Expert Systems With Applications 162 (2020) 113851

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Research trends in text mining: Semantic network and main path analysis of selected journals

Hoon Jung [a], Bong Gyou Lee [b,*]

[a] Hana Institute of Finance, Seoul 07321, South Korea
[b] Yonsei University, Seoul 03722, South Korea

ARTICLE INFO

Keywords:
Text mining
Research trends
Keywords
Semantic network analysis
Main path analysis

ABSTRACT

In this study, network and main path analyses were conducted on 1856 studies related to text mining, by extracting keywords and citation information from the text of each paper. Our findings indicate that research papers on text mining have been published in 45 academic disciplines in the 1980s and 1990s, 105 disciplines in the 2000s, and 171 disciplines in the 2010s. The results show that using text mining as a research topic and method has rapidly increased. We also demonstrate that the main theme of text mining research is discourse and content analysis in the 1980s and 1990s, biology and data mining in the 2000s, and medicine and advanced text mining in the 2010s. Moreover, we examined the main citation path for text mining studies and suggest that the main focus of text mining studies has evolved from information science to information systems and technology management. Additionally, influential papers have been recently published in fields such as architecture and social ecology revealing the wide scope of text mining. This article presents an understanding of previously unexplored research trends in text mining and how these trends shed light on the most influential academic papers in the field.

1. Introduction

Text mining is a technique for extracting meaningful information from data in text form. The targets of text mining range from academic literature to social networking sites, posts and comments about the news, voice of the customer, speech to text (STT) data, and more. Text mining is also actively used as a means of analyzing research trends in various fields of study, such as information systems, technology management, education, library and information science, psychology, sociology, and others. However, there has not been enough study on the research trends of text mining itself, in spite of fact that text mining is being utilized in a variety of fields of study. Therefore, when researchers need to find papers with academic importance and contribution on text mining, they primarily rely on the reputation of the journal in which the paper is published and number of the citations the paper has received. The purpose of this study is to define which papers make a significant academic contribution, what is the main research path of current text mining studies, and to predict future research trends in text mining. To answer these questions, we analyzed 1856 papers about text mining stored in the international academic citation databases, Scopus and Web of Science. To find current trends in text mining, semantic network analysis and main path analysis are implemented as text mining methods in this paper.

2. Theoretical background

2.1. Text mining

"Text" in text mining is defined as a symbol stored in digital form. Images and video files are also defined as objects of text mining (Grimmer & Stewart, 2013). However, in this study, we consider text as data expressed in characters only. Text mining finds new information in human character-based data by extracting context and meaning using natural language and document processing techniques. The typical process of text mining analysis begins with pre-processing the collected text data. Usually at this stage a morphological analysis is performed to sort sentences into parts of speech. The main keywords are extracted based on key topics and words that appear simultaneously in the same paragraphs or sentences. Then the characteristics and frequency of the words are defined and analyzed through a variety of text mining techniques, such as keyword network analysis, association analysis, opinion mining, topic modeling, emotion analysis, and others. Text mining can

* Corresponding author.
E-mail address: bglee@yonsei.ac.kr (B.G. Lee).

https://doi.org/10.1016/j.eswa.2020.113851
Received 29 January 2020; Received in revised form 14 July 2020; Accepted 4 August 2020
Available online 11 August 2020
0957-4174/

focus an analysis on areas of interest to the researcher, but the interpretation by the researcher can be arbitrary. Hence, the generalization of the results of text mining should be done sparingly. Conclusions drawn in certain studies may not be repeated in other studies (DiMaggio, Nag, & Blei, 2013). However, despite some of these limitations, text mining has been actively used as an analysis technique in many academic studies due to the advantages of analyzing large amounts of unstructured data that cannot process during traditional data classification and analysis.

2.2. Network analysis

Network analysis, often called "keyword network analysis" or "semantic network analysis", interprets phenomenon through the use of networks built by linking words and words appearing in the text to create a map of relationships. Network analysis describes the structured results from uncategorized data rather than fully categorized data. In other words, from a relational perspective, not an independent object but the relationships between the objects that help a situation or phenomenon to be understood more clearly. The network consists of nodes corresponding to keywords and lines or links indicating the relationship between them. In this study, we utilize Gephi 0.9.2 and VOSviewer 1.6.9 as the software for network analysis. The measures to determine the centrality of a node, such as degree centrality which measures how many nodes are directly connected, closeness centrality which scores each node based on its closeness to all other nodes within the network, and betweenness centrality which measures the number of times a node lies on the shortest path between other nodes, are used. In this study, network nodes are visualized based on eigenvector centrality which values the relative importance of nodes.

2.3. Main path analysis

Main path analysis establishes the important route among the citation relationships of papers. It is a technique that sheds light on the academic, behind-the-scenes relationship and the broadening path of knowledge by visualizing the use of citations between institutions. In this study, the main path is derived based on the search path count (SPC). The SPC is the total number of times that a link is traversed from the source to the end of the path in which a paper is cited. We used four types of main path analysis; forward local main path, backward main path, global main path, and key-route main path analysis. Forward local main path and backward local main path analyses select and connect the link with the largest SPC at each contact point. The global main path analysis selects the path where the total sum of the SPCs is the largest and the key-route main path analysis selects the path where the largest link is first and combines the important path forward and backward. The key-route main path search solves the problem of missing some routes and includes all important connections. We used the software, Pajek 64–5.07a as the main path analysis tool.

2.4. Literature analysis

There have been several studies on research trends of academic fields such as information science (Lee, Kim, & Kim, 2010), education (Hung, 2012), machine learning (Sharma, Kumar, & Chand, 2018), biomedicine (Zhai et al., 2015), business intelligence (Moro, Carneiro, Cortez, & Rita, 2015), and medical informatics (Kim & Delen, 2018) using quantitative analysis on peer reviewed papers. All of these studies are based on text-mining techniques that examines papers in journal database to reveal the research trends. A text mining-based research trend analysis has been conducted not only for the specific field of study but also for cross-disciplinary studies. Calero-Medina and Noyons (2008) analyzed all the studies having "absorptive capacity" as the keyword of the paper by text mining and revealed which academic fields used the concept of absorption capacity. In the present study, we collected and analyzed the papers that contain "text mining" or "text analysis" in the title or author keyword of the paper to identify the types and trends of the academic fields in which text mining was utilized, the major research topics, and key papers concerning the research on text mining. As mentioned, even though text mining is being utilized in a variety of fields of study, few scholars have examined research trends of text mining itself. This study is apparently the first study to comprehensively analyze the research trends in text mining.

3. Analysis and classification of text mining studies

3.1. Scope of analysis

The data sources used in this study were Web of Science and Scopus. Moro et al. (2015) conducted searches with keywords such as "banking" and "Business Intelligence" to study research trend of Business Intelligence in banking. If the authors judge subjectively whether a paper is a study on "Business Intelligence in banking" or not, it will weaken the reliability of the results. Moreover, if there are thousands of papers on Business Intelligence, it is almost impossible to analyze the relevance respectively. Hence, this study collected all the papers that contains the word "text mining" or "text analysis" in the title or author keyword of the paper through text mining technique to select relevant journals, targeting all the English journals listed in Scopus for the past 40 years.

