A cluster and content analysis of data mining studies in Library and Information Science

Jinxuan Ma and Brady D. Lund

Emporia State University, School of Library and Information Management

Abstract: This study examines the use of data mining strategies in library and information science research journals, including the types of studies that use the strategy as well as most popular journals for publishing article that employ the strategies. A content and cluster analysis was performed with articles published in major LIS journals during the years 2006, 2012, and 2018. The *Journal of the Association for Information Science and Technology* had the most such articles with 42, with *Information Processing and Management* following close behind with 32. A cluster analysis performed based on word frequency in these articles' abstracts identified three unique clusters associated with the topics of Publications/Citations, Consumer Behavior, and Information/Media Use. This analysis indicates a shift away from Publications/Citations towards more Consumer Behavior-based data mining studies. The findings of this study may be significant for current researchers in preparing and publishing their own data mining-based studies and determining avenues for publishing their work.

Keywords: Library and Information Science Research, Data Mining, Data Analysis, Cluster Analysis, Research Methods

1. Introduction

With a vast collection of data continuously generating at increasing "volumes, velocities, and varieties" (Kaur, Sood, & Verma, 2020, p. 1463), big data and analytic implementation emerged in uncovering new information from diversified data sources to support users' better decision-making. Examples of those data sources are information system logs, social media posts, sensor information from detection equipment or IoT (Internet of Things) devices, and more. As one subdomain of artificial intelligence for knowledge discovery, data mining refers to a systematic process of identifying and discovering "new and non-trivial patterns, relations, and trends in large datasets" (Schuh et al., 2019). Compared to conventional data analytic approaches, data mining commonly employs machine learning and/or statistical methods when analyzing large datasets in diversified content and format (e.g., numerical, textual, or multimedia data). Data mining techniques have also been widely applied to

Received: 6.10.2020 / Accepted: 15.2.2021 ISSN 2241-1925

data-informed research areas, such as bioinformatics, consumer preferences, market trends, insurances, and information seeking behavior on online social media. Indeed, data mining has become a compelling and robust research strategy for knowledge discovery and management in library and information science (LIS) (Jones & Salo, 2018; Marchionini, 2017; Puarungroj et al., 2018). Though various data analytic techniques have existed and been utilized in LIS research for decades, its trends as a distinct method in LIS research remain unclear.

With data mining working as "all-encompassing" term for a broad usage of knowledge discovery strategies, this study aims to provide some clarity to the historical use of data mining strategies in LIS, by examining their presence in articles published in major LIS journals over the course of twelve years from 2006 to 2018. It selects a sample of 31 LIS journals based on a set of criteria used in similar studies of the nature of LIS (e.g., Tuomaala, Jarvelin, & Vakkari, 2014): Included in Tuomaala et al.'s 2014 study as one of the "top" journals in LIS and/or ranked in the 2019 Journal Citation Report of Social Science Citation Index, and published consecutively across the years of 2006, 2012, 2018. These years were selected, as opposed to long periods like 20 years (used in the Tuomaala study [refers to Tuomaala et al., 2014]) due to the belief that the rate of the production of scientific publications has grown significantly in recent years, while the topics of interest to researchers have evolved rapidly, making it more significant to illustrate changes over smaller periods of time. Six-year periods were employed in light of the last year examined in Tuomaala study and the last full year before this study began (2018). Three equidistant points—2006, 2012, and 2018—seemed most appropriate to examine shifts in research topics and approaches.

2. Literature Review

2.1. Data Mining Research in LIS

Data comes in many forms. Anything that can be perceived can be considered data and most of that data can be analyzed to produce novel insights about the world. For this literature review, however, we will focus on only two types of data that are most commonly used in LIS studies: text and numerical data.

Recent LIS studies have illustrated the multitude of research strategies that fall under data mining methods. One such strategy is using natural language processing (NLP) applications, such as Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA), which uses statistical modelling to generate emergent topics from a set of text/documents. One common use of LDA, both in research and industry, is in analyzing consumer reviews. Guen and Juyoung (2018) analyze consumer cosmetic reviews, identifying key attributes within these reviews and the relative polarity of sentiment associated to these attributes. Kim and Chun (2019) perform an analysis of car reviews using similar text mining procedures. Zhou et al. (2020) utilize text similarity measures to compare title and content of product reviews and the overall helpfulness of a particular review.

