

## 6.2.B. Specjalizowane archiwa medyczne (bazy danych)

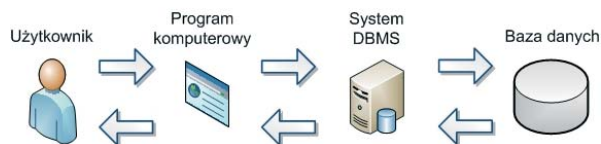
Tomasz Pięciak

### 1. ROLA INFORMACJI CYFROWEJ W DZISIEJSZYM ŚWIECIE

Udostępnienie internetu dla ogółnoświatowego ogółu ludzkości umożliwiło wymianę informacji na niespotykaną dotąd skalę. Z każdym dniem w świecie nauk medycznych i pochodnych ujawnia się kilka tysięcy nowinek i wynalazków. To co przed laty spisywane było w grubych księgozbiorach, a dostępne tylko dla wybranych, powoli odchodzi w zapomnienie. A wszystko to dzięki informacji w postaci cyfrowej. Kilkanaście lat temu wymianę danych poprzez Internet mogły dokonywać tylko ośrodki naukowe na specjalnie dostosowanych ku temu platformach. Wszegobecność baz danych przeszła najśmielsze oczekiwania projektantów z lat 60. ubiegłego stulecia. Początkowo składowanie informacji cyfrowej w wielkich przedsiębiorstwach polegało na wyodrębnieniu oddzielnych kategorii, niejednokrotnie powiązanych z kilkoma oddziałami, takich jak na przykład spis pracowników, spis przysługujących im wynagrodzeń czy historia zatrudnienia. Zmiana danych personalnych osoby wymuszała aktualizację danych w każdej z tych kartotek. Wkrótce jednak okazało się, że tzw. pliki płaskie czyli sekwencyjnie zapisane dane nie wystarczają. Problem jaki wówczas istniał to nie tylko przeszukiwanie danych, wyodrębnianie interesujących fragmentów czy komplikacja związana z aktualnością danych w wielu oddziałach korporacji, ale przede wszystkim ochrona informacji przed niepowołanymi osobami. Przykładem mogą tutaj być wspomniane listy płac, do których oprócz działu kadrowego nie powinien mieć nikt inny dostępu. Z czasem udoskonalono struktury komputerowe pozwalające przechowywać informację, czego skutkiem było powstanie baz danych.

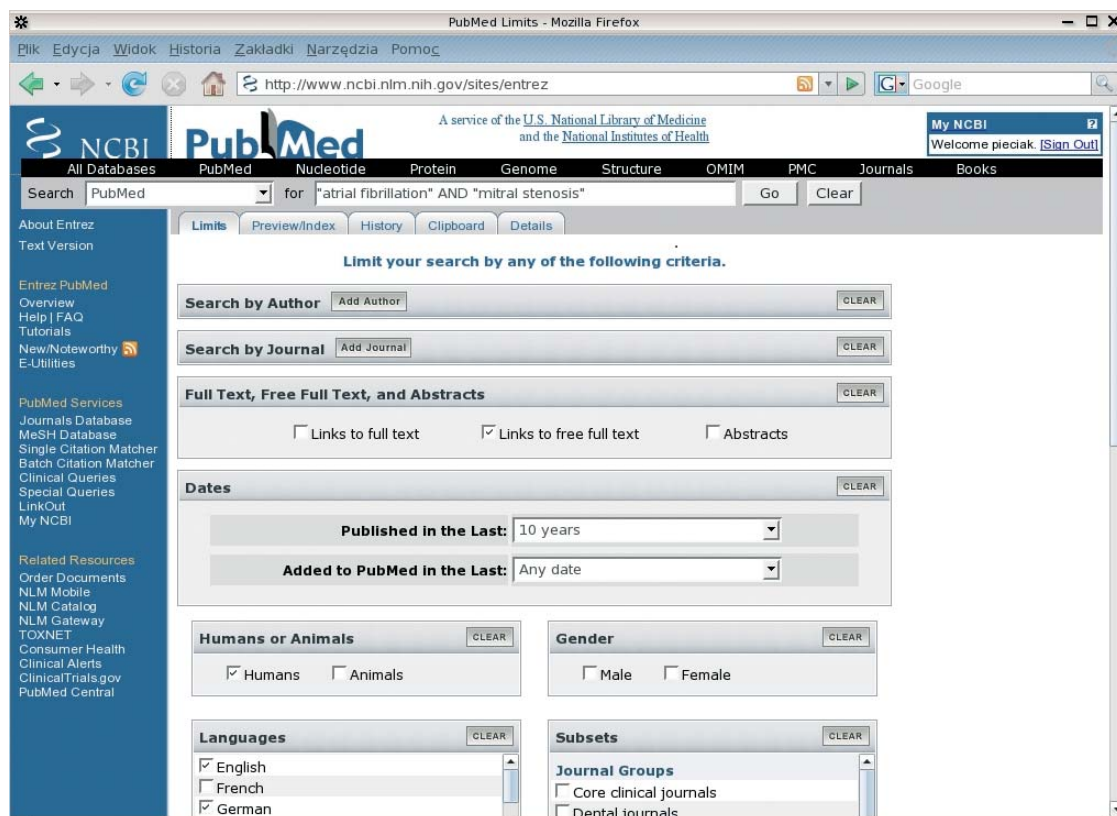
W najprostszym przypadku bazą danych możemy nazwać agregat do przechowywania informacji w postaci tekstu, dźwięku czy obrazów oraz interfejs służący do wymiany w/w dokumentów z otoczeniem. To czy dana osoba będzie uprawniona do korzystania z zasobu, albo jedynie jego części określają tzw. schematy i pod-schematy będące odpowiednio opisem struktury bazy danych użytkowanej przez oprogramowanie oraz konkretnego odbiorcę. Najbardziej powszechnymi w użytku są niewątpliwie relacyjne bazy danych, choć wyko-

