perbaikan pada media Si Kamper perlu dilakukan. Perbaikan-perbaikan tersebut pada aspek isi, penyajian, dan kegrafikaan. Perbaikan dilakukan pada tiap-tiap aspek.

Pada aspek isi, Si Kamper diperbaiki dari segi jumlah kata. Si Kamper pada pembuatan kali pertama masih terdiri atas 5 kosakata. Jumlah itu ditambah menjadi lebih dari 5 kosakata. Selain itu, pada aspek makna kata, Si Kamper juga menambahkan arti kata dan kata padanan dalam bahasa Inggris sehingga siswa dapat membandingkannya dengan padanan kata dalam bahasa Inggris

Pada aspek penyajian, media Si Kamper draf awal tidak menyertakan suara apa pun sebagai pemotivasi atau perangsang semangat belajar. Hal itu mendapat saran perbaikan dari siswa untuk menambahkan unsur suara. Oleh karena itu, ditambahkan suara pada media Si Kamper sehingga Si Kamper tidak lagi hanya berbasis visual.

Perbaikan kedua pada aspek penyajian adalah urutan tampilan kata sebagai soal untuk melanjutkan permainan. Sementara itu, Pada aspek kegrafikaan, desain dibuat lebih menarik.

4.2 Kualitas Si Kamper

Penilaian kualitas media Si Kamper diberikan oleh pengajar lain. Pengajar dipilih berdasarkan kemampuannya pada bidang pengajaran, desain, dan bahasa. Berdasarkan penilaian, nilai terendah adalah dua dengan kualifikasi kurang baik dan skor tertinggi adalah empat dengan kualifikasi sangat baik. Skor terendah pada aspek suara. Media Si Kamper draf awal yang dinilai kualitasnya memang tidak memuat suara. Oleh karena itu, aspek ini telah diperbaiki. Skor terendah kedua pada aspek jumlah kata. Jumlah kata belum sesuai dalam draf awal media Si Kamper.

Secara umum, penilaian terbaik oleh semua validator pada aspek kemudahan untuk diakses karena Si Kamper akan di-*share* di *internet* secara gratis. Berikutnya, pada aspek makna kata yang dapat dimaknai dengan mudah. Pada revisi pertama setelah uji coba terbatas, media Si Kamper dinilai oleh validator. Penilaian validator setelah revisi pertama menjadi poin kualitas media Si Kamper sebesar 89%. Pada tabel konversi oleh Nurgiyantoro, jumlah itu menunjukkan bahwa media Si Kamper sangat layak untuk digunakan dalam pembelajaran.

4.3 Keefektifan Si Kamper

Media Si Kamper diuji tingkat keefektifan pada saat implementasi. Pengujian tingkat keefektifan pada dilakukan pada siswa . Keefektifan media Si Kamper bagi diukur berdasarkan respons pelajar setelah menjadi peserta implementasi dan rata-rata hasil tes. Respons yang diminta berupa respons bebas yang berbentuk uraian. Hal ini bertujuan mendapatkan saran sebagai bahan perbaikan.

Pada tes kosakata, skor rata-rata siswa adalah 82,07. Nilai itu menunjukkan bahwa media Si Kamper terkategori baik. Media Si Kamper pun dapat meningkatkan motivasi dan ketertarikan siswa untuk belajar. Bahkan, seorang siswa menyatakan akan belajar lebih banyak kosakata baru bahasa Indonesia. Mereka menyatakan pula bahwa melalui permainan ini mereka baru mengetahui keberadaan kosakata baru yang berpadanan dengan bahasa Inggris. Saat diminta untuk menggunakan, siswa masih merasa ragu, tetapi dia akan mencoba pada kesempatan yang baik dan ilmiah.

Selain dari nilai siswa, respons atas media Si Kamper juga diberikan oleh siswa. Komentar dan saran itu telah digunakan sebagai perbaikan untuk membuat media Si Kamper sebagai draf akhir. Keefektifan media Si Kamper dinilai dari saran dan komentar itu masih kurang. Oleh karena itu, media Si Kamper perlu dikembangkan menjadi lebih baik.

V. SIMPULAN

Si Kamper yang telah dikembangkan menunjukkan respons positif dari siswa. Siswa menjadi mengetahui dan mengenali kosakata baru bahasa Indonesia sebagai pengganti kosakata bahasa Inggris yang sering mereka gunakan. Meskipun respons positif itu belum diikuti oleh kemauan siswa untuk menggunakan kosakata baru dalam kehidupan sehari-hari dengan alasan kenyamanan dengan lawan bicara, media Si Kamper sudah menjalankan tugasnya untuk mengenalkan kosakata baru bahasa Indonesia. Hal itu sesuai dengan maksud

dari penelitian ini yaitu mengembangkan media untuk mengenalkan kosakata baru bahasa Indonesia sejak dini kepada siswa sekolah menengah pertama.

Media Si Kamper dapat digunakan secara menyeluruh oleh guru bahasa Indonesia di sekolah sebagai langkah awal untuk mengenalkan kosakata baru bahasa Indonesia. Media ini pun perlu lagi dikembangkan dengan menambah dan mengganti kosakata baru. Pengembangan lebih lanjut dari media ini sangat dibutuhkan.

Daftar Pustaka

- Amelia, R. Nugrahadi, D.T., & Budiman, I. (2015). Implementasi “Picture Description” pada Aplikasi Media Pembelajaran Bahasa Inggris. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 2 (1), 22—33
- Arikunto, S. (2010). *Prosedur Penelitian : Suatu Pendekatan Praktik*. Jakarta : Rineka Cipta
- Asadi, F. & Berimani, S. (2015). The Effect of Audio-Visual Materials on Iranian Second Grade High School Students’ Language Achievement. *International Journal of Language and Linguistics*. 3 (2), 69—75
- Asemota, H. E. (2015). The Role of Media in English Language Development. *International Journal of Humanities & Social Sciences Studies (IJHSSS)*. 2 (3), 311—316
- Bakhsh, S.A. (2016). Using Games as a Tool in Teaching Vocabulary to Young Learners. *English Language Teaching*, 9 (7), 121-128
- Derakhshan, A. & Khatir, E.D (2015). The Effects of Using Games on English Vocabulary Learning. *Journal of Applied Linguistics and Language Research*, 2 (3), 39—47
- Gilakjani, A.P. (2012). The Significant Role of Multimedia in Motivating EFL Learners’ Interest in English Language Learning. *I.J. Modern Education and Computer Science*, 4, 57—66
- Kalyuga, M., Mantai, L., & Marrone, M. (2013). Efficient Vocabulary Learning Through Online Activities. *Procedia Social and Behavioral Sciences*, 83 35—38
- Klentien, U. & Kamnungwet, W. (2015). The Impact of Using Electronic Media in English Teaching for Elementary and Secondary Students in Thailand. *International Journal of Information and Education Technology*, 5 (8), 582—586
- Maksudin. (2006). Media Pembelajaran Bahasa Arab. *Al-Arabiyah*, 2(2), 15—28
- Muhson, A. (2010). Pengembangan Media Pembelajaran Berbasis Teknologi Informasi. *Jurnal Pendidikan Akuntansi Indonesia*, 8 (2), 1—10
- Nalliveettil, G.M. & Alenazi, T.H.K. (2016). The Impact of Mobile Phones on English Language Learning: Perceptions of EFL Undergraduates. *Journal of Language Teaching and Research*, 7 (2), 264—272
- Naz, A.A. & Akbar, R.A. (2012). Use of Media for Effective Instruction its Importance: Some Consideration. *Journal of Elementary Education*, 18 (1-2), 35—40
- Nurgiyantoro, B. (2011). *Penilaian Pembelajaran Bahasa Berbasis Kompetensi*. Yogyakarta : BPFE
- Oroujlou, N. (2012). The Importance of Media in Foreign Language Learning. *Procedia Social and Behavioral Sciences*, 51, 24—28

- Rahman, R. A. & Tresnawati, D. (2016). Pengembangan *Game* Edukasi Pengenalan Nama Hewan dan Habitatnya dalam 3 Bahasa sebagai Media Pembelajaran Berbasis Multimedia. *Jurnal STT-Garut*, 13 (1), 184-190, diambil dari <http://jurnal.sttgarut.ac.id>
- Rao, B.M. (2014). Use of Media as an Instructional Tool in English Language Teaching (ELT) at Undergraduate Level. *International Journal of English and Literature*, 5 (6), 141—143
- Sakat, A.A., dkk. (2012). Educational Technology Media: Method in Teaching and Learning Progress. *American Journal of Applied Sciences*, 9 (6), 874—878
- Swiatek, A. (2013). The Impact of Online Media on Second Language Acquisition by Polish Students. *US-China Foreign Language*, 11 (10), 1—7
- Tavakoli, M. & Esmae'li, S. (2013). The effect of Using Print Media on Children's L2 Literacy Development: A Longitudinal Study. *Journal of Language Teaching and Research*, 4 (3), 570—578
- Trianto. (2007). *Model Pembelajaran Terpadu dalam Teori dan Praktek*. Jakarta: Prestasi Pustaka

**PEMANFAATAN NATURAL LANGUAGE PROCESSING (NLP) UNTUK
PERANCANGAN TESAURUS DIGITAL**
*(NATURAL LANGUAGE PROCESSING [NLP] UTILIZATION FOR DIGITAL
THESAURUS DESIGNING)*

Tri Wahyu Retno Ningsih, Debyo Saptono
Universitas Gunadarma
twahyurn@gmail.com, debyosap@gmail.com

Abstrak

Penelitian kebahasaan berbasis teknologi informasi semakin berkembang. Perkembangan ini berdampak positif, seperti banyaknya implementasi penelitian dalam bidang kajian natural language (bahasa alami) dan *speech recognition*. Bidang kajian tersebut memadukan antara bidang kebahasaan dan teknologi informasi. Penelitian ini memfokuskan pada kajian bahasa alami atau *natural language* yang seringkali dikenal dengan istilah *Natural Language Processing* (NLP), yaitu merancang model komputasi bahasa atau tesaurus digital yang memungkinkan terjadinya interaksi antara manusia dan komputer menggunakan perantara bahasa alami. Perancangan tesaurus digital didasarkan atas fitur-fitur yang terdapat pada tesaurus konvensional. Perancangan tesaurus digital menggunakan bahasa pemrograman *Python* yang terdapat pada *framework Tensorflow*. Beberapa tahap yang dilakukan dalam perancangan tesaurus digital adalah tahap perancangan, pengumpulan data korpus, pengecekan ejaan, pengolahan data, dan pemberian makna. Tesaurus digital yang dihasilkan diharapkan dapat ini memberikan aksesibilitas kepada pengguna karena fitur-fitur dalam tesaurus konvensional dapat ditemukan pada tesaurus digital dalam bentuk yang lebih modern.

Kata kunci: perancangan, tesaurus digital, tesaurus konvensional, *natural language processing* (NLP), Python

Abstract

Language research based on information technology is growing. This growing has a positive impact, such as the number of research implementations in the field of natural language studies and speech recognition. The field of the study combines the field of language and information technology. This study focuses on natural language studies or often referred to as Natural Language Processing (NLP), which is designing a model of language computation or digital thesaurus that allows interaction between humans and computers using natural language intermediaries. The design of digital thesaurus is based on features found in conventional thesaurus. The design of digital thesaurus is using Python programming language which contained in framework Tensorflow. Some of the steps done in digital

thesaurus design are designing phase, corpus data collection, spell checking, data processing, and giving meaning. The resulting digital thesaurus is expected to provide accessibility to the user because the features in the conventional thesaurus can be found on the digital thesaurus in a more modern form.

Keywords: *design, digital thesaurus, conventional thesaurus, natural language processing (NLP), Python*

I. PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan bidang elektronika, komputasi, dan informatika, dan berbagai sistem komputer telah menghasilkan sistem otomatisasi di berbagai bidang kehidupan. Perkembangan sistem otomatisasi komputer ini telah dimanfaatkan oleh ahli-ahli teknologi informasi yang menggunakan bahasa sebagai objek penelitiannya. Tak terkecuali, juga pemanfaatan sistem komputerisasi oleh para ahli bahasa yang tengah merancang perencanaan bahasa. Perencanaan bahasa merupakan sejenis *blueprint* program jangka panjang disertai rincian tahapannya yang berorientasi ke masa depan.

Di tengah perkembangan sistem otomatisasi komputer tersebut, keberadaan perencanaan bahasa yang komprehensif terhadap bahasa Indonesia bersifat mutlak. Salah satu pengembangan bahasa yang memanfaatkan sistem komputer adalah perancangan tesaurus bahasa Indonesia berbasis sistem digital. Sejalan dengan pengembangan bahasa, Alwi (2000) menjelaskan fungsi bahasa Indonesia. Ada lima hal yang harus diperhatikan jika akan melaksanakan pengembangan bahasa Indonesia:

1. Bahasa Indonesia harus dikembangkan agar penyelenggaraan kehidupan negara dan pemerintahan dapat berjalan dengan baik.
2. Menjadi bahasa pengantar yang efektif pada semua jenis jenjang pendidikan
3. Menjadi bahasa perhubungan nasional terutama dalam bidang perencanaan dan pelaksanaan pembangunan nasional.
4. Menjadi sarana pengembangan ilmu pengetahuan dan teknologi, dan
5. Menjadi sarana pengembangan kebudayaan

Berdasarkan hal-hal tersebut di atas, pengembangan bahasa Indonesia harus direncanakan sejalan dengan fungsi bahasa Indonesia sebagai bahasa nasional dan bahasa negara karena pengembangan bahasa Indonesia bukanlah pekerjaan yang mudah. Seiring dengan perkembangan komputasi, pengguna dapat mengembangkan berbagai sistem otomatis untuk keperluan analisis bahasa. Sistem otomatisasi komputer yang berobjek material bahasa tersebut dapat dimanfaatkan oleh pengguna (user) untuk menyelesaikan banyak pekerjaan, seperti mencari informasi, melakukan penilaian, dan perkiraan untuk merencanakan atau melakukan analisis bahasa menggunakan komputerisasi.

Kemampuan berbahasa pada komputer tersebut tentu saja diharapkan dapat meningkatkan kualitas komunikasi antara manusia-mesin atau yang sering disebut sebagai *human-machine interfacing*. Agar sistem komputer dapat berkomunikasi dengan manusia, diperlukan landasan teori komputasi yang digunakan manusia untuk mengenali karakteristik suatu bahasa alami. Kaidah tata bahasa yang digunakan manusia untuk membentuk suatu

kalimat yang benar dan memiliki pengertian merupakan suatu bukti bahwa diperlukan landasan teori tertentu. Hal ini lah yang mendorong lahirnya bidang linguistik komputasional (*computational linguistics*), yaitu bidang yang menggabungkan teori linguistik dan informasi dengan teknik komputasi.

Dalam bidang kajian linguistik komputasi, komputer merupakan salah satu alat bantu untuk melakukan data analisis bahasa. Salah satu pemanfaatan komputer untuk menganalisis kebahasaan disebut pemrosesan bahasa alami. Pemrosesan bahasa alami atau disebut *Natural Language Processing* (NLP) bertujuan mengenal secara komputasional berbagai bahasa alami (manusia) sehingga suatu komputer dapat memiliki kemampuan mengerti bahasa manusia. Berkaitan dengan hal tersebut, beberapa ahli NLP merumuskan definisi NLP sebagai serangkaian teknik komputasi yang secara teoretis berfungsi untuk menganalisis teks-teks alami yang didasarkan atas analisis linguistik. Tujuannya adalah mencapai pemrosesan bahasa yang menyerupai manusia untuk menyelesaikan berbagai tugas menggunakan komputer.

Natural Language Processing (NLP) merupakan salah satu sarana untuk menyelesaikan tugas tertentu yang berbasis komputasi, seperti sistem *Information Retrieval* (IR) dan Terjemahan Mesin (MT). Sistem *Information Retrieval* (IR) dan Terjemahan Mesin (MT) berbasis NLP memiliki tujuan untuk menyediakan informasi yang lebih tepat dan lengkap sesuai dengan kebutuhan informasi terhadap pengguna (user). Tujuan sistem NLP di sini adalah untuk merepresentasikan makna dan maksud yang sebenarnya dari permintaan pengguna (user), yang dapat dinyatakan secara alami dalam bahasa sehari-hari.

Beberapa jenis aplikasi dapat dibuat dengan memanfaatkan NLP, seperti *Natural Language text-based application* dan *dialogue-based application*. *Text-based application* mencakup segala macam aplikasi yang melakukan proses terhadap teks tertulis, contohnya mencari topik tertentu pada genre tertentu. Selain itu, NLP dapat berfungsi untuk melakukan analisis bahasa pada level tertentu yang melibatkan aspek-aspek dalam keilmuan linguistik, seperti fonologi dan fonetik, morfologi, sintaksis, dan semantik. Salah satu contoh fungsi NLP adalah untuk menentukan sistem kelas kata dalam gramatika. Jika kita akan mendeskripsikan tata bahasa secara memadai maka klasifikasi kata dalam linguistik modern dianggap sebagai salah satu aspek tata bahasa yang sejajar dengan aspek-aspek lain dan harus mendapat perlakuan yang seimbang (Kridalaksana, 2005: 5).

Menurut Kridalaksana (2005), secara keseluruhan tata bahasa atau gramatika mempunyai komponen-komponen berikut:

1. Struktur gramatikal yang memperlihatkan bagaimana bangun gramatika suatu bahasa sehingga kita dapat melihat konstruksi dan konstituensi dari unsur-unsur gramatikal yang berasal dari leksem, di samping hubungan sintagmatis dan paradigmatis diantaranya.
2. Sistem gramatikal yang memperlihatkan bagaimana unsur-unsur gramatikal berperilaku sebagai satuan yang terorganisir menjadi satu hierarki dari yang terkecil, yaitu morfem sampai dengan wacana.
3. Peran gramatikal yang memperlihatkan bagaimana gramatika menjadi ungkapan dari konfigurasi semantik yang mengkombinasikan konsep-konsep sehingga bahasa menjadi alat komunikasi yang bermakna.

Penjelasan Kridalaksana mengenai satuan gramatika tersebut, digunakan oleh peneliti sebagai acuan untuk menentukan token atau unsur terkecil dari perancangan tesaurus digital.

Hal ini penting diperhatikan, karena perancangan suatu aplikasi atau sistem komputer memerlukan kecermatan dalam metode dan bersesuaian dengan teori yang telah dijabarkan oleh ahli-ahli bahasa. Pendapat Kridalaksana (2005) mengenai satuan gramatikal, juga dijelaskan oleh Lyons (1996) yang membagi kategori gramatikal dalam kategori primer, yakni kelas kata dan kategori sekunder, yang terdiri atas modus, kala, aspek, diatesis, jumlah, dan kasus (Lyons, 1996). Dalam penelitian ini, peneliti membatasi objek penelitian, yaitu salah satu ujud gramatikal berupa kata.

Penelitian kebahasaan yang mengambil objek penelitian bahasa, seperti pemrosesan kata dapat memanfaatkan NLP untuk merancang aplikasi atau sistem komputer. Cara kerja NLP yang menyerupai pola berpikir manusia, relatif membantu dan memudahkan peneliti untuk merancang aplikasi atau sistem komputer yang menggunakan aspek abahasa sebagai objeknya. Contohnya, beberapa jenis pemrosesan bahasa diketahui bekerja, ketika manusia memproduksi atau memahami bahasa. Diperkirakan bahwa manusia menggunakan semua tingkatan bahasa (fonologi, morfologi, sintaksis, dan semantik) dalam memproduksi atau memahami bahasa karena setiap tingkat menyampaikan berbagai jenis makna. Pola berpikir manusia tersebut diadopsi oleh NLP untuk mencapai pemrosesan bahasa yang menyerupai bahasa manusia, salah satunya merancang tesaurus digital. Beberapa contoh kegiatan yang memanfaatkan NLP, antara lain, (1) mengungkapkan data entri (kata), (2) menerjemahkan kata ke kata bahasa lain, (3) menjawab pertanyaan tentang padanan kata, dan (4) membuat kesimpulan tentang makna kata.

Tesaurus secara etimologi didefinisikan ‘khazanah’ yang berasal dari kata ‘*thesauros*’ (bahasa Yunani). Tesaurus berisi seperangkat kata yang saling bertalian maknanya (Tesaurus Indonesia, 2008). Tesaurus pada umumnya disusun berdasarkan gagasan atau tema sehingga tidak hanya berisi entry-entry mengenai persamaan kata, tetapi ditemukan pula antonim atau lawan katanya. Selain itu, dalam tesaurus juga ditemukan akronim yang lazim digunakan, seperti *a* (*adjektiva*), *adv* (*adverbial*), *ant* (*antonym*), *cak* (*cakapan*), *n* (*nomina*), *num* (*numeralia*), *p* (*partikel*), *pron* (*pronominal*), dan *v* (*verba*). Dalam hal ini tesaurus juga berperan dalam temu kembali informasi pada proses penelusuran informasi melalui sistem komputerisasi. Contohnya, ketika pengguna (user) ingin menemukan kata ‘menggunakan’, maka akan diperoleh indeks-indeks pencarian bertema ‘menggunakan’.

Berdasarkan kajian di atas, penelitian ini bertujuan untuk menyampaikan konsep perancangan tesaurus digital dalam domain sinonim dan similaritas kata dengan memanfaatkan NLP. Penelitian ini juga mempertimbangkan basis data yang memadai berupa koleksi data kebahasaan yang disimpan secara elektronik sebagai karakter alfanumerik dalam *files*. Sampel yang digunakan melibatkan kumpulan kata dalam jumlah besar dan diharapkan semua entri yang ada dalam tesaurus itu lengkap dan proporsional.

II. TINJAUAN PUSTAKA

2.1. Pemrosesan Bahasa Alami (*Natural Language Processing*)

Teori linguistik yang berorientasi pada mekanisme logika komputer telah berkembang sejak tahun 1940-an dan mengalami percepatan dengan adanya teori Tata Bahasa Transformasional (Chomsky, 1967). Dalam hal ini, komputer tidak hanya digunakan sebagai

sarana utama banyak pekerjaan, tetapi menjadi infrastruktur bagi interaksi antara manusia. Komputer berperan dalam banyak bidang. Walaupun komputer telah diadaptasi sesuai dengan sifat dan kebutuhan manusia tetapi pada hakikatnya komputer tetap menjadi alat yang terus menerus dipengaruhi oleh lingkungannya (Mahfudz et al., 2013). Salah satunya adalah domain *Natural Language Processing* (NLP)

Natural Language Processing (NLP) merupakan satu pendekatan komputerisasi untuk menganalisis teks yang didasarkan pada serangkaian teori dan seperangkat teknologi. NLP seringkali dipilih oleh ahli bahasa untuk menyelesaikan beberapa pekerjaan yang memerlukan kombinasi antara linguistik dan komputasi, meskipun ada beberapa metode atau teknik lain yang dapat dipilih untuk mencapai jenis analisis bahasa tertentu.

Model komputasi NLP dapat digunakan untuk kepentingan ilmiah misalnya meneliti sifat-sifat dari suatu bentuk bahasa alami yang memudahkan komunikasi antara manusia dan komputer. Dalam beberapa literature dijelaskan bahwa pengolahan bahasa alami atau NLP mengenal beberapa tingkat pengolahan, yaitu (1) fonetik dan fonologi yang berkaitan dengan proses aplikasi menggunakan metode *speech based system*, (2) morfologi, yaitu aplikasi untuk memisahkan aspek satuan gramatikal seperti kata dasar dan kata bentukan, (3) semantik berhubungan dengan makna kata dan bagaimana makna suatu kata membentuk makna suatu kalimat. Contohnya, *semantic analyzer* yang mempunyai himpunan rule dalam basis pengetahuan untuk menginterpretasikan sebuah kalimat.

Hutchins (1993) memberikan ulasannya mengenai fungsi komputer untuk analisis bahasa. Salah satu ulasannya adalah sejarah pentingnya mesin penerjemah (*machine translation*). Selama beberapa tahun pertama muncul *electronic calculator*, dan dilanjutkan oleh penelitian-penelitian lain tentang penggunaan komputer sebagai alat bantu penerjemah bahasa. Pada tahun 1954, kolaborasi antara IBM dan George Town University menghasilkan suatu mesin pertama yang bisa disebut sebagai mesin penerjemah (*machine translation*). Sistem-sistem awal yang dihasilkan berkenaan dengan mesin penerjemah secara garis besar terdiri atas kamus bilingual yang entrinya adalah kata –kata dari *Source of Language* (SL) apabila diberikan suatu masukan kata-kata dari *Target of Language* (TL). Sistem yang seperti itu dianggap terlalu rumit dan hanya dapat digunakan untuk kata-kata dan pola kalimat tertentu. Oleh karena itu, diperlukan adanya suatu metode yang lebih sistematis untuk mengefektifkan kinerja mesin penerjemah. Berbagai upaya penelitian telah dilakukan misalnya ide dari Noam Chomsky mengenai *syntactic transformation* (Hutchins, 1993).

Dalam penelitiannya, Fiederer dan O'Brien (2009) memberikan argumentasinya mengenai mesin penerjemah. Salah satu alasan untuk dibangunnya aplikasi atau sistem komputer untuk mesin penerjemah adalah bertambahnya permintaan penerjemah. Alasan lainnya adalah kelangkaan penerjemah profesional yang berkualitas, keinginan untuk menembus pasar baru, dan adanya persyaratan publikasi dalam beberapa bahasa selain bahasa asal. Ketika mesin penerjemah digunakan untuk keperluan publikasi, beberapa tingkat *postediting* (pasca-editing) masih diperlukan.

Metode lain untuk meningkatkan kualitas luaran dari mesin penerjemah adalah *controlled language* (CL). Pengaplikasian CL pada sumber masukan bertujuan untuk mengurangi kerumitan dan mengurangi ambiguitas sumber masukan dari mesin penerjemah. Fiederer dan O'Brien (2009) meneliti perbandingan antara kualitas terjemahan hasil mesin penerjemah yang telah melewati proses pasca-editing dengan hasil terjemahan manusia. Hasil

penelitiannya menjelaskan bahwa hasil terjemahan manusia lebih diminati. Hasil penelitian Fiederer dan O'Brien (2009) menunjukkan bahwa ketika kualitas digunakan sebagai parameter kejelasan dan akurasi (*clarity* dan *accuration*) maka hasil terjemahan yang menggunakan mesin penerjemah dan telah melalui proses pasca-editing sama baiknya dengan hasil terjemahan manusia.