A popular use of text mining strategies is sentiment analysis. At the most basic level, this type of analysis compares each word in a selection of text to a reference dictionary that associates words with a particular sentiment (such as a "positive" or "negative" sentiment); the sentiment for the entire selection of text is then computed based on these word scores. This method is also valuable for analyzing consumer reviews to efficiently identify what the public thinks about a product/service. Recent contributions to this research area within LIS focus on refining the analysis strategy to assess sentiment more accurately. One approach is to focus on aspects of a specific language, as was done by Song, Park, and Shin (2019). Aspect-based sentiment analysis, used by Song et al. (2019), as well as Tubishat, Idris, and Abushariah (2018) and Yang, Zhang, Jiang, and Li (2019), breaks text selections into aspects of the "thing" being reviewed and then assigns sentiment values. For instance, in a restaurant review, there could be aspects such as "atmosphere," "service quality," and "food quality." This clearly provides more useful outputs than a review-by-review sentiment score.

Sentiment analysis is a type of text mining that falls under a greater topical umbrella of "text classification" studies. As noted by Altinel and Ganiz (2018), text classification is one of the most powerful means by which huge amounts of text-based data can be analyzed. Cluster analysis, discussed in the following section, is another type of analysis that can be used for text classification, by sorting documents into groupings based on similarities identified in their text. Many recent studies propose other methods of text classification. Feng, Guo, Jing, and Hao (2012) propose a Bayesian (variable, dynamic) method for text classification. Rehman, Javed, and Babri (2017) discuss the means through which a classifier of data in a text classification project can be evaluated, contrasting traditional measures like balanced accuracy, which utilizes true positive and false positive rates to evaluate a classifier's efficacy, with their own measure, normalized difference measure (NDM). NDM divides the true positive and false positive rate by minimum document frequency rate (i.e., the relevance/frequency of the classifier to the corpus of documents).

Along with consumer reviews, social media has also become a significant source of data for text-mining studies. With corporations like Facebook being some of the most profitable and powerful in the world, and their entire business model centered on the willingness of users to share information with one another in a public forum, there is plenty of data to be analyzed. That data is highly valuable to industry. Several recent text mining studies, published in LIS journals, have focused on optimizing the extraction and analysis of social media data to create novel consumer insights, such as Balbi, Misuraca, and Scepi (2018), Mongeon (2018), and Celik and Deliz (2018).

Among these text mining studies, only a few focus on specific aspects of libraries, as opposed to the more general "information" context. Al-Daihani and Abrahams (2016) used text mining approaches, including word frequency analysis (WFA)—with one, two, and three word phrases—and semantic and sentiment analysis, to analyze the content of academic libraries tweets. This type of analysis provides insight into the topics of significant interest to academic libraries during the period of time studied. Durr (2020) analyzed relationships between data-science job postings and the data science curriculum offered at iSchools (a consortium of schools dedicated to the study of information, associated with LIS, computer science, information systems, and other information-related disciplines). Using latent semantic analysis, Durr (2020) notes the different emphasis that iSchool syllabi and data science job posting place on certain topics. For instance, "team work" is emphasized in a greater percentage of job posting documents and syllabi, while "customer service" is emphasized more in the syllabi.

Data mining strategies are also commonly used in scientometrics. In these studies, the specific strategy used is typically network analysis, which illustrates the type and strength of relationships among a collection of data. Co-authorship and co-word analysis are two of the many network analysis approaches. These approaches identify relationships among authors and terms, respectively, used in publications. Many recent studies in LIS journals have used these approaches, such as Qiu, Dong, and Yu (2014), Fang (2015), and Franssen and Wouters (2019). Tseng, Wang, Lin, Lin, and Juang (2007) used similar methods to map the content of patent documents. Many studies have utilized data mining strategies to analyze a corpus of topics or the ontology of a specific discipline based on qualities of bibliographic data. This is the case in Joo, Choi, and Choi's (2018) study of the research domain of knowledge organization and Lee, Kim, and Kim's (2010) study of digital library research trends.

Regression analysis, factor analysis, and structural equation modeling are common methods used in quantitative studies employing data mining strategies. All of these methods reveal underlying relationships in a set of data that are more-statistically sophisticated and insightful descriptive statistics or basic correlation or tests of variance. These methods can be seen in recent publications like Islam, Ahmad, Rafi, and Zheng's (2020) study of the use of big data analytics in academic libraries and Carlozzi's (2018) study that profiles the socioeconomic standing of public libraries' service population (such as education level of residents) and funding for the library.

