Evolution of research topics in LIS between 1996 and 2019: an analysis based on latent Dirichlet allocation topic model

Xiaoyao Han[1]

Abstract

This study investigated the evolution of library and information science (LIS) by analyzing research topics in LIS journal articles. The analysis is divided into five periods covering the years 1996–2019. Latent Dirichlet allocation modeling was used to identify underlying topics based on 14,035 documents. An improved data-selection method was devised in order to generate a dynamic journal list that included influential journals for each period. Results indicate that (a) library science has become less prevalent over time, as there are no top topic clusters relevant to library issues since the period 2000–2005; (b) bibliometrics, especially citation analysis, is highly stable across periods, as reflected by the stable subclusters and consistent keywords; and (c) information retrieval has consistently been the dominant domain with interests gradually shifting to model-based text processing. Information seeking and behavior is also a stable field that tends to be dispersed among various topics rather than presented as its own subject. Information systems and organizational activities have been continuously discussed and have developed a closer relationship with e-commerce. Topics that occurred only once have undergone a change of technological context from the networks and Internet to social media and mobile applications.

Keywords Library and information science · LDA · Research trends

Introduction

A notable trend has been observed in library and information science (LIS): the inclusion of "information" as part of the discipline's name. After the foundation of the first library school in 1887, the School of Library Science at the University of Pittsburgh added "information" to its name in 1964, becoming the School of Library and Information Science. By the 1990s, almost all former library schools had followed the University of Pittsburgh's example (Hjørland 2018). In the twenty-first century, the diminishing use of "library" and related terms was found in LIS dissertations (Sugimoto et al. 2010) and a decreased interest in library management was found in LIS publications (Figuerola et al. 2017); the tendency today is to use the label "Information Science" alone (Olson and Grudin 2009).

✉ Xiaoyao Han
  hanxiaoy@hu-berlin.de

[1] School of Library and Information Science, Humboldt Universität zu Berlin, Berlin, Germany

These shifts in nomenclature indicate that the research areas of LIS have changed substantially over time, and the research focus is shifting to informational issues.

To understand the development of the discipline and how its research topics have changed over time, many researchers have explored the changes of research topics based on literature in the field. Bibliometric methods are prevalent approaches in evaluation studies (Zhao and Strotmann 2008, 2014; White and McCain 1998; Chang et al. 2015). Content analysis was also a widely applied approach (Järvelin and Vakkari 1993; Koufogiannakis et al. 2004; Blessinger and Frasier 2007). Today, an increased interest has shown in using model-based approaches to explore the intellectual structure of a domain (Sugimoto et al. 2010; Liu et al. 2015; Figuerola et al. 2017). The method allows researchers to examine large document collections.

Many studies have consulted journal rankings in a single year to compile a journal list for a diachronic analysis. However, such data corpora might have limitations. First, highly cited journals identified in a single year may limit the topic spectrum for the period of study. Nearly all journals are biased towards a certain research area to some extent, which can often be inferred from the name of the journal. The preference of selected journals may heavily influence the results. Second, journals that gained researchers' attention decades ago may no longer be the center of focus due to the rapidly changing environment. Emerging topics may not be fully represented in the analyses. Neglecting these issues may lead to research results that are not representative enough to capture topic changes in the domain.

In response to such limitations, this study applied an improved method for journal selection by consulting all available journal citation reports to generate a dynamic journal list. The analysis was divided into five periods covering the years 1996–2019. Correspondingly, five datasets that include the most influential journals for each period were included in the evolution analysis. Furthermore, to address the research gap that diachronic analyses are rarely conducted based on journal articles using model-based approaches, this study utilized the latent Dirichlet allocation (LDA) topic model to uncover the underlying topics in the text corpora. Many studies have confirmed that LDA can effectively cluster meaningful and interpretable topics from a large number of documents (Blei and Lafferty 2007; Yau et al. 2014; Suominen and Toivanen 2015). The central research questions are as followings. What research topics have been addressed LIS between 1996 and 2019? How the research topics have changed over time? By combining an innovative journal selection method and LDA topic model, this study aims to contribute a new perspective on observing and understanding the research development in LIS.

Literature review

Studies of topic changes and intellectual structure in LIS can be divided into three groups according to the methods used: content analysis, bibliometric methods, and model-based approaches. This introduction organizes the relevant literature based on the method category. A structured review and comparison of the literature with the essential properties of these studies, is provided in the "Appendix 1".

Content analysis

In content analysis, researchers identify the composition of research content and sort articles into classification schemes by analyzing a data corpus with a certain number of

articles. Studies using content analysis to detect research development in LIS have either adopted the classification schemes of other researchers or have devised new schemes (Chang et al. 2015). This method is relatively dated and primarily appears in publications from the 1970s and 1980s (Tuomaala et al. 2014).

Blessinger and Frasier (2007) analyzed 10 influential journals using a combination of content analysis and citation analysis for the period 1994–2004. The study found that in 1994–2004, librarians were still mainly writing about the profession's practical issues; In addition, they found that new technologies in information science, most notably the Internet, had a tremendous impact on almost every aspect of our profession during this decade. Tuomaala et al. (2014) conducted a content analysis of LIS evolution. They examined a total of 42 journals from the years 1965, 1985, and 2005. They identified the four most prominent research areas in LIS: information storage and retrieval, scientific communication, library and information-service activities, and information seeking. They further concluded that information retrieval was the most popular area of research over 1965–2005. The most significant changes in the investigated period were the decreasing interest in library and information-service activities and the growth of research about information seeking and scientific communication.

Bibliometric methods

Bibliometrics comprises various techniques, including keywords analysis, direct citation analysis, co-citation analysis, and bibliographic coupling analysis. Studies of the intellectual structure and development of LIS frequently use these techniques.

Keywords analysis

Onyancha (2018) investigated the evolution of LIS by tracking author-supplied keywords in research articles published between 1971 and 2015. The author found that LIS evolved from information systems design and management in the 1970s to encompass scientific communication, information storage and retrieval, information access, information and knowledge management, and user education in 2015.

Citation analysis

Larivière et al. (2012) presented an encapsulated history of LIS by examining approximately 96,000 papers in 61 journals over the field's first hundred years (1900–2010). Their analysis of lexicon frequency and bibliometric indicators revealed two major structural shifts: in 1960, LIS changed from a professional field focused on librarianship to an academic field focused on information and use, and in 1990, LIS began to receive more citations from outside the field. The study of Åström (2007) examined the most-cited articles from 21 LIS journals to identify the changes in research fronts from 1990 to 2004. The study showed that the main fields in LIS are information seeking and retrieval (ISR) and informetrics. The author also found that changes in the discipline can be seen primarily within these two fields rather than in new fields entering the discipline. The study showed that the IR field had become ISR and that webometrics had grown considerably, to the extent that it has come to dominate LIS research over 2000–2004. Chang et al. (2015) analyzed keywords, bibliographic coupling, and co-citation to track changes in LIS research subjects during four periods between 1995 and 2014. By examining 580 highly cited

LIS articles, they found that the two subjects "information seeking (IS) and information retrieval (IR)" and "bibliometrics" appeared in all four periods. However, they observed that the percentage of articles in which the topics appeared was decreasing for IS and increasing for bibliometrics.

Model-based approaches

Model-based approaches have frequently been employed to detect the intellectual structure of a scientific domain based on the aggregated literature. This approach enables researchers to examine a larger corpus of text data than content analysis and bibliometric methods.

Liu et al. (2015) investigated the intellectual structure of library and information science using the formal concept analysis (FCA) method. By analyzing the papers published in 16 prominent journals in the LIS domain from 2001 to 2013, the authors identified nine main LIS research themes: bibliometrics, scientometrics, and informetrics; citation analysis; information retrieval; information behavior; libraries; user studies; social network analysis; information behavior; and webometrics. Sugimoto et al. (2010) examined 3121 doctoral dissertations using an LDA model to explore the development of LIS from 1930 to 2009. They found that LIS topics have changed substantially over time. Nonetheless, some themes occurred in multiple periods, representing core areas of the field: library history, citation analysis, information-seeking behavior, information retrieval, and information use. The authors noted the diminishing use of the word "library" and related terms. Another study using LDA was conducted by Figuerola et al. (2017), who analyzed title and abstract of academic production to investigate significant trends and subdomains in LIS. They examined in total of 92,705 documents for the period 1978–2014 in the database LISA (Library and Information Science Abstracts). In the results, they identified 19 dominant topics, which were further clustered into four main areas: process, information technology, library and specific areas of information application. Furthermore, they observed a notable growth in the specialized documentation for specific areas of activity (business, health, low, education, media, heritage), a decrease in the relative importance of libraries, and a constant but changing interest in information technology.

