Research Advances in Computer and Information Science over the Past Two Decades: IPM Journal Topic Analysis Postprint

Authors: Li Hanxiao, Du Xingye

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose/Significance] This study conducts a thematic analysis of papers published in Information Processing & Management from 2000 to 2020 to understand the IPM journal' s thematic focus and evolutionary trends, providing references for the development of computer and information science and related research. [Method/Process] First, based on 1,852 research papers from the ScienceDirect full-text database, statistical analysis and visualization of paper titles, abstracts, and keywords were performed to delineate major thematic categories. Second, research topics within each category were systematically reviewed. Finally, research focuses across different periods were compared to analyze thematic evolutionary trends. [Results/Conclusion] IPM primarily focuses on three thematic categories: information retrieval, text analysis, and user studies. Overall, it exhibits evolutionary characteristics including consistently maintaining information retrieval as the core theme, shifting from text and information analysis toward multimedia and knowledge analysis, and conducting in-depth analysis and mining of user sentiment.

Full Text

Preamble

Research Progress in Computer and Information Science in Recent 20 Years: Thematic Analysis of Information Processing and Management

Li Hanxiao[1,2], Du Xingye[1,2]
[1]National Science Library, Chinese Academy of Sciences, Beijing 100190
[2]Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract

[**Purpose/Significance**] This paper analyzes the themes of papers published in *Information Processing and Management* from 2000 to 2020 to understand the thematic focus and evolution trends of IPM, providing references for the development and related research in computer and information science. [**Method/Process**] First, based on 1,852 research papers from the ScienceDirect full-text database, we conducted statistical analysis and visualization of paper titles, abstracts, and keywords to classify major thematic categories. Second, we systematically reviewed the research themes within each category. Finally, we compared research focuses across different periods and analyzed thematic evolution trends. [**Result/Conclusion**] IPM primarily focuses on three themes: information retrieval, text analysis, and user research. The journal exhibits evolution characteristics including: consistently maintaining information retrieval as the core theme, shifting from text and information analysis toward multimedia and knowledge analysis, and conducting in-depth analysis and mining of user sentiments.

Keywords: *Information Processing and Management*; IPM; computer and information science; thematic analysis

Introduction

The advent of the digital, networked, and intelligent era has brought tremendous impact to the development of computer and information science, leading to rapid growth in related research and changes in research themes. As the primary carrier and dissemination platform for scientific research achievements, academic journals bear an important mission in academic exchange [1]. Thematic analysis of journal publications can provide better understanding of research progress and evolution characteristics in a discipline. *Information Processing and Management* (IPM) was founded in 1963, originally named *Information Storage and Retrieval* (ISR), and officially renamed IPM in 1975, which continues to this day. According to SCI-JCR data, the journal had a CiteScore of 8.6 and an impact factor of 6.222 in 2020, ranking in the first quartile in both Computer Science, Information Systems and Information Science & Library Science categories. IPM is dedicated to publishing cutting-edge research results at the intersection of computer and information science, and enjoys high influence and recognition in both the international computer and information systems community and the library and information science community, making important contributions to advancing the field. Therefore, systematically reviewing the publication themes of IPM can to some extent reflect research progress in computer and information science and demonstrate IPM' s academic contributions to field development.

Previous scholars have conducted thematic analyses of IPM. F. E. DeHart analyzed references in papers published in IPM, JASIS, and JD from 1987-1990, focusing on comparing the proportion of monograph citations, finding that the three most cited monograph topics in IPM during 1989-1990 were information storage and retrieval systems, artificial intelligence, and discourse analysis [2]. M. Y. Tsay conducted bibliometric analysis and comparison of JASIST, IPM, and JD from 1998-2008, finding that the three most cited journal paper topics in IPM were search, online information retrieval, and information work, while the three most cited book topics were information storage and retrieval systems, information retrieval, and computer algorithms [3]. Wang Yuefen et al. compared publication characteristics of *Modern Library and Information Technology* and similar international journals including IPM from 2006-2015, finding that information retrieval was the largest research hotspot in IPM, with other hotspots including user behavior analysis, text mining algorithms, text classification, and semantic analysis [4].

