Towards a Semi-Automated Approach for Systematic Literature Reviews

Tim Denzler
Friedrich-Alexander-University Erlangen-Nuremberg, tim.denzler@fau.de

Martin Robert Enders
Friedrich-Alexander-University Erlangen-Nuremberg, martin.enders@fau.de

Patricia Akello
University of Texas, patricia.akello@utsa.edu

Towards a Semi-Automated Approach for Systematic Literature Reviews

Completed Research

Tim Denzler
Friedrich-Alexander-University
Erlangen-Nuremberg
tim.denzler@fau.de

Martin Robert Enders
Friedrich-Alexander-University
Erlangen-Nuremberg
martin.enders@fau.de

Patricia Akello
The University of Texas
at San Antonio
patricia.akello@utsa.edu

Abstract

Given the growing output of scientific literature, researchers are faced with a daunting challenge when it comes to performing systematic literature reviews. Hence, the use of Information Systems to achieve operational excellence has gained increasing importance in systematic literature reviews. However, existing solutions to support systematic literature reviews are often restrained to a single aspect of the process or lack interoperability. As such, researchers may not be able to efficiently leverage recent promising advancements in Machine Learning and Text Analytics. Therefore, we developed a flexible and modifiable artifact that aims to support systematic literature review processes from a holistic point of view. We expect our artifact to be a first step towards semi-automation of systematic literature reviews, which will gain relevance in the near future, as the trend of rising scientific literature output is expected to continue. Our development process follows a Design Science Research approach including continuous evaluation.

Keywords

Literature Review, Design Science, Information Systems, Automation, Text Analytics, Machine Learning

Introduction

As digitalization and globalization continue to advance their impact by making more and more information accessible online, we can observe the same development in scientific literature, as the output of publications has increased significantly in recent years (White 2019). In light of this continuously growing torrent, systematic literature reviews have become even more instrumental, as they guide researchers in assessing existing scholarly publications, preventing duplicate efforts, and identifying research gaps to be addressed in the future (vom Brocke et al. 2009; Webster and Watson 2002). However, given the immense volume of research published today, systematically reviewing literature has become a daunting and time-consuming task, as it requires researchers to search, compile, and assess an extensive number of publications (Webster and Watson 2002).

As such, there has been an increasing interest in using Information Systems to increase efficiency, productivity, and bring about operational excellence in the literature review process (Brendel et al. 2020; Watson and Webster 2020). In particular, recent advancements in Machine Learning and Text Analytics provide promising models, such as GPT-3 developed by OpenAI (Brown et al. 2020), that can further support researchers in conducting high-volume systematic literature reviews. As existing software developed for reviewing literature is usually isolated and lacks interoperability (Beller et al. 2018; Felizardo and Carver 2020), researchers might run the risk of not being able to efficiently leverage emerging advancements in Machine Learning and Text Analytics in their systematic literature reviews and thus fall short of their possibilities.

Following the "Design Science Research Methodology Process Model" proposed by Peffers et al. (2007) this paper aims to answer the following general research question: How can recent Machine Learning and Text Analytics advancements be leveraged efficiently in the systematic literature review process? In particular, we aim to answer this research question by demonstrating the design of a flexible, modifiable, and extendable artifact for holistically supporting intelligent systematic literature reviews based on available open-source tools. The artifact framework and its documentation are accessible under an associated GitHub repository[1].

Related Work

An early definition by Feldman and Dagan (1995, p. 1) describes "Knowledge Discovery in Textual Databases" or Text Analytics as the exploration of structured or unstructured text data in order to discover relevant and interesting patterns. Tandel et al. (2019) describe five overarching tasks of Text Analytics: (1) *information retrieval*, (2) *information extraction*, (3) *summarization*, (4) *clustering*, and (5) *categorization*. *Information retrieval* is the task of discovering documents that contain required information from a large collection of documents (Guo et al. 2019). *Information extraction* describes the automated extraction of useful information based on a well-defined request (Niklaus et al. 2018). Allahyari et al. (2017) describe *text summarization* as the automated task of reducing a text to a concise summary, while at the same time maintaining its key information. As described by Hotho et al. (2005), *clustering* has the task of partitioning a set of documents based on their similarity, and *categorization* assigns keywords or categories to documents based on a predefined set. Text Analytics has been utilized in various applications in Information Systems research, such as analyzing consumer brand engagement online (Chai et al. 2020; Kulkarni et al. 2020), detecting fake news (Ghosh and Shah 2019; Sarin and Kumar 2020), and analyzing government discourse (Cogburn 2020; Cogburn et al. 2020). Furthermore, recent advancements in Text Analytics, such as Bidirectional Encoder Representations from Transformers (BERT) by Google (Devlin et al. 2019) and GPT-3 by OpenAI (Brown et al. 2020), promise improved performances and a wide range of new possibilities for application. Text Analytics techniques are already in use to support literature reviews, for example by performing clustering (Galati and Bigliardi 2019; Marjanovic and Dinter 2018), bibliometric analysis (Chen et al. 2019; Pascal and Renaud 2020; Ribeiro et al. 2020), and more recently question answering (Kierszbaum and Lapasset 2020; Schmidt et al. 2020).