The variation of research methods and data-selection criteria makes it difficult to compare results. However, these studies exhibit the different characteristics of LIS research subjects, enabling researchers to observe the research trends in LIS from different perspectives.

Methodology

In this study, the author devised an improved method for journal selection by consulting all available JCRs for LIS. A dynamic journal list and five datasets were generated for further analysis. The author used LDA topic modeling to detect the underlying topics in the text corpus.

Journal selection and data collection

The author developed an improved data-selection process by consulting all available journal rankings in LIS from 1997 to 2018. JCR is an annual publication that provides the impact factors and rankings of journals based on citations. The journals with the highest

impact factors are often considered core journals in the field and attract the attention of researchers. As a result, journals with high impact factors are often used to study disciplinary development and evolution. Since 1997, JCR has incorporated LIS as a discipline. Until the 2018 report, there are 22 yearly reports for LIS. Based on these reports, the most influential journals for each period between 1996 and 2019 (1996–2000, 2001–2005, 2005–2010, 2011–2015, and 2016–2019) were identified. For each of the five periods, the occurrence of each of the top 20 journals was recorded. In this way, the nine journals with the highest occurrence were selected (Table 1). Articles were sourced from Scopus based on the journal list. For the topic analysis, the title, abstract, and keywords of each document was used. In total 14,053 articles published from 1996 to 2019 were collected (Table 2). The data sets were further processed and analyzed by LDA topic modeling.

LDA

As a result of the information explosion, new algorithmic tools are needed to understand, organize, and search information from large informational corpora. Topic modeling was designed to uncover hidden topical patterns in vast corpora. LDA is a probabilistic model proposed by Blei et al. (2003). The model analyzes the thematic structure of a corpus and can also perform topic clustering or text classification based on the topic distribution. Previous studies (Blei and Lafferty 2007; Yau et al. 2014; Suominen and Toivanen 2015) found that LDA performed well for understanding the rich underlying topical structure of a field.

LDA treats data as from a generative probabilistic process. This approach assumes that a document is composed of a group of words and that there is no sequential relationship between them. Therefore, it represents typical bag-of-words modeling. The intuition behind LDA is that documents include multiple topics, and so there is a probability of topic distribution for each topic. Each topic is depicted as a distribution over terms in a fixed vocabulary, with different topics represented by different probabilities of words within the vocabulary. The LDA topic model can be visualized by a graphical model (Fig. 1). The boxes are plates representing replications. The figure can be explained as follows: There are K topics in the collection. Each topic features a multinomial distribution over the vocabulary and is assumed to have been drawn from a Dirichlet (η). The generative process is performed for each document d as follows: First, select a distribution over topics θ_d from Dirichlet(α). Then, for each word n in the document, draw a topic index $z_{d,n}$ from the topic proportions θ_d. Finally, draw the observed word $w_{d,n}$ from the selected topic $\beta_{z_{d,n}}$.

There are many implementation tools for LDA. This analysis applied the Gensim python library (*Gensim: LDA Model*, n.d.) to perform LDA. Before applying LDA, one must decide the number of topics for the corpus. Perplexity analysis, which estimates the performance of topic clustering based on a smaller set of data, was often used to determine the number of topics. There are also researchers who chose the number of topics based on their judgment or tests (Blei et al. 2003; Newman and Block 2006; Figuerola et al. 2017). In the present study, the author consulted the cluster number used in the studies of Sugimoto et al. (2010) and Yan (2015). (Despite using different approaches to determine the topic number, they both used 50 topics in their studies.) The number of topics for each text corpus was independently decided based on LDA tests on sample texts and ranged from 30 to 50 topics.

After gaining the results based on 14,059 articles, for each period the topics were ranked by probability values and the top ten clusters with their 10 most relevant keywords

Table 1 Journal list

1996–2000	2001–2005	2006–2010	2011–2015	2016–2019
Journal of Documentation	Journal of Documentation	Information & Management	Journal of Informetrics	Journal of Information Technology
JASIS[a]	Information Processing and Management	JASIST	Journal of Information Technology	Government Information Quarterly
Library and information science	JASIST	Information Processing and Management	Information & Management	International Journal of Information Management
College and Research Libraries	Information & Management	Scientometrics	JASIST	Journal of Informetrics
Library Quarterly	Information Society	Journal of Information Technology	Scientometrics	Information Processing and Management
Scientometrics	Journal of Information Science	Journal of Information Science	International Journal of Information Management	JASIST
Library Trends	College and Research Libraries	Journal of Informetrics	European Journal of Information Systems	Telemetrics and Informetrics
Journal of Academic Librarianship	Scientometrics	Government Information Quarterly	Government Information Quarterly	Journal of Knowledge Management
Information Management	Journal of Information Technology	Library and information Science Research	Information and Organization	Information & Management

[a]This journal has been known as the *Journal of the American Society for Information Science* (JASIS before 2000), the *Journal of the American Society for Information Science and Technology* (JASIST 2000–2014), and the *Journal of the Association for Information Science and Technology* (JASIST 2014–present)

Table 2 Number of articles

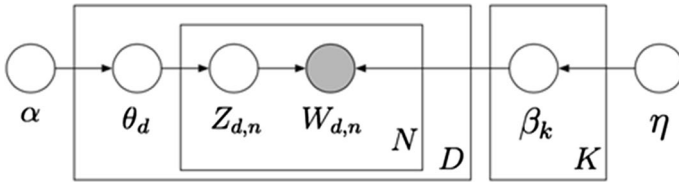|	1996–2000	2001–2005	2005–2010	2011–2015	2016–2019	Total
Number of documents	1509	2029	3223	3840	3452	14,053

Fig. 1 Latent Dirichlet allocation graphical model (Blei 2009). α: Dirichlet parameter. θ_d: The topic distribution of document d generated. β_k: Distribution over terms of topic k. $Z_{d,n}$: Topic assignment of word n in document d. $W_{d,n}$: observed word n in document d. α: delta Dirichlet distribution. η: topic hyperparameter

were selected as being the most representative topics. To facilitate the understanding and analysis of topics, the author manually examined a number of articles: For each corpus, a set of randomly selected articles was generated by searching every 20th document. In this sample texts set, the original text including title, keywords, and abstract along with the topic distribution for each document was examined in detail, enabling the identification of the most representative document and the interpretation of the top topics embedded in different research contexts. The most relevant documents and their source were displayed in "Appendix 2–6".

Results

For a clear overview, each topic was labeled as some keywords. Since the definition of LIS subdomains is still a discussion in the field, the author attempted to label the topics by objectively reflecting their keywords and representative documents rather than pre-defining the domains. This may contribute a straightforward understanding of the topic and an easier comparison with previous studies.

1996–2000

In this period, 1509 documents were included in the corpus. The sample texts set examined for further interpretation contained 76 articles. Ten representative articles and their sources are listed in the "Appendix 2". Table 3 provides an overview of the results.