These previous studies all compared IPM with other journals, mostly using quantitative analysis, but paid less attention to the connotative evolution of each research theme. Therefore, we systematically review the publication themes of IPM over the past 20 years (2000-2020) to understand its thematic focus and evolution trends, providing references for the development of computer and information science and related research, as well as useful insights for library and information science.

Data and Methods

Between 2000 and 2020, IPM published 1,852 research papers. We exported the paper data from the ScienceDirect full-text database to form a data source for statistical analysis. First, we conducted keyword frequency statistics and created a keyword cloud map (see Figure 1 [Figure 1: see original paper]), finding that research hotspots mainly involved information retrieval, machine learning, natural language processing, query expansion, social media, sentiment analysis, text classification, and information seeking. Second, we extracted paper titles and abstracts and used VOSviewer for co-occurrence analysis to obtain a co-occurrence network map (see Figure 2 [Figure 2: see original paper]), which clearly shows that research themes primarily fall into three major categories: information retrieval, text analysis, and user research. Finally, based on the 1,852 research papers, we reviewed IPM' s research themes over the past 20 years from these three categories, compared research focuses across different periods, and identified thematic evolution trends.

Theme 1: Information Retrieval

Information retrieval, which emerged in the 1950s, is a discipline that primarily studies the representation, storage, organization, and access of information [5]. Over the past 20 years, information retrieval has consistently been a key focus of IPM, covering topics such as information retrieval models, search engines, and image retrieval.

3.1.1 Information Retrieval Models

Information retrieval models refer to mathematical models that describe documents, queries, and their relationships (matching functions) in information retrieval [6]. Common retrieval models include Boolean, probabilistic, vector space, language, and ranking models. IPM research primarily focuses on probabilistic models. For example, K. Sparck Jones et al. developed a probabilistic model for information retrieval [7-8], which improved upon the probabilistic model proposed by their team in 1976 [9] and remains the most widely recognized retrieval model today. Many studies have modified this model to form a commonly applied version, shown in Equation (1) [10]:

$$\mathrm{sim}(Q,D)=\sum_{i\in Q}\frac{(k_1+1)f_i}{K+f_i}\cdot\frac{(k_2+1)qf_i}{k_2+qf_i}\cdot\log\frac{(r_i+0.5)/(R-r_i+0.5)}{(n_i-r_i+0.5)/(N-n_i-R+r_i+0.5)}$$

where f_i refers to the frequency of term i in the retrieved document, qf_i refers to the frequency of term i in the query, N refers to the size of the entire document collection, r_i refers to the number of relevant documents containing term i, n_i refers to the number of documents containing term i, R refers to the size of the document set relevant to the query, and K, k_1, and k_2 are hyperparameters set based on experience [10].

Other scholars have constructed information retrieval probabilistic models from multiple perspectives and conducted experimental tests. Z. B. Xu et al. developed a probabilistic information retrieval model based on Dirichlet Compound Multinomial (DCM) distribution that enables efficient retrieval and accurate ranking [11]. F. Dahak et al. established a probabilistic model for XML information retrieval that utilizes user expectations to estimate contextual importance, demonstrating its effectiveness through experiments [12]. Additionally, vector space models and ranking models have received extensive discussion. For instance, X. Y. Tai et al. proposed an information retrieval model based on the Vector Space Model (VSM) that can utilize user-provided relevance information to improve retrieval performance [13]. J. F. Guo et al. analyzed and compared the basic assumptions, design principles, and learning strategies of neural ranking models, discussing future development trends such as learned indexes, learning with external knowledge, learning with visualization techniques, learning with context, and understanding neural ranking models [14]. Overall, related research primarily focuses on the construction, testing, comparison, and improvement of retrieval models, aiming to enhance information retrieval system performance and achieve higher-precision information retrieval.