Proses pasca-editing atau pemeriksaan kualitas dari hasil terjemahan mesin penerjemah oleh manusia memerlukan biaya tinggi. Selain itu, proses pasca-editing juga memerlukan waktu lama. Untuk menjawab permasalahan tersebut, Papineni et al. (2002) mengembangkan suatu metode evaluasi mesin penerjemah otomatis yang cepat, tidak berbiaya mahal, serta digunakan pada berbagai bahasa. Papineni (2002) mengembangkan suatu temuan baru, yaitu *blue metric* yang merupakan *numerical metric* untuk mengukur *translation closeness* (kedekatan terjemahan) dari hasil terjemahan oleh mesin penerjemah.

Berkenaan dengan pemeriksaan kualitas hasil terjemahan mesin terjemahan, Doddington (2002) memberikan alternatif lain mengenai evaluasi otomatis dari kualitas mesin penerjemah dengan menggunakan *N-gram Co-Occurrence Statistics* yang membandingkan keluaran dari mesin penerjemah dengan terjemahannya yang dijadikan referensi dalam domain statistik dari *word N-grams*.

2.2 Konsep Tesaurus

Tesaurus menurut kamus bahasa Inggris-Indonesia (Echols, 2003: 586) berarti kamus, ensiklopedia atau buku lain yang berisi keterangan. Sementara itu, Longman (2001) memberikan pengertian tesaurus sebagai buku yang mendaftar kata-kata secara berkelompok dengan kata-kata yang memiliki kesamaan pengertian. Kamus Besar bahasa Indonesia (Departemen Pendidikan Nasional, 2008) menjelaskan bahwa tesaurus adalah (1) buku referensi berupa daftar kata dengan sinonimnya, (2) buku referensi berupa informasi tentang berbagai konsep atau istilah di pelbagai kehidupan dan pengetahuan.

Kennedy (2000) menjelaskan pengertian tesaurus adalah daftar istilah yang digunakan dan istilah yang tidak digunakan, serta istilah yang terkait yang tersusun secara alfabetis. Dalam tesaurus arti dari istilah ditentukan dan hubungan hirarkies dengan istilah lain dapat dilihat. Tesaurus merupakan suatu kosa kata terkontrol yang diatur secara terurut, yakni relasi ekivalensi, hirarkies, dan asosiasi antar term ditampilkan secara eksplisit dan diidentifikasi dengan indikator relasi standar (NISO, 1999). Berdasarkan ISO 278 tentang *Documentation-Guidelines for the Establishment and Development of Monolingual Thesauri* (1986), tesaurus merupakan kosa kata suatu bahasa indeksasi terkontrol, yang secara formal diatur sedemikian rupa sehingga ada suatu hubungan yang terencana antar konsep yang dibuat secara jelas dan tegas.

Tesaurus memiliki fungsi berdasarkan fungsionalitasnya, yaitu:

1. Tesaurus *stand-alone* yang berfungsi membantu pengguna memahami arti suatu term beserta relasinya, membantu mencari variasi term dan mencari translasi term.
2. Tesaurus terintegrasi memiliki dua fungsi utama, yaitu pengambilan informasi dan pengideksan.

Tesaurus memiliki struktur yang berbentuk pohon, yang terdiri atas beberapa elemen, yaitu (1) term formal (*description*), (2) term non-formal (*non-description*), (3) deskripsi

term, yaitu definisi dan *scope note*, (4) relasi antar term, (5) relasi lain yang dapat ditambahkan sendiri

III. KERANGKA TEORI DAN METODE PENELITIAN

Bahasa pemrograman *Python* menjadi dasar untuk perancangan tesaurus digital. Bahasa ini tidak terlalu populer, jika dibandingkan dengan bahasa pemrograman C atau Java, tetapi *Python* mempunyai beberapa keunggulan. *Python* merupakan bahasa pemrograman yang tidak menggunakan *compiler*. Dan memiliki sifat *open-source* dan dapat digunakan untuk melakukan berbagai hal dari mulai mengembangkan *web*, mengembangkan *video game*, membangun GUI *desktop*, maupun mengembangkan perangkat lunak.

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. *Python* juga didukung oleh komunitas yang besar. *Python* mendukung multi paradigma pemrograman, utamanya, namun tidak dibatasi, pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Salah satu fitur yang tersedia pada *Python* adalah sebagai bahasa pemrograman dinamis yang dilengkapi dengan manajemen memori otomatis. Seperti halnya pada bahasa pemrograman dinamis lainnya.

Python umumnya digunakan sebagai bahasa skrip meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa skrip. *Python* dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform system operasi. Kode *Python* dirancang agar mudah dipelajari dan didesain ulang. Bahasa pemrograman ini juga mendukung pemrograman berorientasi objek dan pemrograman fungsional. Selain itu, program yang ditulis menggunakan *Python* dapat dijalankan hampir semua sistem operasi (Unix, Windows, Mac OS, dan lain-lain, termasuk untuk perangkat mobile. Melalui mekanisme tertentu, kode *Python* dapat diintegrasikan dengan aplikasi yang ditulis dalam bahasa pemrograman lain.

Natural Language Toolkit (NLTK) merupakan salah satu platform yang digunakan untuk mengelola dan mengolah bahasa manusia, NLTK digunakan untuk pemrosesan teks. fasilitas-fasilitas yang ada dalam NLTK antara lain klasifikasi, tokenization, stemming, tagging, parsing, dan penalaran semantik. NLTK menggunakan bahasa pemrograman python, dan digunakan sebagai tools yang bagus untuk pengembangan aplikasi NLP (*Natural Language Processing*). Contohnya, untuk mendapatkan similarity dua kata menggunakan perintah :

```
w1 = wordnet.synset('ship.n.01')
w2 = wordnet.synset('boat.n.01')
print(w1.wup_similarity(w2))
```

Untuk menjalankan fungsi *stemming* pada kata “working” menggunakan perintah :

```
from nltk.stem import PorterStemmer
stemmer = PorterStemmer()
print(stemmer.stem('working'))
```

Untuk mendapatkan sinonim dari kata yang ditentukan ('computer') pada *database* WORDNET menggunakan perintah :

```
from nltk.corpus import wordnet
synonyms = []
for syn in wordnet.synsets('Computer'):
    for lemma in syn.lemmas():
        synonyms.append(lemma.name())
print(synonyms)
```

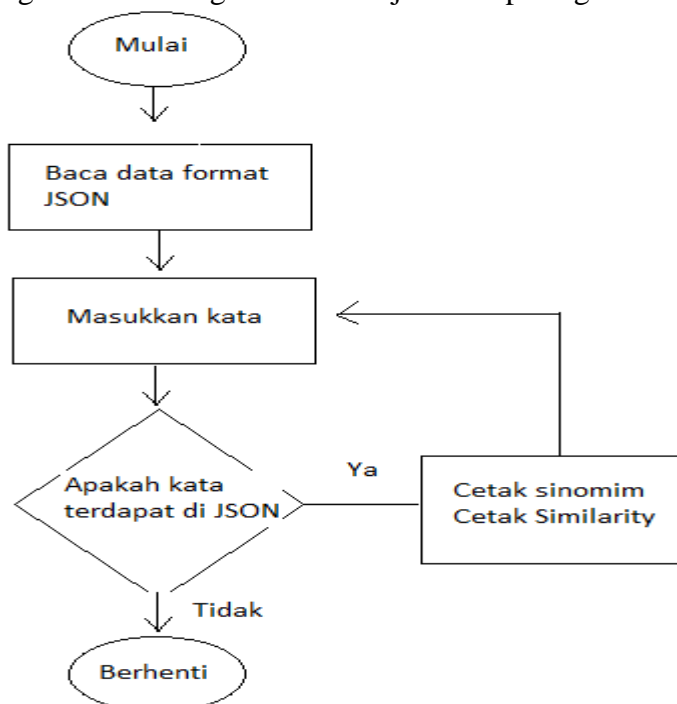
Untuk mendapatkan daftar kata yang tidak penting pada bahasa Inggris pada *database* WORDNET menggunakan perintah :

```
from nltk.corpus import stopwords
stopwords.words('english')
```

TensorFlow adalah sebuah perpustakaan software sumber terbuka yang digunakan untuk mesin belajar dalam berbagai macam tugas pemahaman persepsi dan bahasa. *TensorFlow* adalah API generasi kedua, menggantikan DistBelief (API generasi pertama) yang digunakan oleh 50 tim yang berbeda untuk penelitian dan pengembangan puluhan produk komersial milik Google, seperti *Speech Recognition*, *Gmail*, *Google Image*, dan *Google Search*. *TensorFlow* pada awalnya dikembangkan oleh Brain Google Team untuk tujuan penelitian dan produksi Google, kemudian dirilis dengan sumber terbuka di bawah lisensi Apache 2.0 pada 9 November 2015. TensorFlow berjalan di platform Linux, Windows, dan Mac OS X.

IV. ANALISIS PENELITIAN

Proses perancangan tesaurus digital offline dijelaskan pada gambar 1.



Gambar 1. Cara kerja Tesaurus Digital Offline

Gambar 1 menjelaskan tentang cara kerja perancangan tesaurus digital offline. Langkah-langkah yang dilakukan adalah membaca data format *JavaScript Object Notation* (JSON). JSON adalah suatu format ringkas pertukaran data komputer, mudah dibaca serta mudah diterjemahkan dan dibuat (*generate*) oleh komputer. JSON digunakan untuk merepresentasikan struktur data sederhana dan array. JSON terbuat dari dua struktur, yaitu kumpulan pasangan nama atau nilai (objek, rekaman, struktur, kamus) dan daftar nilai terurutkan (*an ordered list of values*). Pengguna melakukan entri ‘kata’ dan dibaca oleh JSON, jika ditemukan kata yang dimaksud, maka akan muncul sinonim kata dan tingkat similaritas kata tersebut. Perhitungan tingkat similaritas kata menggunakan library NLTK yang diimport ke dalam *Python*. Jika tidak ditemukan kata yang dimaksud proses pencarian kata akan berhenti. Contoh notasi pada JSON, dijelaskan sebagai berikut :

```
"menggunakan": {
  "tag": "v",
  "sinonim": [
    "memakai",
    "memanfaatkan",
    "memerlukan",
    "menghabiskan",
    "mengonsumsi"
  ]
}
```

memanggil library JSON dengan perintah :

```
import json
```

memanggil file JSON menggunakan perintah :

```
mydict = load('dict.json')
```

membaca data dari file JSON menggunakan perintah berikut :

```
def load(filename):
```

```
    with open(filename) as data_file:
```

```
        data = json.load(data_file)
```

```
    return data
```

Untuk mendapatkan sinonim dari kata yang ditentukan (*word*) menggunakan perintah :

```
def getSinonim(word):
```

```
    if word in mydict.keys():
```

```
        return mydict[word]['sinonim']
```

```
    else:
```

```
        return []
```

Berikut ini contoh keluaran dari program, berupa daftar sinonim dari kata yang dipilih (kata ‘**menggunakan**’) beserta tingkat similaritasnya. Melalui proses analisis menggunakan memanfaatkan *Natural Language Processing* (NLP) yang terdiri atas *library NLTK* dan *framework TensorFlow* ditemukan daftar sinonim dan similaritas kata ‘menggunakan’, yang dijelaskan pada tabel 1.

Tabel 1. Contoh Keluaran Program

Kata	Tingkat similaritas	Kelas Kata
memakai	0.008108136721642399	verba
memanfaatkan	0.006617663264965374	verba
memerlukan	0.004505890905284084	verba
menghabiskan	0.00428286400119506	verba
mengkonsumsi	0.003952770315791044	verba

Berdasarkan tabel keluaran program dapat dijelaskan bahwa similaritas kata menggunakan adalah memakai, memanfaatkan, memerlukan, menghabiskan, dan memerlukan. Tingkat similaritas tertinggi terdapat pada kata memakai.

IV. SIMPULAN

Penelitian ini telah memanfaatkan *Natural Language Processing* (NLP) yang terdiri atas *library NLTK* dan *framework TensorFlow* dan telah berjalan dengan baik, tetapi masih dalam tahap pengembangan terutama dalam perhitungan graf similaritas. Keakuratan similaritas perlu ditingkatkan karena sangat bergantung dari banyaknya data di JSON. Oleh karena itu, untuk melakukan pengujian tingkat similaritas yang lebih akurat diperlukan data dalam jumlah besar atau menambah data di JSON.

Referensi

- Alwi, Hasan, dkk. (2000). Bahasa Indonesia: pemakai dan pemakainnya. Jakarta. Badan Pengembangan dan Pembinaan Bahasa. Kementerian Pendidikan dan Kebudayaan.
- Departemen Pendidikan Nasional. (2008). *Kamus Besar Bahasa Indonesia*. Edisi ke-4. Jakarta: PT. Gramedia Pustaka Utama.
- Doddington, George. (2002). Automatic evaluation of machine translation quality using N-gram Co Occurrence Statistics. www ldc.upenn.edu/Projects/TIDES/Translation
- Echols., John & Shadily Hasan. 2003. *Kamus Inggris-Indonesia*. Jakarta: PT. Gramedia Pustaka Utama.
- Fiederer, Rebecca and Sharon O' Brien (2009). Quality and machine translation: a realistic objective?" *The Journal of Specialised Translation* 11, 52-74.
- Hutchins, John. (1993). Latest developments in machine translation technology: beginning a new era in MT research. *In: MT Summit (1993)*, 11-13.
- Hutchins, John. (1999). The development and use machine translation systems and computer-based translation tools. *International Symposium on Machine Translation and Computer Language Information Processing, 26-28 June 1999, Beijing China*.
- Kennedy, Jay, dan Cherryl Schauder. (2000). *Records Management: A Guide for Students a Practioners of Records Management*. Melbourne: Addison, Wesley, Iongman.
- Kishore Papineni, Salim Rouks, Todd Ward, John Henderson, and Florence Reeder. (2002). Corpus-based comprehensive and diagnostic MT evaluation: Initial Arabic, Chinese, French, and Spanish results. *In Proceedings of Human Language Technology. San Diego, CA. To appear*.

- Kridalaksana, H. (2005). *Kelas Kata dalam Bahasa Indonesia*. Jakarta. Gramedia Pustaka Utama.
- Lyons, John (1996). Toward a 'notional' theory of the parts of speech. *Journal of Linguistics*. 2: 209-36.
- Mahfudx, S. Nasution, M.K. dan Nasution, S. *Knowledge Sharing a Model*. Cornell University Library, 2013.
- NISO Standard z39,19. (1999). Documentation Guidelines for the establishment and development of monolingual thesauri. Available form:
<http://www.chin.gc.ca/English/Standards/vocabulary/thesaurus.html>.

RANCANG BANGUN KAMUS HTML 5 BERBASIS WEB (ARCHITECTURE WEB-BASED DICTIONARY of HTML 5)

Ekohariadi ¹⁾, Dodik Arwin Dermawan ²⁾, Setya Chendra Wibawa ³⁾

¹⁾²⁾³⁾Jurusan Teknik Informatika Universitas Negeri Surabaya
e-mail : ekohariadi@unesa.ac.id¹⁾, dodikdermawan@unesa.ac.id²⁾,
setyachendra@unesa.ac.id³⁾

Abstrak

Perkembangan system informasi sebagai aplikasi yang dapat membantu masyarakat dengan format berupa text, gambar dan video mengalami perkembangan sangat pesat, khususnya system informasi berbasis internet. Seiring dengan perkembangan tersebut, maka penelitian ini bertujuan untuk membangun sebuah kamus berkaitan dengan istilah – istilah yang terdapat di dalam HTML5, sehingga dapat membantu programmer pemula dalam memahami HTML5. *Hyper Text Markup Language* atau lebih dikenal dengan HTML, merupakan bahasa markup yang banyak dimanfaatkan dalam proses pembuatan web yang dapat menampilkan text, gambar dan video serta dapat ditambahkan tautan – tautan hingga dapat menampilkan informasi yang interaktif sedangkan HTML 5 merupakan pengembangan dari versi HTML yang memiliki berbagai kelebihan. Kamus HTML5 dibangun menggunakan metode waterfall sehingga dalam proses pembuatannya lebih terstruktur dan menghasilkan sebuah aplikasi kamus yang mudah digunakan khususnya saat proses pencarian istilah – istilah yang terdapat di dalam HTML5.

Abstract

The development of the information system as an application that can help people with the format in the form of text, image and video had developed very rapidly particularly internet-based information system. Along with these developments, this research aims to build a dictionary related to the terms contained in HTML5, so it can help novice programmers in understanding HTML5. Hyper Text Markup Language or better known as HTML, is a markup language that is widely utilized in the process of creating a web that can display text, images and video and can be added links - links to display interactive information, While HTML 5 is the development of the HTML version which has many advantages. HTML5 Dictionary is built using waterfall method so that in the process of making more structured and produce an application dictionary is easy to use especially when the process of the search terms that contained HTML5.

I. PENDAHULUAN

Hyper Text Modelling Language atau yang lebih dikenal dengan HTML merupakan bahasa markup yang umumnya dipakai oleh siswa atau mahasiswa yang baru pertama kali mempelajari tentang pembuatan halaman web [1]. Kesulitan yang sering dihadapi oleh para pemula ketika belajar membuat halaman web menggunakan HTML adalah banyaknya istilah yang terdapat di dalam HTML. Hal tersebut semakin kompleks dengan berkembangnya HTML menjadi HTML 5, sehingga semakin bertambah pula istilah – istilah yang harus dipahami oleh programmer yang baru belajar membuat halaman web.

HTML dibangun sekitar tahun 1990 oleh Tim Bernes-Lee seorang peneliti di Conseil Europeen pour Recherche Nucleaire (CERN) laboratory di Geneva [2]. HTML merupakan bahasa markah atau dalam bahasa Inggris disebut sebagai *markup language* yang digunakan untuk menafsirkan, menulis dan menampilkan teks, gambar maupun video ke dalam halaman web. Perkembangan HTML sangat pesat, pada tahun 1993 diluncurkan HTML+ yang selanjutnya di sempurnakan oleh *Internet Engineering Task Force* (IETF) pada bulan November tahun 1995 menjadi HTML 2.0 yang memiliki kemampuan untuk menampilkan suatu form pada dokumen dan hal tersebut menjadi cikal bakal web interaktif. Pada tahun yang sama menyusul diluncurkannya HTML 3.0 yang memiliki beberapa fasilitas, salah satu diantaranya yakni Figure perkembangan dari Image berfungsi untuk mengatur posisi gambar dan table [3]. HTML 3.0 juga mendukung adanya rumus matematika dalam dokumen yang digunakannya. Pada tahun 1996 W3C (World Wide Web Consortium) ikut berperan dalam pengembangan HTML hingga meluncurkan HTML 3.2 dan diperkenalkan secara resmi pada bulan Januari 1997 [4]. Kelebihan dari HTML 3.2 adalah dapat meletakkan teks di sekeliling gambar serta dapat menggabungkan script di luar script HTML, salah satu diantaranya adalah javascript. Pada bulan Desember 1997 diluncurkan HTML 4.0 dengan banyak perubahan dibandingkan HTML 3.2 khususnya terkait perintah – perintah image, meta, text, table, form dan lain – lain. Selanjutnya perintah – perintah tersebut disempurnakan lagi pada bulan April 1998 menjadi HTML 4.01 dengan kelebihan adanya fasilitas plugin video, untuk menjalankan video dengan berbagai format. Perkembangan terakhir dari HTML adalah HTML 5 yang dirilis sekitar bulan Maret 2010 dengan kelebihan diantaranya adalah *script* lebih sederhana, memiliki standard penulisan script untuk menampilkan file video ke browser tanpa plugin tambahan, mampu untuk membuat media player menggunakan video.js yang terdapat di library java dan css serta proses loading lebih cepat [5]. Perkembangan HTML yang pesat, menjadi latar belakang dibangunnya kamus pemrograman HTML 5. Sedangkan kamus sebagai alat bantu atau media rujukan untuk mencari atau menjelaskan tentang kosa kata maupun istilah, juga mengalami perkembangan yang signifikan. Beberapa peneliti telah mengembangkan kamus online, salah satu diantaranya pembuatan kamus kebidanan menggunakan android yang sangat membantu pengguna khususnya di bidang kebidanan karena lebih praktis jika dibandingkan dengan kamus dalam bentuk buku [6]. Perkembangan kamus online juga merambah ke bahasa daerah yakni kamus Minang – Inggris yang dapat membantu masyarakat dalam memahami padanan bahasa Minang ke bahasa Inggris [7].

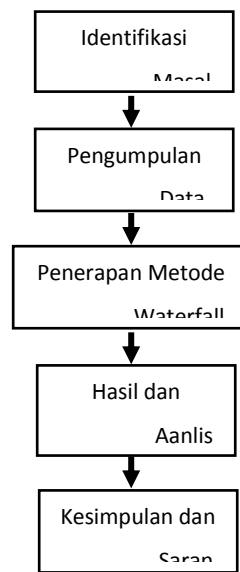
Penelitian ini lebih difokuskan kepada rancang bangun kamus HTML 5 menggunakan web, yang dilengkapi dengan tautan ke dalam beberapa aplikasi *scripting online* yang telah ada di internet. Tujuan dari tautan tersebut adalah mempermudah masyarakat, khususnya siswa atau mahasiswa yang baru pertama kali belajar membuat halaman web. Setelah memahami kata maupun istilah yang terdapat di dalam kamus HTML 5, pengguna bisa

langsung mempraktekkan melalui tautan – tautan yang telah disediakan. Di dalam kamus HTML 5 juga disediakan fasilitas berbagi ilmu yang memudahkan pengguna dalam mencari literature berupa ebook berkaitan dengan pembuatan halaman web, khususnya yang menggunakan HTML.

II. KERANGKA TEORI DAN METODOLOGI PENELITIAN

2.1. Kerangka Pikir

Pengembangan sebuah aplikasi di lakukan melalui beberap tahap, demikian pula dengan proses pengerjaan rancang bangun kamus HTML 5. Tahapan – tahapan yang dilakukan tertuang di dalam kerangka pikir sebagai berikut :



Gambar 1. Kerangka Pikir Rancang Bangun Kamus HTML 5

2.2. Identifikasi Masalah

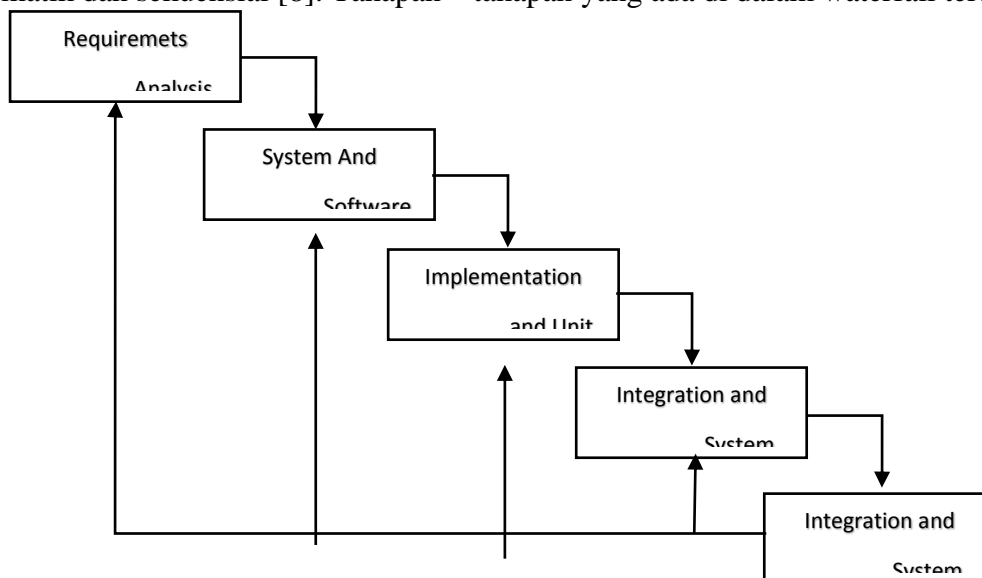
Proses awal dalam pembuatan kamus HTML 5 adalah berupa identifikasi permasalahan yang muncul saat melakukan proses perancangan. Permasalahan yang muncul berkaitan dengan bagaimana memberikan kemudahan bagi pengguna dalam memanfaatkan kamus HTML 5, sebagai rujukan dalam mempelajari, memahami dan mempraktekkan secara langsung proses pembuatan halaman web.

2.3. Pengumpulan Data

Pengumpulan data merupakan proses lanjutan setelah melalui identifikasi permasalahan. Pada kamus HTML 5 yang menjadi data utama adalah kata maupun istilah – istilah yang terdapat di dalam HTML sampai dengan HTML 5. Proses yang dilakukan untuk pengumpulan data dalam pembuatan kamus HTML 5 menggunakan beberapa sumber, diantaranya ebook, modul perkuliahan dan modul pelajaran khususnya di bidang pembuatan web. Di dalam kamus HTML 5 juga diberikan fasilitas bagi pengguna yang ingin berkontribusi secara langsung untuk menambahkan kata ataupun istilah yang berkaitan dengan HTML sampai dengan HTML 5 sehingga semakin dapat menambah perbendaharaan kata dan istilah.

2.4. Penerapan Metode Waterfall

Proses pengerjaan rancang bangun kamus HTML 5 menggunakan metode waterfall, hal tersebut tidak terlepas dari kelebihan yang dimilikinya, salah satu diantaranya adalah sistematis dan sekuensial [8]. Tahapan – tahapan yang ada di dalam waterfall terdiri dari [9] :



Gambar 2. Metode Waterfall

Requirements Analysis

Layanan dan fasilitas yang ada di dalam system kamus HTML 5 berdasarkan pada tujuan yakni kemudahan bagi pengguna dalam mempelajari pembuatan halaman web khususnya menggunakan HTML 5

System dan Software Design

Perancangan kamus HTML 5 melalui beberapa tahapan, khususnya kebutuhan perangkat keras maupun perangkat lunak. Hal tersebut dituangkan dalam bentuk desain arsitektur pembuatan system secara menyeluruh yang melibatkan identifikasi permasalahan serta tujuan awal dari perancangan.

Implementation and Unit Testing

Merupakan tahapan lanjutan setelah melakukan perancangan system secara menyeluruh yakni dengan memimplemetasikan seluruh perangkat keras dan perangkat lunak. Perangkat lunak dibangun berdasarkan fungsi dari masing – masing unit, diantaranya unit untuk melakukan pencarian kata atau istilah, unit untuk melakukan berbagi ilmu dengan fungsi bisa mengambil dokumen berektension pdf, dan unit untuk menambahkan kata atau istilah dalam kamus HTML 5. Pada masing – masing unit perangkat lunak tersebut, dilakuan pengujian dengan tujuan memastikan bahwa setiap unit berjalan sesuai fungsinya.