Library science dominates the period's top two topics, A1 and A2. Public libraries are the main research context in A1. The cluster includes issues such as digital resources, librarians, and technologies. Topic A2 addresses issues including librarianship, collection management, and research activities with an emphasis on academic libraries. Topic A3 outline a research field related to information management, including issues regarding data processing, technologies, and information use. Citation analysis (A7) and relevant clusters were also frequently discussed in this period. Issues related to impact factors (A4) attracted the most attention in the domain. In addition, studies of patent publications (A8)

Table 3 Latent Dirichlet allocation results for the period 1996–2000

Topic	Label	Keywords
A1	Digital library/services	Library, service, digital, public, information, librarian, technology, new, need, access
A2	Academic library/librarianship	Library librarian, academic, faculty, university, article, collection, research, resource,study
A3	Information systems/information management	System, information, model, process, management, data, technology, use, development, decision
A4	Citation analysis/impact factor	Journal, citation, publication, research, impact, scientific, science, factor, number, study
A5	Scientific collaboration	Country, collaboration, research, analysis, international, science, paper, scientific, co, bibliometrics
A6	Information retrieval	System, information, user, retrieval, search, web, database, interface, document, searching
A7	Citation analysis	Research, analysis, journal, study, article, science, field, interdisciplinary, citation, social
A8	Patent analysis	citation, science, patent, paper, scientific, research, technology, indicator, theory, technological
A9	Information- seeking behavior	Information, research, model, professional, seeking, process, article, new, scientist, student
A10	Information systems/design	Information, group, research, quality, decision, study, image, design, process, focus

and scientific collaborations (A5) formed two small clusters that received much attention from researchers. The keywords in the cluster A6 identify the issue of information retrieval, including both information systems and users. Topic A9 represents the field of information seeking and behavior. A review of relevant documents indicates that topic A10 primarily includes studies about information system design and improvement.

2001–2005

The corpus for 2001–2005 included 2029 documents. The sample set included 102 articles. Ten representative texts are listed in the "Appendix 3". An overview of the results is provided in Table 4.

The keywords in B1 provide a rather broad description of the digital information environment. A review of representative documents indicates that the topic is closely relevant to information retrieval. Topic B2 suggests that various methods and algorithms were explored and evaluated based on text sources. The research demonstrated in cluster B3 is closely related to informational activities in organizations and industries, including information technology, information processing, and information systems. The analysis of sample documents relevant to cluster B4 shows that the topic demonstrates the continuity of the study of information systems from the last period (A3). In this period, library-related issues merged into one cluster (B5). The issue relevant to WWW is demonstrated in topic B6. The keywords of topic B7 reveal an interest in networks, communities, and social environments. Topic A8 includes two subjects: the implementation of new systems or models in an organization and testing the acceptance of customers. Topic B9 shows the continuity of the cluster citation analysis (Topic A4) from the last period. In cluster B10, the majority of keywords concentrate on the user studies. A review of sample texts suggests a similar research area to the 1996–2000 period (Topic A9).

2006–2010

The corpus for this period included 3223 documents. The sample set included 162 articles. Ten sample texts and their sources are listed in the "Appendix 4". An overview of the results is provided in Table 5.

Citation analysis (Topic C1) with an emphasis on journal impact factor grew considerably to the extent that it became the most significant research area in LIS over the years 2006–2010. The cluster C2 encompasses documents related to information retrieval, with a preference for documents about text-based retrieval. This keywords in topic C3 clearly identify the field of information seeking and behavior. Knowledge management (Topic C4) first emerged as a major LIS subdomain in 2006–2010. The keywords reveal an essential feature of this topic: knowledge management studies had a close relationship with commercial organizations. The topic of Governmental information management is reflected in cluster C5. A review of representative documents revealed that service, digital resources and e-government popular themes in this topic. The keywords of topic C6 reveal an interest in scientific collaboration. Linguistic analysis based on text mining was identified as an influential topic cluster (C7) in this period. After the first appearance in 1996–2000 as topic A8, patent analysis occurred again in this period (C8). In the research included in cluster C9, the interaction between users and systems was thoroughly investigated to test the acceptance of new technology. A similar cluster was also identified in the last period (Topic B8). Topic C10's most prevalent terms show that its theme is scientific performance. The

Table 4 Latent Dirichlet allocation results for the period 2001–2005

Topic	Label	Keywords
B1	Information retrieval	System, user, data, study, computer, search, design, analysis, based, software
B2	Text processing	Retrieval, document, based, method, text, algorithm, using, system, indexing, set
B3	Organizational information activities	Technology, management, business, electronic, system, decision, communication, organizational, organization, industry
B4	Information systems	System, theory, image, concentration, formula, term, concept, model, paper, intelligent
B5	Digital libraries/services	Library, digital, service, resource, academic, learning, university, student, study, electronic
B6	WWW	Web, search, engine, site, page, link, online, wide, website, study
B7	Networks/communities	Network, community, social, practice, word, ict, competitive, market, architecture, company
B8	Technology application	Model, technology, planning, acceptance, service, customer, uss, 'perceived, factor, strategic
B9	Citation analysis	Journal, citation, research, impact, science, author, china, sci, article, index
B10	Information seeking behavior	Behavior, seeking, research, individual, human, style, study, context, interaction, problem

Table 5 Latent Dirichlet allocation results for the period 2006–2010

Topic	Label	Keywords
C1	Citation analysis/impact factor	Citation, journal, impact, science, article, publication, factor, paper, cited, scientific
C2	Text processing	Retrieval, document, information, query, text, language, term, system, based, method
C3	Information retrieval/user	Information, user, system, study, seeking, behavior, use, design, retrieval, task
C4	Organizational information activities	Knowledge, management, business, system, study, process, organizational, technology, firm, innovation
C5	Government	Government, public, information, service, policy, technology, state, development, paper, study
C6	Scientific collaboration	Scientific, science, paper, research, collaboration, international, publication, scientist, country, field
C7	Semantic analysis/models and algorithms	Semantic, ontology, method, based, mining, word, concept, information, algorithm, extraction
C8	Patent analysis	Patent, analysis, research, technology, chinese, field, paper, technological, study, trend
C9	Technology application	Technology, user, use, model, adoption, acceptance, effect, factor, theory, perceived
C10	Research performance	Research, productivity, china, study, analysis, wikipedia, activity, researcher, factor, economic

sample document, which has the highest reference probability for the topic, also supports this assumption. However, the sample set does not explain the appearance of words like "Wikipedia" and "economic."

2011–2015

In this period, 3840 documents were included in the corpus. The sample set included 192 articles. Ten sample texts together with their source are listed in the "Appendix 5". An overview of the results is provided in Table 6.

   Bibliometrics analysis (Topic D1) was the largest topic in the 2011–2015 period. The keywords reveal that trend analyses of publications within a scientific field or a country were frequently conducted during this period. Including some of the same words as topic D1, Topic D2 seems to be a replicated topic. However, a detailed examination of the keywords reveals an emphasis on "evaluation" and "performance." The research performance in various fields was intensively studied using various methods. The sample set indicates that comparisons between research fields, nations, and regions were also popular during this period. Topic D3 concentrates on the techniques and methods applied for measurement in citation analysis. Topic D4 may be considered a continuation of topic C6, scientific collaboration, from last period. Citation analysis with an emphasis on journal impact factors (D5) appears again as one of the most important topics in this period. The keywords in topic D6 cover information management in multiple contexts. Governmental issues (D7) continued to receive considerable interest from researchers in this period. Topic D8 demonstrates research interest in the online-community environment. The sample texts reveal that various research questions were investigated, including user behavior, business activities, and social media. Topic D9 encompasses studies of informational activities in various types of organizations. Analyses related to ranking activities (D10) are identified as a popular topic in this period.

2016–2019

In this period, the corpus includes 3452 documents. The sample set included 173 articles. Ten sample texts and their sources are listed in the "Appendix 6". An overview of the results is provided in Table 7.