3.1.2 Search Engines

Research on search engines concentrates on retrieval performance evaluation, query expansion and relevance feedback, and web page ranking algorithms. Regarding retrieval performance evaluation, L. Vaughan proposed a set of measurement methods based on precision and recall concepts to evaluate search engine performance and stability, with experimental results showing that these methods can effectively distinguish search engine performance [15]. F. Can et al. argued that manual evaluation of search engine retrieval performance is costly, and therefore introduced an automatic search engine evaluation method, proving through experiments that its evaluation results are consistent with manual assessments [16].

As an important branch of query optimization, query expansion primarily uses information from user query logs and relevance feedback to expand user queries. For example, H. Kim et al. proposed a method based on clustering user query logs that can bridge the vocabulary gap between user queries and retrieval systems to some extent [17]. S. Jung et al. used click data from search engine users as an implicit source of relevance feedback information, discussing the reliability of relevance feedback and its variations [18]. Additionally, query expansion has increasingly focused on semantic-based relevance feedback techniques to address the semantic gap between queries and documents, such as J. M. Wang et al.' s proposal of a pseudo-relevance feedback framework combining relevance matching and semantic matching to improve feedback document quality [19].

In terms of web page ranking algorithms, numerous scholars have developed more efficient ranking algorithms based on mainstream algorithms such as PageRank, HITS, and OPIC. For instance, A. M. Z. Bidoki et al. proposed a DistanceRank algorithm based on reinforcement learning that defines the "average number of clicks" between two web pages as distance, where pages with smaller distances receive higher rankings. Experimental results show this algorithm outperforms others in web page ranking and crawling scheduling [20].

3.1.3 Image Retrieval

Building upon text retrieval, multimedia retrieval technologies using images, audio, and video as retrieval objects have gradually developed. IPM-related research has primarily focused on image retrieval, evolving from Text-based Image Retrieval (TBIR) to Content-based Image Retrieval (CBIR). The foundation of CBIR lies in the selection, extraction, and representation of image content features such as color, texture, and shape [21], and related research has generally approached from this perspective. For example, P. W. Huang et al. proposed two texture feature representation methods (CSG-vector and EDP-string) based on texture similarity and designed an efficient image retrieval system accordingly [22]. T. C. Lu et al. addressed color features in images by representing global features through color distribution, mean values, and standard deviations, while using image bitmaps to represent local features to improve image retrieval accuracy [23].

However, content-based image retrieval also faces a difficult-to-cross semantic gap [24]. To address this, semantic-based image retrieval technology has gradually developed. S. Pandey et al. proposed a semantic and image retrieval system for hierarchical image databases with semantic classification, enabling images to be mapped to multi-dimensional feature spaces while simultaneously representing image semantics through clustering and indexing, ultimately achieving efficient retrieval of desired semantics and corresponding images [25].

Theme 2: Text Analysis

Text analysis involves representing text content and extracting features to enable computer recognition and processing, thereby determining text topics and the attitudes and emotions of text providers. IPM research on text analysis primarily concentrates on text mining, sentiment analysis, and knowledge graphs.

3.2.1 Text Mining

Text mining can extract previously unknown, understandable, and ultimately usable knowledge from large amounts of text data in electronic form, and apply this knowledge to better organize information to support reference and utilization [26-27]. IPM-related research primarily revolves around text classification and text clustering. Text classification refers to organizing documents into predefined categories, typically using machine learning algorithms [28]. For example, A. Elnagar et al. compared commonly used deep learning models for Arabic text classification and proposed a classification method entirely based on deep learning models [29].