This is in line with findings by Felizardo and Carver (2020) describing that researchers have become more invested in the automation of the systematic literature review process in recent years. However, tools are often developed in isolation with limited or no compatibility and interoperability. This slow and fragmented development hinders researchers from fully leveraging the potential of automation in systematic literature reviews. As such, approaches from a more holistic point of view are required. In their research, Tauchert et al. (2020) propose an artifact that points towards this overall goal. However, the presented artifact mainly focuses on the process and the implementation of already established techniques rather than the artifact's extensibility and flexibility. As pointed out by Marshall and Wallace (2019), another problem of these isolated approaches of academic groups is their lack of maintenance and thus the eventual obsoleteness of their artifact. At the same time, however, it is assumed that commercial software companies are slow to integrate emerging Machine Learning and Text Analytics advancements due to limited demand.

The "Systematic Review Toolbox"[2] developed by Marshall and Brereton (2015) provides a community-driven and regularly maintained catalog of 212 tools (as of 02/2021) supporting systematic reviews. In this catalog, 77 tools are listed in the area of text analytics, machine learning, and data mining; 23 tools are identified as holistic support for the entire literature review process. However, the vast majority of these 23 tools mainly focus on supporting the process of manually performed literature reviews.

Methodology

The development of the artifact follows the "Design Science Research Methodology Process Model" proposed by Peffers et al. (2007). It describes six activities that serve as guidelines for carrying out design science research: (1) *identify problem and motivate*, (2) *define objective of a solution*, (3) *design and*

[1] Associated GitHub repository: https://github.com/HoliMoLiRev/HoliMoLiRev

[2] Website of the "Systematic Review Toolbox": http://systematicreviewtools.com

development, (4) *demonstration*, (5) *evaluation*, and (6) *communication*. The *identify problem and motivate* activities entail the identification and definition of the specific research problem to be addressed, which in our research comprises the increasing difficulty in reviewing the growing torrent of scientific publications. The *define objectives of a solution* activity requires researchers to define feasible and meaningful objectives for the proposed solution, which includes holistic process support and leveraging state-of-the-art Machine Learning and Text Analytics techniques. In the *design and development* activity an artifact is created that aims to achieve the objectives of the research, which we realize by developing a modifiable framework. In the *demonstration* activity, the artifact is put into practice and applied to the research problem. In our research, the artifact is used to generate and analyze a literature corpus for a concurrent research project in the Digital Twin domain. In the *evaluation* activity, the obtained results are then assessed based on the predefined objectives and feedback gathered from the project. Finally, the *communication* activity requires researchers to present the artifact in its entirety to a relevant audience, which is realized through this paper and its associated GitHub repository. As proposed by Peffers et al. (2007), we follow activities (3) to (6) iteratively by continuously evaluating and refining the artifact in increments throughout the development process.

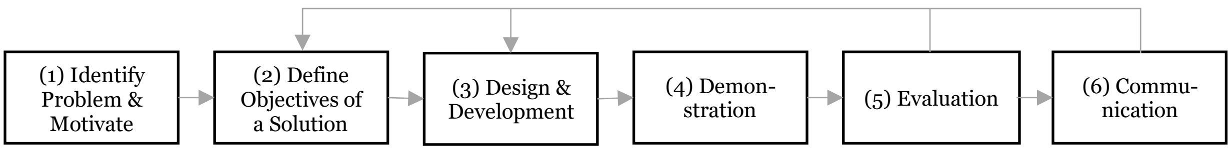

Figure 1. "Design Science Research Methodology Process Model" (Peffers et al. 2007)

Artifact Description

In this section, we describe the artifact's context, its conceptual architecture, its main tools used for instantiation, and its selected integrated features.

Artifact Context

In order to integrate new emerging Machine Learning and Text Analytics technologies efficiently into the artifact, its ability to be modified and extended by researchers has to be emphasized. Therefore, we focused primarily on open-source tools and libraries, as they meet these requirements and allow for extending and modifying the artifact without experiencing lock-in effects in the future. When choosing the programming language to develop the artifact, we selected Python due to its simplicity, its available Text Analytics libraries as well as its extensive documentation and large community. As development and runtime environment for the artifact we chose JupyterLab.