   Citation analysis (Topic E1) with an emphasis on journal impact factor continues to constitute a large volume of today's LIS publications. The continual appearance of this topic in all study periods (topics A4, B9, C1, and D5) proves that citation analysis is a steady and essential field in LIS. Social media (E2) is a fast-growing field that emerges as a significant topic during this period. Information management and processing (E3) is also a stable area appearing across multiple periods covered in this study (B3, C4, and D9). Topic E4 focuses on the applications of various algorithm- and model-based approaches on large document collections. Similar topics were also identified for previous periods (B2, C3, and C8). Knowledge sharing (E5) is a new topic emerging in this period. Studies within this topic usually investigate group communication mechanisms to improve work efficiency. A review of documents representative of topic E6 suggests a continuous interest in e-governance during this period. Topic E7 identifies informational activities within organizations. Studies under this topic investigated various strategies and systems related to information management to improve companies' performance. In topic E8, a series of studies regarding online commercial activities, especially user behavior, are included. The topic of mobile

Table 6 Latent Dirichlet allocation results for the period 2011–2015

Topic	Label	Keywords
D1	Bibliometric analysis	Research, analysis, country, science, study, bibliometric, paper, publication, field, trend
D2	Research performance	Research, field, analysis, evaluation, science, scientific, researcher, performance, method, quality
D3	Citation analysis/measurement	Citation, index, paper, cited, publication, reference, number, analysis, article, author
D4	Scientific collaboration	Collaboration, science, scientific, publication, research, international, author, paper, country, impact
D5	Citation analysis/impact factor	Journal, citation, impact, article, factor, science, year, published, number, paper
D6	Information management	Information, management, study, case, strategy, practice, alignment, issue, new, need
D7	Government	Government, public, policy, citizen, sector, decision, local, open, making, agency
D8	Online/community	Online, community, intention, trust, consumer, effect, satisfaction, commerce, perceived
D9	Organizational information activities	Organizational, management, business, firm, performance, organization, capability, resource, process, relationship
D10	Ranking research	Ranking, indicator, field, based, rank, score, method, performance, normalization, different

Table 7 Latent Dirichlet allocation results for the period 2016–2019

Topic	Label	Keywords
E1	Citation analysis/impact factor	Citation, journal, science, scientific, paper, publication, analysis, research, article, impact
E2	Social Media	Social, medium, networking, online, twitter, use, site, network, facebook, communication
E3	Organizational information activities	Management, process, organization, business, organizational, km, resource, information, system, model
E4	Text processing	Method, based, algorithm, model, approach, proposed, graph, ranking, feature, art
E5	Knowledge sharing	Knowledge, sharing, team, relationship, organizational, finding, effect, implication, purpose, design
E6	Government	Government, public, citizen, open, local, policy, sector, participation, governance, agency
E7	Organizational innovation/performance	Innovation, performance, firm, capability, business, effect, capacity, competitive', relationship, based
E8	E-commerce	Consumer, commerce, customer, online, trust, brand, influence, purchase, effect, social
E9	Mobile application	Mobile, perceived, intention, value, application, factor, apps, app, structural, model
E10	Knowledge management	Knowledge, project, management, approach, paper, transfer, design, research, implication, purpose

application is clearly demonstrated in topic E9. Similar to social media, the social app topic comprises a wide range of issues, such as user behavior, health information, and information privacy. Based on a review of sample documents, topic E10 is considered one branch of studies related to knowledge management.

Discussion

Aggregated results

Aggregated results based on all 50 topics analyzed above are provided in this section for a diachronic track of the research trends in LIS.

Field level

Table 8 provides a summary of all topics from 1996 to 2019. The topics are organized spatially from top to bottom in descending order by their probability value. The topics were grouped into three categories to gain a holistic view of the LIS domain: library science (orange), bibliometrics (green), and information science and related issues (blue). A clear decrease in library science can be observed: During the period 1996–2000, library science still dominated the field with the top two clusters; in 2001–2005, library science shrank to a single cluster with a much lower proportion of total research; and since 2006, there are no clusters representing library issues in the top 10 topics.

Bibliometrics was grouped as another category in LIS for its large number of documents and stable subclusters. In this study, the term "bibliometrics" represents bibliometrics, scientometrics, and informetrics, which share overlapping interests in the dynamics of disciplines as reflected in their literature (Hood and Wilson 2001). Although the number of topic clusters related to bibliometrics fluctuates across the periods, bibliometrics has proven to be a stable area in LIS. Some topics repeatedly occur across periods, and in rare cases, new topics emerged. Although citation analysis comprises only one cluster in the recent period, the field was nonetheless identified as the largest topic in the corpus.

On the contrary, information science and its related fields show a stable number of clusters but reflect intensive changes within the field. Table 9 provides more information about topic-level changes.

Topic level

In Table 9, topics are ranked by stability (the number of occurrences across periods) rather than probability value. Similar clusters are highlighted in the same color for a diachronic overview. The purple color at the bottom includes all topics that occurred less than twice across all periods.

Citation analysis is the most stable cluster and appeared in all periods. In this cluster, the topic of journal impact factors was intensively investigated. Information retrieval appears in all periods except 2011–2015, where bibliometrics thrived and formed more subclusters. Information retrieval has proven to be the most important area within information science, as researchers have focused more on model and algorithm-based text analysis.

The next layer of topics comprises information systems and organizational activities. These two fields have demonstrated a close relationship through time and are sometimes

Table 8 Overview of results (field level). (Color table online)

| 1996-2000	2001-2005	2006-2010	2011-2015	2016-2019
Digital library/ services	Information retrieval	Citation analysis/impact factor	Bibliometric analysis	Citation analysis/impact factor
Academic library/ librarianship	Text processing	Text processing	Research performance	Social Media
Information systems/ information management	Organizational information activities	Information retrieval/user	Citation analysis/measurement	Organizational information activities
Citation analysis/ impact factor	Information systems	Organizational information activities	Scientific collaboration	Text processing
Scientific collaboration	Digital libraries/ services	Government	Citation analysis/Impact factor	Knowledge sharing
Information retrieval	WWW	Scientific collaboration	Information management	Government
Citation analysis	Networks/ communities	Semantic analysis/ models and algorithms	Government	Organizational innovation/performance
Patent analysis	Technology application	Patent analysis	Online/ community	E-commerce
Information-seeking behavior	Citation analysis	Technology application	Organizational information activities	Mobile application
Information systems/design	Information seeking behavior	Research performance	Ranking research	Knowledge management

	Library science
	Bibliometrics
	Information science and related issues

difficult to separate. In the first two periods, information systems were discussed mainly in terms of their application and design, whereas in the last two periods, the topic had a close relationship with commercial activities. It is worth noting that knowledge management and knowledge sharing were discussed intensively from 2016 to 2019.

The field of information seeking and behavior was stable in the first three periods. However, studies conducted since 2011 did not mention the field as frequently as previous studies. This does not necessarily imply that user studies are becoming less prevalent because such studies may be dispersed over other topics, such as social media. Scientific collaboration and research performance appeared in three periods. The review of relevant documents shows that the field has a close relationship with citation analysis.

Table 9 Overview of results (topic level). (Color table online)

1996–2000	2001–2005	2006–2010	2011–2015	2016–2019
Citation analysis/ impact factor	Citation analysis	Citation analysis/ impact factor	Citation analysis/ measurement	Citation analysis/ impact factor
Citation analysis	Information retrieval	Patent analysis	Citation analysis/ impact factor	Text processing
Patent analysis	Text processing	Text processing	Bibliometric analysis	Organizational information activities
Information retrieval	Organizational information activities	Semantic analysis/ models and algorithms	Organizational information activities	Knowledge sharing
Information systems/ information management	Information systems	Organizational information activities	Information management	Knowledge management
Information systems/design	Information-seeking behavior	Information retrieval/ users	Scientific collaboration	Government
Information-seeking behavior	Digital libraries/ services	Scientific collaboration	Research performance	E-commerce
Scientific collaboration	WWW	Research performance	Government	Social media
Digital library/ services	Networks/ communities	Government	Online/ communities	Organizational innovation/ performance
Academic library/ librarianship	Technology applications	Technology applications	Ranking research	Mobile applications

Citation analysis		Scientific collaboration
Information retrieval		Library science
Organizational activities		Government
Information system		Topics occurred less than twice
Information seeking and behavior		

Library-related issues occurred in the first two periods while government issues were studied during the last three. The study of Liu and Yang (2019), which exclusively examined research topics in library journals for the period 2008–2017, reveals the frequent use of the keywords "e-government" and "government" in library science. It is reasonable to assume that governmental issues may have taken over the interest in libraries to some extent.

Topics that occurred less than twice across all periods are listed at the bottom of the table. These topics included WWW, technology applications, networks and online communities, research ranking, social media, organizational innovation, and mobile applications. Topics that occurred only once have undergone a change of technological context from the Internet and networks to social media and mobile applications. These short-lived topics have one common feature: most of them describe a research context and are motivated by the development of technology. Although only appearing for a limited time, these topics have received so much attention that the discipline LIS is recognized as technology-driven.