An emerging area of considerable research interest in LIS is learning analytics, the analysis of data about learners and learning used to inform educational management and technology. Kyle Jones, a professor at Indiana University-Purdue University Indianapolis's School of Informatics and Computing, has published significantly in this area, including both what learning analytics are and the risks associated with the use of these analytics (Jones, 2017; Jones &

Salo, 2018; Jones et al., 2020). Because learning analytics is such a broad and new area of study, a multitude of methods can be used depending on the types of data used and how they are retrieved and analyzed.

2.2. Cluster Analysis in LIS Research

Clustering methods have been utilized in LIS research for several decades (Wu, Fuller, & Wilkinson, 2001). One early case was in the work of Chen and Chen (2006), where the authors proposed a novel clustering algorithm for classifying types of library readers. Using this algorithm, the authors identified five distinct clusters of readers based on library use statistics. They proposed that this algorithm/findings could be used to inform how materials are cataloged by libraries, to make library organization more relevant for users. Karunagnayake and Nagata (2014) conducted a related study that classified undergraduate library users into four clusters based on the variables of search capability, positive image of the library, familiarity with digital resources, usage of professional library assistance, browsing of indexes, and readiness in searching: Ineffective Library Users, Effective Library Users, Ineffective but Positive Users, and Self-Sufficient Users. Ineffective library users rated poorly on all six measures. Effective library users rated highly on all measures except "readiness in searching." Ineffective but positive users rated high in usage of professional library assistance and browsing of indexes, but low in the other four measures. Self-sufficient users, conversely, rated highly on these four measures but low in usage of professional library assistance and browsing of indexes.

Chen (2012) utilized k-means clustering to identify themes in online health discussion groups. Seven themes/categories of discussion were found by the researcher: generic, support, patient-centered, experiential knowledge, treatment/procedures, medications, and condition management. These themes paint a more specific view of the information needs of patients and the types of support sought. Kim (2015) used a similar method to analyze data from the 2009 Annenberg National Health Communication Survey to differentiate "active" and "inactive" seekers of health information and possible determinants of inactive seeking behaviors. Some of the variables that were found to be particularly indicative of inactive seekers were age (younger individuals were more likely to be inactive seekers), sex (male), education (highly-educated), ethnicity (white) and income level (high). Individuals who were exposed to more varied media were more likely to fall into the active seekers group.

Claudio-González, Martin-Baranera, and Villarroya (2016) utilized a cluster-analytic approach to classify the business models used by academic journals. Four clusters were identified: one that has a larger number of subscribers and low amount of external financial support—open access journals that sell advertisements; one that has a large number of subscribers and high amount of external support—journals that rely on institutional funding; one with mixed funding—open access journals that rely on grants/donations in addition to advertisements or institutional funding; and one that relies heavily on commercial transactions to remain financially viable—journal with a traditional publishing model. Koizumi and Widdersheim (2019) utilized a hierarchical cluster analysis method to classify models of hiring and management among academic libraries. Seven distinct models/clusters were identified: strategy of higher than average employment of librarians in all specialties; strategy to employ higher than average numbers of electronic resources librarians; strategy to emphasize employment of research and instructional librarians; strategy to emphasize employment of systems librarians; strategy to emphasize employment of instructional librarians only; strategy to emphasize the employment of metadata librarians; and strategy to strengthen general library systems. The authors of both articles concluded that clustering may be helpful in informing administrative decisions among the entities studied.

Common in bibliometrics research are the methods of keyword and authorship cluster analysis (Dutta, Majumber, & Sen, 2011; Erfanmanesh, Abdollah, & Asnafi, 2014; Qilan & Willet, 2011; Qiu, et al., 2014; Wildegard, 2016). These methods use the inclusion of keywords or authors in scholarly publications to group them together. This method can be used to draw together research in otherwise disparate areas that share certain themes, like social media research. For instance, Gan and Wang (2015) utilized cluster analysis to examine relationships between productivity in research and status on social media. Šubelj, Van Eck, and Waltman (2016) clustered types of scientific publications based on how they were cited, while Tseng and Tsay (2013) classified subfields within LIS based on characteristics of journals.

2.3. Research Problem and Question

Little is known about the evolution of data mining research in library and information science studies. Understanding more about how these research approaches are used in LIS—the subject matter, trends over time, and top journals for publishing this research—may provide valuable orientation for future data mining-based research within the discipline. This study poses one research question: What are the trends in number, subject matter, and top journals for the publication of data mining studies in LIS journals during the years 2006, 2012, and 2018?