Conceptual Architecture

Inspired by the systematic literature review process and focusing on a centralized data storage, the artifact incorporates six different conceptual components, as shown in Figure 2.): (1) *collection*, (2) *import*, (3) *data storage*, (4) *filtering & screening*, (5) *transformation*, and (6) *analysis & visualization*. The *collection* component is used to download PDF documents and metadata of scientific literature (depicted in Figure 2.) as a gray box) from corresponding databases or journal websites. The *import* component converts and loads the PDFs and metadata into the *data storage* component. The *data storage* component serves as a central repository. It follows a document-oriented data model with papers as central entities, including their content (e.g., as PDF or plain text) and metadata (e.g., authors, publication year, or outlet) as attributes. The *filtering & screening* component enables researchers to exclude papers based on their content and metadata. The *transformation* component allows for pre-processing the content and metadata of papers as different features in the *analysis & visualization* component require different input data. Finally, the *analysis & visualization* component supports researchers in generating findings from scientific literature and visualizing them by applying Text Analytics and visualization techniques. This architecture aims to allow researchers to add, modify or remove components around the central data store according to their requirements as well as to extend it with new emerging technologies.

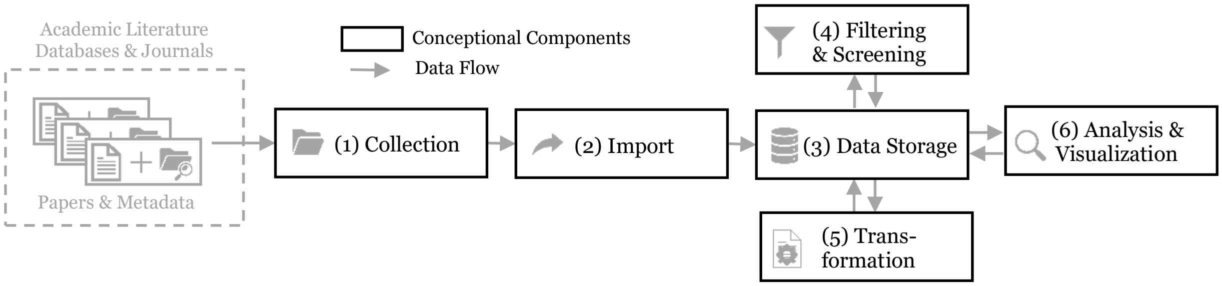

Figure 2. Schema of the artifact architecture and its conceptional components

Tool Selection

There is a variety of tools and libraries in Python related to "natural language"[3] with more than 1,200 projects in the Python Package Index alone (as of 02/2021). As such, it is essential to outline the criteria on which we selected the tools integrated into the artifact. Our tool selection criteria follow the "Vienna Principles" (Beller et al. 2018) in that we aim to use publicly available tools and therefore minimize barriers for researchers to adapt the artifact. However, the focus on open source also requires us to take further aspects into account, such as recent updates and maintenance, comprehensive documentation and community activity, usability, and balance between performance and maturity. As we emphasize compatibility, the tools are not viewed in isolation but assessed based on their architectural fit. Table 1.) provides an overview of the main tools used to build the artifact. It is important to point out that the selection and integration of these tools are not intended to be a final decision. The overview rather provides a first glance on how an artifact instantiation could be put into practice and therefore illustrates how researchers could benefit from the usage of these tools in their systematic literature reviews. As new trends and requirements emerge, tools in the artifact can be removed, substituted, or added by researchers accordingly.

Tool	Task	Access/ Repository
Zotero	Paper Collection; Duplicate Removal	Homepage/ Repository
pyzotero	Data Import from Zotero	Repository
langdetect	Language Detection	Repository
CrossRef API	External Reference Linking	Homepage/ Repository
Gephi	Network Visualization	Homepage/ Repository
MongoDB	Document Storage	Homepage
mongoengine	Access MongoDB	Repository
NLTK	Tokenization; Lemmatization; Stop Word Removal	Repository
spaCy	Named Entity Recognition; Text Classification	Homepage/ Repository
Gensim	LDA Topic Modeling	Homepage/ Repository
pyLDAvis	LDA Visualization	Repository
Transformers	Text Summarization; Question Answering	Repository

Table 1. Overview of main tools integrated in the artifact

Artifact Features

In this section, we describe the artifact features based on its six components: (1) *collection*, (2) *import*, (3) *data storage*, (4) *filtering & screening*, (5) *transformation*, and (6) *analysis & visualization*.