Meanwhile, text features targeted by text classification have evolved from simple words, phrases, and sentences to grammatical and semantic features. For instance, A. Mohasseb et al. proposed a grammar-based classification framework for question classification in question-answering systems that can effectively distinguish different question types [30]. Z. Kastrati et al. proposed a semantically rich document representation model that can automatically classify financial documents [31]. In text clustering, numerous scholars have proposed various clustering algorithms to optimize clustering performance, such as G. B. Hu et al., who developed a semi-supervised clustering method based on the K-Means algorithm that can constrain the clustering process [32]. C. L. Chen et al. proposed a hierarchical clustering method based on frequent fuzzy itemsets aimed at improving hierarchical clustering accuracy [33]. Other scholars have proposed probabilistic models and algorithms for document clustering, proving through experiments that their performance surpasses previously widely used models and algorithms [34-35]. Additionally, some research has explored application methods and practical effects of text mining in information retrieval [36], user services [37], patent analysis [38], and topic identification [39] from an application scenario perspective.

3.2.2 Sentiment Analysis

The development of various social media platforms such as forums and blogs, along with the emergence of review websites like Dianping, has provided the public with open platforms for emotional communication and consumption reviews [40], generating large amounts of opinions, emotions, evaluations, attitudes, and sentiments toward products, services, events, and topics [41]. Sentiment analysis, or opinion mining, uses natural language processing and text mining techniques to analyze, process, and extract these subjective texts with emotional coloring [42].

IPM research on sentiment analysis primarily uses social media as a platform to analyze user viewpoints and emotions from published content or comments. For example, A. Balahur and S. M. Mohammad analyzed emotions, moods, purposes, styles, and corresponding sentiment analysis systems in tweets using Twitter as a case study [43-44]. A. Severyn et al. constructed an opinion mining model capable of handling new domains or languages for the vast amount of user-generated content on YouTube, validating it through experiments [45]. Other scholars have analyzed user evaluations of products or services to mine attitudes and emotions, such as M. Al-Smadi et al., who proposed a supervised machine learning-based method for sentiment analysis of hotel reviews [46].

During sentiment analysis research, scholars have constructed numerous sentiment analysis models for different scenarios. For instance, A. Kumar et al. proposed a deep learning model for fine-grained sentiment analysis in textual and visual social data [47]. Z. Mahmood et al. developed a Roman Urdu corpus and used it as a basis for developing a deep learning model for mining emotions and attitudes [48].

3.2.3 Knowledge Graphs

Knowledge graphs essentially reveal semantic relationships between entities/concepts in semantic networks [49]. IPM-related research primarily involves three aspects: knowledge graph technology, knowledge graph construction, and knowledge graph applications.

In knowledge graph technology, numerous scholars have focused on knowledge entity extraction, discussing knowledge extraction and representation in knowledge graphs. For example, H. C. Cho et al. studied named entity recognition with multiple segment representations [50]. L. Derczynski et al. described a Twitter entity disambiguation dataset and conducted empirical analysis of named entity recognition and disambiguation in tweets [51]. X. Tang et al. proposed a multi-source knowledge representation learning model that combines entity descriptions, hierarchical types, and textual relations to improve knowledge representation effectiveness [52].

In knowledge graph construction, related research primarily builds semantic relationship-based knowledge graphs from corpora. For instance, I. Bounhas et al. constructed a morpho-semantic knowledge graph from Arabic spoken corpora, using contextual knowledge to infer semantic dependencies between entities and evaluating concentrated use cases for document indexing and query expansion [53]. In terms of applications, knowledge graphs can be applied to retrieval systems, question-answering systems, and big data analysis. For example, D. F. Li et al. proposed a cascade model that simultaneously considers semantic and graph features, designing different cascade structures for knowledge inference and retrieval [54]. S. Shin et al. designed a predicate constraint dictionary for the meaning of consultation questions in question-answering systems and proposed a predicate constraint-based question-answering system that can improve search accuracy [55]. F. Janssens et al. conducted bibliometric analysis of nearly a thousand articles published in five library and information science journals between 2002-2004, using knowledge graphs to visualize term networks [56].