3. Methods

From an initial data set of 3422 scholarly articles published in a set of 31 core LIS journals (based on the LIS journal ranking by the 2019 Journal Citation Report of Social Science Citation Index) across 2006, 2012, and 2018, those articles were identified that utilized a text or data mining technique as part of the data collection and/or data analysis (n=301). In 2006, 68 articles were identified, with 121 in 2012, and 112 in 2018. Full abstracts for these articles were imported in an Excel datasheet before being transferred to RapidMiner—a data science platform used for data analytics and machine learning—for a K-Medoid cluster analysis, which was utilized to model clusters among the abstracts based on unique term frequencies in the abstracts. The number of clusters /was set at three, based on the uniqueness and clear divisions between three clusters as opposed to a greater number. Following the identification of the three clusters, the researchers sought to delineate trends in data mining publications across 2006, 2012, and 2018. These trends include the shift in proportion of articles in each cluster for each year, the proportion of articles from each journal that fall in each cluster, and the specific data mining methods employed.

4. Results

Figure 1 displays the three clusters identified from the K-Medoid cluster analysis along with the ten most-unique terms among each clusters. These terms were used to create working names for the clusters: Cluster 1 was named "Publications/Citations" based on the presence of terms associated with ranking and categorization, journals, reviews, science, and citations; Cluster 2 was named "Consumer Behavior" based on the presence of terms associated with consumers, decision, behavior, and awareness; Cluster 3 was named "Information/Media Use" based on the presence of terms associated with people, participation, seeking, motivation, and media.

Figure 1

Unique Terms by Cluster

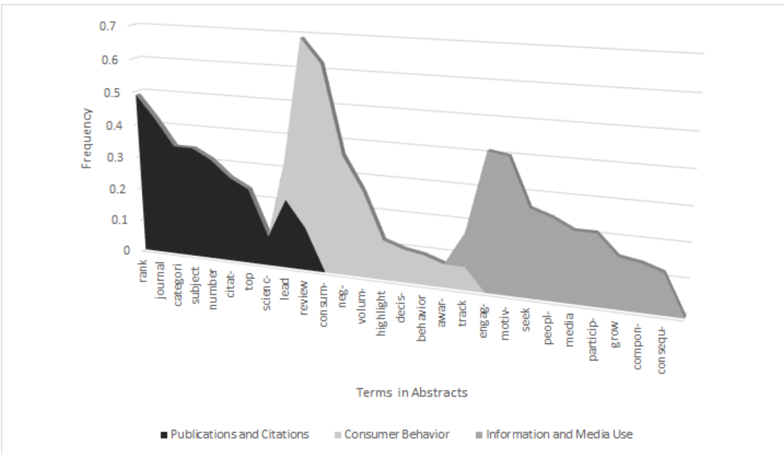


Figure 2 displays the number of articles for each year that were organized into the three clusters. Most of the clusters remain fairly stable over the years, with the exception of Publications/Citations and Consumer Behavior between 2012 and 2018. The proportion of articles on the topic of publishing dropped by ten percent while the proportion on the topic of Consumer Behavior increased twelve percent. This suggests a recent, and fairly pronounced, shift in data mining-based research towards topics associated with consumers rather than Publications/Citations/Scientometrics. The majority of data mining articles remain associated with the topic of Information/Media Use, which evolved to include areas like social media use research over the course of the three years studied.