The *collection* component enables researchers to collect papers as well as to detect and remove duplicate papers. The *import* component converts the data (e.g., PDF to plain text) and loads it from the *collection* component into the *data storage* component. The *data storage* component is the central data repository which can be enriched by researchers with derived attributes that are generated in the *filtering & screening* component, *transformation* component, or *analysis & visualization* component. Thus, for example, the

[3] https://pypi.org/search/?q=%22natural+language%22

results of complex and time-consuming Text Analytics techniques can be stored for later evaluation and visualization.

The *filtering & screening* component implements various features that allow for the exclusion of papers based on predefined criteria. While filtering aims to exclude papers based on their metadata, screening excludes papers by assessing their content. Filtering can be performed based on generic metadata attributes, such as publication date or publication outlet. However, it is also possible to use derived attributes as criteria, such as the number of pages. A more advanced filtering feature is the artifact's ability to capture internal citations (references between papers included in a literature corpus) and external citations (paper in a literature corpus references a paper not included in this literature corpus). This allows researchers to identify relevant papers and authors in a corpus, based on citation networks, co-citation networks, or authorship networks. Furthermore, this can be utilized as semi-automated snowballing and provide researchers with the opportunity to discover papers that were previously overlooked or unintentionally excluded. When it comes to screening, the artifact enables researchers to use basic techniques, such as keyword occurrences or keyword distribution within each paper. The artifact also provides more sophisticated screening approaches, such as language detection, for example, if only papers written in English should be included. As can be seen in the Digital Twin literature corpus visualization in Figure 3.) on the left, using the artifact to generate a citation network capturing included papers (green nodes), excluded papers (red nodes), and external papers (blue nodes) may allow for identifying influential papers and serve as an affirmation that no substantial gaps exist within the literature corpus.

The *transformation* component allows for pre-processing available data, which is essential as different Text Analytics techniques in the *analysis & visualization* component require different input. This includes converting plain text to lowercase, splitting plain text into predefined chunks (tokenization), or storing plain text in a Bag-of-Words model. Furthermore, the *transformation* component features identifying and expanding acronyms and normalizing the different spellings of concepts. This allows researchers to recognize both acronyms and concepts as the same information. Further pre-processing features include the removal of numeric token and single-character tokens, lemmatization of tokens (Bergmanis and Goldwater 2018) as well as the removal of stop words (e.g., “but”, “and”, “we”, and “after”).

The *analysis & visualization* component aims to support researchers in analyzing scientific literature and visualizing the results. This ranges from basic features, such as word co-occurrence matrices or bibliometric analyses, to more advanced features, such as topic modeling, text summarization, and question answering. The artifact allows for topic modeling through Latent Dirichlet Allocation (Blei et al. 2003), which can be described as an unsupervised Text Analytics technique to identify underlying topics from a large collection of papers (Eickhoff and Neuss 2017). A generated topic model can then be visualized using the LDAvis method developed by Sievert and Shirley (2014), as can be seen in Figure 3.) on the right. Integrating text summarization based on pre-trained Transformer models enables researchers to automatically extract key statements in a paper without having to read the paper in its entirety. In addition, more specific inquiries can be answered by integrating extractive question answering based on pre-trained Transformer models. Question answering provides direct answers from a knowledge base to questions posed by users in natural language (Kwiatkowski et al. 2019). This enables researchers to find specific information through questions in natural language and receive answers based on knowledge from the content of papers (Schmidt et al. 2020). The basic and advanced features in the analysis component may be applied and combined synergistically.

Evaluation

In this section, we evaluate the artifact based on three aspects proposed by Gregor and Hevner (2013) for positioning and presenting research in Design Science: (1) validity, (2) efficacy, and (3) utility. In our research, the artifact was used to generate and analyze a literature corpus for a concurrent research project in the Digital Twin domain. As described by Kritzinger et al. (2018), Digital Twins can broadly be defined as “digital counterparts of physical objects” that are connected to each other.

Figure 3. Examples of visualization: Citation network in Gephi used for filtering and screening (left); Topic model visualization used for conceptualization in pyLDAvis (right)[4]

Validity describes whether an artifact contributes to achieving its stated objective (Gregor and Hevner 2013, p. 351). Our artifact aims to enable researchers to efficiently leverage Machine Learning and Text Analytics advancements in systematic literature reviews. When used to generate and analyze a literature corpus in the Digital Twin domain, the artifact reduced the time and effort required by researchers to generate, filter, and screen 1,024 papers in the literature corpus, in particular when compared to performing these tasks manually or being limited by the lack of interoperability of existing tools. Furthermore, semi-automatically conducting this step may have avoided potential human errors and bias. We were able to leverage Machine Learning and Text Analytics techniques, such as question answering, which facilitated information extraction. For example, querying each paper for "What is the definition of a Digital Twin?" provided a list of potential answers from papers and enabled us to rapidly identify and compare commonalities and differences between definitions.