Comparison with previous studies

Sugimoto et al. (2010) conducted a representative study using author-topic model, an extension of LDA, to investigate the evolution of LIS. The authors investigated dominant topics in doctoral dissertations between 1930 and 2009. Table 10 lists the topics identified by Sugimoto et al. (2010) for periods that overlap with this study.

Regarding the areas of information retrieval, information seeking, and library science, the two studies share a consistency in the overlapping period. However, there are two significant differences. First, the present study identifies bibliometrics as a substantial component of LIS, while there is no such topic in the overlapping period in the results of Sugimoto et al. (2010). Sugimoto et al. (2010) discussed a similar issue when comparing their findings to the study of Åström (2007), who used highly cited journals as a data source. This may suggest that bibliometrics was not intensively investigated among LIS dissertations. Secondly, this study presents a broader spectrum of topics than the study of Sugimoto et al. (2010). This may reflect the different data sources applied in the two studies, suggesting that journal articles are more flexible and sensible to the external social circumstances and technological development, whereas dissertations have a relatively narrow research scope concentrating on the core areas of LIS.

The study of Åström (2007) used co-citation analysis to investigate research topics in LIS based on highly cited journal articles. The author's findings regarding bibliometrics and information science are consistent with this study (Table 11). However, Åström (2007) found no clusters relevant to libraries in periods that overlap with this study and fewer subfields overall. This discrepancy may reflect the combined effect of journal selection and research methods. Chang et al. (2015) conducted another representative bibliometric analysis. The authors applied three methods to examine LIS research subjects based on 580 highly cited journal articles (1995–2014). They identified three areas in LIS: bibliometrics, IS, and IR and AIT (application of Internet technology). They found a decreasing trend in IS and IR and an increasing trend in bibliometrics, which can also be found in the

Table 10 Summary of topics during 1990–2009 in Sugimoto et al. (2010, p. 193)

1990–1999	2000–2009
Model development	Information use
Library outreach	Internet
Information-seeking behavior	Information-seeking behavior
Librarianship/management	IR/user-centered
IR/media	IR/classification

Table 11 Summary of topics identified by Åström ([2007](#)) for 1995–2004

1995–1999	2000–2004
Information seeking and retrieval	Webometrics
Informetrics	Information seeking and retrieval
Information technology	Children s information behavior
Experimental information retrieval	Health informatics
Bibliometrics and research performance	

periods that overlap with this study. Across these two studies, bibliometric methods based on highly cited journals rarely identified library science as a dominant LIS topic.

LDA

The interpretation of the topics generated in this study is generally straightforward. LDA proved to be an excellent method for understanding the rich underlying topical structure of a field and on demonstrating emerging and sustained trends. One essential feature of LDA is that it assumes that one document addresses multiple topics. On the one hand, this assumption enables researchers to detect the underlying structure of a corpus more precisely. On the other hand, it provides a detailed perspective on how the topics are combined in the documents, which is especially essential for LIS because it is interdisciplinary and technology-driven. Until now, most studies of interdisciplinary disciplines have been based primarily on indirect indicators such as faulty composition, co-authorship or citations (e.g., Chang and Huang [2011](#); Huang and Chang [2012](#); Prebor [2010](#)). In contrast, topic-level analysis enables the direct inspection of the topic components of a large text corpus and the examination of how the different topics are combined. In this way, LDA may provide a new level of granularity for examining highly interdisciplinary areas of LIS (Yan [2015](#)).

The relationship between technology and LIS can be further elucidated using LDA. For example, today, numerous disciplines take social media as a popular research subject, including psychology, social science, computer science, and economics. In order to contribute to a better understanding of the nature of LIS, it is important to determine which aspects of social media were integrated into the field, rather than merely claiming that social media is a hot topic in the field and placing it alongside other topics such as information retrieval or user studies. Numerous such topics make the development of LIS rather confusing and unclear. LDA enables further exploration of the point where technology and LIS meet.

Without the correlation information between topics, the identification of relationships between clusters largely depends on the manual interpretation of researchers. Regarding the limitation, Blei and Lafferty ([2007](#)) proposed a modified topic model, the correlated topic model (CTM), which "gives a more realistic model of the latent topic structure where the presence of one latent topic may be correlated with the presence of another" (p. 19). Another improvement based on LDA is dynamic topic models (DTM), which was specifically designed for the study of the time evolution of topics. DTM can be used to capture the evolution of topics in a large sequential text corpus, observing how new topics emerge and disappear over time in a field (Blei and Lafferty [2006](#)). Future studies regarding research

trends and the intellectual structure of a domain may also consider utilizing the modified models.

Library and information science

In its early stages, library science focused on the professions of librarianship and collection management. This focus on practical library issues rather than the management of information in books led to the name "library science," which misled some about the nature of the subject, suggesting a science taking the institute library as the primary research subject. This led to various critiques of the field in its early stage. However, what makes the library a distinct organization is the feature of managing a large amount of information. Without the efficient tools available today, collection management at that time largely depended on professional librarians with technical skills. It was not until the transfer of information from paper to digital form that activities and studies related to information management radical changed. The shift of attention to information systems has been accompanied by the fading of library science. Libraries become ordinary organizations, as the large volume of information that they manage is not unusual among organizations equipped with modern technologies. The focus of libraries has gradually moved to user service and governance. In the new digital environment, researchers must acquire skills in information management to solve problems and provide better access to users, much as earlier librarians had to gain professional skills. As a result, LIS and computer science are intimately related. This relationship has been verified by numerous studies of various aspects of LIS, such as journal rankings, university faculty, and research topics.

Today, the scope of the field of information science as an independent domain is larger than information science evolved from LIS. One remarkable feature is the field's close relationship with the economy and information activities within organizations (Stock and Stock 2013), which was also demonstrated by the results in this study during 2016–2019. These subjects are not widely accepted as classic fields in LIS, which is why many researchers exclude certain journals from the LIS categories in JRC. However, given that LIS will develop further towards information science, such merging is inevitable and, in fact, is already reflected in the rankings of influential journals in LIS. Another issue that complicates the definition of information science as a field is its intimate relationship with information technology. The rapid development of technology causes the field's research focus to change constantly. However, the field's changing topics have one feature in common: they all address the properties of the external information environment. Two constant features are information and humans. Saracevic (1999) noted that it is hard to predict the future of LIS because the field is, by its nature, technology driven. However, perhaps precisely because of this nature, the future path of LIS is foreseeable when the true nature of the field is understood.

The third broad cluster included in the field is bibliometrics. Bibliometrics, especially citation analysis, is very stable across the periods. It has developed into one of the most dominant areas of LIS today. According to Hayes (2009), the major impact of the automation of libraries in the period of 1990–2008 was on print journals, which are rapidly disappearing and being replaced by electronic access via the Internet. This might be one of the main reasons for the thriving development of this field in the later periods. Apart from citation analysis, some areas related to scientific communication, research performance and international publications remained popular over the decades.

The data corpus in the period of 2015–2019 shows frequent use of the word "knowledge," partially replacing the word "information" in some circumstances. The Fig. 2 shows the term frequency of the word "library," "information," and "knowledge" over the years in the datasets. The word "information" and "knowledge" have experienced an evident growth, whereas the word "library" has been gradually less used. The findings in the figure consistent well with the assumption concluded in the topic analysis regarding library science and the use of the word knowledge.

Limitations

The topic clusters are sensitive to the amount of data being analyzed. A journal rarely features complete coverage of research areas in LIS. Instead, most journals are, to some extent, biased towards a certain domain. The different number of articles in each journal caused a cluster heterogeneity issue in LDA when all the documents were joined into one corpus. With a larger volume of texts, some topics may generate multiple small clusters that would otherwise be assigned to a single category by manual interpretation. This issue is especially noticeable in the period 2011–2015, where topics about bibliometrics formed more sub-clusters.