Theme 3: User Research

Studying users' information acquisition, seeking, and utilization behaviors helps information service organizations improve information service system performance and service quality more targeted [57]. Information seeking behavior, user generated content, and personalized service & human-computer interaction are the main focuses of user research.

3.3.1 Information Seeking Behavior

Information seeking behavior refers to individuals' purposeful search for information to meet certain goal requirements. Beyond general information searching, information seeking emphasizes the complete process of fulfilling entire information needs, exploring the reasons behind user search behaviors, influencing factors, user characteristics, and individual differences [58]. Many scholars have analyzed information seeking behavior characteristics of different user types. For example, S. Makri et al. analyzed the information seeking behavior of 27 lawyers and proposed improvements to Ellis' s information seeking behavior model [59]. H. R. Jamali et al. investigated the information seeking behavior of physics and astronomy researchers, revealing disciplinary differences in information seeking behavior and finding that interdisciplinary fields were more likely to use general search tools to obtain information [60]. M. Lykke et al. surveyed doctors' information seeking behavior, finding that most doctors could generate well-structured queries using system features and search strategies [61].

Additionally, as social standards continue to improve, people' s demand for health information has gradually increased, and health information behavior has received more extensive attention. Related research has analyzed health information behavior from different perspectives. Regarding health information needs, W. J. Pian et al. systematically reviewed consumer health information need theories, pointing out that future research should focus on social and emotional dimensions [62]. Regarding health information providers, X. F. Zhang et al. explored doctors' motivations for sharing health information on online platforms, finding that beyond material motivations, professional motivations played a primary role [63]. Regarding health information seekers, I. Huvila et al. studied middle-aged and elderly people' s health information behavior preferences and motivations, comparing them with younger and older adults' health information behavior [64]. Other research has analyzed health information avoidance behavior to promote understanding of health information [65].

3.3.2 User Generated Content

User generated content is a network information resource creation and organization model that developed with the rise of Web 2.0, referring to text, images, videos, and other content created by users in various forms on the Internet [66]. IPM-related research rarely discusses the theoretical foundation of user generated content directly, but instead uses it as a data source for opinion mining, sentiment analysis, and public opinion management research. For example, A. Severyn et al. conducted opinion mining on user generated content on YouTube [45]. Y. D. Ge et al. explored the impact of emotions in user generated content on stock markets [67]. L. F. Li et al. studied the public' s negative emotions on social media after natural disasters and the impact of posts from users with large followings on repost numbers [68].

Additionally, user generated content on social media may cause the spread of rumors and fake news, and scholars have studied rumor identification and detection. Y. H. Liu et al. proposed a model based on Long Short-Term Memory (LSTM) and max pooling that identifies rumor propagation processes by capturing dynamic changes in repost content, propagators, and propagation structures, validating it using Sina Weibo data [69]. S. A. Alkhodair et al. proposed a breaking news rumor detection model based on word2vec and LSTM-RNN (see Figure 3 [Figure 3: see original paper]), conducting experiments with Twitter data to prove that the model outperforms others in precision, recall, and F1-score [70].

In this model, first, tweet W is tokenized into a series of words $(W_1, \ldots, W_{T-1}, W_T)$. Then, the word2vec model converts the word sequence into a vector sequence $(X_1, \ldots, X_{T-1}, X_T)$ and passes it to the LSTM-RNN model through weighted connections. Finally, the LSTM-RNN model determines whether the sample category is rumor (R) or non-rumor (NR) and outputs it as the final vector [70].

3.3.3 Personalized Service and Human-Computer Interaction

In personalized service, research primarily focuses on personalization in recommendation systems. J. Wang et al. proposed a social media personalization framework and built a ranking model that integrates users' tagging history in tag and content recommendations, making system suggestions consistent with user preferences [71]. F. M. Belem et al. improved personalized tag recommendation from object-centric and user-centric perspectives and compared it with object-centric recommendation methods [72]. S. Renjith et al. discussed the development of tourism recommendation systems, tracing the evolution from general search engines to personalized recommendation systems and then to context-aware personalized recommendation systems [73].