Efficacy describes the degree of an artifact's contribution to achieving its stated objective, without taking its context into account (Venable et al. 2012). The artifact's interoperability has increased the degree of automation in conducting systematic literature reviews compared to existing isolated tools. It allows for relatively fast execution of multiple analysis types, which is especially utile in rapidly growing domains and high-volume systematic literature reviews. Researchers are able to publish their modified code and parameters, which fosters transparency and to some extent replication in the systematic literature review process through programmatical analysis when compared to manual systematic literature reviews. Its holistic approach, accessibility, and the absence of vendor lock-in allow researchers to modify and extend the artifact based on their requirements, which enables researchers to potentially leverage state-of-the-art tools and techniques, which may not be possible with existing tools. However, human intervention is still required for handling errors, inaccuracies, and leveraging synergies between artifact and researcher. However, it is important to note that the artifact's efficacy may depend on the user type. We expect users with basic programming skills to apply and adapt the artifact with ease and therefore be able to leverage its full functionality. This facilitates conducting systematic literature reviews using Machine Learning and Text Analytics advancements and may lead to new insights in scientific literature through potentially new and innovative analysis approaches.

Utility is achieved if the stated objective of an artifact has value outside its limited development environment objective (Gregor and Hevner 2013, p. 351). The artifact has only been assessed on the Digital Twin literature corpus, therefore making it difficult to evaluate this aspect. As the Digital Twin domain is multidisciplinary, papers were extracted from a variety of journals and conferences, leading to at least partial generalizability. However, its current requirements for technical expertise may limit its use to researchers within fields related to Information Systems and Computer Science. In general, the artifact demonstrates that semi-automation can be used to facilitate the integration of emerging Machine Learning and Text Analytics advancements in the systematic literature review process.

[4] Interactive versions of the images can be found at the following website: https://holimolirev.github.io/HoliMoLiRev/

Discussion, Limitations and Further Research

In the following section, we discuss decisions made during the development of the artifact and their consequences, the limitations of our research, and potential directions for further research and artifact development.

During the development of the artifact, several decisions had to be made regarding design and implementation. However, due to the limited space within this paper, not every decision can be discussed in detail. The open-source reference management tool Zotero provides an intuitive GUI and a browser plugin that allows for relatively easy extraction of PDFs and metadata (Böhner et al. 2020). Through this semi-automated functionality, the artifact can rapidly compile a literature corpus and is less prone to human errors, therefore improving the literature corpus quality. The integration of the document-oriented database MongoDB allows for storing binary data of any size and enables researchers to add derived attributes to each paper, which might require substantial processing time when being created. The artifact's modifiable code components are mostly written in Python. Python's intuitive syntax and the use of well-documented libraries remove entry barriers for researchers that might not necessarily possess comprehensive coding skills. Furthermore, we were able to leverage existing state-of-the-art Machine Learning libraries, such as TensorFlow and PyTorch, and integrate more sophisticated Text Analytics techniques, such as text summarization and question answering. However, as mentioned by Olorisade et al. (2016), researchers and readers could view the artifact features as "magic tools" and hence run the risk of accepting its output without understanding, questioning, or reporting the underlying processes and implementations. The use of JupyterLab as a development and execution environment allows for enriching the code with markdown comments and detailed explanations. This gives researchers the opportunity to illustrate underlying concepts, processes, and implementation rather than just presenting the results. In accordance with vom Brocke et al. (2009) promoting the importance of rigorously documenting the search process, our artifact enables researchers to publish their modified source code, further increasing transparency by facilitating peer reviews and replication studies.

There are some limitations to our research that need to be addressed. As of now, the evaluation of the artifact is limited to the Digital Twin literature corpus. There are no insights from applying the artifact to other literature corpora or extensive user acceptance testing. By focusing on the selection of accessible and modifiable tools, we might have missed tools that have better performance or functionality when integrated.

In the future, the substitution or extension of integrated tools in various components of the artifact may be explored further. In accordance with the proposed "Vienna Principles" (Beller et al. 2018), we intend to make our artifact publicly available and therefore enable researchers to modify and extend the artifact based on their requirements. For example, the artifact could be extended by integrating ASReview, an open-source tool that applies Machine Learning techniques to rank papers based on relevance to the researcher's preferences, which are assessed through a small sample of manually screened abstracts (van de Schoot et al. 2021). The development of an easy-to-use GUI would further improve usability and make the artifact available to a broader audience beyond the Information Systems research community. Moreover, evaluating the artifact's functionality on literature collections from different domains as well as conducting expert interviews with Information Systems researchers and non-Information Systems researchers would further increase the artifact's validity. As proposed in the Task-Artifact Cycle (Carroll et al. 1991), the task of our research to semi-automatically support systematic literature reviews posed the requirements for the development of our artifact. By applying and evaluating our artifact on the Digital Twin literature corpus, the characteristics of the task itself changed as well. For example, by shifting from specific to broader searches in combination with advanced filtering techniques as well as highlighting and differentiating decisions between humans and technology. As more sophisticated technologies and tools emerge in Machine Learning and Text Analytics, our artifact provides researchers with a quick way of integrating new features without having to forego established functionalities. Overall, we understand our artifact as a system in development that serves as starting point towards a more refined proof of concept in the future.