Integration and System Testing

Ketika seluruh unit telah selesai dilakukan uji coba, maka langkah selanjutnya adalah mengintegrasikan atau menggabungkan seluruh unit tersebut hingga menjadi sebuah system. Sistem yang telah digabungkan akan dilakukan uji coba untuk memastikan bahwa system telah berjalan sesuai dengan tujuan awal dari perancangan.

Operation and Maintenance

Tahapan ini adalah langkah terakhir, akan tetapi memerlukan waktu yang cukup panjang. Hal tersebut tidak terlepas dari berjalannya fungsi pada system itu sendiri. Jika sebuah system yang dibangun tidak mengalami kendala saat pengoperasiannya maka dapat dikatakan bahwa system berjalan normal, akan tetapi jika sebaliknya maka perlu dilakukan perbaikan pada system tersebut.

III. HASIL DAN ANALISIS PENELITIAN

3.1. Analisis Sistem

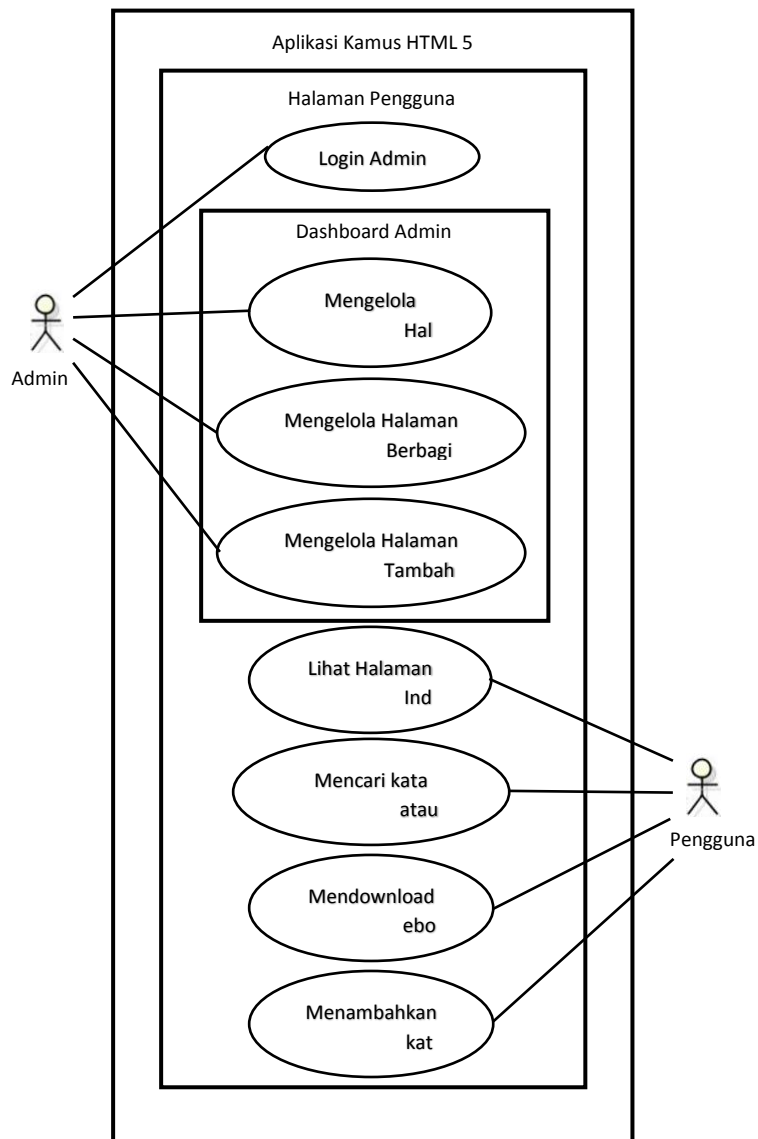
Analisis system merupakan upaya untuk mengidentifikasi dan mengevaluasi berbagai permasalahan yang muncul pada suatu system, sehingga dalam pengembangan system selanjutnya dapat diperbaiki. Analisis system untuk kamus HTML 5 dapat dijabarkan sebagai berikut :

- HTML merupakan bahasa markup yang dapat digunakan untuk membuat halaman web. Untuk membuat tampilan dari halaman web menjadi semakin menarik dan dinamis, HTML masih memerlukan script dari bahasa pemrograman yang lain, salah satunya adalah java script.
- Di dalam kamus HTML 5 terdapat fasilitas yang memberikan kemudahan untuk pengguna jika ingin mencoba secara langsung *scripting online* dengan memanfaatkan tautan pada aplikasi *scripting online* yang telah ada di internet. Hal tersebut menjadikan kamus HTML 5 bergantung pada aplikasi lain sehingga kurang optimal dikarenakan aplikasi *scripting online* bercampur dengan bahasa pemrograman yang lain.
- Sumber data yang berupa kata – kata dan istilah yang di masukkan ke dalam kamus HTML 5 tidak hanya berkaitan dengan HTML 5, akan tetapi juga berkaitan dengan istilah – istilah yang terdapat di HTML versi sebelumnya sehingga mengakibatkan kurang focus pada HTML 5.
- Kamus HTML 5 juga menyediakan fasilitas bagi pengguna yang ingin memberikan kontribusi dengan menambahkan perbendaharaan kata ataupun istilah yang berkaitan dengan HTML 5. Proses filtering dari kontribusi yang diberikan pengguna masih dilakukan secara manual oleh admin kamus HTML 5, sehingga membutuhkan waktu yang lama untuk proses filtering tersebut

3.2. Desain Sistem

- **Use Case Diagram**

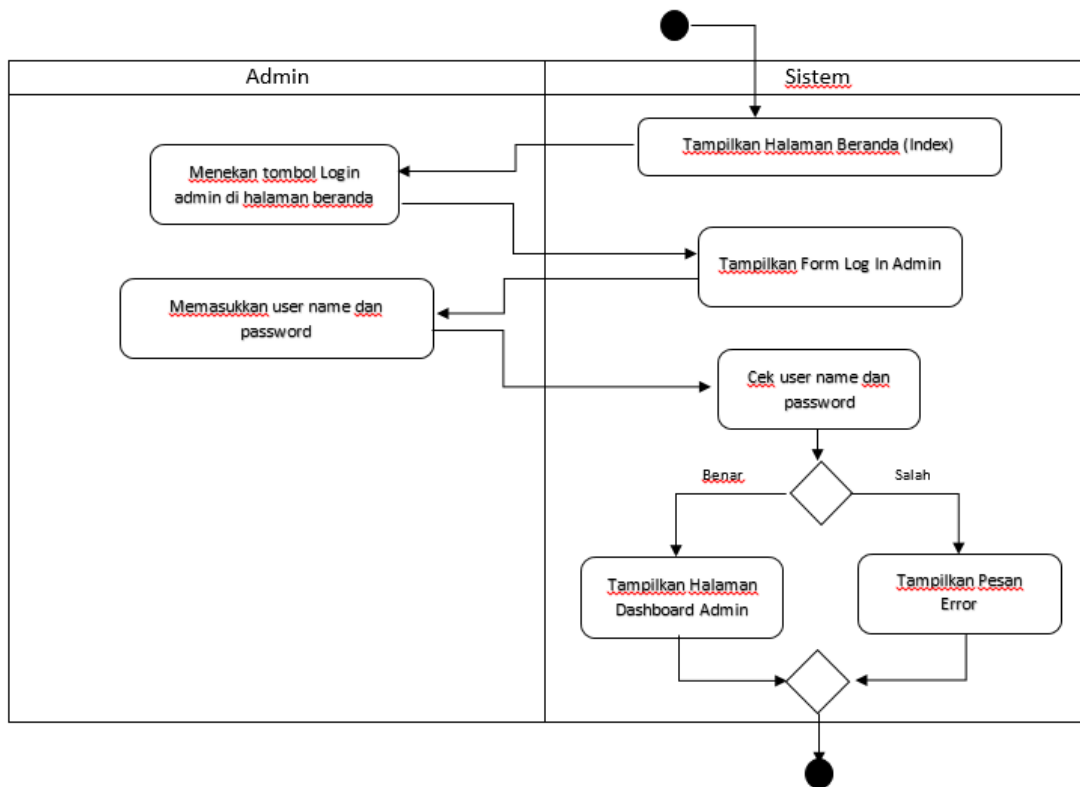
Merupakan diagram yang sangat membantu saat melakukan proses mendesain suatu system, khususnya fungsionalitas dari system yang terdiri dari berbagai unit yang saling bertukar informasi [10]



Gambar 3. Use Case Diagram Kamus HTML 5

- **Activity Diagram**

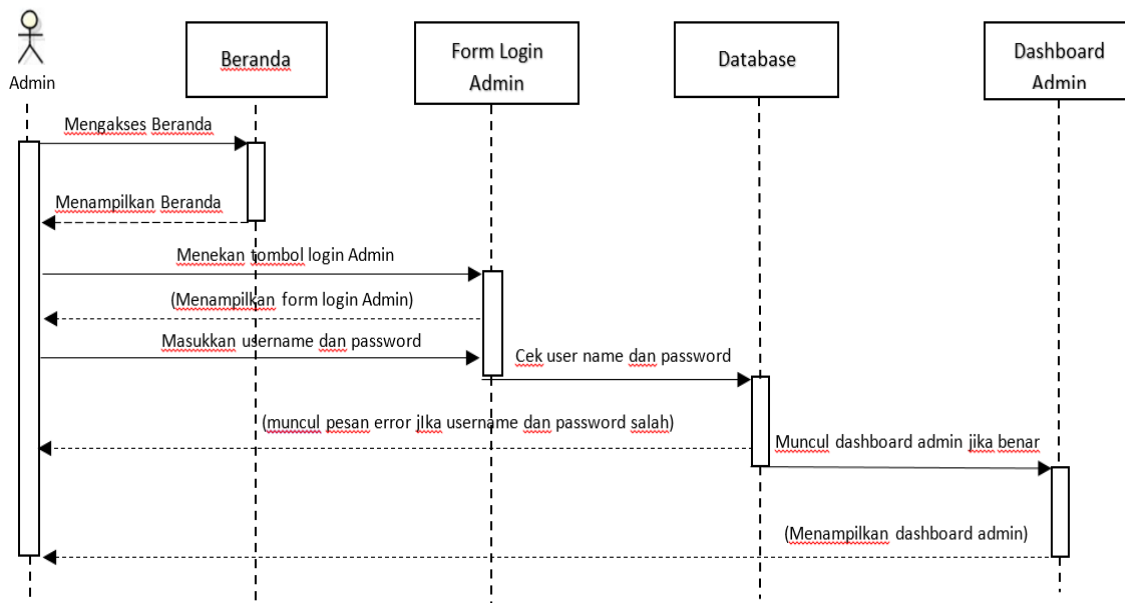
Merupakan diagram yang digunakan untuk menggambarkan urutan aktivitas dari sebuah system. Activity diagram dibuat diawal pemodelan dari proses pengerjaan system, hal tersebut dikarenakan dapat membantu pengguna dalam memahami keseluruhan proses [11].



Gambar 4. Activity Diagram Login Admin

- Sequence Diagram

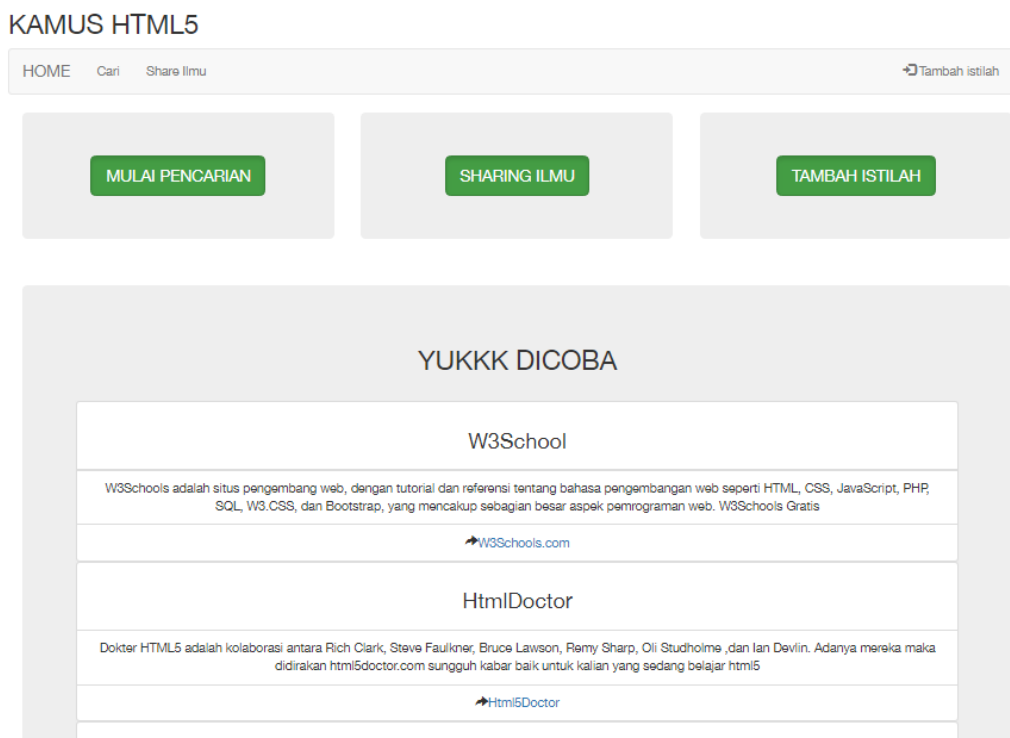
Diagram yang tersusun dari berbagai objek yang saling berkolaborasi secara dinamis. Tujuan pembuatan sequence diagram adalah untuk menampilkan berbagai rangkaian pesan dan interaksi antar objek yang terjadi pada titik tertentu dalam sebuah system [1].



Gambar 5. Sequence Diagram Login Admin

3.3. Hasil

Kamus HTML 5 merupakan aplikasi berbasis web dengan tampilan halaman beranda seperti pada gambar 6 berikut



Gambar 6. Halaman Beranda Kamus HTML 5

Di dalam halaman beranda terdapat tiga tombol yang masing – masing mempunyai fungsi berbeda. Tombol mulai pencarian digunakan untuk mencari kata atau istilah – istilah yang berkaitan dengan HTML 5. Tombol Sharing Ilmu merupakan tombol yang dapat digunakan untuk membaca ataupun mendownload ebook mengenai pembuatan halaman web khususnya menggunakan HTML. Tombol terakhir berupa Tambah Istilah yang berfungsi untuk memberikan kesempatan bagi masyarakat untuk berkontribusi menambah kata atau istilah – istilah tentang HTML 5. Selain ketiga tombol diatas terdapat tautan (link) yang menghubungkan secara langsung ke beberapa aplikasi scripting online, diantaranya W3School dan HTMLDoctor.

KAMUS HTML5

HOME Cari Share Ilmu [Tambah istilah](#) [Log In Admin](#)

masukkan kata kunci

CARI

© kamushtml5 2018

Gambar 7. Halaman Pencarian Kata / istilah Kamus HTML 5

Setelah klik tombol Mulai Pencarian yang terdapat di halaman beranda, maka akan muncul halaman pencarian seperti gambar 7 di atas. Pencarian kata atau istilah berkaitan dengan HTML cukup dengan memasukkan kata atau istilah ke dalam text box dan dilanjutkan dengan klik tombol Cari.

KAMUS HTML5

HOME Cari Share Ilmu [Tambah istilah](#) [Log In Admin](#)

Anda bisa mendownload atau membaca e-book file pdf tutorial HTML 5

SHARE ILMU

- 01 [html & css-thecomplete5thed](#)
[html & css-thecomplete5thed-t.powellmcgraw-hill2010bbs_nxpowerlite.pdf](#)
- 02 [HTML_ The Complete Reference, Second Edition](#)
[HTML_ The Complete Reference, Second Edition.pdf](#)
- 03 [practical_HTML_5_projects](#)
[14259556589146701650.pdf](#)

1
Total e-book : 3 e-book

© kamushtml5 2018

Gambar 8. Halaman Berbagi Ilmu

Gambar 8 merupakan halaman Berbagi Ilmu yang dapat digunakan oleh pengguna untuk mendownload atau membaca ebook yang berkaitan dengan HTML. Halaman tersebut akan tampil ketika pengguna melakukan klik tombol Sharing Ilmu yang terdapat di halaman beranda.

HOME Cari Share Ilmu [Tambah Istilah](#) [Log In Admin](#)

DATA DIRI ANDA

*Nama Lengkap

*Pekerjaan

*Nomor telephone

*Email

*Jenis Kelamin

ARTIKEL

*Judul

*Istilah yang akan di tambah

Gambar 9. Halaman Untuk Tambah Kata / Istilah di Kamus HTML 5

Seorang pengguna kamus HTML 5 juga dapat memberikan kontribusi dengan menambahkan kata atau istilah – istilah HTML yang belum ada di dalam daftar kamus HTML 5. Pengguna cukup dengan melakukan klik di tombol Tambah Istilah yang terdapat di halaman beranda, maka selanjutnya akan muncul halaman Tambah Kata/Istilah seperti pada gambar 9. Pada halaman tersebut, seorang pengguna yang ingin memberikan kontribusi wajib untuk mengisi form yang telah disediakan, dengan tujuan setiap kata atau istilah yang diberikan oleh pengguna dapat diketahui asal usul pengirim yang telah berkontribusi.

Masuk sebagai administrator

Email

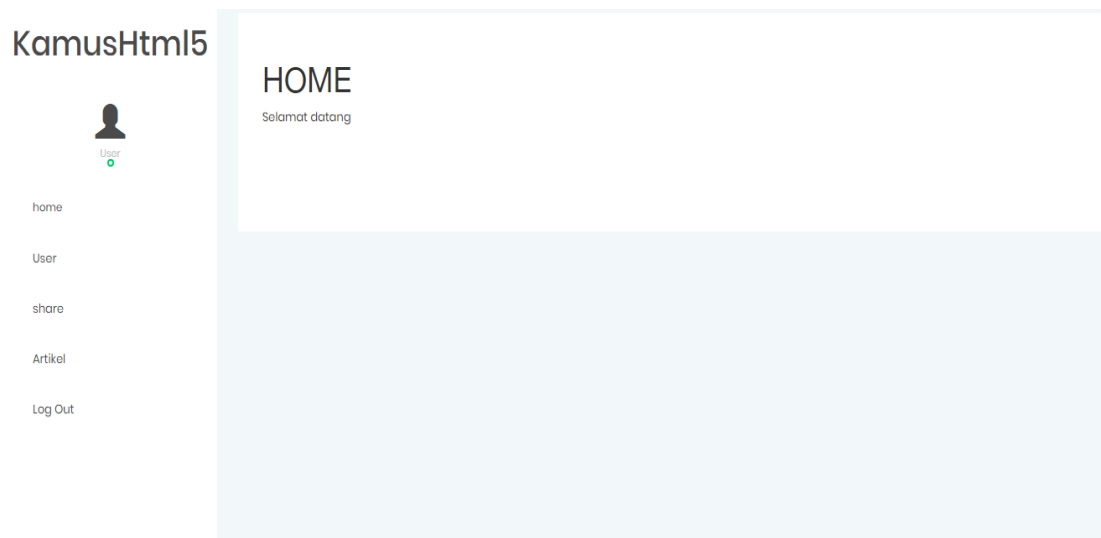
Password

Masuk

Kembali

Gambar 10. Halaman Login Masuk Ke Dashboard Admin

Di dalam kamus HTML 5 juga terdapat halaman khusus admin atau disebut juga dengan dashboard admin. Fungsi dashboard admin adalah untuk melakukan pengaturan konten atau isi dari seluruh halaman yang ada di dalam kamus HTML. Tidak semua dapat mengakses atau masuk ke dalam dashboard admin karena khawatir disalah gunakan. Untuk masuk ke dashboard admin, harus masuk terlebih dahulu ke halaman login seperti pada gambar 10 diatas, dengan cara klik tombol log in admin yang terdapat di halaman beranda. Di dalam halaman login terdapat email dan password sebagai verifikasi hak akses. Jika email dan password yang dimasukkan benar, maka akan muncul halaman dashboard admin seperti pada gambar 11. Jika salah maka akan muncul pesan error.



Gambar 11. Halaman Dashboard Admin

Di dalam kamus HTML 5 terdapat halaman dashboard admin yang memiliki beberapa tombol, diantaranya tombol user akan menampilkan pengguna yang memberikan kontribusi terkait kata atau istilah di dalam HTML. Tombol share berfungsi untuk menampilkan ebook yang bisa diakses oleh pengguna dan tombol artikel berfungsi untuk menampilkan artikel – artikel yang akan di munculkan di halaman beranda. Tombol yang terakhir adalah tombol logout berfungsi untuk keluar dari halaman dashboard admin.

IV. SIMPULAN

Berdasarkan hasil penelitian rancang bangun kamus HTML 5 berbasis web, dapat diambil kesimpulan sebagai berikut :

- HTML 5 adalah versi terbaru dari HTML yang merupakan bahasa markup dan dapat digunakan untuk membuat halaman web.
- HTML 5 dapat digabungkan dengan dengan script dari bahasa pemrograman lain hingga dapat membuat halaman web semakin menarik dan dinamis.

- Kamus HTML 5 dilengkapi dengan beberapa fasilitas yang berupa pencarian kata, sharing ilmu, menambahkan istilah, dan tautan ke aplikasi scripting online, sehingga memberikan kemudahan bagi pengguna.
- Kamus HTML 5 merupakan aplikasi berbasis web yang dibangun sebagai alternatif rujukan bagi masyarakat, khususnya siswa maupun mahasiswa yang baru belajar mengenal pembuatan halaman web.

Daftar Pustaka

1. Wahyu Purnomo. Pemrograman Web 1 Untuk SMK/MAK Kelas X. Hak Cipta @2013, Kementerian Pendidikan & Kebudayaan
2. HTML: The Complete Reference, Second Edition., Thomas A. Powell., Osborne/McGraw-Hill
3. Sejarah Perkembangan HTML., <https://www.ilmukucerdas.com/2017/12/sejarah-perkembangan/#respond>
4. Sejarah HTML & XHTML., Pemrograman Web Teknik Informatika Fakultas Teknologi Industri., Universitas Ahmad Dahlan., <http://eprints.uad.ac.id/585/11/Materi-03b.HTML-Dasar%23sejarah-html-dan-xhtml.pdf>
5. Analisis dan Perbandingan HTML 4.01 dan HTML 5 (Studi Kasus : web Online Music Store) http://repository.uksw.edu/bitstream/123456789/2391/2/T1_672007062_Full%20text.pdf
6. Moh. Jasri., Ahmad Buhari. 2016. Rancang Bangun Kamus Kebidanan Berbasis Android Dengan Eclipse. Systemic Vol. 02, No. 01, Agustus 2016, 39-44
7. Ratih Purwasih, Yessy Prima Putri, Aulia Prima. 2017. Rancang Bangun Kamus Minag – Inggris Menggunakan Bahasa Pemrograman Hypertext Processor (PHP). Jurnal Gramatika. ISSN : 2442-8485. E-ISSN : 2460-6319
8. Pressman, R.S. 2002. Rekayasa Perangkat Lunak : Pendekatan Praktisi (Buku Dua). Yogyakarta. Penerbit Andi
9. Sommerville, I. (2011). Software Engineering 9th Edition. Addison-Wesley.
10. Sugiarti, Yuni. 2013. Analisis dan Perancangan UML (Unified Modelling Language). Yogyakarta: Graha Ilmu.
11. Astria Hijriani, S. Kom., M.Kom. Tim asistem Rekayasa Perangkat Lunak. 2017. Modul Responsi Rekayasa Perangkat Lunak. Edisi 1(2017). Laboratorium Komputasi Dasar Jurusan Ilmu Komputer FMIPA Universitas Lampung.

**PENYUSUNAN KAMUS DIGITAL IELTS
BERBASIS ANDROID**
*(COMPILING IELTS DIGITAL DICTIONARY
USING ANDROID-BASED)*

Homsatun Nafiah¹, Ahmad Afandi²
TEST English School¹, Virtual Reality Academy²
homsatunnafiah78@gmail.com¹, achaffandhy@gmail.com²

Abstrak

Kemajuan teknologi informasi mempengaruhi banyak aspek kehidupan, termasuk bidang leksikografi. Digitalisasi merupakan proses mengubah informasi dari format analog menjadi format digital dan merupakan produk utama dari pengembangan teknologi informasi yang memiliki daya tarik tersendiri bagi penyusun dan pengguna kamus. Crystal (1995/2003: 446) menyatakan bahwa era ini memberikan pengaruh besar terhadap jenis kamus yang terjadi sebagai hasil dari perkembangan yang pesat dalam bidang leksikon dengan memanfaatkan perangkat korpus. Akan tetapi, perkembangan kamus di Indonesia perlu senantiasa ditingkatkan, khususnya kamus bahasa Inggris bagi pemelajar IELTS (International English Language Testing System). Oleh karena itu, penelitian ini bertujuan untuk merumuskan cara membuat kamus elektronik bagi pemelajar IELTS berbasis android.

Keberadaan kamus menempati posisi penting dalam proses pemerolehan Bahasa Inggris. Dalam penelitian ini, beberapa perangkat korpus dimanfaatkan dalam menyusun data leksikon seperti AntWordProfiler, Familizer, dan AntConc (Kwary and Jurianto, 2017). Kemudian, prinsip-prinsip dalam kosakata dari Graves (2016) digunakan dalam merumuskan fitur-fitur kata dalam kamus. Untuk menyajikan data kamus tersebut menjadi perangkat digital berbasis android, Android Studio digunakan sebagai salah satu pengembang terpadu untuk aplikasi android dimana alat pengembang ini dimanfaatkan untuk memasukkan kata-kata pilihan beserta fiturnya. Aplikasi ini dapat dioperasikan dalam sebuah perangkat elektronik berbasis android.

Dengan adanya kamus IELTS berbasis android ini diharapkan dapat meningkatkan peran kamus dalam pembelajaran IELTS di era modern saat ini. Kamus ini tidak hanya dilengkapi dengan definisi, tetapi juga sinonim, contoh kalimat, dan keluarga kata. Pada akhirnya tujuan utama dari pengembangan ilmu pengetahuan dapat tercapai yaitu dengan menghadirkan inovasi dalam bidang leksikografi dalam bentuk kamus IELTS ekabahasa berbasis android yang *user-friendly*. Sehingga, para pemelajar IELTS dapat ikut andil dalam memanfaatkan teknologi yang ada untuk menguasai kompetensi Bahasa Inggris demi masa depan yang lebih baik.