However, when searching for two consecutive words like "text mining" or "text analysis", papers such as, "Affect analysis of text using fuzzy semantic typing" (Subasic & Huettner, 2001), "A machine learning approach to sentiment analysis in multilingual web texts" (Boiy & Moens, 2009), and "Text and structural data mining of influenza mentions in web and social media" (Corley, Cook, Mikler, & Singh, 2010), would not be found. Therefore, we searched for studies which included "text" and "mining" or "text" and "analysis" in the title. We also considered searching for "content analysis" as a keyword phrase which is typically used in the same way as "text analysis", but when "content analysis" was used, a rhetorical or contextual study on lyrics, "A Content-based Analysis of Shahriar's Azerbaijani Turkish Poem Getmə Tərsa Balası in Terms of Religious Images and Interpretations", (Mozaheb, Shahiditabar, Monfared, & Mirzapour, 2016), qualitative research papers on specific topics, "The Rap on Chicano and Black Masculinity: A Content Analysis of Gender Images in Rap Lyrics" (Baker-Kimmons & McFarland, 2011), "Experiences of living with dementia: Qualitative Content Analysis of semi-structured interviews" (Mazaheri et al., 2013), were included in the search results. Hence, when both "text" and "mining" are included in the titles of literature, or both "text" and "analysis" are included, we considered them as meeting the criteria for this study. We also decided to analyze papers published since 1980 that received more than one citation by the end of December 2019. We categorized the selected journals into (1) the 1980s and 1990s, when text mining was rarely used; (2) the 2000s, when the internet and digital literature were spreading; (3) the 2010s, when text mining was expanding in various disciplines and smart phones and social networking services became popular. The purpose of this analysis is to

1980-1990s	2000s	2010s
Computer Science Artificial Intelligence	Computer Science Artificial Intelligence	Computer Science Artificial Intelligence
Computer Science Information Systems	Computer Science Information Systems	Computer Science Information Systems
Information Science Library Science	Computer Science Theory Methods	Computer Science Interdisciplinary Applications
Computer Science Interdisciplinary Applications	Computer Science Interdisciplinary Applications	Mathematical Computational Biology
Engineering Electrical Electronic	Mathematical Computational Biology	Education Educational Research
Computer Science Theory Methods	Biochemical Research Methods	Information Science Library Science
Computer Science Software Engineering	Biotechnology Applied Microbiology	Engineering Electrical Electronic
Medical Informatics	Information Science Library Science	Linguistics
Computer Science Hardware Architecture	Computer Science Software Engineering	Language Linguistics
Health Care Sciences Services	Language Linguistics	Multidisciplinary Sciences

Fig. 1. Top 10 academic fields of text mining studies by period.

identify the characteristics and implications of recent research trends in text mining and predict the future of main path analysis in text mining research.

3.2. The academic fields of text mining studies

Appendices 1, 2, and 3 show the classification of papers containing "text" and "mining" or "text" and "analysis" in the title based on the Web of Science data. Text mining studies were published in 45 academic fields in the 1980s and 1990s (1980–1999), 105 in the 2000s (2000–2009), and 171 in the 2010s (2010–2019). In other words, the papers related to text mining are widely applied to various academic studies and the quantitative trend is increasing.

Fig. 1 shows the categories of computer science artificial intelligence and computer science information systems held the top 10 places from 1980 to 2019 and information science library science dropped slightly from 3rd place in 1980s to 6th place in 2010. The category of engineering electrical electronic fell from fifth place in the 1980s and 1990s to seventh place in the 2010s. Computer science software engineering fell from 7th place in the 1980s to 9th place in the 2000s and dropped

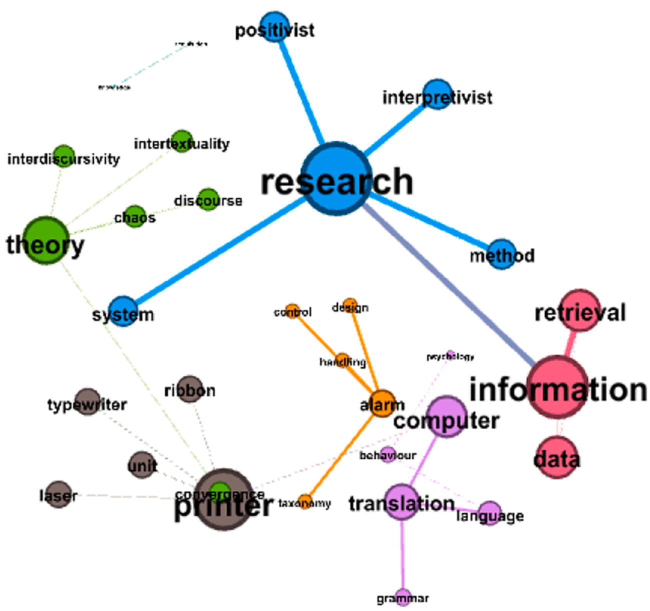

Fig. 2. Keyword network for 1980s and 1990s (author's keyword).

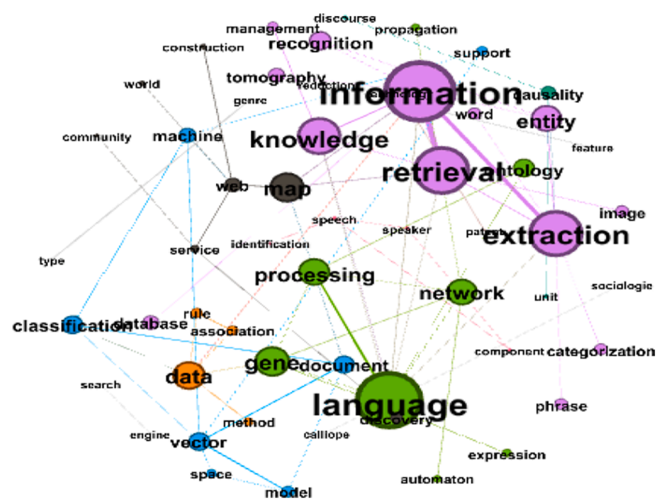

Fig. 3. Keyword network for 2000s (author's keyword).

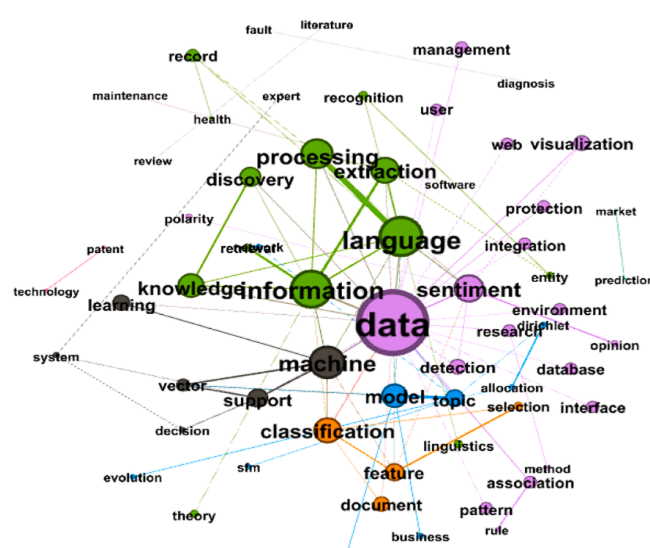

Fig. 4. Keyword network for 2010s (author's keyword).

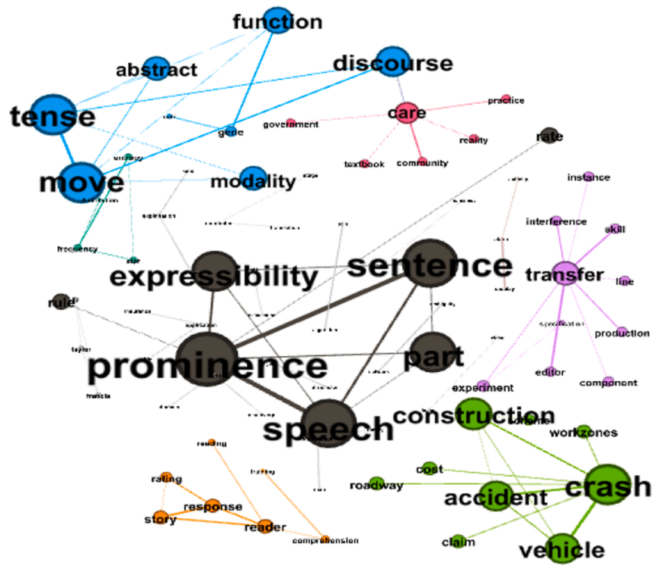

Fig. 5. Keyword network from abstract of text mining studies in 1980s and 1990s.

again to 13th place in the 2010s. On the other hand, medical informatics, computer science hardware architecture, and health care sciences services, which ranked 8th, 9th and 10th respectively in the 1980s and 1990s, have been pushed out of the top 10 since the 2000s.