To ensure that the research trends and the most prevalent topics in each period could be sensitively detected, this study applied a journal selection method based on all available JCRs of LIS. This limited the study period to 1996–2019. The exclusion of the period before 1996 is due to the lack of a consistent journal selection criteria, leading to a gap period that is worthy of further investigation.
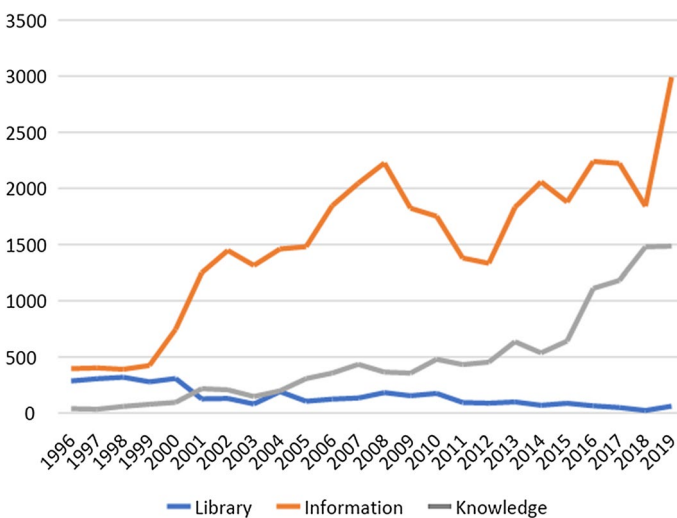

Fig. 2 Term frequency

Conclusion

This study used an improved method of journal selection to investigate the evolution of research topics in LIS based on LDA modeling. The analysis was divided into five periods covering the years from 1996 to 2019. A dynamic journal list with the most influential journals of each period was generated. In total,14,053 journal articles were included in the analysis, and their titles, abstracts, and keywords were used as the text corpus for LDA to identify underlying topics. For each period, the top 10 topics and their keywords were identified for further analysis.

This diachronic analysis shows that library science is gradually losing its dominance within LIS. One of the most remarkable indicators is the decrease in clusters related to library issues. In information science, information retrieval has consistently been the dominant domain, with its interests gradually shifting towards model- or algorithm-based text processing. Information seeking and behavior is also a stable field which tends to disperse among various topics rather than be identified as a distinct topic. Information systems and organizational activities have been discussed continuously and have developed a close relationship with e-commerce. The short-lived topics (those that appear in only one period) evidence a shift in technological context from the Internet and networks to social media and mobile applications. Bibliometrics has proven to be a stable area in LIS. Some topics, including citation analysis, scientific collaboration, and research performance, repeatedly occur across periods, and in rare cases, new topics emerged.

This study presents front research topics in LIS for the period 1996–2019. By combining a unique journal selection method and LDA topic modeling, the research contributes to a new perspective on observing the evolution of the domain. This study exhibits a diversity of research topics in LIS and reveals some research diachronic trends. In future work, the research topics in LIS journals before 1996 may still be worth to be explored using LDA. Furthermore, the structure and development of interdisciplinarity could be further examined by analyzing topic distribution in documents. From a holistic view, the author argues for further classification of research topics and the establishment of a systematic topic frame, for example, distinguishing topics by different attributes like method, research context, content, and user group.

Appendix 1: Literature overview

Title	Authors	Methodology	Conclusion	Source	Period coverage
Visualizing library and information science concept spaces through keyword and citation based maps and clusters	Fredrik Astrom (2002)	Co-citation, keywords	Three clusters were formed by keywords and keywords plus author analysis: Library science, information retrieval and bibliometrics	Articles from nine LIS journals, the four highest ranked general IS journals and the five highest ranked LS journals were selected	1998–2000
Analysis of a decade in library literature: 1994–2004	Kelly Blessinger and Michele Frasier (2007)	Content analysis. citation analysis	The analysis illustrated that librarians are still largely writing about the practical issues that face the profession. New technologies in information science, most notably the Internet, had a tremendous impact on almost every aspect of our profession during this decade	10 influential journals. Journals selection consulted JCR and exit in both Library Literature and SSCI. Ulrich periodical directory was consulted to eliminate the journals that focus mainly on information science	1994–2004

Title	Authors	Methodology	Conclusion	Source	Period coverage
Changes in the LIS research front: time-sliced cocitation analyses of LIS journal articles, 1990–2004	Fredrik Astrom (2007)	Co-citation analysis	The results show a stable structure of two distinct research fields: informatics and information seeking and retrieval (ISR). The focus on the internet is an important change	21 LIS Journals. 39–65 most cited documents were used for each time slice	1990–2004

Title	Authors	Methodology	Conclusion	Source	Period coverage
The shifting sands of disciplinary development: Analyzing north American library and information science dissertations using latent Dirichlet allocation	Cassidy Sugimoto, Daifeng Li, Terrell Russell, Craig finly and Ying Ding (2011)	LDA	The main topics in LIS have changed substantially from those in the initial period (1930–1969) to the present (2000–2009). Some themes occurred in multiple periods, representing core areas of the field: library history, citation analysis, and information-seeking behavior. Two topics occurred in three of the five periods: information retrieval and information use	3121 doctoral dissertations at North American library and information science programs	1930–2009

Title	Authors	Methodology	Conclusion	Source	Period coverage
A bibliometric Chronicling of Library and Information Science s first hundred years	Vincent Lariviere, Cassidy R Sugimoto and Blaise Cronin (2012)	Lexicon frequency and bibliometric indicators	Two major structural shifts are revealed: in 1960, LIS changed from a professional field focused on librarianship to an academic field focused on information and use, and in 1990, LIS began to receive a growing number of citations from outside the field	160 journals, in a total of approximately 96,000 papers having the classification "Information Science & Library Science" created by CHO research and used by the US National Science Foundation	1900–2010
Evolution of Library and information science, 1965–2005: Content analysis of journal articles	Otto Tuomaala,Kalervo Jarvelin, and Pertti Vakkari (2014)	Content analysis	Information retrieval has been the most popular area of research over the period studied. The most significant changes are the decreasing interest in library and informa- tion- service activities and the growth of research into information seeking and scientific communication	42 core journals of LIS.	1965–2005

Title	Authors	Methodology	Conclusion	Source	Period coverage
Detecting the intellectual structure of library science based on formal concept analysis	Ping Liu, Qiong Wu, Xiangming Mu (2015)	Formal concept analysis (FCA)	They identified nine major research themes of LIS: bibliometrics, scientometrics, and informetrics; citation analysis; information retrieval; information behavior; libraries; user studies; social network analysis; information visualization; webometrics	10,684 documents from 16 "world's leading journals of LIS", downloaded from SCI and SSCI. Including reviews, journal articles, and proceeding papers	2001–2013

Title	Authors	Methodology	Conclusion	Source	Period coverage
Evolution of research subjects in library and information science based on keyword, bibliographical coupling and co-citation analyses	Yu-Wei Chang, Mu_Hsuan Huang, Chiao wen Lin (2015)	Keyword, bibliographical coupling, and co-citation	The results revealed that the two subjects "information seeking (IS) and information retrieval (IR)" and "bibliometrics" appeared in all 4 phases. However, a decreasing trend on IS and a increasing trend on bibliometric in percentage of articles.	580 highly cited LIS articles in ten journals with highest impact factors. To ensure that the selected journals were LIS-oriented, they defined LIS journals as those that were indexed simultaneously in three LIS databases of LISA; Library Literature and Information Science; and Library, Information Science and Technology Abstracts	1995–2014
Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA	Carlos G. Figuerola, Francisco Javier Garcia Marco, Mari Pinto (2017)	LDA	The results show 19 topics, which can be further grouped into four areas: processes, information technology, library and special areas of information application	92,705 documents in the database of LISA	1978–2014