Conclusion

In light of the continuously growing torrent of scientific literature, the goal of our research was to demonstrate how current Machine Learning and Text Analytics advancements can be leveraged efficiently to support researchers in systematic literature reviews. In accordance with the "Design Science Research

Methodology Process Model" proposed by Peffers et al. (2007), we developed and continuously evaluated a flexible, modifiable, and extendable artifact for holistically supporting intelligent systematic literature reviews based on open-source tools. The artifact in its current form allows researchers to increase transparency, repeatability, efficiency, and productivity in systematic literature reviews. Nevertheless, we see its core strength in its extensibility and thus, its ability to facilitate the use of existing and emerging advancements in Machine Learning and Text Analytics, particularly in high-volume systematic literature reviews (>1000 publications). Our study contributes to the body of knowledge in the areas of systematic literature review process automation and optimization by proposing an artifact that enables researchers to efficiently automate systematic literature reviews to a certain degree, as well as demonstrating its current possibilities and limitations. As such, it is a first step towards a more flexible and holistic approach of semi-automating systematic literature reviews and leveraging emerging advancements in Machine Learning and Text Analytics. We can expect this will provide a valuable aid to researchers that have to work under time pressure, such as the current COVID-19 vaccine research being done as well as other disciplines that face the challenge of evaluating large volumes of scientific papers within a given period. As we expect the trend of a rising number of scientific publications to continue, the support of Information Systems in conducting systematic literature reviews will become increasingly important.

REFERENCES

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., and Kochut, K. 2017. "Text Summarization Techniques: A Brief Survey," *ArXiv:1707.02268 [Cs]*. (http://arxiv.org/abs/1707.02268).

Beller, E., Clark, J., Tsafnat, G., Adams, C., Diehl, H., Lund, H., Ouzzani, M., Thayer, Kristina, Thomas, J., Turner, T., Xia, J., Robinson, K., Glasziou, P., Adams, C., Ahtirschi, O., Beller, E., Clark, J., Christensen, R., Diehl, H., Elliott, J., Glasziou, P., Graziosi, S., Kuiper, J., Lund, H., Moustgaard, R., O'Connor, A., Ouzzani, M., Riis, J., Robinson, K., Soares-Weiser, K., Thayer, Kris, Thomas, J., Turner, T., Tsafnat, G., Vergara, C., Wedel-Heinen, I., Xia, J., and On behalf of the founding members of the ICASR group. 2018. "Making Progress with the Automation of Systematic Reviews: Principles of the International Collaboration for the Automation of Systematic Reviews (ICASR)," *Systematic Reviews* (7:1), p. 77. (https://doi.org/10.1186/s13643-018-0740-7).

Bergmanis, T., and Goldwater, S. 2018. "Context Sensitive Neural Lemmatization with Lematus," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana: Association for Computational Linguistics, June, pp. 1391–1400. (https://doi.org/10.18653/v1/N18-1126).

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," *Journal of Machine Learning Research* (3:Jan), pp. 993–1022.

Böhner, D., Stöber, Thomas, and Teichert, Astrid. 2020. *Reference Management Software Comparison*, p. 19.

Brendel, A. B., Trang, S., Marrone, M., Lichtenberg, S., and Kolbe, L. M. 2020. "What to Do for a Literature Review? – A Synthesis of Literature Review Practices," *AMCIS 2020 Proceedings*. (https://aisel.aisnet.org/amcis2020/meta_research_is/meta_research_is/2).

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," in *ECIS*.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. 2020. "Language Models Are Few-Shot Learners," *ArXiv:2005.14165 [Cs]*. (http://arxiv.org/abs/2005.14165).

Carroll, J. M., Kellogg, W. A., and Rosson, M. B. 1991. "The Task-Artifact Cycle," in *Designing Interaction: Psychology at the Human-Computer Interface*, USA: Cambridge University Press, pp. 74–102.

Chai, M., Luo, J., and Lan, J. 2020. "Brand User Attention Model Based on Online Text Reviews: An Empirical Study of New Energy Automobile Brands," *Big Data and Analytics*, p. 13.