4. Semantic network analysis of selected journals

4.1. Network analysis using Gephi

We conducted a semantic network analysis based on the keywords created by the author(s) in the abstract. Python software was used to extract keywords from each paper from Scopus and to derive the network between simultaneously occurring keywords in the same abstract. During data extraction and analysis, the threshold, which represents the frequency of pairs of keywords presented at the same time, was set at 2 from 1980 to 1999, 4 from 2000 to 2009, and 7 from 2010 to 2019. The network data between keywords in each paper was derived as shown in Figs. 2–4. These are visualized by Gephi 0.9.2. The size of the node represents the value of eigenvalue centrality and the unique color of the node signifies differences in modularity.

The analysis of the keyword network showed that the number of studies about text mining had increased along with the broadening of the scope of the studies so that the inter-keyword network was gradually becoming more complex. However, when the number of studies was smaller, such as in the 1980s and 1990s, words that might come from any paper such as “research”, “information”, and “theory” were highlighted as the main keywords, making it difficult to infer any specific implications. Text Mining is a study of morpheme, words, sentences, paragraphs, corpus, and documents. In this sense, the keywords presented by the author in the paper’s abstract are not in the form of a sentence but an unstructured set of words. Because the findings from the

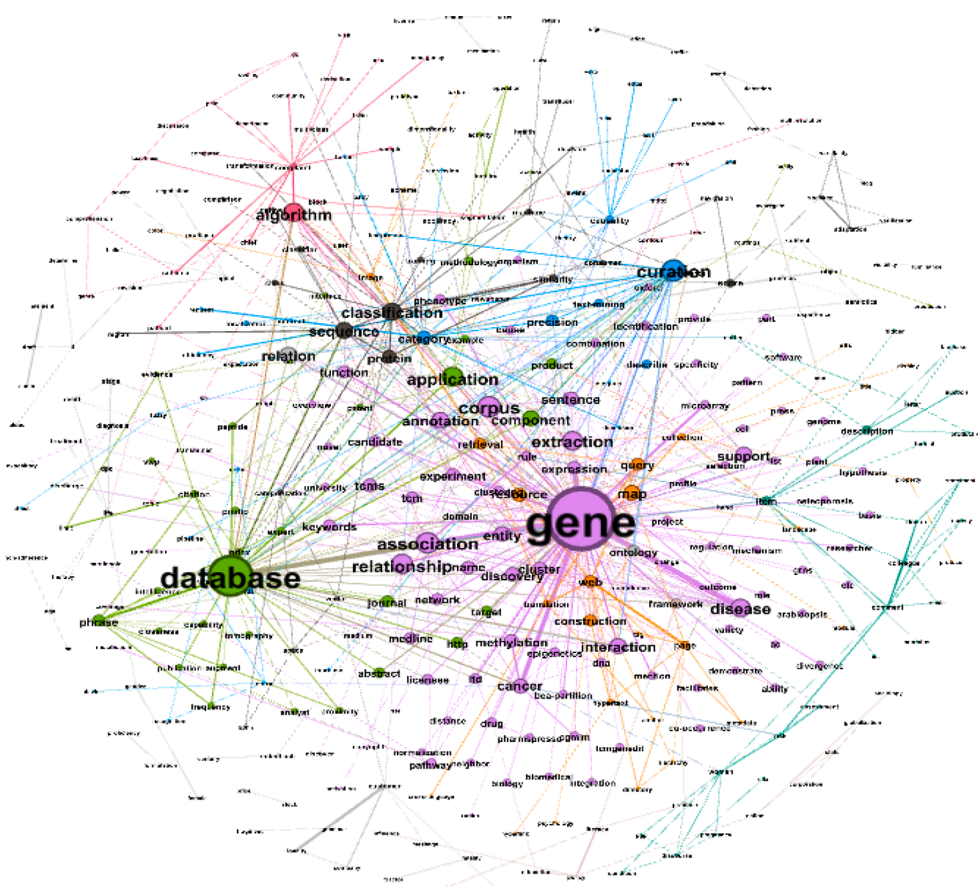

Fig. 6. Keyword network from abstract of text mining studies in 2000s.

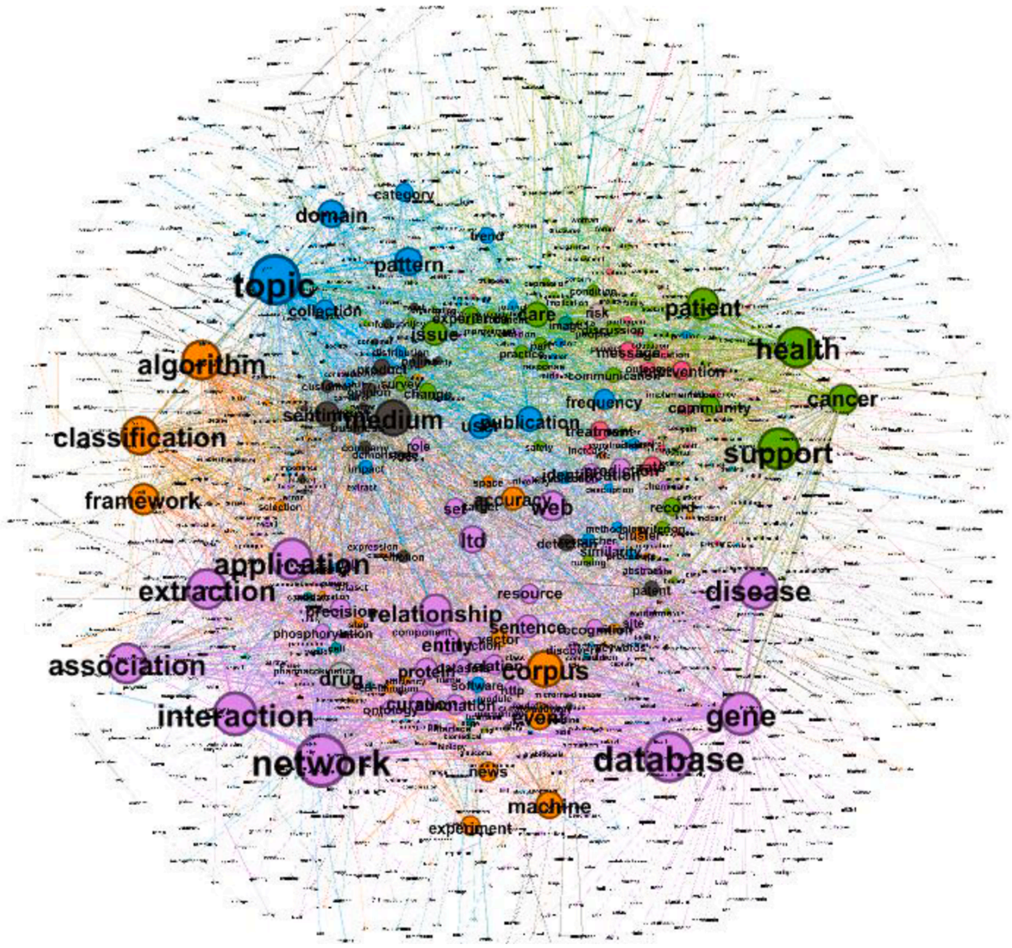

Fig. 7. Keyword network from abstract of text mining studies in 2010s.

network analysis for the keywords chosen by the author are limited in some ways, it is necessary to analyze the data at the paragraph level rather than the unstructured set of words. Therefore, we conducted an additional keyword network analysis based on the abstract for each paper. In addition, to exclude words that appear frequently in most papers regardless of their area of focus, we applied the ”stop word“ function of Python. Excluded words were: “article”, “result”, “information”, “system”, “study”, “research”, “knowledge”, “paper”, “data”, “document”, “approach”, “method”, “problem”, “system”, “tool”, “literature”, “task”, “technique”, “feature”, “design”, “language”, “structure”, “program”, “process”, “case”, “model”, “process”, “representation”, “word”, “time”, “year”, “work”, “use”, “term”, “search”, “author”, “field”, “level”, “area”, “development”, “source”, “concept”, “context”, “review”, “conclusion”, “group”, “help”, “quality”, “value”, “number”, “performance”, “student”, “science”, “test”, “aim”, “keyword”, “form”, “measure”, “report”, “type”, “processing”, “factor”, “effect”, “difference”, “content”, “amount”, “strategy”, “way”, “management”, “background”, “show”, “technology”, “question”, “measure”, “theme”, “challenge”, “interest”, “question”, “order”, “evaluation”, “service”, “world”.