Abstract

The development of information technology affects many aspects of life, including in the field of lexicography. Digitalization is the process of converting information from analog format to digital format and is the main product of the development of information technology that has its own fascination for compiler and dictionary user. Crystal (1995/2003: 446) reveals that this era greatly influences the kind of dictionary that occurs as a result of the rapid development in the field of lexicon by utilizing the corpus device. However, the development of dictionaries in Indonesia should always be improved, especially English dictionaries for IELTS (International English Language Testing System) learners. Therefore, this study aims to formulate how to create an electronic dictionary for IELTS learners on android-based.

The existence of the dictionary occupies an important position in the process of obtaining English language. In this study, some corpus devices are utilized in compiling lexicon data such as AntWordProfiler, Familizer, and AntConc (Kwary and Jurianto, 2017). Then, the vocabulary principles by Graves (2016) are used in formulating word features in the dictionary. To present the data dictionary into an android-based digital device, Android Studio is used as one of the integrated developers for android applications where the developer tool is used to enter the selected words along with their features. This app can be operated in an android based electronic device.

By having the IELTS android-based dictionary, it is expected to enhance the role of the dictionary in IELTS learning in today's modern era. This dictionary is not only equipped with definitions, but also synonyms, sentence examples, and word families. Ultimately the main goal of science development can be achieved by presenting innovations in the field of lexicography in the form of a monolingual IELTS dictionary using android-based which is user-friendly. Therefore, IELTS learners can take part in utilizing the existing technology to master the competence of English for a better future.

I. PENDAHULUAN

Kemajuan teknologi informasi tentunya dapat mempengaruhi banyak aspek kehidupan, termasuk dalam bidang leksikografi. Di era modern saat ini, teknologi merupakan sebuah bukti dari kemajuan dan pengembangan pengetahuan manusia yang terwujud dalam media digital seperti telepon pintar, internet, dan komputer. Digitalisasi merupakan proses mengubah informasi dari format analog menjadi format digital yang memiliki daya tarik tersendiri bagi para penyusun dan pengguna kamus.

Crystal (1995/2003: 446) menyatakan bahwa era ini memberikan pengaruh besar terhadap jenis kamus yang terjadi sebagai hasil dari perkembangan yang pesat dalam bidang leksikon dengan memanfaatkan perangkat korpus. Akan tetapi, perkembangan kamus di Indonesia perlu senantiasa ditingkatkan, khususnya kamus bahasa Inggris bagi para pengambil tes kemampuan bahasa Inggris seperti IELTS (International English Language Testing System).

Mengingat kebutuhan akan penguasaan bahasa Inggris khususnya di Indonesia mengalami peningkatan. Hal ini salah satunya merupakan sebuah upaya pemerintah

Indonesia dalam meningkatkan kualitas sumber daya manusia melalui pendidikan lanjut baik dalam negeri maupun luar negeri. Sedangkan, perkembangan kamus bagi pemelajar bahasa Inggris telah tersedia dengan berbagai tujuan dan fitur yang ditawarkan baik kamus cetak maupun digital. Dengan kelebihan dan kekurangan yang dimiliki oleh setiap kamus, maka pengguna dituntut untuk pandai dalam memilih kamus yang sesuai serta efektif sesuai dengan tujuan pembelajaran. Untuk tujuan penguasaan kosakata, Kwary and Jurianto (2017) memberikan saran untuk membuat daftar kata dengan menggunakan teks bacaan yang digunakan siswa termasuk buku materi IELTS menggunakan perangkat korpus. Di sisi lain, Graves (2016) merekomendasikan bahwa dalam pembelajaran kosakata memerlukan deskripsi lebih tidak hanya definisi, tetapi juga sinonim, kelas kata, maupun imbuhan.

Pengetahuan kosakata telah mempengaruhi tingkat pemahaman pemelajar bahasa Inggris, karena kosakata menjadi komponen yang menonjol dalam memperoleh dan mengajarkan bahasa Inggris. Nation (2001: 492) menegaskan bahwa kosakata bukanlah tujuan dalam belajar bahasa, melainkan membantu siswa untuk menguasai empat keterampilan dalam bahasa termasuk membaca secara lebih efektif.

Membaca merupakan satu dari empat keahlian yang diujikan dalam tes IELTS yang memiliki jumlah kosakata terbanyak. Dengan kata lain, pengetahuan akan kosakata sangat menunjang pelajar dalam memahami teks yang ada. Chalmers & Walkinshaw (2014) menginvestigasi apakah dan sejauh mana para peserta tes IELTS Academic Reading menggunakan strategi membaca yang efektif dan apakah penggunaannya memberi dampak pada hasil tes mereka. Hasil penelitiannya menunjukkan bahwa peserta didik dengan tingkat pemahaman leksikon yang rendah merupakan aspek paling sulit yang harus diatasi pada strategi tes.

Terdapat korelasi yang erat antara pengetahuan kosakata dan pemahaman bacaan, jika siswa tidak secara teratur mengembangkan pengetahuan kosakata mereka, maka itu akan mempengaruhi pemahaman bacaan mereka (Chall & Jacobs, 2003). Hal ini tentunya akan mempengaruhi nilai yang didapat oleh peserta tes. Oleh karena itu, penting bagi pelajar bahasa Inggris khususnya bagi pemelajar IELTS untuk mempelajari kosakata yang digunakan dalam teks Reading IELTS. Salah satunya yaitu melalui media daftar kata atau kamus yang berisi kosakata pilihan yang dilengkapi dengan fitur penjelasan dari setiap kata.

Perangkat korpus dan sistem digital dimanfaatkan untuk mendapat data yang komprehensif. Sehingga, penelitian ini bertujuan untuk merumuskan cara membuat kamus elektronik bagi pemelajar IELTS berbasis android sebagai inovasi dalam dunia pendidikan yang sejalan dengan kemajuan dan perkembangan teknologi saat ini.

II. KERANGKA TEORI DAN METODE PENELITIAN

2.1 Kerangka Teori

2.1.1 Korpus

Untuk mendapatkan daftar kata pilihan, korpus digunakan sebagai media yang membantu proses penyortiran kata dalam jumlah besar. Korpus adalah kumpulan teks tertulis atau teks lisan yang dapat digunakan sebagai perwakilan dari berbagai bahasa,

jumlah konotasi, bentuk yang dapat dibaca mesin, sempel dan representasi, ukuran terbatas, dan gagasan (Lüdeling & Merja, 2009).

Dalam perkembangannya, korpus dimanfaatkan dalam bidang linguistik. Istilah untuk linguistik korpus mengacu pada studi data bahasa dalam skala besar yang menggunakan komputer untuk menganalisis banyak ucapan tertulis atau teks tertulis dalam studi bahasa, prosedur, metode, untuk mempelajari bahasa (McEnery & Hardie, 2011). Sehingga, dalam penelitian ini, pendekatan korpus digunakan untuk mendapatkan kata-kata pilihan dalam konteks Reading IELTS.

2.1.2 Prinsip-Prinsip Kosakata

Setelah mendapatkan daftar kata melalui pendekatan korpus, tahap selanjutnya adalah mencantumkan fitur-fitur dari tiap kata. Graves (2016: 5) merumuskan 4 prinsip dalam kosakata, diantaranya (1) *frequent, varied, and extensive language experiences*, (2) *teaching individual words*, (3) *teaching word-learning strategy*, and (4) *fostering word consciousness*.

Dari keempat prinsip diatas, penulis telah merangkum beberapa poin utama yang berfokus pada fitur dari sebuah kata. Graves (2016: 7) menyatakan bahwa dalam pembelajaran kosakata faktor yang sangat efektif yang perlu disertakan dari sebuah kata adalah definisi dan informasi konteks. Untuk mendapatkan pemahaman yang baik pemelajar diharuskan untuk secara aktif berlatih kosakata yang dipelajari baik dalam bentuk lisan maupun tulis.

Selain definisi dan konteks, Graves juga menyarankan untuk melengkapi sebuah kata dengan *word part*. Istilah *word part* mengacu pada bagian dari sebuah kata yang ada di awal, tengah atau diakhir kata. Dalam bahasa Inggris *word part* dapat terdiri dari *inflection, derivational suffixes, prefixes*, dan *Latin or Greek roots* (Graves, 2004).

Infleksi merujuk pada akhiran kata yang merubah bentuk dasar kata ke bagian lain seperti *tense, number*, dan *aspect*. Contohnya yaitu kata *houses* terdiri dari kata *house* dan infleksi *-s*, *wanted* mengarah ke *want* dan *-ed*. Sehingga dapat disimpulkan bahwa infleksi adalah bentuk morfem yang tidak mengubah kelas kata *base* (bentuk dasar) yang ditempelinya menjadi kelas kata lain. Sedangkan derivasional mengacu pada bentuk morfem yang dapat mengubah makna dan kelas kata *base* yang ditempelinya. Contohnya yaitu kata *hopeless* yang terdiri dari *hope* dengan status kata sebagai kata kerja atau kata benda, kemudian di tambahkan *-less* membentuk kelas kata baru yaitu kata sifat.

Strategi mempelajari kosakata dapat juga di lakukan dengan menggunakan kamus, *glossary* atau *thesaurus*. Kamus merupakan sebuah buku referensi yang berisi definisi yang digunakan untuk mengetahui makna kata. Dalam proses pembelajaran, akses semantis dapat juga disertakan dalam fitur kata seperti sinonim atau antonym (Graves: 2016). Sehingga, peneliti menggunakan 4 fitur kata yang telah diusulkan oleh Graves (2016), dan menggunakan beberapa perangkat lunak korpus dalam menyusun daftar kata (Kwary and Jurianto, 2017) untuk membantu pemelajar dalam memahami makna dan penggunaan kata dalam konteks Reading IELTS yaitu definisi, sinonim, konteks (terwakili oleh contoh kalimat yang langsung diambil dari sumber buku), dan *word part*

(dipelajari melalui oleh *word family* yang secara lengkap mencantumkan bentuk perubahan tiap kepala kata).

2.2 Metode Penelitian

2.2.1 Teknik Pengumpulan Korpus

Penelitian ini menggunakan pendekatan korpus sebagai metode dalam mendapatkan daftar kata pilihan yang disortir melalui beberapa perangkat korpus. Dalam hal ini, Biber (2011) menyatakan bahwa korpus menyediakan penggunaan bahasa yang aktual yang dapat di gunakan sebagai sebuah alat dalam penelitian, salah satunya dalam bidang linguistik. Beberapa perangkat lunak yang digunakan yaitu:

AntFileConverter 1.2.0 (Anthony, 2015), sebuah perangkat lunak yang berfungsi untuk mengubah data dengan format pdf ke .txt. Program komputer ini direkomendasikan oleh Kwary and Jurianto (2017) dalam proses pembuatan *word list* yang dapat di akses tanpa membayar www.laurenceanthony.net/software/antfileconverter/. Dalam penelitian ini menggunakan semua teks Reading IELTS dari 10 buku Cambridge Practice Test dari tahun 1996 – 2015.

AntWordProfiler 1.4.0w (Anthony, 2013), sebuah perangkat lunak yang berfungsi untuk menyortir kata sesuai dengan tujuan penelitian. Dalam membuat daftar kata pilihan, peneliti menggunakan kategori kata yang memiliki frekuensi tertinggi. Langkah yang dapat diambil yaitu dengan mengklik “User File(s)”, dan “choose” atau pilih berkas, kemudian mengeluarkan Academic Word List dengan mengeceklis “3_awl_570.txt” dimenu pilihan “Level List(s)” yang bertujuan untuk membuat daftar kata akademik yang baru, pilih semua item di menu opsi “Output Settings”, untuk memproses pilih “start”, dan simpan hasil data di folder.

Karena daftar kata yang tersedia di General Service List (GSL) 1 atau GSL 2 menampilkan daftar kata yang umum, maka peneliti menggunakan pencarian “not found” pada file yang sudah di simpan. Langkah ini tidak hanya untuk menciptakan daftar kata di luar GSL, tetapi juga untuk menghitung jarak (*range*) dan frekuensi (*frequency*) kata.

Microsoft Excel digunakan untuk mengurutkan dan menstabilkan jumlah *range* dan *frequency* yang diperlukan (Coxhead, 2000; Gardner & Davies, 2013) dari yang tertinggi ke terendah. Penelitian ini membatasi jumlah frekuensi 7 atau lebih dari 7, dan jumlah jarak sebesar 3.

Familizer, sebuah perangkat lunak yang bermanfaat untuk menyusun *headword* dan memberikan *word family* dari setiap kata (Cobb, 2016) yang telah di dapat melalui beberapa langkah diatas. Perangkat ini juga direkomendasikan oleh Kwary and Jurianto (2017) yang menyatakan bahwa perangkat dapat mengakses data korpus sebesar 10MB dengan format .txt atau sekitar satu juta kata. Perangkat komputer ini dapat secara otomatis mengklasifikasikan lebih dari satu kata yang memiliki kesamaan infleksi ke dalam satu grup kepala kata (*headword*) dengan mengakses halaman <https://www.lex Tutor.ca/familizer/>.

Setiap kata dilengkapi dengan definisi yang diambil dari rujukan kamus Cambridge Advanced Learner’s Dictionary, 3rd edition (Walter, 2008).

Peneliti menggunakan Skell Sketchengine untuk mencari sinonim dari setiap kata yang berfungsi untuk memberikan keterangan terhadap ekspresi lain dari kata pada alamat web <http://skell.sketchengine.co.uk/run.cgi/skell#>.

Untuk mendapatkan informasi konteks, peneliti memanfaatkan perangkat lunak AntConc 3.4.4w (Anthony, 2014) untuk mendapatkan contoh kalimat yang secara aktual digunakan dalam berkas yang telah di masukkan dengan memilih opsi Corcordance. Langkah ini berfungsi untuk memberi contoh penggunaan kata dalam sebuah kalimat, selengkapnya pemelajar juga dapat memperkirakan makna kata melalui kata-kata yang digunakan dalam sebuah kalimat.

2.2.2 Teknik Pembuatan Kamus Elektronik Berbasis Android

Dalam penelitian ini tidak hanya menyajikan daftar kata yang nantinya dapat dibuka di laptop melalui Ms Word atau dengan cara di cetak, tetapi dapat di akses melalui sistem android. Sehingga dapat memberikan kemudahan dan keefektifan karena dapat di operasikan di ponsel pintar (*smartphone*). Android Studio telah dipilih sebagai perangkat yang membantu mengubah bentuk format Ms Word ke dalam sebuah aplikasi Android. Android Studio adalah alat pengembang dan editor kode IntelliJ yang menawarkan berbagai fitur dalam meningkatkan produktivitas dan kreativitas seseorang ketika membuat aplikasi Android (Developer.android). Adapun langkah-langkah dalam pembuatan aplikasi menggunakan Android Studio memiliki sebagai dua tahapan utama yaitu:

- 1) Membuat rancangan tampilan dari aplikasi yang diinginkan tampilan awal dapat terlihat di gambar \ berikut ini.

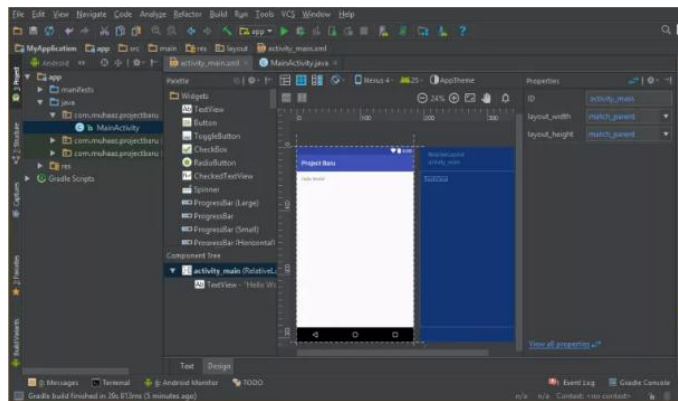
Gambar \ Tampilan Awal Android Studio



Tahapan pembuatannya yaitu: (1) menekan opsi “Start a new Adroid Studio project”, (2) mengisi identitas dari proyek yang akan di kerjakan di kolom *New Project* seperti nama aplikasi dan lainnya, (3) memilih target device android (perangkat yang nantinya digunakan misalnya *Phone and Tablet* dibagian atas), kemudian menentukan SDK minimal agar aplikasi bisa berjalan, apabila kita memilih APK 10 berarti APK yang ada dibawahnya tidak akan berjalan, (4) menambah aktivitas “Add an Activity to Mobile” yang berfungsi untuk memilih tampilan aplikasi atau *template activity*, (5) mengisi *Customize the Activity*, apabila ingin mengisi nama aktivitas dan

layout sesuai yang tertera, maka langkah selanjutnya adalah klik “Finish”, kemudian akan muncul tampilan editor Java dan XML yang telah dibuat seperti contoh gambar berikut ini:

Gambar 3 Tampilan Editor Java Dan XML

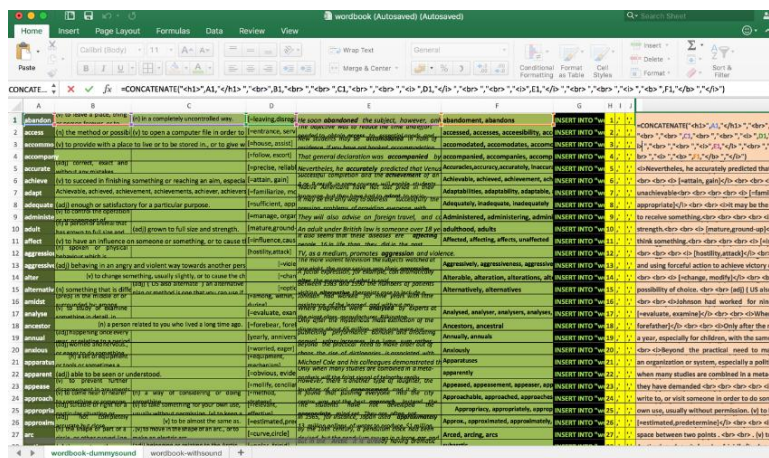


Langkah selanjutnya yaitu dengan membuka “Edit Configuration”, dan buka bagian app. Kemudian atur sesuai keterangan yang diminta dan klik Apply atau Ok. Kemudian klik Run pada tombol segitiga hijau atau dengan klik Run, dan (7) Pilih perangkat (Device) Android yang sudah terhubung USB, atau dapat menggunakan Android Virtual Device, lalu klik Ok. Selanjutnya, langkah berikutnya adalah peng-codingan kamus digital dengan menggunakan bahasa Java dan Android Studio.

2) Tahap pembuatan dan penginputan database kosakata.

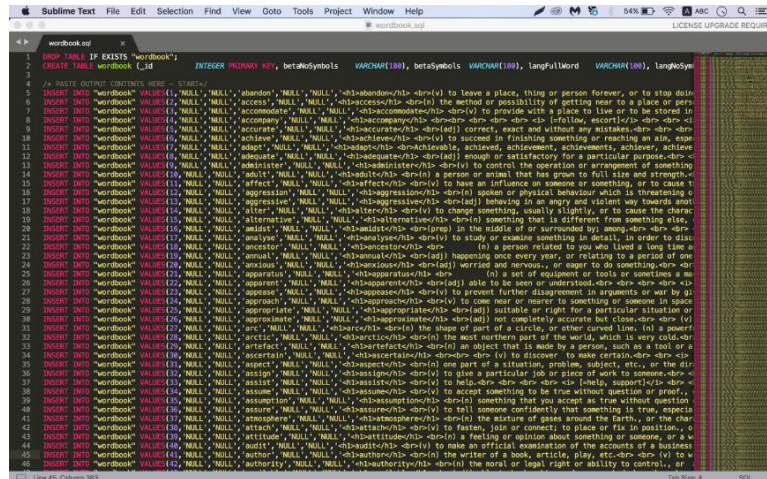
Untuk memudahkan dalam membuat database, maka hal pertama yang dilakukan adalah menginput semua kata dan fitur yang akan disertakan dalam ke database kamus elektronik. Dengan sedikit bantuan dari formula “concatenate” excel dan syntax “html”, maka, kata dan fitur yang telah diinputkan akan otomatis ter “output” di kolom yang telah kita sediakan.

Gambar 4 Kolom Input Data Menggunakan Excel



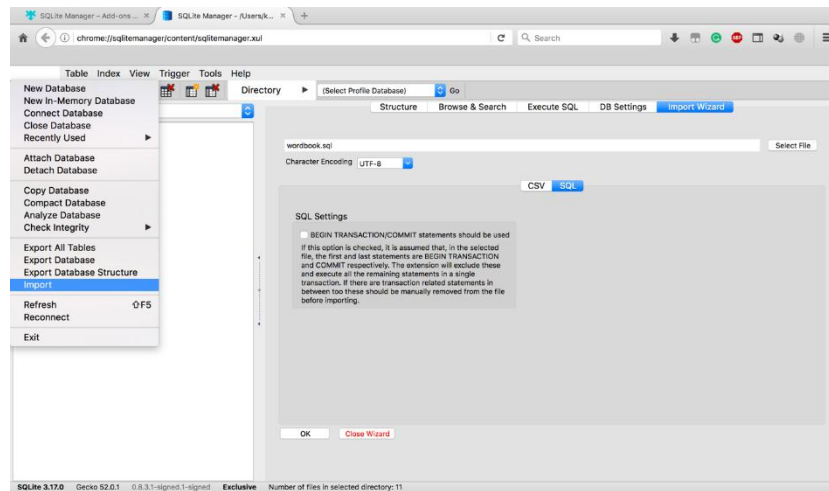
Setelah itu, salin hasilnya ke file workbook.sql dengan bantuan aplikasi “sublime text” seperti gambar 5 dibawah ini.

Gambar 4 Penyalinan Berkas ke Format .sql



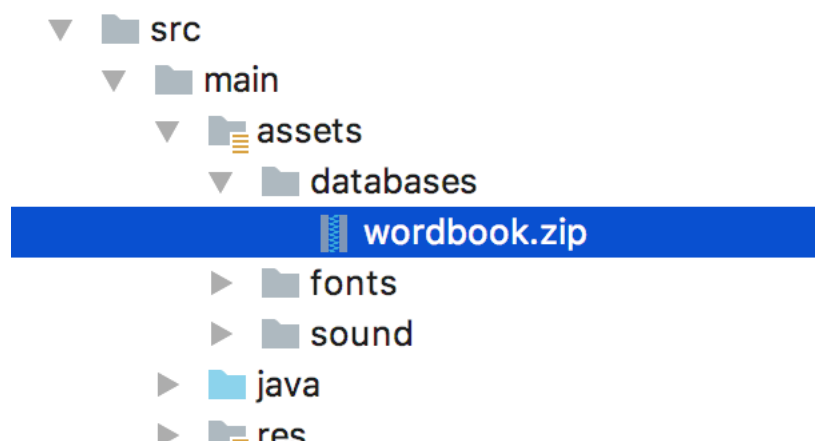
Setelah file `wordbook.sql` telah tersimpan, maka, tahap selanjutnya ialah mengkonversi file `wordbook.sql` ini ke `wordbook.db`. Konversi ini diperlukan karena file dengan format `.db` sangat mendukung apabila digunakan di banyak *mobile devices* seperti android, windows mobile, dan ios. Kemudian, caranya yaitu dengan menimport file sql tersebut ke file db dengan bantuan `sllite manager` di browser firefox seperti gambar dibawah ini.

Gambar 5 menimport file .sql ke file .db



Setelah selesai mengimport, tahap selanjutnya yang dilakukan adalah mengkompres file `wordbook.db` kedalam ekstensi zip dan memindahkannya ke folder database “Comprehensive IELTS Dictionary” atau nama lain dari proyek yang telah dibuat sebelumnya di tahapan *layout* Android Studio.

Gambar 6 mengkompres file wordbook.db kedalam ekstensi zip



Setelah semua tahap diatas selesai. Maka aplikasinya pun siap di *BUILD*.

III. ANALISIS PENELITIAN

Pengetahuan kosakata merupakan bekal yang sangat mendukung seseorang ketika memahami sebuah bahasa. Hal ini didukung oleh pernyataan Graves (2016) bahwa kosakata menjadi bagian yang sangat diperlukan dalam pedagogi bahasa. Oleh karena itu, perbendaharaan kosakata mempengaruhi tingkat kemampuan berbahasa seseorang termasuk kemampuan membaca, karena teks bacaan terdiri dari banyak kata yang digunakan untuk mendeskripsikan ide.

Pengajaran dan pembelajaran kosakata menjadi bagian yang paling menantang karena berkaitan dengan pemilihan kosakata yang efektif untuk dipelajari oleh siswa (Coxhead, 2000). Di lain pihak, Kwary and Jurianto (2017) menyarankan untuk membuat daftar kata (*word list*) menggunakan teks bacaan yang digunakan siswa dalam belajar. Hal ini dirumuskan untuk menyediakan bahan materi pelajaran yang lebih terpusat selama aktivitas belajar mengajar kosakata. Mengingat kebutuhan akan kosakata pemelajar bahasa Inggris khususnya dalam konteks IELTS perlu ditingkatkan, maka penelitian ini mencoba merumuskan kosakata pilihan untuk membantu pemelajar dalam memahami teks Reading IELTS.

Untuk mengetahui jumlah tipe kata (*word types*) dan token kata (*word tokens*), perangkat lunak AntConc (Anthony, 2014) digunakan untuk mengidentifikasi jumlah tersebut tiap sumber buku terbitan Cambridge. Hasilnya dapat di lihat di tabel berikut ini:

Table V Hasil AntConc dari 10 Sumber Buku

No	Tahun	Word Types	Word Tokens
1.	1996	3,502	20,884
2.	2000	3,567	19,226
3.	2002	3,772	21,987
4.	2005	3,038	12,996
5.	2006	4,137	21,310
6.	2007	3,564	18,679
7.	2009	3,934	21,233
8.	2011	3,208	14,885
9.	2014	2,996	13,599
10.	2015	3,527	19,638
Total		35,245 (13,489)	184,437

Tabel di atas menunjukkan hasil jumlah *word type* dan *word tokens* yang berbeda-beda. *Word type* mengacu pada jumlah kata yang berbeda yang terjadi dalam sebuah teks atau pidato, sedangkan *word token* adalah jumlah kata dalam bahasa tulis atau lisan tanpa mempertimbangkan seberapa sering kata tersebut digunakan (Scott, 2008). Contohnya adalah, kalimat “a good friend is a friend who respects you” terdiri dari 7 *word types* and 9 *word tokens*, dimana kata “a” atau “friend” diulang sebanyak dua kali, sehingga mereka terhitung sebagai satu kata.