rzystuje się również modele obiektowe czy rozwiązania hybrydowe. Ponieważ architektura oprogramowania korzystającego z baz danych najczęściej jest rozwiązaniem wielowarstwowym (rys. 1), dlatego każda z nich komunikuje się jedynie z sąsiednią (sąsiednimi). Użytkownik końcowy nie jest zmuszony do zapoznawania się z pozoru ze skomplikowanym językiem zapytań SQL (Structured Query Language), a jedynie przygotowanym do tego programem komputerowym. Odpowiednio wysłane zapytanie przez użytkownika poprzez aplikację (tutaj wypełnienie formularza) trafia do systemu zarządzania bazą danych (ang. *Database Management System*, DBMS), skąd po interakcji z bazą danych jest zwracana żądana odpowiedź zawierająca poszukiwane treści do użytkownika (rys. 1).



Rys. 1. Schemat działania zapytanie – odpowiedź

Również projektanci aplikacji komputerowej nie są zmuszeni do zaznajomienia się z fizyczną budową baz danych, korzystając jedynie z odpowiednio przygotowanego interfejsu wymiany informacji przygotowanego przez architektów bazy danych, najczęściej za pomocą wspomnianego języka SQL. Naturalnie wykorzystując kwerendy w języku zapytań możemy filtrować wyniki poszukiwań wedle upodobań, nie mniej jednak jest to okupione podstawową jego znajomością oraz elementarnymi wiadomościami z zakresu algebry Boole'a. W niektórych przypadkach usługodawcy baz danych umożliwiają w okrojonym zakresie wykorzystanie bezpośrednich zapytań do bazy danych i/lub uwzględniając zasady algebry Boole'a. Objawia się to między innymi możliwością selekcjonowania wyników poszukiwania uwzględniając założone kryteria. Aby użytkownik nie zablądził w terminologii, na przykładowej rycinie (rys. 2) wprowadzono dwa hasła rozdzielone logicznym operatorem AND, dzięki któremu wyszukana zostanie lista z abstraktami zawierającymi jedno jak i drugie sformułowanie.



Rys. 2. Wyszukiwane uwzględniające zadane kryteria

## 2. UOGÓLNIONE MEDYCZNE BAZY DANYCH

Nie inaczej ma się to w medycynie, gdzie istnieją tysiące rozproszonych baz danych, czyli takich, w których poszczególne jej części znajdują się w różnych miejscach, zaś sama baza sprawia dla odbiorcy końcowego wrażenie zintegrowanej w jedną całość. Mając odpowiednią autoryzację możemy pobierać interesujące nas informacje na bieżąco o światowych odkryciach, bez względu na miejsce zamieszkania. A wszystko to dzięki wspomnianemu, łatwemu na dzień dzisiejszy dostępowi do ogólnoświatowej sieci Internet.

Najbardziej rozpowszechnioną i znaną dziś internetową bazą danych dokumentów medycznych jest niewątpliwie baza Medline, utrzymywana przez U.S. National Library of Medicine, mieszcząca się w Stanach Zjednoczonych. Baza przetrzymuje blisko 11 mln wpisów pochodzących z 7300 czasopism, które są wydawane w kilkudziesięciu krajach na świecie. Dostęp do bazy może być realizowany na wiele sposobów. Jednym z najbardziej rozpowszechnionych jest dostęp poprzez serwis PubMed (<http://www.ncbi.nlm.nih.gov/sites/entrez/>) będący darmowym silnikiem wyszukiwań artykułów w bazie Medline jak również abstraktów pochodzących ze stron wydawców (rys. 2).