We analyzed the text of abstracts for 1,856 papers and used Python to extract the frequency of keyword pairs appearing together within the

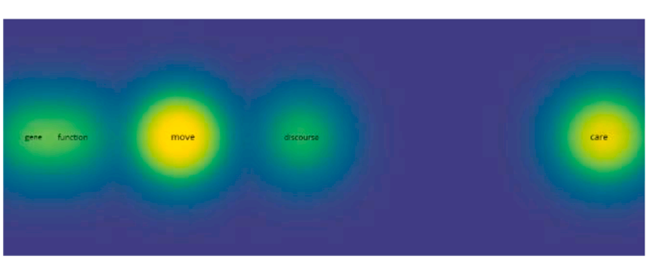

Fig. 8. Density visualization for abstract-based keywords in the 1980s and 1990s.

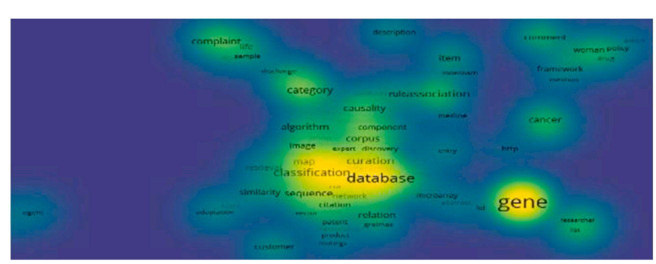

Fig. 9. Density visualization for abstract-based keywords in 2000s.

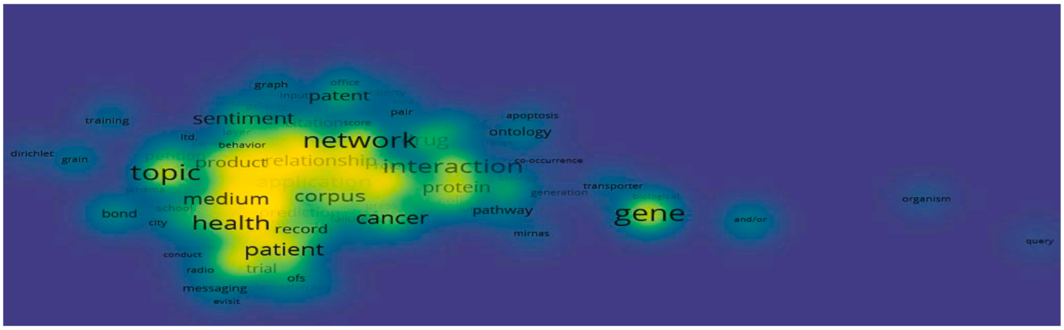

Fig. 10. Density visualization for abstract-based keywords in 2010.

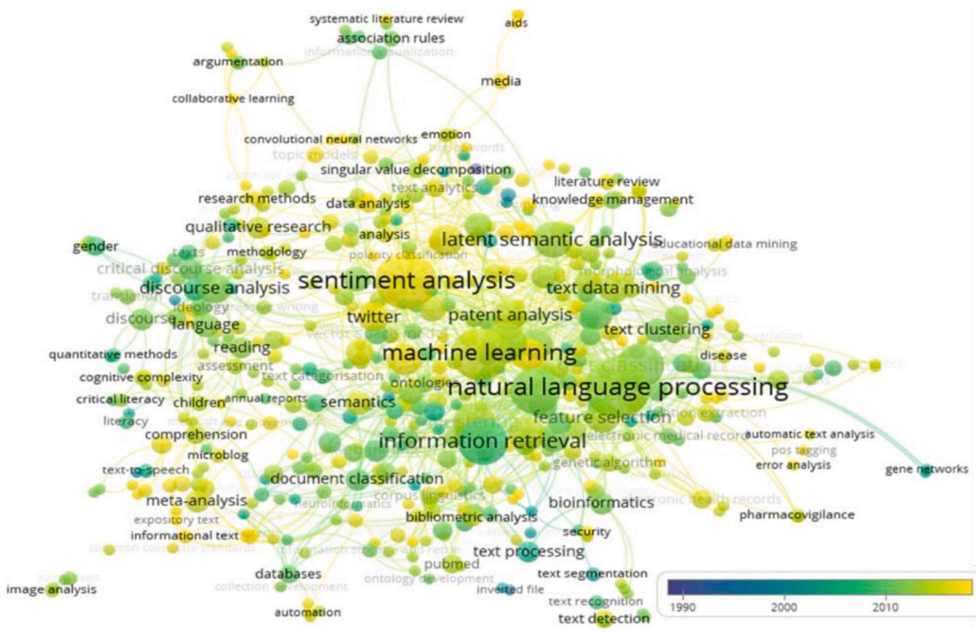

Fig. 11. Year-by-year visualizations of the keyword network (from 1980s to 2010s).

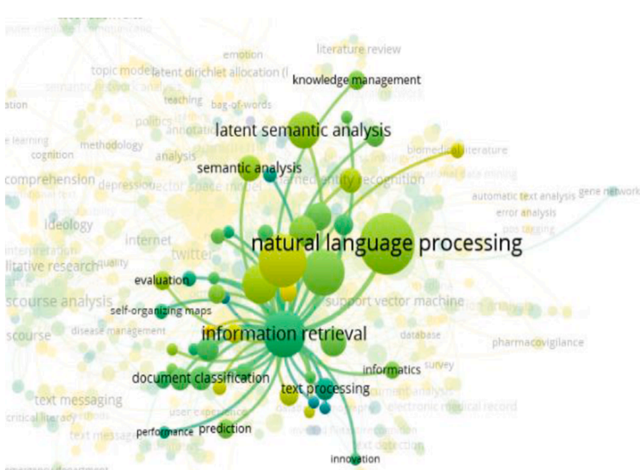

Fig. 12. Year-by-year visualizations of the keyword network (focused on 2000s).

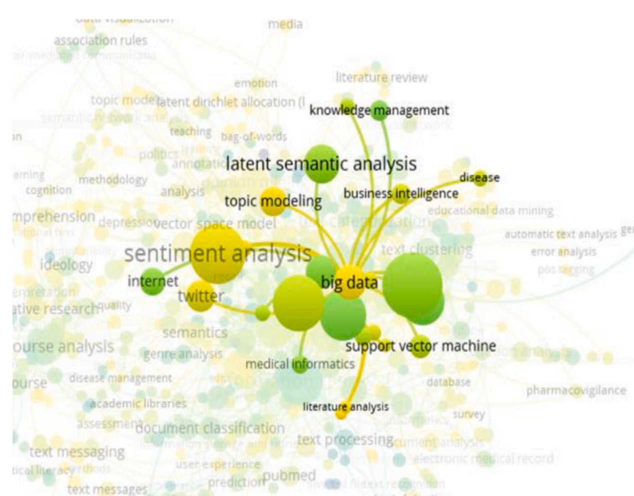

Fig. 13. Year-by-year visualizations of the keyword network (focused on 2010s).