Setelah mendapatkan jumlah *word types* dan *word tokens*, langkah selanjutnya adalah mengurutkan kata tersebut berdasarkan frekuensi melalui perangkat AntWordProfiler 1.4.0w (Anthony, 2013) dengan mengeluarkan Academic Word List 570 oleh Coxhead (2000) untuk mendapatkan AWL baru khusus Reading IELTS. Proses ini memperoleh hasil sebagai berikut:

Tabel ^ Hasil AntWordProfiler dari 10 Sumber Data

Level	File	Token	Token%
1	1_gsl_1st_1000.txt	138,869	75.41%
2	2_gsl_2nd_1000.txt	12,471	6.77%
0	-	32,819	17.82%
Total		184,159	100%

Persentase kumulatif dari tiap berkas terlihat pada tabel diatas. Tingkatan tertinggi diduduki oleh General Service List (GSL) 1000 pertama sebesar 75.41%, and 138,869 *word tokens*. Tingkatan selanjutnya adalah GSL 1000 kedua sebesar 12,471 *word tokens*, atau 6.77% dari seluruh data. Kemudian, tingkatan terendah yaitu jumlah di luar GSL 1 dan GSL 2 sebanyak 17.82% atau 32,819 *word tokens*. Selengkapnya, West (1953) menyatakan bahwa GSL merupakan daftar kata umum yang tersusun oleh 2000 kata tertinggi yang sering digunakan dalam berbagai disiplin ilmu dalam teks bahasa Inggris melalui proses penyortiran berdasarkan frekuensinya.

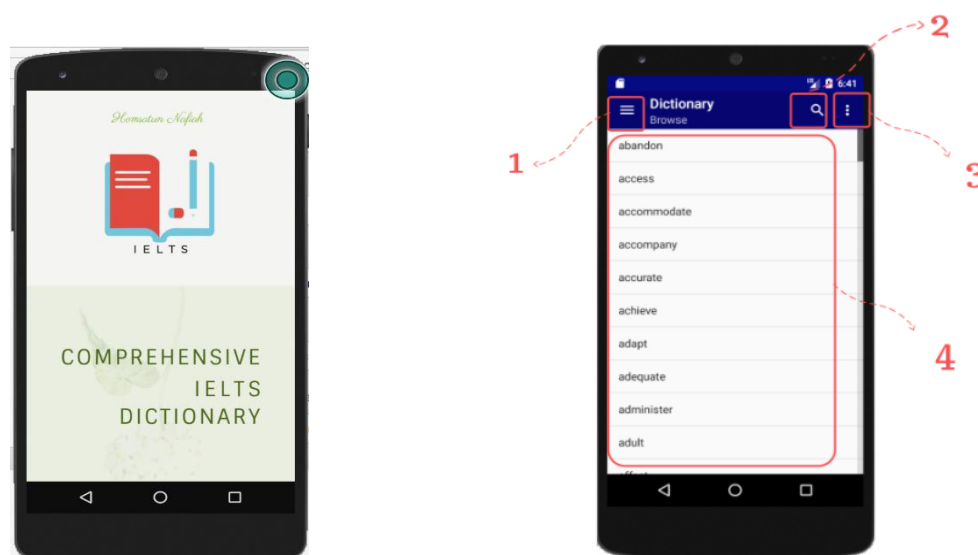
Dalam penelitian ini, peringkat terakhir dari keseluruhan data diatas yaitu 32,819 *word tokens* di gunakan sebagai sumber data to membuat daftar kata untuk pemelejar Reading IELTS. Hal ini berdasarkan pada pertimbangan akan pembuatan kamus yang berfokus pada konteks Reading IELTS, yang tentunya daftar kata yang dipilih pun di luar dari GSL ataupun AWL yang sudah ada. Sehingga, di harapkan dapat melengkapi bank kata yang sudah tersedia dalam bahasa Inggris.

Untuk membuat daftar kata yang sesuai dengan tujuan utama pembutan yaitu untuk memfokuskan kosakata yang secara efektif dipelajari, maka maka peneliti membatasi jumlah kata dari 32,819 *word tokens* menggunakan *frequency* dan *range*. Metode ini juga diterapkan oleh Coxhead (2000) yang memformulasikan AWL yang terdiri dari 570 kata dengan proses penyortiran jumlah *frequency* dan *range*.

Penelitian ini memperoleh daftar kata pilihan yang terdiri dari 380 kepala kata (*headwords*) atau sekitar 1.965 kata dari jumlah keseluruhan *word families*. Selama proses analisis data, terdapat dua tahapan yang digunakan untuk menyaring daftar kata tersebut. Yang pertama yaitu mengeliminasi singkatan atau kata yang tidak memiliki arti seperti *de*, *vfr*, *st*, *et*, dan *al*, karena kata tersebut bukan merupakan tujuan utama dalam pembelajaran kosakata dalam penelitian ini. Tahap kedua yaitu mengeluarkan kewarganegaraan, nama negara, dan nama orang, karena aspek tersebut tidak secara formal diajarkan atau dipelajari oleh siswa, dengan kata lain siswa dapat mempelajari kata tersebut di lain subjek.

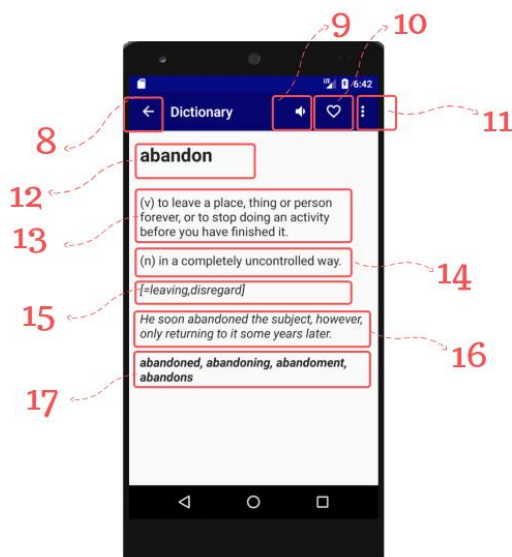
Setelah daftar kata diperoleh melalui perangkat komputer korpus, selanjutnya adalah memasukkan data tersebut ke dalam sistem digital melalui perangkat Android Studio. Sesuai alur pembuatan yang telah dijelaskan di bagian metode, maka hasil akhir dari aplikasi Android untuk IELTS adalah seperti yang terlihat pada gambar berikut ini:

Gambar 1 Tampilan Depan dari Aplikasi



Peneliti memberi nama aplikasi ini dengan “Comprehensive IELTS Dictionary”, karena diharapkan para pemelajar dapat memanfaatkan kamus ini sebagai bahan rujukan untuk membantu memahami kosakata dalam IELTS khususnya tes Reading. Pada keterangan nomor 1 berfungsi sebagai jendela baru untuk melihat sejarah pencarian (*history*), *browse*, dan *favorite*. Nomor 2 berfungsi sebagai tombol pencarian kata yang diinginkan. Nomor 3 berfungsi sebagai menu lain dengan pilihan *clear favorites*, *clear history*, *settings*, dan *help*. Nomor 4 adalah representasi dari daftar kata yang ada dalam aplikasi ini.

Gambar 1. Tampilan dari Deskripsi Kata



Gambar di samping ini merupakan tampilan dari deskripsi kata “abandon” (nomor 12), tombol kembali (nomor 8), suara pelafalan (nomor 9), tombol favorite (nomor 10), opsi pengaturan dll (nomor 11), definisi berserta kelas kata (nomor 13 & 14), sinonim (nomor 15), contoh kalimat (nomor 16), dan *word family* (nomor 17).

IV. SIMPULAN

Dengan adanya penyusunan kamus digital ini, diharapkan dapat membantu proses pembelajaran yang lebih mudah, akurat dan cepat untuk di akses. Sehingga, dalam era digital saat ini, dunia pendidikan dapat ikut merasakan perkembangan teknologi informasi yang semakin pesat, termasuk bagi pemelajar bahasa Inggris seperti IELTS, guna mencapai tujuan dari pendidikan itu sendiri. Untuk kedepannya, para peneliti dapat mengembangkan produktivitas dalam teknologi informasi digital berbasis Android atau lainnya pada disiplin ilmu yang berbeda seperti kamus kesehatan.

Daftar Pustaka

Anthony, L. 2013. AntWordProfiler (Version 1.4.0w) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/>

Anthony, L. 2014. AntConc (Version Version 3.4.4w) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/>

- Anthony, L. 2015. AntFileConverter (Version Version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/>
- Biber, D. 2011. Corpus linguistics and the study of literature. *Scientific Study of Literature*, 1(1), 15–23. <https://doi.org/10.1075/ssol.1.1.02bib>
- Chall, J. S., & Jacobs, V. A. 2003. Poor children's fourth-grade slump. *American Educator*, 27, 14–15.
- Chalmers, J., & Walkinshaw, I. 2014. Reading strategies in IELTS tests: Prevalence and impact on outcomes. *Research Strategies*, 30(1), 12–14.
- Cobb, T. 2016. 20 K Familizer Proto (Version 0.6 [Software]).
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Crystal, D. 1995. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.
- Gardner, D., & Davies, M. 2013. A New Academic Vocabulary List. *Applied Linguistics Research Gate*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>
- Graves, M. F. 2004. Teaching prefixes: as good as it gets? In *J. Baumann and E. Kame'enui (Eds.), Vocabulary instruction: Research to practice* (pp. 81–99). New York: Guilford Press.
- Graves, M. F. 2016. *The Vocabulary Book: Learning and Instruction* (Second). New York: Teachers College Press.
- Kwary, D. A., & Jurianto. 2017. Selecting and creating a word list for English language teaching. *Teaching English with Tachnology*, 17(1), 60–72.
- Lüdeling, A., & Merja, K. 2009. *Corpus linguistics: an international handbook* (Vol. 1). Berlin: Walter de Gruyter.
- McEnery, T., & Hardie, A. 2011. *Corpus Linguistics Method, Theory and Practice* (1st ed.). UK: Cambridge University Press.
- Nation, I. S. P. 2001. *Learning vocabulary in another language* (1st Edition). Cambridge: Cambridge University Press.
- Scott, M. 2008. WordSmith Tools (Version 5). Liverpool: Lexical Analysis Software. Retrieved from http://www.lexically.net/downloads/version5/HTML/index.html?type_token_ratio_proc.htm
- Walter, E. 2008. Cambridge Advanced Learner's Dictionary (3rd ed.).
- West, M. 1953. *A general service list of English words*. London: Longman.

**APLIKASI KAMUS ELEKTRONIK
YANG RAMAH BAGI PENGGUNA PONSEL PINTAR
PENYANDANG TUNANETRA**
*(ELECTRONIC DICTIONARY APPLICATION WHICH IS ACCESSIBLE
FOR BLIND USERS OF SMARTPHONE)*

Aptia Ardiasri
Politeknik Negeri Malang
aptiaardiasri@gmail.com

Abstrak

Digitalisasi masuk ke dalam bidang leksikografi. Hal tersebut dibuktikan dengan adanya kamus yang berbasis elektronik. Kamus elektronik yang sekarang populer digunakan adalah aplikasi kamus elektronik yang ada di ponsel pintar. Kamus yang ada di ponsel pintar masih berupa kamus dengan tulisan alfabetis seperti pada umumnya. Hal tersebut membuat penyandang tunanetra kesulitan untuk menggunakan kamus pada ponsel pintar. Berdasarkan hal tersebut, salah satu solusi yang ditawarkan adalah membuat kamus audio yang bisa digunakan oleh penyandang tunanetra. Penulisan makalah ini bertujuan memaparkan gagasan pengembangan aplikasi kamus elektronik yang dapat digunakan oleh penyandang tunanetra. Prosedur pengembangan diadaptasi dari model pengembangan 4D. Melalui prosedur pengembangan tersebut, aplikasi yang dihasilkan didesain sesuai spesifikasi produk yang ditetapkan, yaitu isi, sistematika, perangkat lunak, bahasa, dan tampilan. Penyusunan aplikasi ini memerlukan kolaborasi dari tiga pihak, yaitu pakar leksikografi, pakar teknologi, dan pakar ilmu tunanetra.

Kata kunci: aplikasi *mobile*, kamus elektronik, ponsel pintar, tunanetra

Abstract

Digitalization enters the field of lexicography. This is evidenced by the existence of an electronic-based dictionary. The most popular electronic dictionary used today is the electronic dictionary application available on smartphones. The dictionary on the smart phone is still a dictionary with alphabetical writing as in general. It makes blind users difficult to use dictionary on smartphone. Based on this, one of the solutions offered is to create an audio dictionary that can be used by people with visual impairment. This paper aims to explain the idea of developing an electronic dictionary application that can be used by people with visual impairment. The development procedure is adapted from the 4D development model. Through these development procedures, the resulting applications are designed to specific product specifications, ie content, systematics, software, language, and

display. The preparation of this application requires collaboration from three parties, lexicographers, technologists, and blind scientists.

Key word: *mobile application, electronic dictionary, smartphone, blind users*

I. PENDAHULUAN

Era digital telah memasuki berbagai bidang, tidak terkecuali leksilografi, cabang ilmu bahasa yang fokus pada teknik penyusunan kamus. Kamus Besar Bahasa Indonesia yang dahulu dicetak bentuk buku yang berukuran tebal, sekarang kamus tersebut sudah tersedia dalam bentuk perangkat lunak atau aplikasi kamus elektronik yang dapat diunduh dengan gratis dan dapat dibaca di ponsel pintar (*smartphone*). Kemajuan tersebut tentunya sangat memudahkan pengguna/ masyarakat yang membutuhkan kamus secara praktis dan cepat di segala keadaan, mengingat manusia di era digital saat ini sangat dekat dengan keberadaan ponselnya. Era *mobile* membawa kemudahan dalam berbagai bidang karena dukungan teknologi dapat membuat pekerjaan diberbagai bidang menjadi relatif mudah (Prayitno 2013:2). Meskipun aplikasi kamus elektronik di ponsel pintar tersebut sangat memudahkan pengguna, tetapi kemudahan yang ditawarkan aplikasi tersebut ternyata belum sepenuhnya dirasakan oleh pengguna ponsel pintar penyandang tunanetra.

Fisik ponsel pintar yang saat ini diproduksi mengunggulkan fitur layar sentuh. Fitur layar sentuh pada ponsel pintar yang tidak memiliki tekstur tersebut akan menyulitkan pengguna tunanetra untuk menggunakan kamus elektronik di ponsel pintarnya. Oleh karena itu, di era digital ini sebaiknya dikembangkan aplikasi kamus elektronik yang tidak hanya ditujukan bagi pengguna pada umumnya, tetapi juga dikembangkan aplikasi kamus elektronik di ponsel pintar bagi penyandang disabilitas, khususnya tunanetra.

Aplikasi kamus elektronik (KBBI) yang ada di *Playstore* rata-rata merupakan kamus yang berisi kosakata dalam bentuk tulisan. Hal tersebut kurang ramah bagi penyandang tunanetra karena penyandang tunanetra tidak bisa mengetik dan melihat tulisan yang ada di kamus elektronik tersebut. Sebaiknya, aplikasi KBBI elektronik di ponsel pintar juga dilengkapi dengan sensor suara yang sekarang sudah marak digunakan oleh berbagai aplikasi perkamusan lain, misalnya *Voice Translator* yang tersedia di *Playstore*.

Aplikasi *Voice Translator* merupakan kamus dwibahasa yang dilengkapi dengan fitur *input-output* kosakata dalam bentuk suara. Aplikasi tersebut dapat dioperasikan dengan cara: pengguna mengucapkan kata yang diinginkan, lalu akan muncul beberapa kosakata yang mirip penulisannya dengan suara yang diucapkan tersebut. Kemudian, pengguna memilih kosakata yang sesuai dengan yang diucapkan. Setelah itu, aplikasi akan membaca membacakan tulisan itu kepada pengguna.

Meskipun aplikasi *Voice Translator* sudah dilengkapi dengan fitur *input-output* kosakata dalam bentuk suara, tetapi pada proses penggunaannya masih ada tahap yang memerlukan indera penglihatan, yaitu pada tahap konfirmasi kebenaran tulisan yang telah diucapkan. Penyandang tunanetra tidak bisa melihat konfirmasi bentuk tulisan yang ditampilkan pada aplikasi tersebut sehingga aplikasi tersebut perlu dimodifikasi supaya lebih cocok digunakan untuk penyandang tunanetra.

Modifikasi yang dimaksud, yaitu pada tahap konfirmasi akan ada suara yang mengeja per huruf kata yang telah diucapkan pengguna tunanetra. Apabila konfirmasi tersebut benar

sesuai dengan maksud pengguna maka aplikasi akan dilanjutkan untuk mengetahui makna kata sesuai kamus. Akan tetapi, apabila konfirmasi tersebut tidak sesuai dengan maksud pengguna maka aplikasi akan meminta pengguna untuk mengucapkan lagi kata yang dimaksud sampai benar terkonfirmasi.

KBBI elektronik yang dilengkapi suara diharapkan akan memudahkan pengguna ponsel pintar bagi penyandang tunanetra. Kemajuan teknologi diharapkan dapat diaplikasikan untuk mewujudkan kamus tersebut. Oleh sebab itu, perlu dikembangkan aplikasi kamus elektronik yang sesuai dengan kebutuhan penyandang tunanetra.

1.1 Masalah

Masalah yang dibahas pada makalah ini terdiri atas dua hal sebagai berikut.

- 1) Bagaimana spesifikasi aplikasi kamus elektronik yang ramah bagi pengguna ponsel pintar penyandang tunanetra?
- 2) Bagaimana prosedur pengembangan aplikasi kamus elektronik yang ramah bagi pengguna ponsel pintar penyandang tunanetra?

1.2 Tujuan

Tujuan yang dibahas pada makalah ini terdiri atas dua hal sebagai berikut.

- 1) Mengetahui spesifikasi aplikasi kamus elektronik yang ramah bagi pengguna ponsel pintar penyandang tunanetra.
- 2) Mengetahui prosedur pengembangan aplikasi kamus elektronik yang ramah bagi pengguna ponsel pintar penyandang tunanetra.

1.3 Tinjauan Pustaka

1.3.1 Leksikografi

Terdapat tiga subsistem bahasa, yaitu fonologi, gramatika, dan leksikon. Fonologi merupakan sistem bahasa yang berkaitan dengan tata bunyi bahasa. Gramatika merupakan sistem bahasa yang berkaitan dengan bentuk dan susunan kata maupun kalimat. Leksikon merupakan sistem bahasa yang berkaitan kosakata atau perbendaharaan kata. Berdasarkan ketiga subsistem tersebut, pembahasan makalah ini difokuskan pada subsistem leksikon.

Ilmu yang membahas subsistem leksikon adalah leksikologi. Leksikologi sebagai ilmu yang membahas kosakata atau perbendaharaan kata pada suatu bahasa tentunya harus mampu menampung berbagai kosakata atau perbendaharaan kata yang ditemukan pada suatu bahasa sehingga leksikologi memerlukan proses pencatatan kosakata atau perbendaharaan kata yang ditemukan tersebut supaya semuanya dapat terkumpul dan tersusun secara efektif. Pencatatan kosakata atau perbendaharaan kata tersebut dinamakan leksikografi.

Pencatatan kosakata atau perbendaharaan kata yang dilakukan oleh leksikograf, pakar leksikografi, memerlukan media sebagai tempat pengumpulan kosakata perbendaharaan kata yang telah ditemukan, yaitu kamus. Kamus tidak hanya sebagai media pengumpulan kosakata atau perbendaharaan kata, tetapi kamus juga dilengkapi dengan definisi atau arti kosakata

yang terkumpul di dalamnya (Chaer 2007:177). Kerja leksikograf dalam mewujudkan sebuah kamus memerlukan berbagai tahapan mulai dari pengumpulan data yang berupa kata, penyeleksian kata yang sesuai dengan jenis kamus yang akan dibuat, sampai penentuan kata yang baku dan tidak baku, baik dari sisi penulisan maupun pembentukan kata (Setiawan 2015:2).

1.3.2 Jenis-jenis Kamus

Jenis-jenis kamus yang dibahas dibedakan menjadi empat jenis, yakni kamus berdasarkan bahasa sasaran, kamus berdasarkan ukuran, kamus berdasarkan isi, dan kamus berdasarkan media.

Pertama, berdasarkan bahasa sasaran. Kamus berdasarkan bahasa sasaran dibedakan menjadi tiga, yaitu kamus ekabahasa, kamus dwibahasa, dan kamus anekabahasa (Chaer 2007:196).

Kedua, berdasarkan ukuran. Kamus berdasarkan ukuran dibedakan menjadi dua, yakni kamus besar dan kamus terbatas (Chaer 2007:198).

Ketiga, berdasarkan isi. Kamus berdasarkan isinya dibedakan menjadi dua, yakni kamus umum dan kamus khusus (Chaer 2007:202).

Keempat, berdasarkan jenis media. Kamus berdasarkan jenis media dibedakan menjadi dua, yakni kamus kertas dan kamus elektronik.

1.3.3 Kamus yang Ideal

Kamus yang ideal menurut Chaer (2007:206) memiliki kriteria sebagai berikut (1) kelengkapan lema; (2) kemudahan sistematika penyusunan lema; (3) kelengkapan, ketepatan, dan kejelasan makna; (4) pemberian petunjuk lafal dan ejaan; (5) pemberian informasi kategori kata; (6) pemberian informasi variasi kata; (7) pemberian informasi asal-usul kata; (8) pemberian informasi bidang pemakaian kata; (9) pemberian informasi wilayah pemakaian kata; (10) pemberian informasi kelas sosial kata; dan (11) pemberian informasi kata-kata baku.

1.3.4 Aplikasi pada Ponsel Pintar

Aplikasi *mobile* adalah perangkat lunak yang dapat dioperasikan pada ponsel pintar, tablet, dan perangkat sejenisnya. Tolle, dkk. (2017:16-18) membagi aplikasi *mobile* pada ponsel pintar menjadi dua jenis, yaitu aplikasi berdasarkan teknologinya dan aplikasi berdasarkan fungsinya. Kedua jenis tersebut dijelaskan sebagai berikut.

Jenis aplikasi pada ponsel pintar ditinjau berdasarkan teknologinya antara lain sebagai berikut.

- 1) *Short message service* (SMS) merupakan aplikasi untuk mengirim pesan. Aplikasi ini paling banyak digunakan pada ponsel pintar. Perkembangan aplikasi SMS yang populer digunakan saat ini adalah aplikasi notifikasi. Contohnya, Gmail merupakan aplikasi yang memiliki fitur notifikasi ketika ada email yang masuk.

- 2) Halaman *web mobile* merupakan halaman web yang dirancang khusus untuk penggunaan pada perangkat *mobile*. Halaman pada *web mobile* menyesuaikan dengan ukuran layar ponsel.
- 3) *Mobile web widget* merupakan aplikasi web berukuran kecil atau berupa sebagian isi web yang dibuat dengan format HTML, tetapi tidak dibuka *browser*, melainkan aplikasi lain untuk mengeksekusinya, misalnya *Opera Widget*.
- 4) Aplikasi *web mobile* merupakan aplikasi yang tidak membutuhkan instalasi aplikasi karena pengguna akan selalu mendapat versi terbaru.
- 5) Aplikasi *native* merupakan aplikasi yang memerlukan instalasi sebelum digunakan. Aplikasi ini dikembangkan secara khusus pada platform tertentu, misalnya untuk Android.

Jenis aplikasi pada ponsel pintar ditinjau berdasarkan fungsinya antara lain sebagai berikut.

- 1) Internet, aplikasi *web browser* digunakan untuk mencari informasi dunia maya.
- 2) Hiburan, aplikasi *mobile* yang digunakan untuk mencari hiburan, misalnya menyaksikan film secara *streaming* maupun mendengarkan musik.
- 3) Permainan, aplikasi jenis ini digunakan untuk membentuk suatu permainan yang dapat digunakan oleh pengguna.
- 4) Navigasi, aplikasi jenis ini menggunakan fitur GPS yang ditanam pada perangkat *mobile* sehingga dapat digunakan untuk menyediakan informasi lokasi.
- 5) Media sosial, aplikasi jenis ini dapat membantu pengguna untuk berbagi cerita, foto, dan informasi lain dengan pengguna perangkat *mobile* lainnya.
- 6) Fotografi, aplikasi jenis ini digunakan untuk menonjolkan kemampuan kamera yang saat ini populer digunakan.
- 7) *Utility*, aplikasi jenis ini memiliki fungsi untuk membantu kebutuhan sehari-hari penggunanya.

1.3.5 Karakteristik Berbahasa Penyandang Tunanetra

Penyandang tunanetra dapat berbahasa secara lisan dan tulis. Penyandang tunanetra berbahasa verbal seperti manusia pada umumnya, yaitu mereka dapat berbicara dan mendengar ucapan verbal sehari-hari layaknya percakapan manusia pada umumnya. Pada proses komunikasi, mereka menggunakan rangsang suara/audio sehingga tidak berpengaruh terhadap rangsang penglihatan. Akan tetapi, pada saat mereka hendak berbahasa tulis, mereka menggunakan simbol bahasa yang khusus untuk penyandang tunanetra, yaitu huruf Braille. Huruf Braille adalah tanda-tanda yang didasarkan atas penempatan titik-titik pada enam posisi yang tersusun vertikal, masing-masing tiga titik (Pradopo, Suharto, dan Tobing 1977). Sistem huruf Braille meliputi penulisan abjad Braille, tanda baca, tanda-tanda lain, dan singkatan.

Penulisan titik-titik huruf Braille dicetak timbul karena penyandang tunanetra menggunakan indera peraba untuk membacanya. Berdasarkan hal tersebut, huruf Braille memiliki ciri khas yang berbeda dengan huruf pada umumnya, yaitu penulisannya membutuhkan waktu yang cukup lama karena mereka menggunakan alat khusus untuk

menulis dan tulisan huruf Braille tidak dapat diperkecil sehingga memerlukan media tulis yang lebih banyak (Pradopo, Suharto, dan Tobing 1977).

II. SPESIFIKASI PRODUK DAN METODE PENELITIAN PENGEMBANGAN

2.1 Spesifikasi Produk

Spesifikasi aplikasi yang akan dibuat terdiri atas empat hal utama, yaitu isi, sistematika penyajian, bahasa, dan tampilan. Keempat hal tersebut dibahas sebagai berikut.