Figure 2

Trend in Number of Articles/Abstracts Falling in Each Cluster by Year

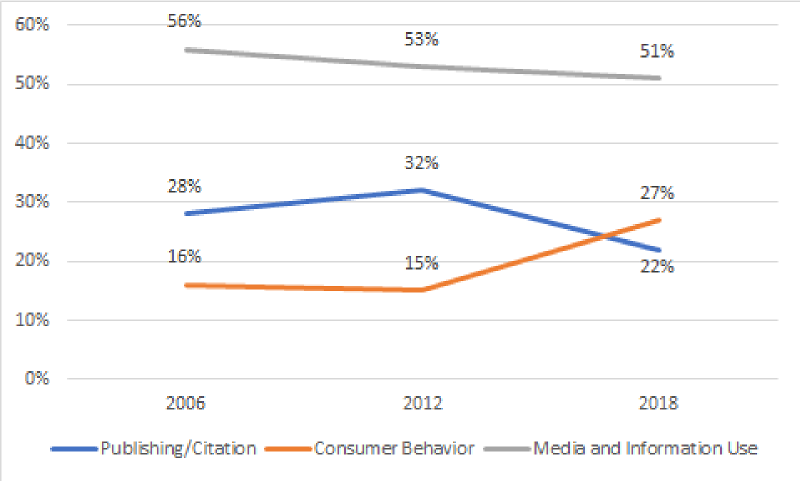


Table 1 displays the frequency of data mining articles published by each of the 31 journals (based on the LIS journal ranking by the 2019 Journal Citation Report of Social Science Citation Index) across 2006, 2012, and 2018, as well as the percentages of articles for each journal that fell into the three clusters. *Journal of the Association for Information Science and Technology* and *Information Processing and Management* were the two journals with the most data mining articles. *Scientometrics* had the largest number of articles in the cluster of Publications/Citations (n=13), *Information Processing and Management* has the most in Consumer Behavior (n=8), and *Journal of the Association for Information Science and Technology* has the most in information/media use (n=22).

Table 1

Data Mining Articles by Journal and Cluster Frequency

Journal	Frequency	Cluster 1 Publications/Citations (%)	Cluster 2 Consumer Behavior (%)	Cluster 3 Information /Media Use (%)
ACM Transactions on Information Systems	3	0	0	100
College & Research Libraries	10	50	0	50
Information & Management	18	6	33	61
Information Processing &	32	34	25	41

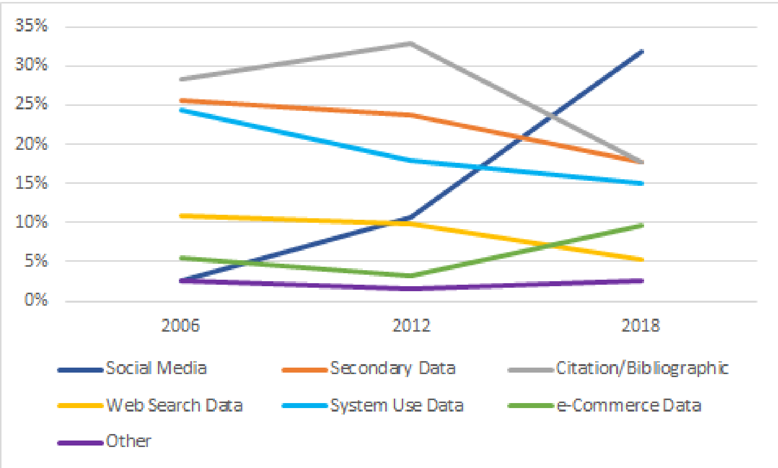
Management				
Information Research	6	17	0	83
Information Retrieval Journal	9	22	33	45
Information Systems	18	11	11	78
Information Systems Research	12	25	33	42
Information Technology & Libraries	5	60	40	0
International Journal of Information Management	18	17	39	44
International Journal on Digital Libraries	6	0	0	100
Journal of Documentation	10	20	20	60
Journal of Education for Library & Information Science	2	0	0	100
Journal of Information Science	13	46	38	16
Journal of Librarianship & Information Science	1	0	0	100
Journal of Management Information Systems	10	40	10	50
Journal of the Association for Information Science & Technology	42	31	17	52
Library & Information Science Research	7	14	0	86
Library Collections, Acquisitions, and Technical Services	2	0	0	100
Library Quarterly	2	0	100	0
Library Resources & Technical Services	6	17	0	83
Library Trends	4	50	25	25
Libri: International Journal of Libraries & Information Services	5	20	20	60
MIS Quarterly	12	8	32	60
Online Information Review	15	27	20	53
Reference & User Services Quarterly	2	50	0	50
Research Evaluation	1	0	100	0
Scientometrics	20	65	0	35
Social Science Computer Review	10	20	0	80

This cluster analysis indicates compelling relationships and trends among LIS articles that employ data mining methods. Clearly, these methods are employed more frequently in journals associated with "Information Science" (e.g., *Information Processing and Management, Journal of the Association for Information Science and Technology, Scientometrics*) than "Library Science" (e.g., *Library Quarterly, College and Research Libraries, Library Trends*). The topicality of these articles may also vary based on publications, as *Scientometrics* journals had the most articles in the Publications/Citations cluster and information systems journals had the most articles in the Consumer Behavior cluster. Changes over the past decade-and-a-half indicate a shift in LIS data mining articles towards a greater focus on Consumer Behavior, away from Publications/Citations behavior and trends.