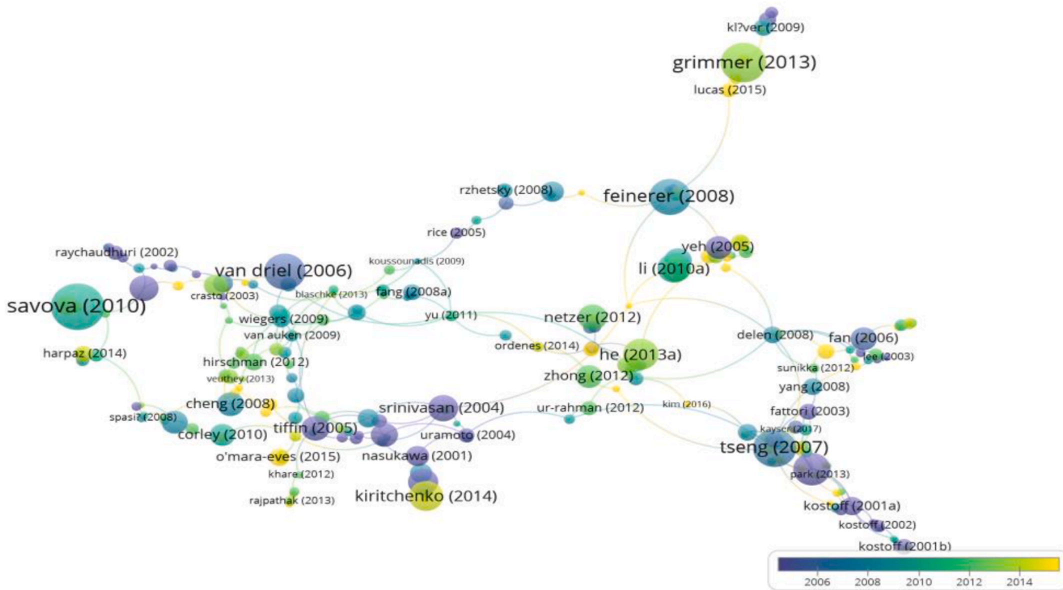

Fig. 14. Citation network of the seminal studies of text mining.

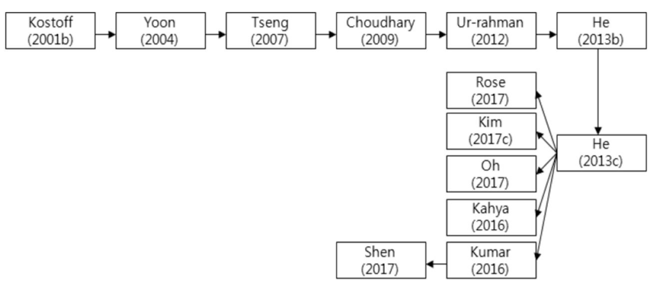

Fig. 15. Forward local main path.

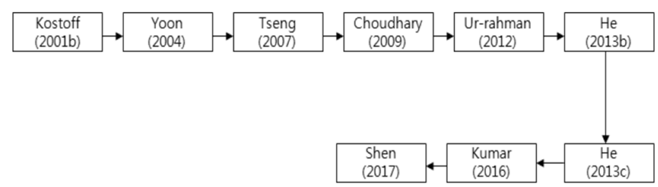

Fig. 17. Standard global main path.

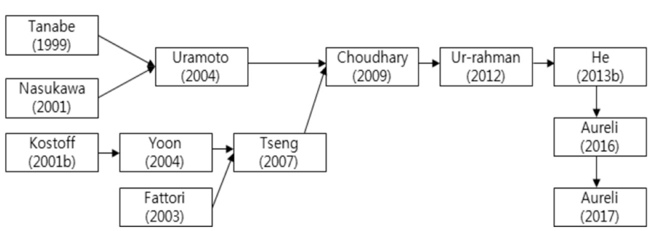

Fig. 16. Backward local main path.

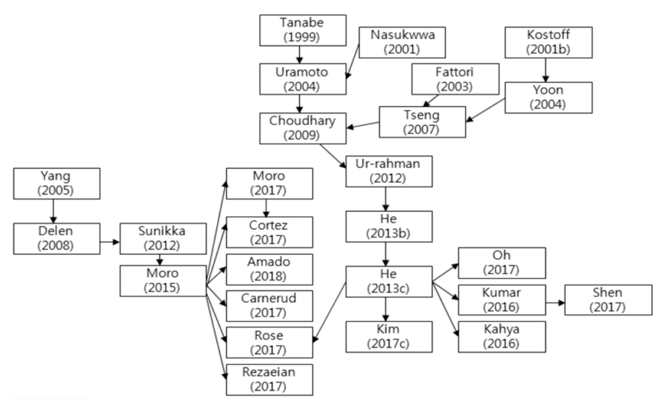

Fig. 18. Key-route local main path.

same abstract and visualized them through Gephi. Keyword networks derived for each period are shown in Figs. 5–7. Expanding the target of the analysis from the author's keywords to the text of the abstract, the visualized links between the keywords provided meaningful insight into the content of the papers. Additionally, the actual number of keyword pairs increased significantly with a threshold of 20 for all three time periods. The size of the nodes in this analysis is also based on eigenvalue centrality and the color of the nodes varies according to the modularity criteria. The keywords of the 1980s and 1990s imply content and intentions in dialogue or sentences, such as "sentence", "tense", and "speech". In contrast, biological and genetic keywords such as "gene" are prominent in the 2000s, and medical and health-related terms such as "health", "cancer", and "patient" appear as the main keywords in the 2010s.

4.2. Keyword network analysis using VOSviewer

A network analysis software, VOSviewer, was chosen for the simultaneous mapping and a clustering of nodes and the density visualization which can make it intuitive to identify critical areas. In VOSviewer, similarity between two words is proportional to the number of times they appear simultaneously and words with high similarity are placed close together. We designated the minimum total link strength of an

Table 1
Key thesis of text mining based on main path analysis.

Author	Title	Academic Field
Tanabe (1999)	MedMiner: An Internet text-mining tool for biomedical information, with application to gene expression profiling	Biology
Kostoff et al. (2001)	Text mining using database tomography and bibliometrics: A review	Technology Assessment and Forecasting
Fattori, Pedrazzi, and Turra (2003)	Text mining applied to patent mapping: A practical business case	Intellectual Property Information
Yoon and Park (2004)	A text-mining-based patent network: Analytical tool for high-technology trend	High Technology Management
Tseng et al. (2007)	Text mining techniques for patent analysis	Information Management
Choudhary, Oluikpe, Harding, and Carrillo (2009)	The needs and benefits of text mining applications on post-project reviews	Information and Communication Technology
Ur-Rahman and Harding (2012)	Textual data mining for industrial knowledge management and text classification	Information Systems
He (2013)	Improving user experience with case-based reasoning systems using text mining and Web 2.0	Information Systems
He et al. (2013)	Social media competitive analysis and text mining: A case study in the pizza industry	Information Management
Moro et al. (2015)	Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation	Information Systems
Shravan Kumar and Ravi (2016)	A survey of the applications of text mining in financial domain	Information Systems
Aureli (2016)	Sustainability disclosure after a crisis: A text mining approach	Social Ecology
Kim, Park, Yun, and Yun (2017)	What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management	Technology Assessment and Forecasting
Shen et al. (2017)	An integrated system of text mining technique and case-based reasoning for supporting green building design	Building Science

item as one.