Pertama, isi. Isi kamus yang akan dihasilkan mengacu pada pendetailan jenis-jenis kamus yang telah dijelaskan sebelumnya. Berdasarkan bahasa sasaran, kamus ini adalah kamus ekabahasa. Berdasarkan isi, kamus ini adalah kamus umum yang mencakup seluruh kosakata di KBBI dan tidak menjelaskan kosakata pada bidang tertentu saja.

Kedua, sistematika penyajian. Sistematika penyajian kamus yang akan dihasilkan ini mengacu pada sistematika urutan kosakata yang terdapat di KBBI, yaitu disusun berdasarkan alfabetis A-Z.

Ketiga, perangkat lunak. Perangkat lunak yang digunakan adalah android studio 2.1-8.1 dan *voice search*.

Keempat, bahasa. Bahasa yang digunakan pada aplikasi ini adalah bahasa Indonesia yang benar, baik, efektif, dan komunikatif.

Kelima, tampilan. Berdasarkan salah satu jenis media yang telah dijelaskan, kamus ini merupakan kamus elektronik yang berupa aplikasi *mobile*.

2.2 Metode Penelitian Pengembangan

Metode penelitian pengembangan yang digunakan untuk membuat kamus elektronik ini mengacu pada model 4D (*four D model*) yang terdiri atas empat tahapan, yakni *define* (penetapan), *design* (perancangan), *development* (pengembangan), dan *diseminate* (penyebaran) (Thiagarajan, Semmel, dan Semmel 1974:6—9).

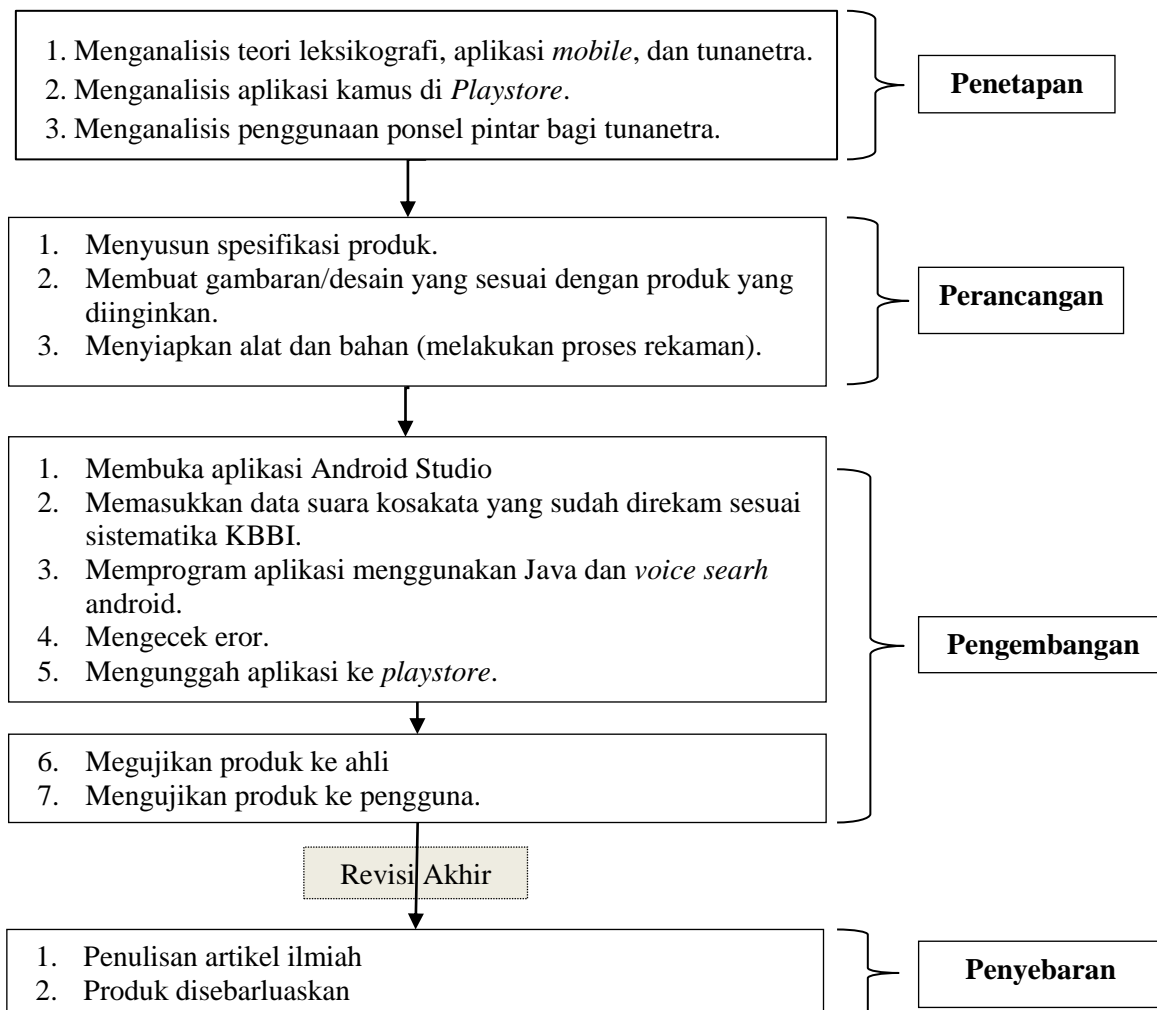
Pertama, penetapan. Tujuan yang ingin dicapai pada tahap ini adalah mendefinisikan dan menetapkan kebutuhan-kebutuhan yang diperlukan sebagai bahan pengembangan produk. Kebutuhan tersebut didapatkan dari dua cara, yakni (1) analisis konsep dan (2) analisis fakta, contohnya masalah yang dihadapi pengguna.

Kedua, perancangan. Tujuan yang ingin dicapai pada tahap ini adalah mendesain draf produk. Pendesaian produk terdiri atas tiga langkah, yaitu (1) menyusun spesifikasi produk, (2) membuat gambaran/desain yang sesuai dengan produk yang diinginkan, dan (3) memilih alat dan bahan.

Ketiga, pengembangan. Tujuan yang ingin dicapai pada tahap ini adalah pengembangan produk. Pada tahap ini dikemukakan dua kegiatan, yaitu (1) pembuatan produk dan (2) pengujian produk.

Keempat, penyebaran. Tujuan yang ingin dicapai pada tahap ini adalah menyebarluaskan produk. Pada tahap ini dilakukan (1) publikasi karya ilmiah dan (2) penyebaran produk.

Prosedur pengembangan ini digunakan sebagai acuan untuk mengembangkan produk mulai tahap penetapan hingga tahap pengembangan. Prosedur penelitian ini disajikan dalam bentuk Bagan 1 berikut.



Bagan 1 Prosedur Penelitian Pengembangan

III. PROSEDUR PENGGUNAAN PRODUK YANG DIHARAPKAN

Aplikasi kamus elektronik yang ramah bagi pengguna ponsel pintar penyandang tunanetra tentunya memiliki prosedur penggunaan. Prosedur penggunaan produk yang diharapkan dijelaskan sebagai berikut.

- 1) Pengguna membuka aplikasi kamus yang telah dikembangkan.
- 2) Aplikasi memberikan informasi bahwa siap digunakan.
- 3) Pengguna mengucapkan kosakata yang diinginkan, kemudian kosakata yang diucapkan tertulis di aplikasi.
- 4) Aplikasi mengeja kosakata yang tertulis per huruf.
- 5) Aplikasi mengonfirmasi ketepatan huruf yang dimaksud oleh pengguna “Apakah kosakata yang dimaksud sudah sesuai?”.

- i. Jika kata yang dieja sudah tepat, pengguna mengatakan “Ya” dan aplikasi langsung mencocokkan kata yang dimaksud kumpulan kosakata bahasa Indonesia sesuai KBBI. Jika kata yang dimaksud ada, aplikasi langsung membacakan makna kata tersebut. Akan tetapi, apabila kosakata yang dicocokkan tidak terdapat pada KBBI, maka aplikasi akan mengeluarkan suara “kosakata tidak ada”, lalu kembali ke menu utama mendapat suara dari pengguna lagi
- ii. Jika kata yang dieja belum tepat, pengguna mengatakan “Belum”, lalu aplikasi akan kembali ke menu utama untuk mendapat suara dari pengguna lagi.

IV. SIMPULAN

Pengembangan aplikasi kamus elektronik yang ramah bagi pengguna ponsel pintar penyandang tunanetra perlu dilakukan. Pengembangan aplikasi tersebut memerlukan kolaborasi dari tiga ahli dengan bidang yang berbeda, yaitu ahli bidang leksikografi, ahli bidang teknologi, dan ahli bidang ilmu tunanetra. Ketiganya perlu berkolaborasi dengan baik supaya aplikasi yang dikembangkan ini dapat menjadi aplikasi yang sesuai dengan kebutuhan penyandang tunanetra.

Daftar Pustaka

- Chaer, A. 2007. *Leksikologi dan Leksikografi Indonesia*. Jakarta: PT Rineka Cipta.
- Pradopo, S., Suharto, dan Tobing, L. 1977. *Pendidikan Anak-anak Tunanetra*. Jakarta: Departemen Pendidikan dan Kebudayaan.
- Prayitno. 2013. *Pengembangan Aplikasi Bahan Ajardalam Perangkat Bergerak*. Jakarta: JARC Seamolec.
- Setiawan, T. 2015. *Leksikografi*. Yogyakarta: Penerbit Ombak.
- Thiagarajan, S., Semmel, D., dan Semmel, M. 1974. *Instructional Development for Training Teacher of Exceptional Children*. Minnesota: The Council for Exceptional Children.
- Tolle, H., Pinandito, A., Kharisma, A. P., dan Dewi, R. K. 2017. *Pengembangan Aplikasi Perangkat Bergerak (Konsep & Implementasi)*. Malang: UB Press.

ALGORITMA PELESAPAN SISIPAN PADA PROSES *STEMMING* BAHASA INDONESIA UNTUK PEMBENTUKAN KATA DASAR (*INFIX STRIPPING ALGORITHM FOR STEMMING INDONESIAN TO FORM BASE WORD*)

Yeni Anistyasari¹, Ekohariadi², Ricky Eka Putra³

Jurusan Teknik Informatika, Universitas Negeri Surabaya

¹yenian@unesa.ac.id, ²ekohariadi@unesa.ac.id, ³rickyeka@unesa.ac.id

Abstrak

Salah satu faktor penting dalam pembentukan kamus elektronik adalah pembentukan kata dasar melalui proses *stemming*. *Stemming* adalah proses pengubahan sebuah kata ke bentuk dasarnya dengan menghilangkan imbuhan yang terdiri dari awalan, akhiran, awalan dan akhiran, dan sisipan. *Stemming* digunakan sebagai proses awal temu kembali informasi. Beberapa algoritma *stemming* untuk Bahasa Indonesia telah dikembangkan, diantaranya adalah algoritma Nazief-Adriani (NA) dan *Confix-Stripping* (CS). Kedua algoritma ini telah teruji dalam proses *stemming* Bahasa Indonesia untuk membentuk kata dasar dengan menghilangkan awalan, akhiran, dan awalan-akhiran. Namun, algoritma NA dan CS tidak banyak mengeksplorasi proses pelepasan sisipan dalam pembentukan kata dasar. Penelitian ini mengajukan ide untuk memperbaiki algoritma *stemming* yang telah ada dengan menambahkan proses pelepasan sisipan. Untuk menguji algoritma yang diajukan, sebuah aplikasi *stemming* berbasis web dikembangkan. Aplikasi ini divalidasi oleh pakar teknologi informasi. Masukan yang digunakan untuk pengujian algoritma pelepasan sisipan adalah berita dari portal daring. Kata dasar yang dihasilkan aplikasi dari proses *stemming* divalidasi oleh pakar Bahasa Indonesia untuk menentukan efektifitas algoritma yang diajukan. Hasil validasi dari pakar teknologi informasi menunjukkan bahwa aplikasi layak digunakan untuk proses uji coba. Sedangkan hasil validasi pakar Bahasa Indonesia membuktikan bahwa algoritma pelepasan sisipan dapat menghasilkan kata dasar yang baik dan sesuai dengan aturan Bahasa Indonesia.

Abstract

One important factor in the development of an electronic dictionary is the formation of base words or root form through stemming. Stemming is the process of converting a word to its root form by removing affixes including prefix, suffix, prefix and suffix, and infix. Stemming is utilized as the initial process of information retrieval. Several stemming algorithms for Indonesian have been developed, including the Nazief-Adriani (NA) and Confix-Stripping (CS) algorithms. Both of these algorithms have been tested in Indonesian stemming process to form basic words by removing prefix, suffix, and prefix. However, NA and CS algorithms have not explored deeply about removing infix. This research therefore proposes an idea for

improving existing stemming algorithms by adding an infix stripping process. To test the proposed algorithm, a web-based stemming application was developed. This application is validated by information technology experts. The input used for testing the algorithm is news from online portal. The root forms generated by the application of the stemming process is validated by Indonesian experts to determine the effectiveness of the proposed algorithm. Validation results from information technology experts indicate that the application is eligible to be used. While the validation results of Indonesian experts prove that infix stripping algorithm can generate appropriate root forms and in accordance with the rules of Indonesian.

I. PENDAHULUAN

Bahasa Indonesia adalah bahasa utama yang digunakan orang Indonesia untuk berkomunikasi baik secara lisan maupun tulisan. Pemilihan kata dalam pengucapan dan dalam menulis kalimat adalah elemen penting. Dalam hal tata Bahasa Indonesia baku, Kamus Besar Bahasa Indonesia adalah referensi utama dalam penggunaan kata dasar. Selain untuk berkomunikasi lisan, penggunaan bahasa Indonesia yang tepat diperlukan saat menulis dokumen formal, jurnal, laporan, dan sebagainya. Dokumen formal yang baik adalah dokumen yang ditulis dalam Bahasa Indonesia dengan tingkat formalitas yang tinggi, menggunakan kosa kata yang tepat, dan menghindari kesalahan tata bahasa dan ejaan (Widjaja and Hansun 2015).

Penelitian tentang aplikasi teknologi informasi untuk Bahasa Indonesia banyak dilakukan, antara lain sistem temu kembali informasi, penarikan kesimpulan dari isi dokumen, kompresi teks, dan klasifikasi teks. Proses penting yang terlibat dalam mengembangkan aplikasi tersebut adalah proses *stemming* yaitu proses yang dilakukan untuk mengembalikan kata berimbuhan ke bentuk dasarnya dengan cara menghilangkan imbuhan. Imbuhan tersebut berupa awalan, akhiran, awalan dan akhiran, dan sisipan (Winarti, Kerami, and Arief 2017). Dalam klasifikasi teks, *stemming* bertugas menyederhanakan kata-kata tanpa menghilangkan makna sehingga ukuran kumpulan data akan berkurang. Dalam penarikan kesimpulan dari isi teks, *stemming* dapat meningkatkan efisiensi. Sedangkan untuk temu kembali informasi, *stemming* mengurangi jumlah indeks dokumen sehingga meningkatkan kinerja sistem temu kembali informasi, menyediakan varian morfologi dalam mencari istilah, dan meningkatkan akurasi. Sedangkan dalam kompresi teks, *stemming* memaksimalkan penggunaan kamus dengan menyimpan hanya kata dasar dalam kamus (Widayanto and Huda 2017).

Stemming Bahasa Indonesia pertama dikembangkan oleh Nazief-Adriani kemudian Jelita Asia memperbaiki algoritma tersebut dan menyebut algoritmanya dengan nama *confix stripping* (CS) *stemmer*. Ada banyak perbaikan yang dilakukan oleh CS *stemmer* sehingga menjadi algoritma dengan akurasi *stemmer* tertinggi (Widayanto and Huda 2017). Pembentukan kata dasar pada kedua algoritma tersebut dilakukan dengan menghilangkan imbuhan. Namun, kedua algoritma tidak membahas secara mendalam tentang sisipan. Seperti diketahui, imbuhan pada Bahasa Indonesia terdiri dari awalan, akhiran, awalan dan akhiran, serta sisipan. Oleh karena itu, penelitian ini mengajukan ide untuk mengembangkan algoritma pelepasan sisipan untuk memperbaiki algoritma *stemming* yang telah ada. Sumber

dokumen masukan untuk proses uji coba algoritma berasal dari portal berita daring. Sebelum dilakukan proses *stemming*, dokumen yang diunduh diproses tokenisasi untuk mendapatkan tiap kata dari kumpulan kalimat yang panjang.

II. KERANGKA TEORI

2.1. *Stemming*

Stemming adalah proses untuk menggabungkan atau memecahkan setiap varian-varian suatu kata menjadi kata dasar. Proses *stemming* pada kata Bahasa Indonesia berbeda dengan *stemming* Bahasa Inggris. Proses *stemming* pada kata Bahasa Inggris adalah proses untuk mengeliminasi akhiran pada kata, sementara proses *stemming* Bahasa Indonesia adalah proses untuk mengeliminasi awalan, akhiran, awalan dan akhiran, dan sisipan (Wibowo 2016; Setiawan et al. 2016).

2.2. Algoritma Nazief-Adriani

Algoritma *stemming* Nazief dan Adriani dikembangkan adalah berdasarkan aturan morfologi Bahasa Indonesia yang mengelompokkan imbuhan menjadi awalan (*prefix*), sisipan (*infix*), akhiran (*suffix*) dan gabungan awalan akhiran (*confixes*). Algoritma ini menggunakan kamus kata dasar dan mendukung *recoding*, yakni penyusunan kembali kata-kata yang mengalami proses *stemming* berlebih (Nazief 2000). Aturan morfologi Bahasa Indonesia mengelompokkan imbuhan ke dalam beberapa kategori sebagai berikut:

1. *Inflection suffixes* yakni kelompok akhiran yang tidak merubah bentuk kata dasar. Sebagai contoh, kata “duduk” yang diberikan akhiran “-lah” akan menjadi “duduklah”. Kelompok ini dapat dibagi menjadi dua:
 - a. *Particle* (P) atau partikel yakni termaksud di dalamnya “-lah”, “-kah”, “-tah” dan “-pun”.
 - b. *Possessive pronoun* (PP) atau kata ganti kepemilikan, termaksud di dalamnya “-ku”, “-mu” dan “-nya”.
2. *Derivation suffixes* (DS) yakni kumpulan akhiran asli Bahasa Indonesia yang secara langsung ditambahkan pada kata dasar yaitu akhiran “-i”, “-kan”, dan “-an”.
3. *Derivation prefixes* (DP) yakni kumpulan awalan yang dapat langsung diberikan pada kata dasar murni, atau pada kata dasar yang sudah mendapatkan penambahan sampai dengan 2 awalan termaksud di dalamnya adalah:
 - a. Awalan yang dapat bermorfologi (“me-“, “be-“, “pe-“ dan “te”).
 - b. Awalan yang tidak bermorfologi (“di-“, “ke-“ dan “se-“).

Berdasarkan pengklasifikasi imbuhan-imbuhan di atas, maka bentuk kata berimbuhan dalam

$$[DP + [DP + [DP]]] KATA DASAR [+DS] + PP$$

Keterangan :

DP : *Derivation prefixes*

DS : *Derivation suffixes*

PP : *Possessive pronoun*

Dengan model Bahasa Indonesia di atas serta aturan-aturan dasar morfologi Bahasa Indonesia, aturan yang digunakan dalam proses algoritma Nazief dan Adriani sebagai berikut:

1. Tidak semua kombinasi awalan dan akhiran diperbolehkan. Kombinasi-kombinasi imbuhan yang tidak diperbolehkan, yaitu “be-i”, “ke-i”, “ke-kan”, “me-an”, “se-i”, “se-kan” dan “te-an”.
2. Penggunaan imbuhan yang sama secara berulang tidak diperkenankan.
3. Jika suatu kata hanya terdiri dari satu atau dua huruf, maka proses tidak dilakukan.
4. Penambahan suatu awalan tertentu dapat mengubah bentuk asli kata dasar, ataupun awalan yang telah diberikan sebelumnya pada kata dasar bersangkutan. Sebagai contoh, awalan “me-“ dapat berubah menjadi “meng-“, “men-“, “meny-“, dan “mem-“. Oleh karena itu diperlukan suatu aturan yang mampu mengatasi masalah morfologi ini.

Algoritma Nazief dan Adriani memiliki tahap-tahap sebagai berikut:

1. Cari kata dalam kamus jika ditemukan maka diasumsikan bahwa kata tersebut adalah kata dasar. Algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 2.
2. Hilangkan *inflectional suffixes* bila ada. Dimulai dari *inflectional particle* (“-lah”, “-kah”, “-tah” dan “-pun”), kemudian *possessive pronoun* (“-ku”, “-mu” dan “-nya”). Cari kata pada kamus jikaditemukan algoritma berhenti, jika kata tidak tidak ditemukan dalam kamus lakukan langkah 3.
3. Hilangkan *derivation suffixes* (“-an”, “-i” dan “-kan”). Jika akhiran “-an” dihapus dan ditemukan akhiran “-k”, maka akhiran “-k” dihapus. Cari kata pada kamus jika ditemukan algoritma berhenti, jika kata tidak tidak ditemukan maka lakukan langkah 4.
4. Pada langkah 4 terdapat tiga iterasi.
 - a. Iterasi berhenti jika:
 - 1) Ditemukannya kombinasi awalan yang tidak diizinkan berdasarkan awalan.

Tabel 1. Kombinasi awalan dan akhiran yang tidak diizinkan

Awalan	Akhiran yang tidak diizinkan
be-	-i
di-	-an
ke-	-i, -kan
Me	-an
Se	-i,kan

- 2) Awalan yang dideteksi saat ini sama dengan awalan yang dihilangkan sebelumnya.
 - 3) Tiga awalan telah dihilangkan.
- b. Identifikasikan tipe awalan dan hilangkan. Awalan terdiri dari dua tipe:
- 1) Standar (“di-“, “ke-“, “se-“) yang dapat langsung dihilangkan dari kata.
 - 2) Kompleks (“me-“, “be-“, “pe-“, “te-“) adalah tipe-tipe awalan yang dapat bermorfologi sesuai kata dasar yang mengikutinya. Oleh karena itu dibutuhkan aturan pada tabel berikut untuk mendapatkan hasil pemenggalan yang tepat.

- c. Cari kata yang telah dihilangkan awalnya. Apabila tidak ditemukan, maka langkah 4 diulang kembali. Apabila ditemukan, maka algoritma berhenti.

2.3. Algoritma *Confix Stripping*

Algoritma *confix stripping* (CS) (Adriani et al. 2007) dikembangkan untuk memperbaiki kinerja algoritma Nazief dan Adriani. Perbaikan *stemming* Bahasa Indonesia yang digunakan algoritma CS terhadap algoritma Nazief dan Adriani sebagai berikut:

1. Penggunaan kamus kata dasar yang lengkap
2. Penambahan aturan untuk kata ulang, seperti “buku-buku”
3. Penambahan aturan untuk awalan dan akhiran
 - a. Akhiran “pun”
 - b. Awalan “ter”
 - c. Awalan “pe”
 - d. Awalan “me”
4. Penambahan proses pengecekan *rulePrecedence*

III. METODE PENELITIAN

Metode yang digunakan untuk penelitian ini terdiri dari analisa kebutuhan, desain algoritma, implementasi algoritma, dan evaluasi. Tahap analisa kebutuhan yang dilakukan yaitu studi literatur tentang metode *stemming* Bahasa Indonesia yang telah dikembangkan oleh peneliti lain. Temuan yang diperoleh dari tahap analisa kebutuhan adalah hingga saat ini metode *stemming* Bahasa Indonesia yang menyertakan pelesapan sisipan belum diteliti secara mendalam. Tahap analisa kebutuhan juga mempelajari aturan sisipan pada Bahasa Indonesia.

Setelah melalui tahap analisa kebutuhan dilakukan desain algoritma. Algoritma yang dikembangkan dinamakan *infix-stripping* yaitu perbaikan dari algoritma Nazief dan Adriani dan algoritma CS, seperti ditunjukkan Gambar 1.



Gambar 1. Algoritma Infix-stripping

Tabel 2. Hasil Uji Coba Algoritma *Infix-stripping*

No	Kata Asli	Hasil <i>stemming</i>	Penilaian Validator		
			1	2	3
1	seruling	suling	1	1	1
2	kerudung	kudung	1	1	1
3	serabut	sabut	1	1	1
4	terperanjat	panjat	1	1	1
5	gerigi	gigi	1	1	1
6	genderang	gendang	1	1	1
7	reruntuhan	runtuh	1	1	1
8	reramuan	ramu	1	1	1
9	telunjuk	tunjuk	1	1	1
10	gelegar	gegar	1	1	1

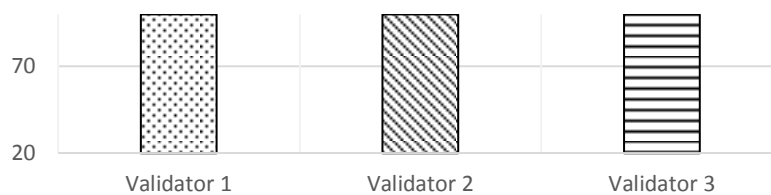
11	telapak	tapak	1	1	1
12	penyelidikan	sidik	1	1	1
13	gelembung	gembung	1	1	1
14	jelajah	jajah	1	1	1
15	pelatuk	patuk	1	1	1
16	leluhur	luhur	1	1	1
17	melaju	maju	1	1	1
18	telangkup	tangkap	1	1	1
19	gemerlap	gerlap	1	1	1
20	gemetar	getar	1	1	1
21	gemilang	gilang	1	1	1
22	gemuruh	guruh	1	1	1
23	gemilap	gilap	1	1	1
24	kemelut	kelut	1	1	1
25	kemilap	kilap	1	1	1
26	kemilau	kilau	1	1	1
27	kemuning	kuning	1	1	1
28	semerbak	serbak	1	1	1
29	seminar	sinar	1	1	1
30	temali	tali	1	1	1
31	temaram	taram	1	1	1
32	temurun	turun	1	1	1
33	kemuncup	kuncup	1	1	1
34	jemari	jari	1	1	1
35	semilir	silir	1	1	1
36	gementar	gentar	1	1	1
37	gemertak	gertak	1	1	1
38	cemerlang	cerlang	1	1	1
39	peranjat	panjat	1	1	1
40	gemulung	gulung	1	1	1

Setelah desain algoritma, dilakukan implementasi dan uji coba algoritma. Data untuk uji coba diambil dari kalimat-kalimat yang ditemukan pada portal berita daring. Berita dari portal daring diunduh dan disimpan otomatis. Kemudian dilakukan proses pemisahan antar kalimat dengan mendeteksi tanda baca titik (.), tanda seru (!), dan tanda tanya (?). Setelah setiap kalimat terpisah, dilakukan pemisahan tiap kata dengan mendeteksi spasi dan tanda koma (,). Kata yang terbentuk ini akan dilakukan proses *stemming*. Hasil implementasi algoritma dan uji coba ditunjukkan di Tabel 2. Empat puluh kata diuji coba untuk menguji kinerja algoritma *infix-stripping*. Keempat puluh kata tersebut mengandung sisipan el, em, dan er. Kemudian, hasil *stemming* dievaluasi oleh tiga orang validator Bahasa Indonesia. Validator diminta mengisi angka 0 jika hasil *stemming* tidak ditemukan pada Bahasa Indonesia baku dan angka 1 jika hasil *stemming* termasuk kata Bahasa Indonesia baku.