Chen, W., Xu, Y., Jin, R., and Wanatowski, D. 2019. "Text Mining–Based Review of Articles Published in the *Journal of Professional Issues in Engineering Education and Practice*," *Journal of*

Professional Issues in Engineering Education and Practice (145:4), American Society of Civil Engineers, p. 06019002. (https://doi.org/10.1061/(ASCE)EI.1943-5541.0000425).

Cogburn, D. L. 2020. *Analyzing Trends and Topics in Internet Governance and Cybersecurity Debates Found in Twelve Years of IGF Transcripts*, p. 10.

Cogburn, D. L., Thomas, D. K. S., and Lai, J. 2020. *Using Computational Text Mining to Understand Public Priorities for Disability Policy Towards Children in Canadian National Consultations*, p. 10.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *ArXiv:1810.04805 [Cs]*. (http://arxiv.org/abs/1810.04805).

Eickhoff, M., and Neuss, N. 2017. *TOPIC MODELLING METHODOLOGY: ITS USE IN INFORMATION SYSTEMS AND OTHER MANAGERIAL DISCIPLINES*, p. 22.

Feldman, R., and Dagan, I. 1995. "Knowledge Discovery in Textual Databases (KDT)," in *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, KDD'95, Montréal, Québec, Canada: AAAI Press, August 20, pp. 112–117.

Felizardo, K. R., and Carver, J. C. 2020. "Automating Systematic Literature Review," in *Contemporary Empirical Methods in Software Engineering*, M. Felderer and G. H. Travassos (eds.), Cham: Springer International Publishing, pp. 327–355. (https://doi.org/10.1007/978-3-030-32489-6_12).

Galati, F., and Bigliardi, B. 2019. "Industry 4.0: Emerging Themes and Future Research Avenues Using a Text Mining Approach.," *Comput. Ind.* (109), pp. 100–113. (https://doi.org/10.1016/j.compind.2019.04.018).

Ghosh, S., and Shah, C. 2019. "Towards Automatic Fake News Classification," *Proceedings of the Association for Information Science and Technology* (55:1), pp. 805–807. (https://doi.org/10.1002/pra2.2018.14505501125).

Gregor, S., and Hevner, A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37), pp. 337–356. (https://doi.org/10.25300/MISQ/2013/37.2.01).

Guo, J., Fan, Y., Pang, L., Yang, L., Ai, Q., Zamani, H., Wu, C., Croft, W. B., and Cheng, X. 2019. "A Deep Look into Neural Ranking Models for Information Retrieval," *ArXiv:1903.06902 [Cs]*. (http://arxiv.org/abs/1903.06902).

Hotho, A., Nürnberger, A., and Paass, G. 2005. "A Brief Survey of Text Mining," *LDV Forum*.

Kierszbaum, S., and Lapasset, L. 2020. "Applying Distilled BERT for Question Answering on ASRS Reports," in *2020 New Trends in Civil Aviation (NTCA)*, , November, pp. 33–38. (https://doi.org/10.23919/NTCA50409.2020.9291241).

Kritzinger, W., Karner, M., Traar, G., Henjes, J., and Sihn, W. 2018. "Digital Twin in Manufacturing: A Categorical Literature Review and Classification," *IFAC-PapersOnLine* (51:11), 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018, pp. 1016–1022. (https://doi.org/10.1016/j.ifacol.2018.08.474).

Kulkarni, U., Deokar, A. V., and Ajjan, H. 2020. "Mining Online Reviews to Uncover Consumer Brand Engagement," in *40th International Conference on Information Systems, ICIS 2019*, Association for Information Systems, January 1. (https://asu.pure.elsevier.com/en/publications/mining-online-reviews-to-uncover-consumer-brand-engagement).

Kwiatkowski, T., Palomaki, J., Redfield, O., Collins, M., Parikh, A., Alberti, C., Epstein, D., Polosukhin, I., Devlin, J., Lee, K., Toutanova, K., Jones, L., Kelcey, M., Chang, M.-W., Dai, A. M., Uszkoreit, J., Le, Q., and Petrov, S. 2019. "Natural Questions: A Benchmark for Question Answering Research," *Transactions of the Association for Computational Linguistics* (7), MIT Press, pp. 453–466. (https://doi.org/10.1162/tacl_a_00276).

Marjanovic, O., and Dinter, B. 2018. "Learning from the History of Business Intelligence and Analytics Research at HICSS – A Semantic Text Mining Approach," *Communications of the Association for Information Systems* (43), pp. 775–791. (https://doi.org/10.17705/1CAIS.04340).