In Figs. 8–10, we show the density visualization of keywords through VOSviewer, but no specific differentiated findings are identified compared to network visualization through Gephi. However, the overlay visualization feature of VOSviewer shows keyword trends by year and it can identify additional effects and characteristics of research trends. The overlay visualization visualizes word pairs that appeared together in text mining studies from 1980 to 2019. The keywords with a minimum number of concurrent appearances (more than 2) are visualized and the analysis are derived as shown in Figs. 11–13. In Figs. 11–13, the color of the node indicates the age of the study. The closer to yellow the node is, the more recent it is. On the other hand, the closer to blue the node is, the older it is, which means close to 1980. Fig. 11 is the result of an analysis of the all papers we targeted. Figs. 12 and 13 are also visualizations that highlight the 2000s and 2010s, respectively. In the 2000s, as we see in Fig. 12, the network is formed around "information retrieval", "natural language processing", and "document classification". In Fig. 13**,** which highlights keywords in relatively recent papers in the 2010s, the latest research trends such as "big data", "twitter", "topic modeling", and "sentiment analysis" are emphasized.

5. Main path analysis of text mining studies

We conducted a main path analysis on two citation databases. The targeted set of studies of the main path analysis are the same as the set used in the previous network analysis. The software used for the main path analysis was VOSviewer, Gephi, and Pajek. First, we used VOSviewer and Gephi to create citation network relationships between studies on text mining and utilized Pajek to calculate the SPC and visualize the main path. Since the paper's publication year is displayed in the visualization results, we implemented the main path analysis both by separating the time periods and by combining the time periods (Fig. 14).

In this study, four types of main path analysis and visualization were performed; forward local main path, backward local main path, key route local main path, and standard global main path. The final results of the main path analyses are shown in Figs. 15–18. From the analysis, it is possible to identify seminal papers on the citation path and to find academic fields in which text mining-related papers were published for each year. Moreover, we can establish the order and direction of the development of the main route. In Fig. 15, the forward local main path analysis, "Social media competitive analysis and text mining" (He, Zha, & Li, 2013), a study on social media and text mining published in the *International Journal of Information Management*, is followed by a paper on architecture, "An integrated system of text mining technique and case-based reasoning for supporting green building design" (Shen, HangYan, Ya, & Zhang, 2017). In Fig. 16, unlike the forward local main path, the path starting with Tanabe (1999) and Nasukawa (2001) is shown as an additional main path. Fig. 18 shows the key-route local main path analysis which minimizes the omission of certain routes, indicates that He et al. (2013) has influenced various academic fields. Also, Moro, Cortez, and Rita (2015), which uses the latent Dirichlet allocation (LDA) technique to analyze the literature related to the banking industry's business intelligence system, has been cited in various academic fields. In particular, "Text mining using database tomography and bibliometrics: A review" (Kostoff, Toothman, Eberhart, & Humenik, 2001) is a study of academic literature databases using text mining, This paper is in the early stages of the all four main path analysis results. Two papers which are studies on patent analysis, "A text-mining-based technology network: Analytical tool for high-technology trend" (Yoon & Park, 2004) and "Text mining techniques for patent analysis" (Tseng, Lin, & Lin, 2007) are also present in the all four main path analysis results. Comparing the results of the main path analysis, significant studies on text mining are produced in the field of literature information (Kostoff et al., 2001) in the early 2000s, but later appeared in a number of studies related to technology management, patents, and information systems. Recently, the scope of text mining studies has expanded to the analysis of social media and other social phenomena. Other current seminal studies in the main path can be found in social ecology and architecture. We also evidence that He et al. (2013) and Moro et al. (2015) authored important papers that have contributed to the broad spread of knowledge and played an important role in each field. In each of the above main path analyses, the papers occurring more than twice are shown in the order of the year of publication in Table 1.

These outcomes suggest that these papers are important studies in the main path of text mining studies. In most cases, the academic journals in which these papers are published are related to information systems or information management, technical management or technological management. The results reveal that papers with highly significant contributions in the field of text mining have been published in domains which were rarely related to text mining in the past. This effect corresponds to the overall expansion of the domain in which the paper was written. The results also show that most important papers use text mining as tools of research rather than focusing on text mining itself and this trend has recently increased.

6. Conclusion

We have analyzed studies on text mining from different time periods and derived research trends from the databases of peer-reviewed literature, Web of Science and Scopus. The results reveal that the number of academic fields where text mining is utilized has increased significantly and specifically identify in which areas of study text mining is being actively applied. In addition, we have extracted keywords occurring simultaneously from the abstracts of text mining papers to analyze network paths and identify the major keywords for each time period based on eigenvector centrality. Our findings indicate that conversational and speech-related keywords such as “discourse” and “speech” in the 1980s and 1990s, biomedical words like “gene” in the 2000s, and medical-related keywords such as “cancer”, and keywords related to advanced analytical techniques such as “topic” and “algorithm” are prominent in the 2010s. In addition, we specifically demonstrate the changes in keywords by year suggesting that research on big data, social media analysis, and emotion analysis (“big data,” “twitter,” “sentiment analysis”) are emerging as the latest research trends. Based on the results of keyword analysis of the abstracts of papers, we can expect that the academic fields publishing papers related to text mining will be steadily expanded in the future and new analysis techniques will also continue to be developed.

We also examined the main path of citation networks among 1,856 studies on text mining and presented evidence regarding influential authors and important contributing papers to the advancement of knowledge and development in the field of text mining. To sum up the results of the four main path analyses, the papers contributing to the academic development of text mining were produced in the information science literature until the early 2000s and in information system and technology management literature in 2010s. In addition, recent important research concerning text mining has been published, surprisingly, in social ecology and architecture and we observed the widespread spread use of text mining throughout various academic fields. We also highlight which studies have had an extensive impact on various academic fields. Recently it has been shown that text mining is increasingly used as a means of research rather than as the purpose of research. Based on the international databases of academic literature, we extracted and preprocessed citation and cited data between the studies. The contribution of this study is that it unearthed research trends on text mining from 1980 to the present and derives the implications of these trends by analyzing semantic networks and main paths within these networks. A future extension of this research would be to analyze research trends of text mining comparing “text mining as a means of study” with “text mining as a subject to study”.

CRediT authorship contribution statement

Hoon Jung: Visualization, Conceptualization. **Bong Gyou Lee:** Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2020.113851.

Appendix B

See Tables B1–B3

Table B1
Academic fields of journals that published papers about text mining in the 1980s and 1990s.

Academic Fields, Number of Papers, Percentage of Papers		%	Academic Fields, Number of Papers, Percentage of Papers		%
Computer Science Artificial Intelligence	**22**	**26.2**	Audiology Speech Language Pathology	1	1.2
Computer Science Information Systems	**20**	**23.8**	Biochemistry Molecular Biology	1	1.2
Information Science Library Science	**14**	**16.7**	Biology	1	1.2
Computer Science Interdisciplinary Applications	**13**	**15.5**	Biotechnology Applied Microbiology	1	1.2
Engineering Electrical Electronic	**12**	**14.3**	Chemistry Analytical	1	1.2
Computer Science Theory Methods	**9**	**10.7**	Chemistry Multidisciplinary	1	1.2
Computer Science Software Engineering	**8**	**9.5**	Education Special	1	1.2
Medical Informatics	**6**	**7.1**	Engineering Industrial	1	1.2
Computer Science Hardware Architecture	**5**	**6.0**	Genetics Heredity	1	1.2
Health Care Sciences Services	**5**	**6.0**	Linguistics	1	1.2
Telecommunications	4	4.8	Management	1	1.2
Computer Science Cybernetics	3	3.6	Mathematical Computational Biology	1	1.2
Engineering Biomedical	3	3.6	Mathematics Applied	1	1.2
Operations Research Management Science	3	3.6	Mathematics Interdisciplinary Applications	1	1.2
Statistics Probability	3	3.6	Medicine General Internal	1	1.2
Biochemical Research Methods	2	2.4	Medicine Legal	1	1.2
Ergonomics	2	2.4	Physics Fluids Plasmas	1	1.2
History Philosophy of Science	2	2.4	Psychology	1	1.2
Multidisciplinary Sciences	2	2.4	Psychology Applied	1	1.2
Neurosciences	2	2.4	Psychology Experimental	1	1.2
Optics	2	2.4	Psychology Multidisciplinary	1	1.2
Social Sciences Mathematical Methods	2	2.4	Rehabilitation	1	1.2
SOCIAL Sciences Interdisciplinary	1	1.2			

Table B2
Academic fields of journals that published papers on text mining in the 2000s.