IV. ANALISIS PENELITIAN

Dari Tabel 2 dapat dilakukan analisa mengenai kata masukan yang dijadikan data uji coba, hasil penilaian oleh validator, dan kesimpulan algoritma *infix-stripping* yang telah diimplementasikan. Kata yang digunakan sebagai masukan uji coba berjumlah empat puluh kata dan semuanya mengandung kata sisipan. Beberapa kata mengandung imbuhan. Hasil *stemming* menunjukkan bahwa kata yang bersisipan dapat diubah menjadi kata dasar dengan menghilangkan sisipan el, em, atau er. Sedangkan kata yang masih berimbuhan, akan dihilangkan imbuhanannya terlebih dahulu lalu dihilangkan sisipannya.

Tiga orang pakar Bahasa Indonesia diminta untuk memvalidasi hasil algoritma *infix-stripping*. Hasil validasi validator menunjukkan bahwa kata yang dihasilkan dari proses *stemming* merupakan kata baku Bahasa Indonesia. Validator 1 menilai seluruh (100%) kata dasar dari hasil *stemming* adalah kata dasar baku yang ditemukan dalam Kamus Besar Bahasa Indonesia. Validator 2 menilai seluruh (100%) kata dasar dari hasil *stemming* adalah kata dasar baku yang ditemukan dalam Kamus Besar Bahasa Indonesia. Demikian pula dengan validator 3, seluruh (100%) kata dasar dari hasil *stemming* adalah kata dasar baku yang ditemukan dalam Kamus Besar Bahasa Indonesia. Grafik yang menunjukkan hasil penilaian validator ditunjukkan di Gambar 2.



Gambar 2. Grafik penilaian validator Bahasa Indonesia

V. KESIMPULAN

Penelitian ini mengajukan ide untuk memperbaiki algoritma Nazief dan Adriani dan algoritma *confix-stripping* pada proses *stemming* Bahasa Indonesia dengan menambahkan fitur pelepasan sisipan. Beberapa kata bersisipan tanpa imbuhan dan berimbuhan diuji coba untuk mengetahui kinerja algoritma yang diajukan. Hasil penilaian yang dilakukan oleh ahli Bahasa Indonesia terhadap hasil *stemming* menunjukkan bahwa algoritma yang diajukan mampu mengembalikan kata sisipan ke bentuk aslinya. Penelitian selanjutnya yang dapat dikembangkan sebagai tindak lanjut dari studi ini adalah membandingkan hasil algoritma *infix-stripping* dengan algoritma Nazief dan Adriani dan algoritma *confix-stripping*.

Daftar Pustaka

- Adriani, Mirna, Jelita Asian, Bobby Nazief, and Hugh E Williams. 2007. "Stemming Indonesian : A Confi X-Stripping Approach." *ACM Transactions on Asian Language Information Processing (TALIP)* 6 (4):1-33.
- Nazief, Bobby. 2000. "Development of Computational Linguistics Research: A Challenge for Indonesia." *Proceedings of the 38th Annual Meeting on Association for Computational*

Linguistics - ACL '00, 1–2.

- Setiawan, Reina, Aditya Kurniawan, Widodo Budiharto, Iman Herwidiana Kartowisastro, and Harjanto Prabowo. 2016. "Flexible Affix Classification for Stemming Indonesian Language." In *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2016*.
- Wibowo, Julianto. 2016. "Pada Kalimat Bahasa Indonesia Dengan Algoritma Stemming." *Jurnal Riset Komputer (JURIKOM)* 3 (5):346–50.
- Widayanto, Hari, and Arief Fatchul Huda. 2017. "Comparison Nazief Adriani And CS Stemmer Algorithm For Stemm Real Data." *E-Proceeding of Engineering* 4 (3):5215–22.
- Widjaja, Marsel, and Seng Hansun. 2015. "Implementation of Porter's Modified Stemming Algorithm in an Indonesian Word Error Detection Plugin Application." *International Journal of Technology* 6 (2):139–50.
- Winarti, Titin, Jati Kerami, and Sunny Arief. 2017. "Determining Term on Text Document Clustering Using Algorithm of Enhanced Confix Striping Stemming." *International Journal of Computer Application* 157 (9):6.

**APLIKASI TRANSKRIPSI FONETIK BAHASA INDONESIA BERBASIS AI
(ARTIFICIAL INTELLIGENCE) UNTUK PEMBELAJAR BIPA**
*(AI [ARTIFICIAL INTELLIGENCE]-BASED APPLICATION OF BAHASA INDONESIA
PHONETIC TRANSCRIPTION FOR BIPA LEARNERS)*

Endang Sartika¹

Anggy Eka Pratiwi²

¹Sastra Inggris, IAIN Purwokerto

²Informatika, Surya University

^{1,2}(endangsartika.ens@gmail.com ; angyekap@gmail.com)

Abstrak

Penelitian ini dimaksudkan untuk menyajikan rancangan aplikasi kamus transkripsi fonetik Bahasa Indonesia berbasis AI (*Artificial Intelligence*) untuk pembelajar BIPA (Bahasa Indonesia bagi Penutur Asing). Aplikasi ini dirancang menggunakan metode *Artificial Intelligence* dengan *Natural Language Processing* (NLP) teknik. Penarikan data dilakukan dari data tertulis dan data suara yang mengacu pada prinsip fonetik Bahasa Indonesia sesuai dengan standard IPA (*International Phonetic Association*). Aplikasi ini menyediakan fitur pencarian untuk pelafalan dan transkripsi fonetik seperti huruf vokal, diftong dan konsonan disertai dengan contoh kalimat dan penggunaannya. Dengan rancangan berbasis AI, aplikasi ini akan memprediksi pencarian kata yang diinginkan pengguna secara cerdas, akurat dan otomatis. Aplikasi ini diharapkan dapat menjadi penunjang pembelajaran Bahasa Indonesia bagi Penutur Asing (BIPA). Dengan Aplikasi ini pembelajar asing akan dengan mudah mempelajari bahasa Indonesia sesuai dengan pelafalan sesungguhnya. Aplikasi ini dirancang berbasis web dengan harapan dapat dikonfigurasi dengan website maupun aplikasi kamus daring KBBI.

Kata kunci: *Transkripsi Fonetik, Pelafalan, AI, NLP, BIPA*

Abstract

This research is intended to present the design of the Artificial Intelligence (AI) based phonetic transcription application of Bahasa Indonesia for BIPA learners (Bahasa Indonesia for Foreign Speakers). This application is designed using Artificial Intelligence method with Natural Language Processing (NLP) technique. The data is collected from written and voice data referring to the phonetic principle of Bahasa Indonesia in accordance with the International Phonetic Association (IPA) standard. The app provides search features for alphabetical pronunciation and transcription such as vowels, diphthong and consonants along with sentence examples and the usage. Applying AI-based design, this app will predict the user's search terms intelligently, accurately and automatically. This application is expected to support the learning of Bahasa Indonesia

for Foreign Speakers (BIPA). Using this application, foreign learners will easily learn Bahasa Indonesia in accordance with actual pronunciation. This app is a web-based designed app that can be configured with KBBI website and online dictionary app.

Keywords: *Phonetic Transcription, pronunciation, AI, NLP, BIPA*

I. Pendahuluan

1.1. Latar Belakang Masalah

Sebagai bahasa dengan penutur terbanyak ke-lima di dunia dan terbanyak pertama di kawasan ASEAN, Bahasa Indonesia menjadi salah satu bahasa yang dipandang penting dan banyak dipejari oleh masyarakat internasional.ⁱ BIPA merupakan program pembelajaran keterampilan Bahasa Indonesia baik menulis, membaca maupun berbicara bagi penutur asing. Peminat pembelajar BIPA semakin meningkat secara signifikan. Bahasa Indonesia kini telah dipelajari di 45 negara dengan seribu lebih pembelajar BIPA setiap tahunnya yang tersebar di 250 lembaga penyelenggara BIPA baik negeri maupun swasta.ⁱⁱ

Peningkatan tersebut dilatarbelakangi oleh beberapa faktor diantaranya adalah bahwa secara umum, Bahasa Indonesia dinilai sebagai bahasa yang mudah untuk dipelajari. Seperti halnya pendapat Andreev yang dikutip oleh Stokhof (1975: 254) menyatakan bahwa *“In general the phonology of the Indonesian language is simple and clear”*, dikarenakan dalam Bahasa Indonesia tidak terdapat banyak perbedaan antara penulisan dan pelafalan kata yang diproduksi seperti bahasa Inggris atau bahasa lainnya. Selain itu, program Darmasiswa juga menjadi salah satu daya tarik tersendiri bagi masyarakat internasional untuk mempelajari Bahasa Indonesia.

1.2. Masalah

Meskipun Bahasa Indonesia dinilai sebagai bahasa yang mudah untuk dipelajari, penuturan dan pengucapan Bahasa Indonesia masih menjadi kendala bagi penutur asing yang menempatkan Bahasa Indonesia sebagai bahasa kedua mereka. Hal tersebut dikarenakan, Bahasa Indonesia memiliki ragam variasi yang dipengaruhi oleh bahasa daerah, kelompok sosial dan dialek individu. Sebagai contoh terdapat variasi bunyi e dalam pengucapan kata *kerā/kəra*, *monyet /mɔŋɛt/*, *sate /saʔe/* (Chaer, 1994: 109). Hal tersebut juga diungkapkan Stokhof dalam artikelnya yang berjudul *On the Phonology of Bahasa Indonesia* (1974:255) bahwa:

“BI is a quickly changing language which functions as communication in a complex multilingual society. The continuous mutual interference between regional languages, dialects and BI necessarily results in a gamut of local, social and individual variants”

Selain adanya ragam variasi pengucapan dalam Bahasa Indonesia, belum tersedianya transkripsi fonetik dalam Kamus Besar Bahasa Indonesia juga menjadi kendala bagi pembelajar BIPA untuk dapat menguasai pengucapan Bahasa Indonesia

secara komprehensif. Tulisan fonetik merupakan sistem lambang yang dipergunakan untuk mewakili bunyi bahasa (Vehraar, 2006: 40). Tulisan fonetik terdiri atas huruf-huruf aksara latin yang ditambah dengan sejumlah tanda diakritik dan modifikasi (Chaer, 1994:109). Tanda diakritik adalah tanda baca tertentu yang ditambahkan pada lambang bunyi yang mengubah nilai fonetis simbol tersebut baik sedikit maupun banyak sebagaimana diungkapkan Roach bahwa *Diacritics is marks which modify the symbol in some way* (2009: 34), misalnya tanda ['] pada huruf e menjadi [é] (Chaer, 1994).

Menurut pendapat Roach (2009: 33-34) terdapat perbedaan antara transkripsi fonetik dan transkripsi fonemis, transkripsi fonemis merujuk pada transkripsi sejumlah lambang bunyi suatu bahasa tanpa harus menunjukkan bunyi-bunyi tersebut mempunyai fungsi sebagai pembeda makna atau tidak, umumnya transkripsi fonemik ditulis diantara dua garis miring // contohnya *enak* /*enak*/. Sementara transkripsi fonetik mengacu pada transkripsi fonemis yang lebih komprehensif dengan menunjukkan fungsi bunyi tersebut sebagai pembeda makna, umumnya ditulis di dalam tanda []. Misalnya bunyi *p* beraspirasi dilambangkan [p^h] pada kata Bahasa Inggris *Pot* [pɒt] 'panci', berbeda dengan suara *p* tidak beraspirasi pada kata *spot* [sp^hɒt] 'tempat'. Perbedaan penggunaan tanda tersebut digunakan untuk merujuk pada perbedaan bunyi suara yang menyebabkan perbedaan makna antara dua atau lebih ujaran. Dalam bahasa Indonesia tidak terdapat bunyi bahasa yang beraspirasi. Namun dalam judul makalah ini penulis menggunakan istilah transkripsi fonetik untuk merujuk pada penulisan fonemis ujaran Bahasa Indonesia mengikuti standar *International Phonetic Association* (IPA) dimana transkripsi bunyi bahasa umumnya dikenal dengan transkripsi fonetik dari pada transkripsi fonemis.

1.3. Tujuan

Penelitian ini ditujukan untuk mengembangkan aplikasi transkripsi fonetik Bahasa Indonesia untuk memudahkan pembelajar Asing dalam pengucapan Bahasa Indonesia berbasis AI (*Artificial Intelligence*) dengan menggunakan *Natural Language Processing* (NLP) tehnik. Setyowati, Bertalya dan Ningsih (2014) sebelumnya telah terlebih dahulu menggarap aplikasi transkripsi fonetik Bahasa Indonesia dengan memanfaatkan bahasa pemrograman C# dan Microsoft Visual Studio, berpendapat bahwa terdapat beberapa alasan mengapa transkripsi fonetik Bahasa Indonesia belum sepenuhnya digarapⁱⁱⁱ. Salah satunya adalah dikarenakan ciri vokal Bahasa Indonesia yang tidak menunjukkan perbedaan pada pelafalan dalam kata yang diproduksi. Dalam hal ini kami kurang setuju dengan pendapat tersebut, karena pada kenyataannya terdapat bunyi-bunyi Bahasa Indonesia yang berbeda dengan penulisannya. Misalnya huruf J pada kata *Jarak* memiliki bunyi yang sama dengan bunyi G dalam kata bahasa Inggris *Age* 'umur' yang dilambangkan dengan lambang bunyi /dʒ/ berdasarkan IPA sehingga apabila ditulis dengan transkripsi fonetik *jarak* menjadi /dʒara/ dan *age* menjadi /eidʒ/.

Selain belum mengikuti kaidah fonetik sesuai standar IPA secara khusus untuk Bahasa Indonesia, aplikasi yang telah dirancang oleh Setyowati, Bertalya dan Ningsih (2014) tersebut belum dapat diakses oleh khalayak umum karena masih dirancang dengan memanfaatkan bahasa pemrograman C#. Meskipun aplikasi tersebut dapat digunakan pada seluruh perangkat computer dengan beragam operasi *windows*, serta bersifat

portable (dimana pengguna tidak perlu meng-instal ulang aplikasi tersebut ketika ingin digunakan), perangkat tersebut dinilai belum memadai untuk memfasilitasi pembelajar BIPA yang semakin meningkat, terutama bagi mereka yang ingin belajar Bahasa Indonesia dari jauh.

Oleh karena itu, penelitian ini dimaksudkan untuk menyajikan rancangan aplikasi fonetik Bahasa Indonesia berbasis *web* yang dapat diakses oleh masyarakat luas baik lokal maupun internasional. Harapan kami, rancangan aplikasi ini dapat dikonfigurasi dengan kamus KBBI daring baik yang berbasis *web* maupun aplikasi *portable* untuk *telepon seluler*. Aplikasi ini dimaksudkan untuk memudahkan pembelajar BIPA dalam mempelajari pengucapan Bahasa Indonesia secara komprehensif.

1.4. Tinjauan Pustaka

Transkripsi fonetik Bahasa Indonesia bukan tidak digarap sepenuhnya, beberapa pakar linguistik telah terlebih dahulu membahas tentang fonetik dan fonologi Bahasa Indonesia seperti Stokhof (1975), Lapoliwa (1988), Chaer (1994), Verhaar (2006) dan yang terbaru Sodenberg dan Olson (2008). Lapilowa, Chaer dan Verhaar dalam pembahasannya mengenai fonetik dan fonologi Bahasa Indonesia belum menjelaskan konsep fonologi Bahasa Indonesia secara khusus melainkan memperkenalkan tentang ilmu fonetik dan fonologi secara umum yang mencakup bahasa lainnya di dunia. Sementara Stokhof (1975) dalam artikennya membahas tentang *optionalities*, *neutralization* dan *doublets (triplets)* dalam Bahasa Indonesia.^{iv}

Sedangkan Sodenberg dan Olson (2008) dalam artikelnya berjudul *Indonesian* yang dipublikasi oleh *The Journal of International Phonetic Association* menjelaskan konsep fonetik dan fonologi Bahasa Indonesia secara jelas dan terperinci menyesuaikan dengan standar Bahasa Indonesia yang telah disempurnakan. Dalam penelitian ini, kami akan lebih banyak merujuk pada teori fonetik yang diusulkan oleh Sodenberg dan Oslon (2008) untuk lebih menyesuaikan kaidah fonetik Bahasa Indonesia yang telah disempurnakan dengan standar *Internasional Phonetic Association (IPA)*. Tulisan fonetik berbeda dengan tulisan ortografi. Secara umum Bahasa Indonesia memiliki 26 abjad yang digunakan dalam tulisan ortografi. Akan tetapi, ke-26 huruf tersebut tidak merepresentasikan bunyi bahasa yang dihasilkan secara fonetis. Berdasarkan teori fonetik sesuai standar IPA yang disarankan oleh Sodenberg dan Olson (2008), Bahasa Indonesia memiliki 31 bunyi bahasa yang terdiri atas 22 bunyi konsonan, 6 bunyi vokal dan 3 diftong.

1.4.1 Konsonan

Konsonan merupakan bunyi yang dihasilkan dengan mempergunakan artikulasi pada salah satu bagian alat-alat bicara (Vehraar, 2006: 33). Terdapat 22 bunyi konsonan dalam Bahasa Indonesia meskipun secara umum Bahasa Indonesia memiliki 21 huruf ortografi.^v

Untuk mempermudah pemahaman fungsi dan lambang fonetik konsonan Bahasa Indonesia tersaji tabel daftar bunyi konsonan Bahasa Indonesia menurut Soderberg and Olson (2008) sebagai berikut:

Tabel 1.1 Bunyi Konsonan Bahasa Indonesia

	Bilabial	Labio-dental	Dental	Alveolar	Post-Alveolar	Palatal	Velar	Glotal
Plosive & affricative	p b		t̚	D	tʃ dʒ		k g	(ʔ)
Nasal	m			N		ɲ	ŋ	
Flap/Trill				R				
Fricative		(f)		s (z)	(ʃ)			h
Approximant	w					j		
Lateral approximant				L				

Sumber: Soderberg and Olson (2008:210)

1.4.1.1 Plosive & Affricate^{vi}

Plosive & affricative dikenal juga dengan konsonan letupan dalam istilah fonetik Bahasa Indonesia. Konsonan letupan merupakan bunyi konsonan yang dihasilkan dengan adanya hambatan arus udara secara keseluruhan di tempat artikulasi tertentu kemudian dilepaskan kembali (Vehraar, 2006: 34).

Tabel 1.1.1 Bunyi Konsonan *Plosive & Affricate*

	Lambang Fonetik	Lambang Ortografi	Kata	Transkripsi Fonetik
Bilabial	/p/	P	Panjang	/paŋdʒaŋ/
	/b/	B	Baju	/baʒu/
Dental	/t̚/	T	Tampan	/t̚ampan/
Alveolar	/d/	D	Datang	/daʒaŋ/
Post-Alveolar	/tʃ/	C	Cakap	/tʃakap/
	/dʒ/	J	Jarum	/dʒarum/
Velar	/k/	K	Kata	/kaʒa/
	/g/	G	Gabus	/gabus/
Glotal	/ʔ/	K	Masuk	/masuʔ/

Bunyi /t̚/ dalam Bahasa Indonesia merupakan bunyi dental sehingga dilambangkan dengan lambang fonetik /t̚/. Sementara bunyi /t/ pada umumnya adalah bunyi alveolar yang ditulis dengan lambang fonetik /t/ seperti dalam kata bahasa Inggris *time* /taim/ ‘waktu’. Bunyi /t̚/ dalam kata *time* berbeda sumbernya dengan bunyi /t/ dalam kata *tidur* /t̚idur/.

Lambang /ʔ/ merupakan lambang bunyi hamzah. Bunyi hamzah digunakan ketika:

- a. Bunyi /k/ merupakan bunyi silabel terakhir suatu kata, maka bunyi /k/ diganti dengan bunyi hamzah/ʔ/. Misalnya: *rusak* /rusaʔ/.

- b. Terletak di antara dua bunyi vokal pada kata-kata serapan dari Bahasa Arab, seperti pada kata *maaf* /maʔaf/.
- c. Terletak di antara awalan yang berakhir dengan huruf vokal dan kata dasar yang berawalan huruf vokal seperti dalam kata *seorang* /seʔoran/ atau *keenam* /keʔenam/.
- d. Terletak di antara kata dasar berakhiran bunyi /a/ dengan kata dasar atau imbuhan yang dimulai dengan bunyi /a/. Contoh: *keadaan* /keʔadaʔan/. (Soderberg and Olson, 2008).

1.4.1.2 Nasal

Nasal disebut juga konsonan sengau merujuk pada bunyi konsonan yang dihasilkan dengan mengalirkan arus udara melalui rongga hidung dengan menutup arus keluar melalui rongga mulut (Vehraar, 2006: 35)

Tabel 1.1.2 Bunyi Konsonan *Nasal*

	Lambang Fonetik	Lambang Ortografi	Kata	Transkripsi Fonetik
Bilabial	/m/	M	Meja	/medʒa/
Alveolar	/n/	N	Nama	/nama/
Palatal	/ɲ/	Ny	Nyaman	/ɲaman/
Velar	/ŋ/	Ng	Panjang	/paŋdʒaŋ/

1.4.1.3 Flap/ Trill

Flap atau *Trill* disebut juga konsonan getaran yang dihasilkan dengan menggetarkan ujung lidah dengan menyentuh gusi sebentar lalu dilepaskan dan diulang kembali (Vehraar, 2006).

Tabel 1.1.3 Bunyi Konsonan Flap/Trill

	Lambang Fonetik	Lambang Ortografi	Kata	Transkripsi Fonetik
Alveolar	/r/	R	Rabu	/rabu/

1.4.1.4 Fricative

Fricative merupakan konsonan geseran dimana bunyi dihasilkan melalui arus yang sangat sempit sehingga sebagian besar arus udara terhambat di tempat pengartikulasiannya (Vehraar, 2006).

Tabel 1.1.4 Bunyi Konsonan *Fricative*

	Lambang Fonetik	Lambang Ortografi	Kata	Transkripsi Fonetik
Labio-dental	/f/	F dan V	Fajar	/faɖʒar/
Alveolar	/s/	S	Sulit	/sulit/

	/z/	Z	Zat	/zaʈ/
Post-alveolar	/ʃ/	Sy	syarat	/ʃaraʈ/
Glotal	/h/	H	rumah	/rumah/

1.4.1.5 *Approximant*

Tabel 1.1.5 Bunyi Konsonan *Approximant*

	Lambang Fonetik	Lambang Ortografi	Kata	Transkripsi Fonetik
Bilabial	/w/	W	Warna	/warna/

1.4.1.6 *Lateral Approximant* atau disebut juga konsonan sampingan dimana bunyi suara dihasilkan dengan menghalangi arus udara sehingga hanya bisa melalui samping lidah baik kanan maupun sebelah kiri (Vehraar, 2006).

Tabel 1.1.6 Bunyi Konsonan *Lateral Approximant*

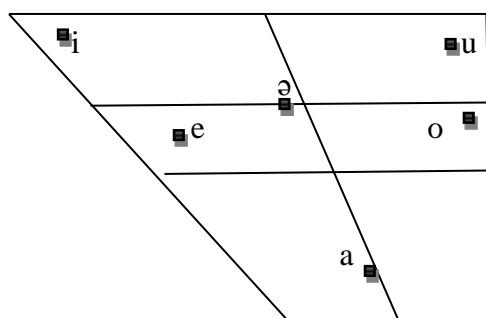
	Lambang Fonetik	Lambang Ortografi	Kata	Transkripsi Fonetik
Alveolar	/l/	L	Libur	/libur/

1.4.2 Vokal

Bunyi vokal merupakan bunyi bahasa yang dihasilkan dengan melibatkan pita-pita suara tanpa adanya penyempitan atau penutupan apapun pada tempat pengartikulasiannya (Vehraar, 2006: 38).