Marshall, C., and Brereton, P. 2015. *Systematic Review Toolbox: A Catalogue of Tools to Support Systematic Reviews*. (https://doi.org/10.1145/2745802.2745824).

Marshall, I. J., and Wallace, B. C. 2019. "Toward Systematic Review Automation: A Practical Guide to Using Machine Learning Tools in Research Synthesis," *Systematic Reviews* (8:1), p. 163. (https://doi.org/10.1186/s13643-019-1074-9).

Niklaus, C., Cetto, M., Freitas, A., and Handschuh, S. 2018. "A Survey on Open Information Extraction," *ArXiv:1806.05599 [Cs]*. (http://arxiv.org/abs/1806.05599).

Olorisade, B. K., de Quincey, E., Brereton, P., and Andras, P. 2016. “A Critical Analysis of Studies That Address the Use of Text Mining for Citation Screening in Systematic Reviews,” in *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering*, EASE ’16, New York, NY, USA: Association for Computing Machinery, June 1, pp. 1–11. (https://doi.org/10.1145/2915970.2915982).

Pascal, A., and Renaud, A. 2020. “15 Years of Information System Design Science Research: A Bibliographic Analysis,” *Hawaii International Conference on System Sciences 2020 (HICSS-53)*. (https://aisel.aisnet.org/hicss-53/os/design_science_research/2).

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. “A Design Science Research Methodology for Information Systems Research,” *Journal of Management Information Systems* (24:3), pp. 45–77. (https://doi.org/10.2753/MIS0742-1222240302).

Ribeiro, M. I., Fernandes, A. J., and Lopes, I. 2020. “The use of Information and Communication Technologies in Agriculture: a bibliometric analysis,” *CAPSI 2020 Proceedings*, p. 14.

Sarin, G., and Kumar, P. 2020. “ConvGRUText: A Deep Learning Method for Fake Text Detection on Online Social Media,” *PACIS 2020 Proceedings*. (https://aisel.aisnet.org/pacis2020/60).

Schmidt, L., Weeds, J., and Higgins, J. P. T. 2020. “Data Mining in Clinical Trial Text: Transformers for Classification and Question Answering Tasks,” *ArXiv:2001.11268 [Cs]*. (http://arxiv.org/abs/2001.11268).

van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., and Oberski, D. L. 2021. “An Open Source Machine Learning Framework for Efficient and Transparent Systematic Reviews,” *Nature Machine Intelligence* (3:2), Nature Publishing Group, pp. 125–133. (https://doi.org/10.1038/s42256-020-00287-7).

Sievert, C., and Shirley, K. 2014. “LDAvis: A Method for Visualizing and Interpreting Topics,” in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, Baltimore, Maryland, USA: Association for Computational Linguistics, June, pp. 63–70. (https://doi.org/10.3115/v1/W14-3110).

Tandel, S. S., Jamadar, A., and Dudugu, S. 2019. “A Survey on Text Mining Techniques,” in *2019 5th International Conference on Advanced Computing Communication Systems (ICACCS)*, , March, pp. 1022–1026. (https://doi.org/10.1109/ICACCS.2019.8728547).

Tauchert, C., Bender, M., Mesbah, N., and Buxmann, P. 2020. “Towards an Integrative Approach for Automated Literature Reviews Using Machine Learning,” *Hawaii International Conference on System Sciences 2020 (HICSS-53)*. (https://aisel.aisnet.org/hicss-53/cl/text_analytics/4).

Torraco, R. J. 2005. “Writing Integrative Literature Reviews: Guidelines and Examples,” *Human Resource Development Review* (4:3), SAGE Publications, pp. 356–367. (https://doi.org/10.1177/1534484305278283).

Venable, J., Pries-Heje, J., and Baskerville, R. 2012. “A Comprehensive Framework for Evaluation in Design Science Research,” in *Design Science Research in Information Systems. Advances in Theory and Practice*, Lecture Notes in Computer Science, K. Peffers, M. Rothenberger, and B. Kuechler (eds.), Berlin, Heidelberg: Springer, pp. 423–438. (https://doi.org/10.1007/978-3-642-29863-9_31).

Watson, R. T., and Webster, J. 2020. “Analysing the Past to Prepare for the Future: Writing a Literature Review a Roadmap for Release 2.0,” *Journal of Decision Systems* (29:3), Taylor & Francis, pp. 129–147. (https://doi.org/10.1080/12460125.2020.1798591).

Webster, J., and Watson, R. T. 2002. “Analyzing the Past to Prepare for the Future: Writing a Literature Review,” *MIS Quarterly*, pp. xiii–xxiii.

White, K. 2019. “Publications Output: U.S. Trends and International Comparisons,” National Science Foundation. (https://ncses.nsf.gov/pubs/nsb20206/).