The sources of data for these data mining studies have shifted significantly across the twelve years, as shown in Figure 3. Social media was used as a source of data for only about 2% of articles in 2006, compared to 33% of articles in 2018. Twitter alone was the data source for 18% of articles in 2018. Bibliographic data has experienced a decline, which aligns with the decrease in the size of the scientometrics cluster for the topic of data mining studies. Secondary and system use data have also experienced modest declines in share of study percentage. These trends illustrate the growing importance of social media as a data source in LIS research.

Figure 3

Data Sources for Data Mining Studies from 2006 to 2018



5. Discussion

Over the three data points from 2006 to 2018, there is a notable shift in data mining studies from a focus on scientometrics to one on consumer behavior. Consumer behavior is specifically focused on how consumers (including users of library and information organizations) look for information and interact with information systems. However, information/media use studies, focused on how diverse populations of people use various sources of information, remains the most common category across all three data points. This may suggest that data mining methods have further diffused into LIS research over time, transitioning from the most quantitative-heavy areas of the discipline (scientometrics) to a broader array of areas.

Journal-level findings indicate that more information science (IS)-oriented publications employ data mining methods than library science (LS) publications. The three journals with the most data mining articles—JASIST, IP&M, and Scientometrics—are generally considered "IS" as opposed to "LS" publications (Huang & Chang, 2012). In terms of a relatively equal balance of articles across the three identified research areas, IP&M offers a great diversity of data mining articles with a 34%–25%–41% split across citations/publications, consumer behavior, and information and media use studies. This may make it an ideal outlet for a variety of data mining-based studies.

In terms of library-focused publications, College and Research Libraries is an anomaly as to the quantity of data mining articles. These articles largely focused on learning analytics applied to academic library operations. Learning analytics is the collection and analysis of data about learners and learning (from sources including learning management systems) and, at the university level, this learner data has implications for academic library services. Several articles in journals of Information Technology and Libraries (ITAL) and Library and Information Science Research (LISR) had a similar emphasis on learning analytics research.

There are several opportunities for further study based on the findings of this study. The rapid change in the sources of data—with emphasis on social media––encourages continued study of data sources. With data collected annually from 2010 to 2020, for instance, researchers may be able to illustrate the growth of specific social media sites like Facebook and Twitter as data sources. Acknowledging that not all LIS journals were selected for this study (as noted in the methodology section, selection was based on journals included in the Social Science Citation Index as well as Jarvelin studies that refer to the two studies conducted by Järvelin and Vakkari [1990; 1993] of the structure of the LIS discipline), it is possible that a researcher could examine a broader group of journals, perhaps including all journals in the Emerging Sources Citation Index as well. This may present more insight across a large and diverse group of publications. Researchers may also look at broader trends over ten- or twenty-year intervals, as with the Jarvelin studies. While this may not provide insight into the more rapid shifts in the field, it would cover a larger period of time than only twelve years from 2006–2018.

6. Conclusion

Revealed in this study are the significant contributions to data mining research, the top journals and themes, among LIS journals. Information science-focused journals like JASIST and IP&M appear most likely to employ this analytical strategy, though the topic of learning analytics is popular among several of the top library science journals. While data mining is not as common as methods like experimentation, questionnaires, and interviews within LIS research, it is certainly a prominent method, which has experienced growth in its use in recent years. This study provides orientation to the evolving nature of data mining studies within the field of LIS research.

References

Al-Daihani, S. M., & Abrahams, A. (2016). A text mining analysis of academic libraries' tweets. Journal of Academic Librarianship, 42(2), 135–143. https://doi.org/10.1016/j.acalib.2015.12.014

Altinel, B., & Ganiz, M. C. (2018). Semantic text classification: A survey of past and recent advances. Information Processing and Management, 54, 1129–1153.

Balbi, S., Misuraca, M., & Scepi, G. (2018). Combining different evaluation systems on social media for measuring user satisfaction. Information Processing and Management, 54, 674–685.

Carlozzi, M. (2018). The socioeconomic profile of well-funded public libraries: A regression analysis. Evidence Based Library and Information Practice, 13(2), 13–26.

Celik, M., & Dokuz, A. S. (2018). Discovering socially similar users in social media datasets based on their socially important locations. Information Processing and Management, 54, 1154–1168.