Terdapat 6 bunyi vokal dalam Bahasa Indonesia sesuai dengan standard IPA yang diusulkan oleh Soderberg and Olson (2008) yaitu /a/, /i/, /u/, /e/, /o/ dan /ə/.^{viii} Bunyi vokal umumnya digambarkan dalam diagram berdasarkan letak sumber suara di dalam lidah seperti diagram berikut:

Diagram 1.1 Bunyi Vokal Bahasa Indonesia



Sumber: Soderberg and Olson (2008: 211)

Tabel 1.2 Bunyi Vokal Bahasa Indonesia^{viii}

Lambang Fonetik	Lambang Ortografi	Kata	Transkripsi Fonetik
/a/	A	Fajar	/fadʒar/
/i/	I	Sulit	/sulit/
/u/	U	Jurus	/dʒurus/
/o/	O	Koran	/koran/
/e/	E	Kecap	/ketʃap/
/ə/	E	Sekarang	/səkaraŋ/

Menurut Vehrar (2006), dalam Bahasa Indonesia terdapat bunyi yang letaknya seakan-akan berada di antara bunyi vokal dan konsonan. Bunyi tersebut disebut ‘*semivokal*’. Bunyi semivokal umumnya dikenal juga dengan istilah *diphthong* dalam Bahasa Inggris. Dalam Bahasa Indonesia terdapat dua bunyi semivokal yaitu bunyi /w/ yang mendekati bunyi /u/ dan bunyi /j/ yang dekat sekali dengan bunyi /i/. Sebagai contoh dapat dijelaskan secara singkat dalam tabel berikut:

Tabel 1.2.1 Semivokal Bahasa Indonesia

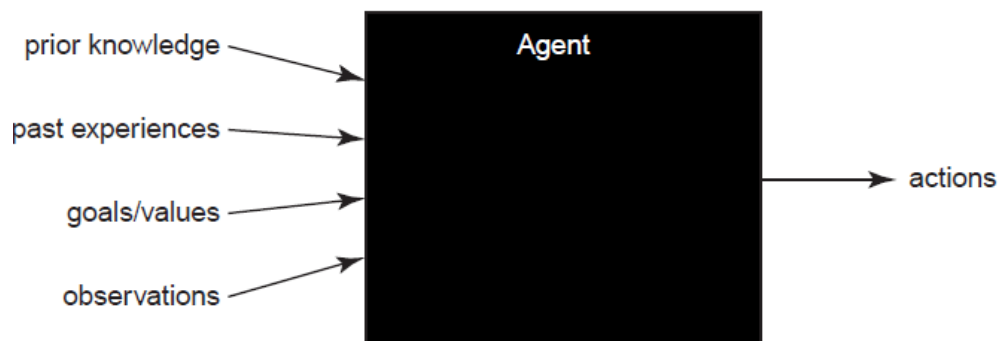
Lambang Fonetik	Lambang Ortografi	Kata	Transkripsi Fonetik
/aj/	Ai	Pantai	/panʈaj/
/aw/	Au	Pisau	/pisaw/
/oj/	Oi	Sepoi- sepoi	/səpoj-səpoj/

II. Kerangka Teori dan Metode Penilitaian

2.1. Kerangka Teori

Aplikasi fonetik Bahasa Indonesia ini akan didesain dengan menggunakan mesin cerdas yang disebut Kecerdasan Buatan (*Artificial Intelligence*). Dalam ilmu computer, kecerdasan buatan disebut juga Intelegensi Artifisial. AI didefinisikan sebagai studi tentang "agen cerdas" yaitu perangkat apa pun yang menyiratkan bahasa sehari-hari. Istilah "kecerdasan buatan" diterapkan ketika mesin didesain untuk meniru fungsi "kognitif" manusia seperti "belajar" dan "pemecahan masalah" (Poole, Mackwort and Goeble, 1998). Diagram dibawah ini menunjukkan sistem kerja Kecerdasan Buatan secara sederhana:

Diagram 3.1 Sistem Kerja Kecerdasan Buatan



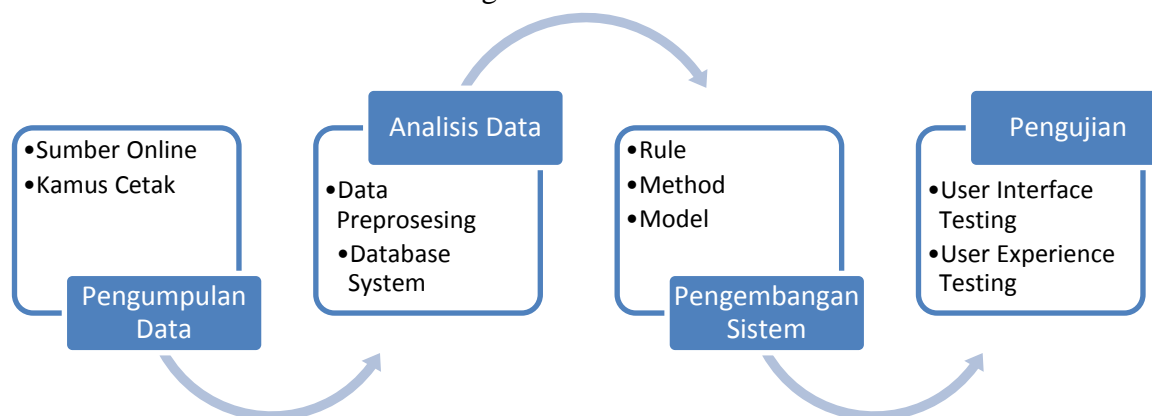
Sumber: Poole, David; Mackworth, Alan; Goebel, Randy.1998. *Computational Intelligence: A Logical Approach*. New York: Oxford University Press. ISBN 0-19-510270-3.

Aplikasi berbasis AI ini didesain dengan menggunakan *Natural Language Processing* (NLP) tehnik. NLP (*natural language processing*) merupakan cabang ilmu komputer dan linguistik yang mengkaji interaksi antara komputer dengan bahasa alami manusia. NLP sering dikaitkan sebagai cabang dari kecerdasan buatan yang bidang kajiannya bersinggungan dengan linguistik komputasional. Kajian NLP antara lain mencakup segmentasi tuturan (*speech segmentation*), segmentasi teks (*text segmentation*), penandaan kelas kata (*part-of-speech tagging*), serta pengawataksaan makna (*word sense disambiguation*). Meskipun kajiannya dapat mencakup teks dan tuturan, pemrosesan tuturan (*speech processing*) telah berkembang menjadi suatu bidang kajian terpisah (Charniak, 1984).

2.2. Metode Penelitian

Diagram dibawah ini menunjukkan metode penelitian yang digunakan dalam penelitian ini.

Diagram 3.2 Metode Penelitian



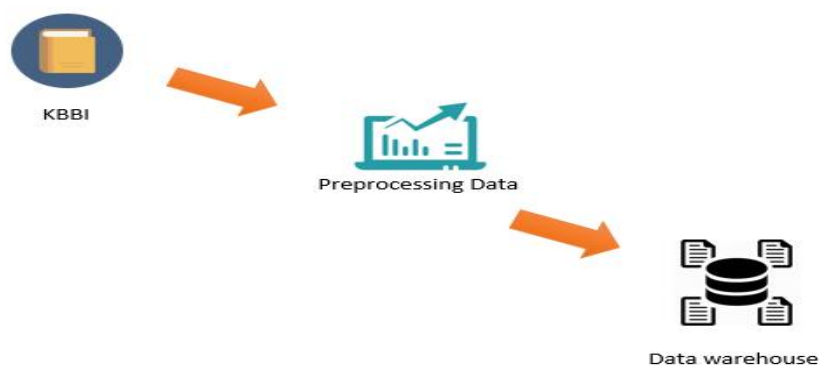
2.3. Metode Pengumpulan Data

Pada penelitian ini, pengumpulan data dilakukan dengan mengambil data primer berupa beberapa contoh kata dari kamus besar bahasa Indonesia (KBBI Online). Pada halaman website: <https://kbbi.kemdikbud.go.id/>. Selain itu data sekunder juga dikumpulkan untuk memperoleh data tambahan yaitu berupa data dari website resmi *International Phonetik Alphabet* (IPA), beberapa buku, dokumentasi dan informasi terkait linguistik Bahasa Indonesia.

2.4. Metode Analisis Data

Pada penelitian ini data yang diambil dari KBBI diproses untuk membuat kumpulan data yang sesuai dengan aturan fonemik. Diagram dibawah ini menunjukkan proses penarikan dan analisis data yang digunakan dalam penelitian ini.

Diagram 3.3 Metode Analisis Data



Pada proses *preprocessing*, data akan disaring dan ditingkatkan sesuai dengan proporsi yang akan digunakan untuk penulisan transkripsi fonetik bahasa Indonesia dan akan dijadikan data baru dan disimpan ke dalam data *warehouse*.

2.5. Metode Pengembangan Sistem

Pada penelitian ini, sistem dikembangkan dengan metode *agile*. *Agile* adalah sebuah metode untuk mengembangkan sistem dengan memecah dari suatu sistem yang besar, menjadi sistem yang kecil dan mudah untuk diatur. *Agile* juga lebih fleksibel karena tidak perlu mengulang proses dari awal. Metode ini diawali dengan analisis kebutuhan, desain sistem, implementasi sistem dan pengujian. Masing-masing tahap bisa diulang dan disesuaikan dengan kebutuhan saat pengembangan.

2.6. Metode Evaluasi Sistem

Pada penelitian ini, sistem dievaluasi berdasarkan data yang dihasilkan oleh sistem. Data tersebut dibandingkan dengan data asli yang diperoleh dari KBBI cetak atau KBBI online dan disesuaikan dengan aturan fonetik Bahasa Indonesia

berdasarkan IPA. Hasil perbandingan yang didapat di catat akurasi. Setelah itu, disimpulkan faktor-faktor yang menyebabkan penurunan akurasi.

III. Pembahasan

Hasil dari rancangan aplikasi fonetik Bahasa Indonesia ini kami beri nama **BI-Phon** yang merupakan singkatan dari *Bahasa Indonesia-Phonetic Tramscription* yang berfungsi untuk membantu cara pelafalan Bahasa Indonesia secara tepat dengan menggunakan system penulisan transkripsi fonetik.

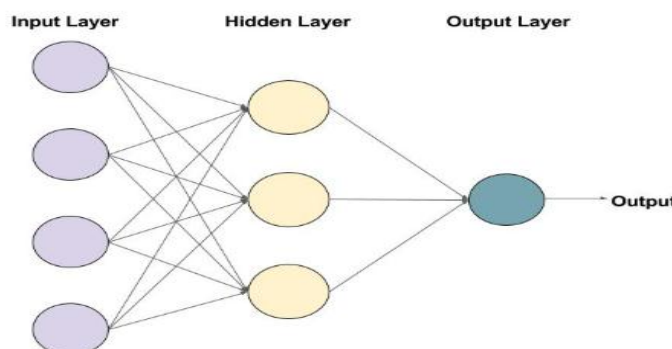
Aplikasi fonetik berbasis AI ini disusun dengan menerapkan tehnik NLP. Proses NLP dilakukan menggunakan metode segmentasi yaitu metode pengelompokan yang disesuaikan berdasarkan kaidah fonetik Bahasa Indonesia yang telah disempurnakan dengan standar *Internasional Phonetic Association* (IPA). Seperti Tabel dibawah ini:

Tabel 4.1 Proses penyusunan rule menggunakan tehnik NLP

LambangFonetik	LambangOrtografi	Kata	TranskripsiFonetik	Rule Type
/p/	P	Panjang	/paŋdʒaŋ/	Rule 1
/b/	B	Baju	/baʒu/	Rule 1
/m/	M	Meja	/medʒa/	Rule 2
/n/	N	Nama	/nama/	Rule2
/s/	S	Sulit	/sulit/	Rule 3
/z/	Z	zat	/zaʔ/	Rule 3

Konsep *Artificial Intelligent* kemudian diolah menggunakan metode *Neural Network* (Jaringan saraf tiruan) yang merupakan sistem adaptif yang dapat mengubah struktur untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut. Secara sederhana, JST merupakan sebuah alat pemodelan data statistik non-linier. JST dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pola-pola pada data (Hornik, Stinchombe, and White, 1989). Hasil dari segmentasi NLP digunakan sebagai sumber data untuk *training*, *testing* serta validasi dengan persentase 70% - 30%. Untuk memvalidasi data yang digunakan benar atau tidak data dibuat sendiri dengan memberikan *missing value* seperti diagram dibawah ini:

Diagram 4.1 Pemrosesan data dengan NLP dan Neutral Network



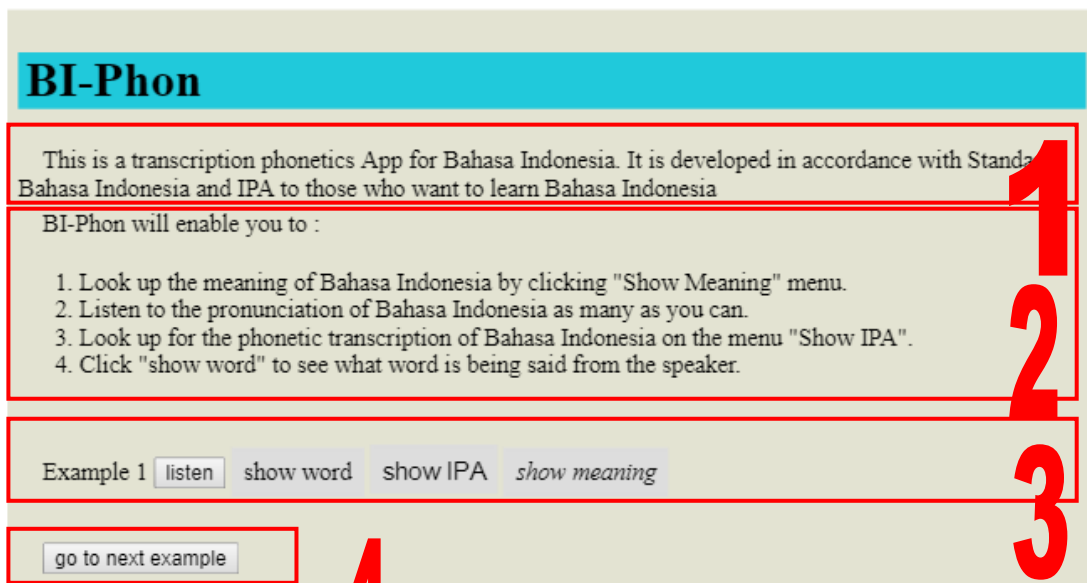
Input layer dalam penelitian ini menggunakan data yang telah diproses berdasarkan segmentasi NLP sehingga terdapat 4 node utama yang akan digunakan sebagai input parameter. Pada *hidden layer* menggunakan satu *hidden* karena jumlah *iterasi* yang sedikit, sehingga *ouput* yang dihasilkan sesuai atau cocok dengan prediksi yang diharapkan.

Aplikasi transkripsi fonetik Bahasa Indonesia ini merupakan aplikasi berbasis website yang rencananya akan diluncurkan secara daring sehingga dapat diakses oleh pembelajar BIPA secara luas. Aplikasi ini terdiri dari beberapa halaman yaitu halaman antarmuka, halaman menu dan halaman keluaran.

1. Halaman Antarmuka

Tampilan aplikasi BI-Phon seperti tampak pada gambar dibawah ini. Terdapat empat bagian utama pada halaman antarmuka yaitu:

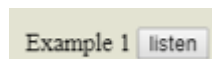
- Bagian 1: Berisi tentang uraian singkat aplikasi BI-Phon
- Bagian 2: Berisi tentang informasi tombol menu yang digunakan sebagai navigasi.
- Bagian 3: Berisi tombol menu dari contoh transkripsi phonetic dan hasil dari itersi dari program.
- Bagian 4: Berisi tombol untuk melanjutkan ke contoh berikutnya.



2. Halaman Menu

Halaman menu berisi beberapa sub menu sebagai berikut:

2.1 Tombol *Listen*



Tombol "*Listen*" digunakan untuk mendengarkan cara pengucapan yang menghasilkan suara sesuai dengan kaidah yang telah diterapkan.

2.2 Tombol *Show word*



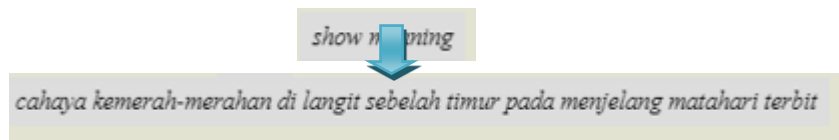
Tombol *show word* digunakan untuk menampilkan tulisan sesuai dengan sound yang didengarkan.

2.3 Tombol Show IPA



Tombol *Show IPA* digunakan untuk menampilkan tampilan ejaan penulisan sesuai standart IPA yang telah terkoneksi dengan sound dan word yang akan diginakan.

2.4 Tombol Show meaning



Tombol *show meaning* digunakan untuk melihat arti kata yang telah diucapkan pada pelafalan phonetic sesuai contoh.

1.2 Halaman Keluaran

A screenshot of the BI-Phon app interface. At the top, there is a blue header with the text 'BI-Phon'. Below the header, there is a paragraph of introductory text: 'This is a transcription phonetics App for Bahasa Indonesia. It is developed in accordance with Standard Bahasa Indonesia and IPA to those who want to learn Bahasa Indonesia'. Below this, it says 'BI-Phon will enable you to :'. A numbered list follows: 1. Look up the meaning of Bahasa Indonesia by clicking "Show Meaning" menu. 2. Listen to the pronunciation of Bahasa Indonesia as many as you can. 3. Look up for the phonetic transcription of Bahasa Indonesia on the menu "Show IPA". 4. Click "show word" to see what word is being said from the speaker. At the bottom, there is a row of buttons: 'Example 2', 'listen', 'show word', 'show IPA', and 'show meaning'. Below this row is a button labeled 'go to next example'.

Pada tampilan awal program akan terlihat seperti gambar diatas. Semua tombol menu yang sejajar dengan Example 2 masih tertutup akan tetapi ketika di klik akan muncul seperti gambar dibawah ini.

BI-Phon

This is a transcription phonetics App for Bahasa Indonesia. It is developed in accordance with Standard Bahasa Indonesia and IPA to those who want to learn Bahasa Indonesia

BI-Phon will enable you to :

1. Look up the meaning of Bahasa Indonesia by clicking "Show Meaning" menu.
2. Listen to the pronunciation of Bahasa Indonesia as many as you can.
3. Look up for the phonetic transcription of Bahasa Indonesia on the menu "Show IPA".
4. Click "show word" to see what word is being said from the speaker.

Example 2 Fajar [fadʒar]

cahaya kemerah-merahan di langit sebelah timur pada menjelang matahari terbit

IV. Simpulan

Aplikasi transkripsi fonetik Bahasa Indonesia (yang selanjutnya kami sebut BI-Phon) ini merupakan aplikasi berbasis AI yang didesain dengan menggunakan *Natural Language Processing* tehnik. Aplikasi ini merupakan aplikasi yang disusun berbasis web yang akan diluncurkan secara daring (online). BI-Phon masih dalam tahap pengembangan dan penyempurnaan dengan harapan kelak dapat dikonfigurasi dengan kamus KBBI daring untuk memudahkan pembelajar BIPA mempelajari pengucapan Bahasa Indonesia sesuai dengan kaidah Bahasa Indonesia yang telah disempurnakan. Aplikasi BI-Phon masih sangat perlu untuk dikembangkan terutama dalam kaitannya dengan kaidah penekanan kata (*stress*) pada pengucapan kata Bahasa Indonesia sehingga diperlukan penelitian lanjutan untuk pengembangan aplikasi ini.

Notes

ⁱBahasa Indonesia menempati urutan ke-lima sebagai bahasa dengan penutur terbanyak di dunia berdasarkan data Kementerian Luar Negeri pada tahun 2012 dengan penutur sebanyak 4.463.950 orang yang tersebar di luar negeri. Bahkan Pada tahun 2007, bahasa Indonesia secara resmi ditempatkan sebagai bahasa asing kedua oleh pemerintah daerah Ho Chi Minh City, Vietnam sejajar dengan Bahasa Inggris, Prancis dan Jepang.
Sumber: <https://edukasi.kompas.com/read/2013/10/23/1253102/BIPA.Tingkatkan.Fungsi.Bahasa.Indonesia.Menjadi.Bahasa.Internasional>.

ⁱⁱ Berdasarkan penuturan Kepala Badan Pengembangan dan Pembinaan Bahasa, Dadang Sunendar pada acara Bulan Bahasa dan Sastra di Kemendikbud pada tanggal 6 Oktober 2017 "Bahasa Indonesia telah dipelajari di 45 negara yang tersebar di 250 lembaga di luar negeri yang terafiliasi dengan pemerintah dan swasta". Dalam upayanya menjadikan Bahasa Indonesia menjadi bahasa internasional, Pemerintah Indonesia melalui Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan telah mengirimkan lebih dari 150 pengajar BIPA ke berbagai negara setiap tahunnya.

Sumber: <https://news.okezone.com/read/2017/10/07/18/1790706/banqqa-menuju-level-internasional-bahasa-indonesia-diajarkan-di-45-negara>

ⁱⁱⁱ Aplikasi Transkripsi Fonetik Bahasa Indonesia sebelumnya telah dikembangkan oleh Setyowati; Bertalya dan Ningsih pada tahun 2014 dalam makalahnya yang berjudul *Aplikasi Fonetik Bahasa Indonesia Berdasarkan IPA (International Phonetic Alphabet) untuk BIPA*. Namun demikian aplikasi yang mereka kembangkan lebih mengacu pada fonetik Bahasa Inggris dari pada Bahasa Indonesia secara khusus.

^{iv} Berdasarkan teori fonetik yang diusulkan oleh Ebeling (1967), Stokhof (1975) dalam artikelnya yang berjudul *On the Phonology of Bahasa Indonesia* menyarankan ada empat aturan dasar yang dapat diterapkan dalam fonetik dan fonologi Bahasa Indonesia. Aturan-aturan tersebut adalah opsionalitas, netralisasi, doublet (triplet) dan variasi bebas.

a. *Opsionalities* adalah istilah yang digunakan untuk menunjukkan unit fonemik yang terdiri dari 'bunyi dasar' dan satu atau lebih "bunyi opsional". Opsionalitas berlaku apabila:

- Penekanan bunyi opsional dalam fonem dasar tidak menyebabkan salah tafsir bagi pendengar, sehingga fonem dasar dinilai identik dengan fonem opsional. Contoh: bunyi huruf F dan V diidentifikasi memiliki bunyi yang sama dalam Bahasa Indonesia seperti dalam kata Fajar /fadʒar/ dan vena /fena/
- Salah satu dari dua atau lebih kemungkinan realisasi format x dapat diterima pula sebagai format y. Misalnya: suara /h/ bersifat opsional. Dapat diucapkan atau tidak seperti di lihat /lihat/ atau /liat/

b. *Neutralization* terjadi apabila terdapat dua fonem (atau lebih) yang bertentangan dapat saling menggantikan dalam posisi tertentu, namun tidak berlaku dalam (satu atau lebih) posisi lain. Suara /d/ dan /t/ dalam Bahasa Indonesia adalah dua set bunyi bahasa yang berbeda seperti dalam kata tutup /tuʈup/ dan datang /daʈaŋ/. Netralisasi dalam Bahasa Indonesia dapat mengganti bunyi /d/ menjadi /t/ tanpa merusak identitas linguistik dari bunyi dasar tersebut. Seperti dalam maksud /maʔsut/ 'meaning'. Suara /d/ hanya dapat diganti dengan /t/ ketika ditempatkan di akhir kata.

c. *Doublet* adalah kondisi ketika dua bunyi suara memiliki fungsi fonetis yang serupa dan tidak merubah makna apapun. Contoh: adik /adiʔ/ dan adek /adɛʔ/; atau senin /sɛnɪn/ atau senen /sɛnɛn/; atau lubang /lubanʒ/ atau lobang /lobanʒ/ substitusi suara sebelumnya karena yang terakhir tidak berubah makna.

^v Bahasa Indonesia tidak memiliki bunyi /v/ karena huruf v dalam Bahasa Indonesia berbunyi sama dengan huruf f yang dilambangkan dengan tulisan fonetik /f/. Bahasa Indonesia juga tidak memiliki bunyi /x/ dikarenakan kebanyakan kata dalam Bahasa Indonesia yang mengandung huruf x berbunyi /k/ dan berbunyi /s/ apabila terletak di awal kata. Selain itu Bahasa Indonesia tidak memiliki kata atau istilah asli berawalan x kecuali istilah serapan dari bahasa lain. Bahasa Indonesia memiliki huruf q namun berbunyi sama dengan bunyi /k/.

^{vi} Terdapat beberapa huruf konsonan plosive berdasarkan sumber suaranya yaitu Bilabial merupakan konsonan letupan yang dihasilkan di antara bibir yaitu bunyi /p/ dan /b/. Dental yaitu konsonan letupan yang dihasilkan diantara ujung lidah dengan gigi yaitu bunyi /t/. Alveolar dihasilkan di antara ujung lidah dengan lengkung kaki gigi seperti bunyi /d/. Post-Alveolar dihasilkan di antara tengah lidah dengan langit-langit keras seperti bunyi /tʃ/ dan /dʒ/. Bunyi konsonan Velar dihasilkan antara pangkal lidah dengan langit-langit tekak misalnya bunyi /k/ dan /g/. Bunyi Konsonan glottal dihasilkan di antara pita suara sebagai contoh yaitu bunyi hamzah /ʔ/ (Vehraar, 2006: 34).

^{vii} Sedangkan Stokhof (1975) berpendapat bahwa terdapat tiga variasi bunyi e pada Bahasa Indonesia yaitu /e/ seperti pada kata eja /edʒa/; /ə/ pada kata sedang /sɛdan/; dan /ɛ/ pada kata monyet /moŋɛt/.

^{viii} Soderberg and Olson menerangkan bahwa bunyi vokal Bahasa Indonesia dituliskan dengan huruf vokal cardinal atau dasar yaitu /a, i, u, e, o/ dikarenakan pada umumnya bunyi vokal Bahasa Indonesia lebih rendah dan tipis dari pada bunyi /ɪ, ɛ, ɔ, ʊ/.

Daftar Pustaka

- Chaer, Abdul. 1994. *Linguistik Umum*. Jakarta: PT Rineka Cipta
- Charniak, Eugene. 1984. *Introduction to artificial intelligence*. Addison-Wesley)
- Hornik, Kurt; Stinchcombe, Maxwell; and White, Halbert. 1989. *Multilayer feedforward networks are universal approximators*, *Neural Networks*. Volume 2, Issue 5, 1989, Pages 359-366, ISSN 0893-6080
- Jones, Daniel. 2011. *Cambridge English Pronouncing Dictionary*. New Delhi: Cambridge University Press
- Lapoliwa, Hans. 1988. *Pengantar Fonologi I: Fonetik*. Jakarta :Pusat Pembinaan dan Pengembahnya Bahasa Departemen Pendidikan dan Kebudayaan
- Poole, David; Mackworth, Alan; and Goebel, Randy. 1998. *Computational Intelligence: A Logical Approach*. New York: Oxford University Press. ISBN 0-19-510270-3.
- Roach, Peter. 2009. *English Phonetics and Phonology*. Delhi: Cambridge University Press
- Setyowati, Lilis; Bertalya dan Ningsih, Tri Wahyu Retno. 2014. *Aplikasi Fonetik Bahasa Indonesia Berdasarkan IPA (International Phonetic Alphabet) untuk BIPA*. Depok: Prosiding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2014) Universitas Gunadarma Vol. 8 Oktober 2014 [7 Juni 2018]
- Soderberg, Craig D and Oslen, Kenneth S. 2008. *Indonesian on the Journal of the International Phonetic Association*. United Kingdom: Cambridge University Press pp. 209-213 <https://www.jstor.org/stable/44526493> [7 Juni 2018]
- Stokhof, W.A.L. 1975. *On the Phonology of Bahasa Indonesia*. Published by: KITLV, Royal Netherlands Institute of Southeast Asian and Caribbean Studies pp. 254-269 <http://www.jstor.org/stable/27862982> [7 Juni 2018]
- The International Phonetic Association. 1999. *Handbook of the International Phonetic Association; A Guide to Use the International Phonetic Alphabet*. Cambridge: Cambridge University Press
- Verhaar, J.W.M. 2006. *Asas-asas Linguistik Umum*. Yogyakarta: Gajah Mada University Press
- <https://www.internationalphoneticassociation.org/content/full-ipa-chart>
- <http://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/>