Aby zrozumieć dalszą część niniejszego opracowania wprowadźmy podstawowe pojęcie relacyjnych baz danych jakim jest krotka (rekord), posiadająca unikalne pole identyfikujące. Jest ona niczym innym jak tablicą (w tabeli 1 zobrazowaną w postaci pojedynczego wiersza zawierającego atrybuty), w której w odpowiednie pola są wpisane interesujące informacje o publikacji, takie jak autorzy, tytuł, rok wydania, słowa kluczowe, nazwa instytucji, która wydała publikację czy wreszcie krótkie streszczenie. Ewidentnie atrybutów może być więcej, jednak co do niektórych jak na przykład słów kluczowych istnieją ściśle określone restrykcje, aby nie wprowadzono do bazy niejednoznaczności. Do tego celu jest przeznaczony moduł MeSH (*Medical Subject Headings*) zawierający ok. 20 tys. słów zestawionych w hierarchiczne kategorie. Uzbrojeni ww. informacje możemy uzyskać dostęp do odpowiedniego artykułu, jednak uważny czytelnik dostrzeże, że posiadanie dla własnego użytku literatury w takiej formie najczęściej wiąże się z odpowiednimi kosztami, które musimy uiścić. Ponadto niejednokrotnie właściciele internetowych elektronicznych zbiorów wymagają danych osobowych użytkowników. Oczywiście łączy się to ze społecznym problemem baz danych, a co za tym idzie udostępniania danych personalnych osobom trzecim w celach marketingowych. Odbiegając od właściwej treści skupmy się jednak na temacie przewodnim niniejszego rozdziału, którym są zasoby dokumentów medycznych.

**Tabela 1**  
Przykładowe rekordy bazy danych abstraktów

Id	Autor	Tytuł	Rok wydania	Słowa kluczowe	Nazwa instytucji	Streszczenie
1	Katarzyna Nowak	Niedomykalność zastawki mitralnej	1995	zastawka mitralna, wskazania do leczenia operacyjnego	ABC	...
2	Jan Kowalski	Stymulacja w arytmiiach przedsionkowych	2007	stymulacja, arytmia przedsionkowa	XYZ	...

Nim uzyskamy dostęp do wybranej publikacji naukowej, najczęściej do dyspozycji mamy jedną z trzech wersji: dokumenty elektroniczne PDF, PS lub publikacja bezpośrednio na stronie internetowej. Pierwszy z nich, format PDF (ang. Portable Document Format) jest najczęściej udostępnianą formą prezentacji artykułów naukowych. Jego zaletą jest względnie niewielka objętość i możliwość otworzenia go na dowolnym komputerze wyposażonym w specjalną do tego przeglądarkę jak Adobe Reader, Xpdf, Kpdf, itd. Drugi z nich, PS (ang. PostScript) przeznaczony jest do bezpośredniego przekierowania dokumentu elektronicznego na urządzenie drukujące. Rzecz jasna istnieje możliwość przeglądania również i tych dokumentów za pomocą odpowiedniego oprogramowania jakim jest na przykład Ghostview. Wybór formatu dokumentu nie gra większej roli. Ważnym faktem jest natomiast to, że artykuły nie są przechowywane przez serwis wyszukiwawczy, a jedynie składowane są hiperłącza do odpowiednich lokalizacji, skąd w przypadku artykułów płatnych po uregulowaniu należności można je pobrać dla własnego użytku. Pozyskiwanie wiedzy w takiej formie bardzo szybko wypiera tradycyjną informację w postaci papierowych czasopism naukowych. Co więcej zastosowanie nośników danych DVD, HD DVD czy nowatorskiej technologii Blu-ray opartej na niebieskim laserze eliminuje problem składowania bezcennych i pożytecznych kilogramów papieru.

### 3. SPECJALISTYCZNE BAZY DANYCH

Szczególnym przypadkiem tekstowych baz danych są niewątpliwie bazy związków chemicznych, zachodzących między związkami reakcji czy informacje o wpływie ich na otoczenie, a w szczególności na zdrowie człowieka. Warto tutaj przytoczyć nazwę jednej z nich, a mianowicie baza Beilstein posiadająca w swych archiwach blisko 25 mln rekordów będących informacjami o substancjach, reakcjach oraz publikacjach z ich zakresu. Wyciągając dane z bazy możemy dowiedzieć się absolutnie wszystkiego na temat poszukiwanej substancji. Pożądanym byłoby nadmienić jeszcze o innej bazie – PubChem (<http://pubchem.ncbi.nlm.nih.gov/>). O ile użytkowanie poprzedniej wymaga wykupienia rocznego abonamentu i wykorzystania specjalnego oprogramowania CrossFire Beilstein Commander, o tyle korzystanie z bazy

PubChem jest darmowe, co więcej oferuje ona atrakcyjną z punktu widzenia bioinformatyka możliwość eksportu opisu danego związku chemicznego do niezależnego od platformy i oprogramowania formatu XML.