Chen, A. P., & Chen, C. C. (2006). A new efficient approach for data clustering in electronic library using ant colony clustering algorithm. Electronic Library, 24(4), 548–559. https://doi.org/10.1108/02640470610689223

Chen, A. T. (2012). Exploring online support spaces: Using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient Education & Counseling, 87(2), 250–257. https://doi.org/10.1016/j.pec.2011.08.017

Claudio-González, M. G., Martín-Baranera, M., & Villarroya, A. (2016). A cluster analysis of the business models of Spanish journals. Learned Publishing, 29(4), 239–248. https://doi.org/10.1002/leap.1046

Durr, A. K. (2020). A Text Analysis of Data-Science Career Opportunities and US iSchool Curriculum. Journal of Education for Library & Information Science, 61(2), 270–300. https://doi.org/10.3138/jelis.2018-0067

Dutta, B., Majumder, K., & Sen, B. K. (2011). Study of subject domain by keyword cluster analysis based on research articles: A case study from physics. Information Studies, 17(4), 195–210.

Erfanmanesh, M., Abdollah, A., & Asnafi, A. (2014). A Scientometric and Social Network Analysis. (English). Journal of Information Processing & Management, 29(2), xv–xvi.

Fang, Y. (2015). Visualizing the structure and the evolving of digital medicine: A scientometrics review. Scientometrics, 105(1), 5–21. https://doi.org/10.1007/s11192-015-1696-1

Feng, G., Guo, J., Jing, B., & Hao, L. (2012). A Bayesian feature selection paradigm for text classification. Information Processing and Management, 48, 283–302.

Franssen, T., & Wouters, P. (2019). Science and its significant other: Representing the humanities in bibliometric scholarship. Journal of the Association for Information Science & Technology, 70(10), 1124–1137. https://doi.org/10.1002/asi.24206

Gan, C., & Wang, W. (2015). Research characteristics and status on social media in China: A bibliometric and co-word analysis. Scientometrics, 105(2), 1167–1182.

Guen, K. S., & Juyoung, K. (2018). Analyzing the discriminative attributes of products using text mining focused on cosmetic review. Information Processing and Management, 54, 938–957.

Huang, M. H., & Chang, Y. W. (2012). A comparative study of interdisciplinary changes between information science and library science. Scientometrics, 91(3), 789–803.

Islam, A. Y., Ahmad, K., Rafi, M. & Zheng. J. (2020). Performance-based evaluation of academic libraries in the big data era. Journal of Information Science. https://doi.org/10.1177/0165551520918516

Jarvelin, K., & Vakkari, P. (1990). Content analysis of research articles in library and information science. Library and Information Science Research, 12, 395–421.

Jarvelin, K., & Vakkari, P. (1993). The evolution of library and information science 1965–1985: A content analysis of journal articles. Information Processing & Management, 29(1), 129–144. https://doi.org/10.1016/0306-4573(93)90028-C

Jones, K. M. L. (2017). Learning analytics, the academic library, and positive intellectual freedom. Journal of Intellectual Freedom & Privacy, 2(2), 7–10. https://doi.org/10.5860/jifp.v2i2.6305

Jones, K. M. L., & Salo, D. (2018). Learning analytics and the academic library: Professional ethics commitments at a crossroads. College & Research Libraries, 79(3), 304–323. https://doi.org/10.5860/crl.79.3.304

Jones, K. M. L., Briney, K. A., Goben, A., Salo, D., Asher, A., & Perry, M. R. (2020). A comprehensive primer to library learning analytics practices, initiatives, and privacy issues. College & Research Libraries, 81(3), 570–591. https://doi.org/10.5860/crl.81.3.570

Joo, S., Choi, I., & Choi, N. (2018). Topic analysis of the research domain in knowledge organization: A latent Dirichlet allocation approach. Knowledge Organization, 45(2), 170–183. https://doi.org/10.5771/0943-7444-2018-2-170

Kim, E., & Chun, S. (2019). Analyzing online car reviews using text mining. Sustainability, 11, 1611.

Kim, J., & Diesner, J. (2015). The effect of data pre-processing on understanding the evolution of collaboration networks. Journal of Informetrics, 9(1), 226–236. https://doi.org/10.1016/j.joi.2015.01.002

Kim, S. (2015). An exploratory study of inactive health information seekers. International Journal of Medical Informatics, 84(2), 119–133. https://doi.org/10.1016/j.ijmedinf.2014.10.003

Koizumi, M., & Widdersheim, M. (2019). Specialties and strategies in academic libraries: A cluster analysis approach. Library Management, 40(1), 45–58. https://doi.org/10.1108/LM-10-2017-0114

Kaur, N., Sood, S. K., & Verma, P. (2020). Cloud resource management using 3Vs of Internet of Big data streams. Computing, 102(6), 1463–1485. https://doi.org/10.1007/s00607-019-00732-5

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(2), 144–161.