Title	Authors	Methodology	Conclusion	Source	Period coverage
Forty-five years of LIS Research Evolution,1971–2015: An informetrics study of the author-supplied keywords	Omwoyo Bosire Onyancha (2018)	Author-supplied keywords analysis	LIS has evolved from information systems design and management in the 1970s to scientific communication, information storage and retrieval, information access, information and knowledge management, and user education in 2015	A total of 26,492 research articles. Data was extracted from Thomson Reuters' citation mainstream indexes	1971–2015
Popular research topics in the recent journal publications of library and information science	Guoying Liu, Le Yang (2019)	Author-supplied keywords analysis	The most popular research topics are Social Media, Data, Web, Information Retrieval, Information Literacy, Students, Evaluation, Collaboration, Knowledge Management, User Studies, and Information Management	LIS journal list developed by Judith Nixon, which focus more on library science and librarianship	2008–2017

Appendix 2: Representative documents for top ten topics in 1996–2000

Topic	Source (authors, year, journal)	Title
A1 Digital library/services	Clay III E.S., Bangs P.C. 2000 *Library Trends*	Entrepreneurs in the public library: Reinventing an institution
A2 Academic library/librarianship	Jaguszewski J.M., Probst L.K. 2000 *Library Trends*	The impact of electronic resources on serial cancellations and remote storage decisions in academic research libraries
A3 Information systems/information management	Wells J.D., Fuerst W.L., Choobineh 1999 *J. Information and Management*	Managing information technology (IT) for one-to-one customer interaction
A4 Citation analysis/impact factor	Haiqi Z. 1996 *Scientometrics*	Research performance in key medical universities in China observed from the scientific productivity
A5 Scientific collaboration	Leta, J., De Meis, L. 1996 *Scientometrics*	A profile of science in Brazil
A6 Information retrieval	Park S. 2000 *Journal of the American Society for Information Science and Technology*	Usability, user preferences, effectiveness, and user behaviors when searching individual and integrated full-text databases: implications for digital libraries
A7 Citation analysis	Glänzel W., Schubert A., Schoepflin U., Czerwon H.-J. 1999 *Scientometrics*	An item-by-item subject classification of papers published in journals covered by the SSCI database using reference analysis
A8 Patent analysis	Tijssen R.J.W., Buter R.K., Van Leeuwen Th.N. 2000 *Scientometrics*	Technological relevance of science: An assessment of citation linkages between patents and research
A9 Information-seeking behavior	Large A., Beheshti J. 2000 *Journal of the American Society for Information Science and Technology*	The Web as a classroom resource: Reactions from the users
A10 Information systems/design	Crabtree A., Nichols D.M., O'Brien J., Rouncefield M., Twidale M.B. 2000 *Journal of the American Society for Information Science and Technology*	Ethnomethodologically informed ethnography and information system design

Appendix 3: Representative documents for top ten topics in 2001–2005

Topic	Source (authors, year, journal)	Title
B1 Information retrieval	Jantz R 2003 *College and Research Libraries*	Information retrieval in domain specific databases: An analysis to improve the user interface of the alcohol studies database
B2 Text processing	Gao J., Zhang J. 2005 *Information Processing and Management*	Clustered SVD strategies in latent semantic indexing
B3 Organizational information activities	Attaran M. 2004, *Information and Management*	Exploring the relationship between information technology and business process reengineering
B4 Information systems	Fortunati L. 2005 *Information Society*	Is Body-to-body communication still the prototype?
B5 Digital libraries/services	Brennan M.J., Hurd J.M., Blecic D.D., Weller A.C. 2002 *College and Research Libraries*	A snapshot of early adopters of e-journals: Challenges to the library
B6 WWW	Nicholas D., Huntington P., Williams P. 2002 *Journal of Information Science*	Evaluating metrics for comparing the use of web sites: A case study of two consumer health web sites
B7 Networks/communities	Aerts A.T.M., Goossenaerts J.B.M., Hammer D.K., Wortmann J.C. 2004 *Information and Management*	Architectures in context: On the evolution of business, application software, and ICT platform architectures
B8 Technology application	Lee M.K.O., Cheung C.M.K., Chen Z. 2005 *Information and Management*	Acceptance of Internet-based learning medium: The role of extrinsic and intrinsic motivation
B9 Citation analysis	Sombatsompop N., Markpin T., Premkamolnetr N. 2004 *Scientometrics*	A modified method for calculating the impact factors of journals in ISI Journal Citation Reports: Polymer Science
B10 Information seeking behavior	Spink A., Wilson T.D., Ford N., Foster A., Ellis D. 2002 *Journal of the American Society for Information Science and Technology*	Information-seeking and mediated searching: Part 1. Theoretical framework and research design

Appendix 4: Representative documents for top ten topics in 2006–2010

Topic	Source (authors, year, journal)	Title
C1 Citation analysis/impact factor	Yu G., Wang L. 2007 *Scientometrics*	The self-cited rate of scientific journals and the manipulation of their impact factors
C2 Text processing	Klein S.T. 2009 *Information Processing and Management*	On the use of negation in Boolean IR queries

Topic	Source (authors, year, journal)	Title
C3 Information retrieval/user	Hertzum M. 2008 *Information Processing and Management*	Collaborative information seeking: The combined activity of information seeking and collaborative grounding Abstract:
C4 Organizational information activities	Nottelmann H., Fischer G. 2007 *Information Processing and Management*	A study of B2B e-market in China: E-commerce process perspective
C5 Government	Dorner D.G. 2009 *Government Information Quarterly*	Public sector readiness for digital preservation in New Zealand: The rate of adoption of an innovation in records management practices
C6 Scientific collaboration	Sooryamoorthy R. 2009 *Scientometrics*	Do types of collaboration change citation? collaboration and citation patterns of South African science publications
C7 Semantic analysis/models and algorithms	Tsuji K., Kageura K. 2006 *Journal of the American Society for Information Science and Technology*	Automatic generation of Japanese-English bilingual thesauri based on bilingual corpora
C8 Patent analysis	Bassecoulard E., Lelu A., Zitt M. 2007 *Scientometrics*	Patent-bibliometric analysis on the Chinese science—Technology linkages
C9 Technology application	Yi M.Y., Jackson J.D., Park J.S., Probst J.C. 2006 *Information and Management*	Understanding information technology acceptance by individual professionals: Toward an integrative view
C10 Research performance	Lercher A. 2010 *Journal of the American Society for Information Science and Technology*	Efficiency of scientific communication: A survey of world science

Appendix 5: Representative documents for top ten topics in 2011–2015

Topic	Source (authors, year, journal)	Title
D1 Bibliometric analysis	Lv P.H., Wang G.-F., Wan Y., Liu J., Liu Q., Ma F. 2011 *Scientometrics*	Bibliometric trend analysis on global graphene research
D2 Research performance	Kao C., Liu S.-T., Pao H.-L. 2012 *Scientometrics*	Assessing improvement in management research in Taiwan
D3 Citation analysis/measurement	Zhai L., Yan X., Zhu B. 2014 *Scientometrics*	The Hl-index: Improvement of H-index based on quality of citing papers
D4 Scientific collaboration	Zhao Q., Guan J. 2011 *Scientometrics*	International collaboration of three 'giants' with the G7 countries in emerging nanobiopharmaceuticals

Topic	Source (authors, year, journal)	Title
D5 Citation analysis/impact factor	Solomon D.J., Björk B.-C. 2012 *Journal of the American Society for Information Science and Technology*	A study of open access journals using article processing charges
D6 Information management	Hullavarad S., O'Hare R., Roy A.K., 2015 *International Journal of Information Management*	Enterprise Content Management solutions—Roadmap strategy and implementation challenges Enterprise Abstract
D7 Government	Janssen K. 2011 *Government Information Quarterly*	The influence of the PSI directive on open government data: An overview of recent developments
D8 Online/community	Kuo Y.-F., Feng L.-H. 2013 *International Journal of Information Management*	Relationships among community interaction characteristics, perceived benefits, community commitment, and oppositional brand loyalty in online brand communities
D9 Organizational information activities	Wang E.T.G., Chou F.K.Y., Lee N.C.A., Lai S.Z. 2014, *Information and Management*	Can intrafirm IT skills benefit interfirm integration and performance?
D10 Ranking research	De Witte K., Hudrlikova L. 2013, *Scientometrics*	What about excellence in teaching? A benevolent ranking of universities