Academic Fields, Number of Papers, Percentage of Papers		%	Academic Fields, Number of Papers, Percentage of Papers		%
Computer Science Artificial Intelligence	**138**	**30.2**	Integrative Complementary Medicine	2	0.4
Computer Science Information Systems	**104**	**22.8**	Literary Theory Criticism	2	0.4
Computer Science Theory Methods	**65**	**14.2**	Literature Romance	2	0.4
Computer Science Interdisciplinary Applications	**43**	**9.4**	Medieval Renaissance Studies	2	0.4
Mathematical Computational Biology	**37**	**8.1**	Ophthalmology	2	0.4
Biochemical Research Methods	**36**	**7.9**	Pharmacology Pharmacy	2	0.4
Biotechnology Applied Microbiology	**36**	**7.9**	Philosophy	2	0.4
Information Science Library Science	**31**	**6.8**	Psychology Applied	2	0.4
Computer Science Software Engineering	**30**	**6.6**	Transportation	2	0.4
Language Linguistics	**23**	**5.0**	Agriculture Multidisciplinary	1	0.2
Linguistics	20	4.4	Asian Studies	1	0.2
Engineering Electrical Electronic	19	4.2	Chemistry Analytical	1	0.2
Religion	18	3.9	Chemistry Multidisciplinary	1	0.2
Statistics Probability	18	3.9	Developmental Biology	1	0.2
Social Sciences Interdisciplinary	15	3.3	Ecology	1	0.2
Genetics Heredity	12	2.6	Education Special	1	0.2
Biochemistry Molecular Biology	11	2.4	Electrochemistry	1	0.2
Literature	11	2.4	Emergency Medicine	1	0.2
Operations Research Management Science	11	2.4	Endocrinology Metabolism	1	0.2
Psychology Experimental	11	2.4	Energy Fuels	1	0.2
Medical Informatics	10	2.2	Engineering Aerospace	1	0.2
Communication	8	1.8	Engineering Civil	1	0.2
Health Care Sciences Services	8	1.8	Ethics	1	0.2
Sociology	8	1.8	Geriatrics Gerontology	1	0.2
Management	7	1.5	History Philosophy of Science	1	0.2
Education Educational Research	6	1.3	Horticulture	1	0.2
Psychology Educational	6	1.3	Humanities Multidisciplinary	1	0.2
Psychology Mathematical	6	1.3	Law	1	0.2
Psychology Multidisciplinary	6	1.3	Literature German Dutch Scandinavian	1	0.2
Engineering Multidisciplinary	5	1.1	Materials Science Ceramics	1	0.2
Neurosciences	5	1.1	Materials Science Composites	1	0.2
Biology	4	0.9	Materials Science Multidisciplinary	1	0.2
Computer Science Cybernetics	4	0.9	Medicine Research Experimental	1	0.2
Computer Science Hardware Architecture	4	0.9	Music	1	0.2
Ergonomics	4	0.9	Neuroimaging	1	0.2
History	4	0.9	Nuclear Science Technology	1	0.2
Imaging Science Photographic Technology	4	0.9	Nursing	1	0.2
Multidisciplinary Sciences	4	0.9	Optics	1	0.2
Political Science	4	0.9	Physics Fluids Plasmas	1	0.2
Telecommunications	4	0.9	Physics Multidisciplinary	1	0.2
Acoustics	3	0.7	Physiology	1	0.2
Automation Control Systems	3	0.7	Planning Development	1	0.2
Business	3	0.7	Plant Sciences	1	0.2
Engineering Industrial	3	0.7	Psychiatry	1	0.2
Mathematics Interdisciplinary Applications	3	0.7	Psychology Biological	1	0.2
Psychology	3	0.7	Psychology Clinical	1	0.2
Public Environmental Occupational Health	3	0.7	Psychology Psychoanalysis	1	0.2
Anthropology	2	0.4	Psychology Social	1	0.2
Area Studies	2	0.4	Radiology Nuclear Medicine Medical Imaging	1	0.2
Cell Biology	2	0.4	Rehabilitation	1	0.2
Chemistry Physical	2	0.4	Remote Sensing	1	0.2
Education Scientific Disciplines	2	0.4	Transportation Science Technology	1	0.2
Food Science Technology	2	0.4			

Table B3
Academic fields of journals that published papers on text mining in the 2010s.

Academic Field, Number of Papers, Percentage of Papers		%	Academic Field, Number of Papers, Percentage of Papers		%
Computer Science Artificial Intelligence	**139**	**10.1**	Nuclear Science Technology	4	0.3
Computer Science Information Systems	**135**	**9.8**	Nutrition Dietetics	4	0.3
Computer Science Interdisciplinary Applications	**112**	**8.1**	Obstetrics Gynecology	4	0.3
Mathematical Computational Biology	**94**	**6.8**	Optics	4	0.3
Education Educational Research	**93**	**6.7**	Psychology	4	0.3
Information Science Library Science	**91**	**6.6**	Psychology Developmental	4	0.3
Engineering Electrical Electronic	**88**	**6.4**	Public Administration	4	0.3
Linguistics	**85**	**6.2**	Transportation	4	0.3
Language Linguistics	**83**	**6.0**	Anthropology	3	0.2
Multidisciplinary Sciences	**57**	**4.1**	Archaeology	3	0.2
Medical Informatics	52	3.8	Audiology Speech Language Pathology	3	0.2
Biochemical Research Methods	51	3.7	Biodiversity Conservation	3	0.2
Computer Science Software Engineering	50	3.6	Engineering Environmental	3	0.2
Operations Research Management Science	48	3.5	Food Science Technology	3	0.2
Communication	47	3.4	Green Sustainable Science Technology	3	0.2
Biotechnology Applied Microbiology	45	3.3	Imaging Science Photographic Technology	3	0.2
Management	44	3.2	Literature Slavic	3	0.2
Computer Science Theory Methods	40	2.9	Physics Applied	3	0.2
Health Care Sciences Services	37	2.7	Plant Sciences	3	0.2
Public Environmental Occupational Health	36	2.6	Psychology Psychoanalysis	3	0.2
Humanities Multidisciplinary	34	2.5	Remote Sensing	3	0.2
Statistics Probability	34	2.5	Social Sciences Biomedical	3	0.2
Business	31	2.2	Surgery	3	0.2
Psychology Multidisciplinary	25	1.8	Toxicology	3	0.2
Engineering Multidisciplinary	24	1.7	Cardiac Cardiovascular Systems	2	0.1
Social Sciences Interdisciplinary	23	1.7	Chemistry Analytical	2	0.1
Political Science	20	1.4	Classics	2	0.1
Biochemistry Molecular Biology	19	1.4	Criminology Penology	2	0.1
Medicine General Internal	19	1.4	Dentistry Oral Surgery Medicine	2	0.1
Psychology Experimental	16	1.2	Education Special	2	0.1
Telecommunications	14	1.0	Energy Fuels	2	0.1
Computer Science Hardware Architecture	13	0.9	Ethics	2	0.1
Engineering Industrial	13	0.9	Ethnic Studies	2	0.1
Literature	13	0.9	Geography	2	0.1
Medicine Research Experimental	13	0.9	Geosciences Multidisciplinary	2	0.1
Chemistry Multidisciplinary	12	0.9	History of Social Sciences	2	0.1
Environmental Sciences	12	0.9	Infectious Diseases	2	0.1
Mathematics Interdisciplinary Applications	12	0.9	Mathematics	2	0.1
Psychology Educational	12	0.9	Meteorology Atmospheric Sciences	2	0.1
Automation Control Systems	11	0.8	Orthopedics	2	0.1
Economics	11	0.8	Philosophy	2	0.1
Health Policy Services	11	0.8	Physics Fluids Plasmas	2	0.1
Physics Multidisciplinary	11	0.8	Physiology	2	0.1
Education Scientific Disciplines	10	0.7	Psychology Mathematical	2	0.1
Genetics Heredity	10	0.7	Radiology Nuclear Medicine Medical Imaging	2	0.1
History	10	0.7	Rheumatology	2	0.1
Planning Development	10	0.7	Social Work	2	0.1
Psychiatry	10	0.7	Theater	2	0.1
Psychology Clinical	10	0.7	Veterinary Sciences	2	0.1
Engineering Civil	9	0.7	Water Resources	2	0.1
Integrative Complementary Medicine	9	0.7	Agriculture Dairy Animal Science	1	0.1
Pharmacology Pharmacy	9	0.7	Anatomy Morphology	1	0.1
Neurosciences	8	0.6	Area Studies	1	0.1
Acoustics	7	0.5	Behavioral Sciences	1	0.1
Biology	7	0.5	Chemistry Applied	1	0.1
Business Finance	7	0.5	Clinical Neurology	1	0.1
Computer Science Cybernetics	7	0.5	Demography	1	0.1
Environmental Studies	7	0.5	Ecology	1	0.1
Law	7	0.5	Electrochemistry	1	0.1
Literature Romance	7	0.5	Endocrinology Metabolism	1	0.1
Nursing	7	0.5	Engineering Aerospace	1	0.1
Oncology	7	0.5	Engineering Mechanical	1	0.1
Psychology Applied	7	0.5	Family Studies	1	0.1
Psychology Social	7	0.5	Film Radio Television	1	0.1
Religion	7	0.5	Folklore	1	0.1
Sociology	7	0.5	Forestry	1	0.1
Chemistry Medicinal	6	0.4	Geography Physical	1	0.1
Construction Building Technology	6	0.4	Literature German Dutch Scandinavian	1	0.1
Ergonomics	6	0.4	Mathematics Applied	1	0.1
Instruments Instrumentation	6	0.4	Medical Laboratory Technology	1	0.1
International Relations	6	0.4	Medieval Renaissance Studies	1	0.1
Primary Health Care	6	0.4	Metallurgy Metallurgical Engineering	1	0.1
Rehabilitation	6	0.4	Microbiology	1	0.1
Respiratory System	6	0.4	Music	1	0.1