Marchionini, G. (2017). Information science roles in the emerging field of data science. Journal of Data and Information Science, 1(2), 1–6. https://doi.org/10.20309/jdis.201609

Mongeon, P. (2018). Using social and topical distance to analyze information sharing on social media. Proceedings of the Annual Meeting of the Association for Information Science and Technology, 81, 397–403.

Puarungroj, W., Pongpatrakant, P., Boonsirisumpun, N., & Phromkhot, S. (2018). Investigating factors affecting library visits by university students using data mining. Libres, 28(1), 25–33.

Qiu, J., Dong, K., & Yu, H. (2014). Comparative study on structure and correlation among author co-occurrence networks in bibliometrics. Scientometrics, 101(2), 1345–1360. https://doi.org/10.1007/s11192-014-1315-6

Rehman, A., Javed, K., & Babri, H. A. (2017). Feature selection based on a normalized difference measure for text classification. Information Processing and Management, 53, 473–489.

Schuh, G., Reinhart, G., Prote, J-P., Sauermann, F., Horsthofer, J., Oppolzer, F., & Knoll, D. (2019). Data mining definitions and applications for the management of production complexity. Procedia CIRP, 81, 874–879.

Song, M., Park, H., & Shin, K. (2019). Attention-based long short-term memory using sentiment lexicon embedding for aspect-level sentiment analysis in Korean. Information Processing and Management, 56, 637–653.

Šubelj, L., Van Eck, N. J., & Waltman, L. (2016). Clustering scientific publications based on citation relations: A systematic comparison of different methods. PLoS ONE, 11(4), 1–23. https://doi.org/10.1371/journal.pone.0154404

Tseng, Y., Wang, Y., Lin, Y., Lin, C., & Juang, D. (2007). Patent surrogate extraction and evaluation in the context of patent mapping. Journal of Information Science, 33(6), 718–736. https://doi.org/10.1177/0165551507077406

Tseng, Y., & Tsay, M. (2013). Journal clustering of library and information science for subfield delineation using the bibliometric analysis toolkit: CATAR. Scientometrics, 95(2), 503–528. https://doi.org/10.1007/s11192-013-0964-1

Tubishat, M., Idris, N., & Abushariah, M. A. (2018). Implicit aspect extraction in sentiment analysis: Review, taxonomy, opportunities, and open challenges. Information Processing and Management, 54, 545–563.

Tuomaala, O., Jarvelin, K., & Vakkari, P. (2014). Evolution of library and information science, 1965–2005: Content analysis of journal articles. Journal of the Association for Information Science and Technology, 65, 1446–1462. https://doi.org/10.1002/asi.23034

Wildegard, L. (2016). A critical cluster analysis of 44 indicators of author-level performance. Journal of Informetrics, 10(4), 1055–1078. https://doi.org/10.1016/j.joi.2016.09.003

Wu, M., Fuller, M., & Wilkinson, R. (2001). Using clustering and classification approaches in interactive retrieval. Information Processing and Management, 37(3), 459–484.

Yang, C. C., Zhang, H., Jiang, B., & Li, K. (2019). Aspect-based sentiment analysis with alternating coattention networks. Information Processing and Management, 56, 463–478.

Zhou, Y., Yang, S., Li, Y., Chen, Y., Yao, J., & Qazi, A. (2020). Does the review deserve more helpfulness when its title resembles the content? Locating help reviews by text mining. Information Processing and Management, 57, 102179, 1–11.

Appendix A

List of Journals Examined in 2006, 2012, and 2018

Journal Name:

ACM Transactions on Information Systems
Canadian Journal of Information and Library Sciences

College and Research Libraries Information Management Information Processing and Management Information Research Information Retrieval Journal Information Systems Management Information Systems Research Information Technology and Libraries International Journal of Information Management International Journal on Digital Libraries Journal of Documentation Journal of Education for Library and Information Science Journal of Information Science Journal of Librarianship and Information Science Journal of Management Information Systems Journal of the Association for Information Science and Technology Library and Information Science Research Library Collections, Acquisitions and Technical Services Library Quarterly Library Resources and Technical Services Library Trends Libri MIS Quarterly Online Information Review Program Reference and User Services Quarterly Research Evaluation Scientometrics Social Science Computer Review