Appendix 6: Representative documents for top ten topics in 2016–2019

Topic	Source (authors, year, journal)	Title
E1 Citation analysis/impact factor	Zhang P., Wang P., Wu Q. 2018 *Journal of the Association for Information Science and Technology*	How are the best JASIST papers cited?
E2 Social Media	Hand L.C., Ching B.D. 2019 *Government Information Quarterly*	Maintaining neutrality: A sentiment analysis of police agency
E3 Organizational information activities	Soto-Acosta P., Placer-Maruri E., Perez-Gonzalez D. 2016 *International Journal of Information Management*	A case analysis of a product life-cycle information management framework for SMEs
E4 Text processing	Al-Salemi B., Ayob M., Kendall G., Noah S.A.M. 2019 *Information Processing and Management*	Multi-label Arabic text categorization: A benchmark and baseline comparison of multi-label learning algorithms
E5 Knowledge sharing	Rao Jada U., Mukhopadhyay S.,Titiyal R. 2019 *Journal of Knowledge Management*	Empowering leadership and innovative work behavior: a moderated mediation examination

Topic	Source (authors, year, journal)	Title
E6 Government	Lee-Geiller S., Lee T.D. 2019 *Government Information Quarterly*	Using government websites to enhance democratic E-governance: A conceptual model for evaluation
E7 Organizational innovation/performance	Chi M., Zhao J., George J.F., Li Y., Zhai S. 2017 *International Journal of Information Management*	The influence of inter-firm IT governance strategies on relational performance: The moderation effect of information technology ambidexterity
E8 E-commerce	Wang W.-T., Wang Y.-S., Liu E.-R. 2016 *Information and Management*	The stickiness intention of group-buying websites: The integration of the commitment–trust theory and e-commerce success model
E9 Mobile application	Fang J., Zhao Z., Wen C., Wang R. 2017 *International Journal of Information Management*	Design and performance attributes driving mobile travel application engagement
E10 Knowledge management	Downes T., Marchant T. 2016 *Journal of Knowledge Management*	The extent and effectiveness of knowledge management in Australian community service organisations

References

Åström, F. (2007). Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. *Journal of the American Society for Information Science and Technology, 58*(7), 947–957. https://doi.org/10.1002/asi.20567.

Blei, D. M. (2009, September 1). *Generative model* [Graph]. Retrieved from http://videolectures.net/site/normal_dl/tag=50740/mlss09uk_blei_tm.pdf.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In: *Proceedings of the 23rd International Conference on Machine Learning—ICML'06* (pp. 113–120). https://doi.org/10.1145/1143844.1143859.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics, 1*(1), 17–35. https://doi.org/10.1214/07-aoas114.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(4–5), 993–1022.

Blessinger, K., & Frasier, M. (2007). Analysis of a decade in library literature: 1994–2004. *College & Research Libraries, 68*(2), 155–169. https://doi.org/10.5860/crl.68.2.155.

Chang, Y.-W., & Huang, M.-H. (2011). A study of the evolution of interdisciplinarity in library and information science: Using three bibliometric methods. *Journal of the American Society for Information Science and Technology, 63*(1), 22–33. https://doi.org/10.1002/asi.21649.

Chang, Y.-W., Huang, M.-H., & Lin, C.-W. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics, 105*(3), 2071–2087. https://doi.org/10.1007/s11192-015-1762-8.

Figuerola, C. G., García Marco, F. J., & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. *Scientometrics, 112*(3), 1507–1535. https://doi.org/10.1007/s11192-017-2432-9.

Gensim. (n.d.). *gensim: LDA model*. Retrieved April 2, 2020, from https://radimrehurek.com/gensim/auto_examples/tutorials/run_lda.html#sphx-glr-auto-examples-tutorials-run-lda-py.

Hayes, R. M. (2009). Library automation: history. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences* (3rd ed., pp. 3326–3337). Routledge: Taylor & Francis. https://doi.org/10.1081/e-elis3-120044024.

Hjørland, B. (2018). Library and information science (LIS), part 1. *Knowledge Organization, 45*(3), 232–254. https://doi.org/10.5771/0943-7444-2018-3-232.

Hood, W. W., & Wilson, C. S. (2001). The Literature of Bibliometrics, Scientometrics, and Informetrics. *Scientometrics, 52*(2), 291–314. https://doi.org/10.1023/a:1017919924342.

Huang, M.-H., & Chang, Y.-W. (2012). A comparative study of interdisciplinary changes between information science and library science. *Scientometrics, 91*(3), 789–803. https://doi.org/10.1007/s11192-012-0619-7.

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

Koufogiannakis, D., Slater, L., & Crumley, E. (2004). A content analysis of librarianship research. *Journal of Information Science, 30*(3), 227–239. https://doi.org/10.1177/0165551504044668.

Larivière, V., Sugimoto, C. R., & Cronin, B. (2012). A bibliometric chronicling of library and information science's first hundred years. *Journal of the American Society for Information Science and Technology, 63*(5), 997–1016. https://doi.org/10.1002/asi.22645.

Liu, P., Wu, Q., Mu, X., Yu, K., & Guo, Y. (2015). Detecting the intellectual structure of library and information science based on formal concept analysis. *Scientometrics, 104*(3), 737–762. https://doi.org/10.1007/s11192-015-1629-z.

Liu, G., & Yang, L. (2019). Popular research topics in the recent journal publications of library and information science. *The Journal of Academic Librarianship, 45*(3), 278–287. https://doi.org/10.1016/j.acalib.2019.04.001.

Newman, D. J., & Block, S. (2006). Probabilistic topic decomposition of an eighteenth-century American newspaper. *Journal of the American Society for Information Science and Technology, 57*(6), 753–767. https://doi.org/10.1002/asi.20342.

Olson, G. M., & Grudin, J. (2009). TIMELINES The information school phenomenon. *Interactions, 16*(2), 15. https://doi.org/10.1145/1487632.1487636.

Onyancha, O. B. (2018). Forty-five years of LIS research evolution, 1971–2015: An informetrics study of the author-supplied keywords. *Publishing Research Quarterly, 34*(3), 456–470. https://doi.org/10.1007/s12109-018-9590-3.

Prebor, G. (2010). Analysis of the interdisciplinary nature of library and information science. *Journal of Librarianship and Information Science, 42*(4), 256–267. https://doi.org/10.1177/0961000610380820.

Saracevic, T. (1999). Information science. *Journal of the American Society for Information Science, 50*(12), 1051–1063. https://doi.org/10.1002/(SICI)1097-4571(1999)50:12%3c1051:AID-ASI2%3e3.0.CO;2-Z.

Stock, M., & Stock, M. (2013). *Handbook of information science*. Berlin, Germany: Walter de Gruyter.

Sugimoto, C. R., Li, D., Russell, T. G., Finlay, S. C., & Ding, Y. (2010). The shifting sands of disciplinary development: Analyzing North American Library and Information Science dissertations using latent Dirichlet allocation. *Journal of the American Society for Information Science and Technology, 62*(1), 185–204. https://doi.org/10.1002/asi.21435.

Suominen, A., & Toivanen, H. (2015). Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *Journal of the Association for Information Science and Technology, 67*(10), 2464–2476. https://doi.org/10.1002/asi.23596.

The Editors of Encyclopaedia Britannica. (n.d.). Library science. Retrieved February 4, 2020, from https://www.britannica.com/science/library-science.

Tuomaala, O., Järvelin, 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*(7), 1446–1462. https://doi.org/10.1002/asi.23034.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*(4), 327–355.

Yan, E. (2015). Research dynamics, impact, and dissemination: A topic-level analysis. *Journal of the Association for Information Science and Technology, 66*(11), 2357–2372. https://doi.org/10.1002/asi.23324.

Yau, C.-K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics, 100*(3), 767–786. https://doi.org/10.1007/s11192-014-1321-8.

Zhao, D., & Strotmann, A. (2008). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology, 59*(13), 2070–2086. https://doi.org/10.1002/asi.20910.

Zhao, D., & Strotmann, A. (2014). The knowledge base and research front of information science 2006–2010: An author cocitation and bibliographic coupling analysis. *Journal of the Association for Information Science and Technology, 65*(5), 995–1006. https://doi.org/10.1002/asi.23027.