(*continued on next page*)

Table B3 (*continued*)

Academic Field, Number of Papers, Percentage of Papers		%	Academic Field, Number of Papers, Percentage of Papers		%
Asian Studies	5	0.4	Ophthalmology	1	0.1
Cell Biology	5	0.4	Otorhinolaryngology	1	0.1
Engineering Biomedical	5	0.4	Pediatrics	1	0.1
History Philosophy of Science	5	0.4	Physics Condensed Matter	1	0.1
Hospitality Leisure Sport Tourism	5	0.4	Physics Mathematical	1	0.1
Materials Science Multidisciplinary	5	0.4	Physics Nuclear	1	0.1
Social Sciences Mathematical Methods	5	0.4	Reproductive Biology	1	0.1
Substance Abuse	5	0.4	Sport Sciences	1	0.1
Transportation Science Technology	5	0.4	Tropical Medicine	1	0.1
Art	4	0.3	Women'S Studies	1	0.1
Engineering Manufacturing	4	0.3	Zoology	1	0.1
Literary Theory Criticism	4	0.3			

References

Aureli, S. (2016). Sustainability disclosure after a crisis: A text mining approach. *International Journal of Social Ecology and Sustainable Development, 7*(1), 35–49.

Baker-Kimmons, L., & McFarland, P. (2011). The rap on Chicano and black masculinity: A content analysis of gender images in rap lyrics. *Race, Gender & Class, 18*(1/2), 331–344.

Boiy, E., & Moens, M.-F. (2009). A machine learning approach to sentiment analysis in multilingual Web texts. *Information Retrieval, 12*(5), 526–558.

Calero-Medina, C., & Noyons, E. C. M. (2008). Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field. *Journal of Informetrics, 2*(4), 272–279.

Choudhary, A. K., Oluikpe, P. I., Harding, J. A., & Carrillo, P. M. (2009). The needs and benefits of Text Mining applications on Post-Project Reviews. *Computers in Industry, 60*(9), 728–740.

Corley, C. D., Cook, D. J., Mikler, A. R., & Singh, K. P. (2010). Text and structural data mining of influenza mentions in web and social media. *International Journal of Environmental Research and Public Health, 7*, 596–615.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U. S. government arts funding. *Poetics, 41*(6), 570–606.

Fattori, M., Pedrazzi, G., & Turra, R. (2003). Text mining applied to patent mapping: A practical business case. *World Patent Information, 25*(4), 335–342.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*, 267–297.

He, W. (2013). Improving user experience with case-based reasoning systems using text mining and Web 2.0. *Expert Systems with Applications, 40*(2,1), 500–507.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management, 33*(3), 464–472.

Hung, J. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology, 43*(1), 5–16.

Kim, Y. M., & Delen, D. (2018). Medical informatics research trend analysis: A text mining approach. *Health Informatics Journal, 24*(4), 432–452.

Kim, K., Park, O.-J., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change, 123*, 362–369.

Kostoff, R. N., Toothman, D. R., Eberhart, H. J., & Humenik, J. A. (2001). Text mining using database tomography and bibliometrics: A review. *Technological Forecasting and Social Change, 68*(3), 223–253.

Lee, J. Y., Kim, H., & Kim, P. J. (2010). Domain analysis with text mining: Analysis of digital library research trends using profiling methods. *Journal of Information Science, 36*, 144–161.

Mazaheri, M., Eriksson, L. E., Heikkilä, K., Nasrabadi, A. N., Ekman, S.-L., & Sunvisson, H. (2013). Experiences of living with dementia: Qualitative content analysis of semi-structured interviews. *Journal of Clinical Nursing, 22*, 3032–3041.

Moro, S. M. C., Cortez, P. A. R., & Rita, P. M. R. F. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications, 42*(3), 1314–1324.

Mozaheb, M. A., Shahiditabar, M., Monfared, A., & Mirzapour, F. (2016). A content-based analysis of Shahriar's Azerbaijani Turkish poem Getmə Tərsa Balası (A Christian Child) in terms of religious images and interpretations. *International Journal of Applied Linguistics & English Literature, 5*(2), 159–163.

Sharma, D., Kumar, B., & Chand, S. (2018). Trend analysis in machine learning research using text mining. *International Conference on Advances in Computing, Communication Control and Networking*, 136–141.

Shen, L., HangYan, H. F., Ya, W.u., & Zhang, Y.u. (2017). An integrated system of text mining technique and case-based reasoning (TM-CBR) for supporting green building design. *Building and Environment, 124*(1), 388–401.

Shravan Kumar, B., & Ravi, V. (2016). A survey of the applications of text mining in financial domain. *Knowledge-Based Systems, 114*(15), 128–147.

Subasic, P., & Huettner, A. (2001). Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems, 9*(4), 483–496.

Tanabe, L. (1999). MedMiner: An Internet text-mining tool for biomedical information, with application to gene expression profiling. *BioTechniques, 27*(6), 1210–1217.

Tseng, Y.-H., Lin, C.-J., & Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Information Processing & Management, 43*(5), 1216–1247.

Ur-Rahman, N., & Harding, J. A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. *Expert Systems with Applications, 39*(5), 4729–4739.

Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research, 15*(1), 37–50.

Zhai, X., Li, Z., Gao, K., Huang, Y., Lin, L., & Wang, L. (2015). Research status and trend analysis of global biomedical text mining studies in recent 10 years. *Scientometrics, 105*, 509–523.