Innym przykładem mogą być konglomeraty farmaceutyczne. Łatwy dostęp do tego rodzaju zbiorów może zapewnić szybką pomoc w zakresie podstawowych informacji na temat koincydencji podawanego farmaceutyku z innymi przyjmowanymi przez pacjenta. Bezspornie pozostaje fakt wiarygodności takich informacji, do której bez kozery wydaje się być pomoc agencji rządowych i dużych ośrodków naukowych udostępniających aktualne wyniki badań w tym zakresie.

Nie sposób na zakończenie nie wspomnieć zarówno o bazach danych sekwencji DNA i RNA których przykładem może być GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>). Początki tego projektu sięgają wczesnych lat 80. ubiegłego stulecia, w których przedsięwzięcie zbiegło się równocześnie z powstaniem bioinformatyki, interdyscyplinarnej dziedziny łączącej w sobie m.in. informatykę, matematykę, genetykę oraz biologię molekularną, której celem jest akwizycja i analiza danych biologicznych, np. kodu genetycznego. Rozwój Internetu zapewnił szybki wzrost bazy, do której oprócz pojedynczych laboratoriów napływają informacje z dużych ośrodków, osiągając łącznie dane o organizmach w ilości ponad 100 tys!

Nieodzowną pomocą dzisiejszych nauk biomedycznych jest gigantyczna moc obliczeniowa urządzeń elektronicznych, potrzebna aby przetworzyć pojawiające się w niemal wykładniczym tempie dane. Powoli współczesne komputery osiągają kres swoich możliwości. Czy to oznacza, że powinna już nadejść era komputerów kwantowych? Na te z pewnością jeszcze przyjdzie nam poczekać, jednak krokiem w przód na dzień dzisiejszy mogłaby być konsolidacja wielu odosobnionych baz danych w jedną całość, choć z przyczyn komercyjnych niekoniecznie możliwa.

### 4. PRZYSZŁOŚĆ WYMIANY INFORMACJI CYFROWEJ

Bez wątpienia XXI wiek będzie zdominowany przez względnie łatwy dostęp do informacji w postaci cyfrowej. Włączając w to najnowsze metody nauczania jak chociażby e-learning da to możliwość kształcenia wy-

soko wykwalifikowanej kadry medycznej niezależnie od miejsca zamieszkania. To co poprzednim pokoleniom wydawało się trudne do wyobrażenia, dla dzisiejszego społeczeństwa jest podstawowym nurtem w kształceniu i pomocą szeroko pojętej diagnostyki. Czy jednak rola cyfrowego świata zostanie na zawsze nieodłączną pomocą nauk medycznych? Odpowiedź na to pytanie zweryfikuje najbliższe kilkadziesiąt lat, lat które miejmy nadzieję zmienią dotychczasowe poglądy wobec sztuki lekarskiej.

### Literatura

- [1] Brookshear J.G., *Informatyka w ogólnym zarysie*, WNT, Warszawa, 2003.
- [2] Kawalec P., *Medyczne bazy danych, Przewodnik lekarza*, 5, 2002, Termedia Wydawnictwa Medyczne, Poznań.
- [3] Kawalec P., *Medyczne bazy danych, cz. 1*, Służba zdrowia, nr 75–76, wrzesień 2002, Warszawa.
- [4] Kawalec P., *Medyczne bazy danych, cz. 2*, Służba zdrowia, nr 77–78, październik 2002, Warszawa.
- [5] Kaćki E., Kulikowski J.L., Nowakowski A., Waniewski E., *Biocybernetyka i Inżynieria Biomedyczna 2000* pod redakcją Macieja Nałęcz, Tom 7, Systemy komputerowe i teleinformatyczne w służbie zdrowia, Exit, Warszawa, 2002.
- [6] Lastówka M., *Relacyjna baza danych dokumentów medycznych*, Elektronika – konstrukcje, technologie, zastosowania, nr 3, 2007, SIGMA-NOT.
- [7] Zajdel R., Kaćki E., Szczepaniak P.S., Kurzyński M., *Kompendium informatyki medycznej*, AlfaMedica Press, Bielsko-Biała, 2003.