In human-computer interaction, besides discussing interactions between users and Internet resources [74], information retrieval systems [75], and emotions in human-computer interaction [76], eye-tracking technology has also been a widely discussed topic. For example, M. J. Cole et al. used eye-tracking technology to model users' interactive information acquisition processes to predict user knowledge levels [77]. M. Clark et al. analyzed users' interaction patterns with email text based on eye-tracking data [78]. B. Hilberink-Schulpen et al. investigated through eye-tracking methods whether the use of foreign languages in job advertisements affects user attention and viewing patterns [79].

Theme Evolution Trends

IPM' s research themes have had different focuses across different time periods. To understand research priorities in different periods and analyze overall thematic evolution trends, we divided the papers from 2000-2020 into four time periods, conducted keyword frequency statistics and visualization for each period, and created keyword cloud maps for different periods (see Figure 4 [Figure 4: see original paper]). Across all four periods, information retrieval remained the most important research theme, closely related to IPM's journal positioning. However, over time, during 2016-2020, research heat on information retrieval declined, while frequencies of keywords such as social media, sentiment analysis, deep learning, machine learning, natural language processing, and text mining increased. Among them, social media appeared 33 times, surpassing information retrieval (31 times) to become the most frequent keyword.

We used CiteSpace software for burst detection to reveal the development trajectory and evolution trends of research themes and predict future research directions. Table 1 shows IPM' s burst terms, burst intensity, and start/end years from 2000-2020. The table shows 24 burst terms over the past 20 years. The research heat of "information retrieval" lasted from 2000 to 2012. Other burst terms with longer duration (over 5 years) include world wide web, retrieval, relevance, design, cocitation, and user. Current research hotspots and frontiers include ranking, graph, impact, knowledge, Twitter, sentiment analysis, word of mouth, and emotion. This largely matches previous analysis results: information retrieval has long been IPM' s main focus, and due to the inseparable relationship between search engine development and web page ranking algorithm updates and optimization, ranking algorithms remain a research frontier after sustained research heat. Additionally, thematic evolution shows characteristics of shifting from text to images and from information to knowledge. In terms of representation forms, information retrieval objects have gradually broken through text limitations to focus on multimedia information with richer content, which can convey more content features beyond basic text information for easier identification, retrieval, and mining. In terms of organization methods, knowledge is more structured and valuable than information. China' s *New Generation Artificial Intelligence Development Plan* also mentions the need to focus on breakthroughs in core technologies for knowledge processing, deep search, and visual interaction [80], implying that knowledge-based analysis, mining, and knowledge graph construction will become key future development directions in computer and artificial intelligence.

Social media and sentiment analysis are also important trends. The massive text information and behavioral data generated by users on social media provide necessary data foundations for sentiment analysis and user behavior research, enabling rapid development of related research under the application of computer and artificial intelligence technologies.

Journal special issues also reflect IPM' s key themes and their changes. Before 2010, special issue themes basically centered on information retrieval. For example, in the 2000 special issue "Web-based information retrieval research," the editors noted that Internet development expanded the scope of information retrieval research, and researchers began to gradually focus on web information retrieval and information retrieval system interaction [81]. Meanwhile, between 2000-2004, the ACM Special Interest Group on Information Retrieval (ACM SIGIR) consecutively held five workshops on mathematical/formal methods in information retrieval, demonstrating the importance of using mathematical and formal methods in information retrieval. IPM therefore selected relevant papers to form the special issue "Model design, formulation and explanation in information retrieval using mathematics."

The rapid development of the Internet has also brought explosive growth in information resources, and complex document information may contain many potentially valuable insights. When traditional text processing techniques and tools cannot meet new user needs, artificial intelligence-based text mining methods have emerged to effectively mine and utilize vast text resources [82]. Meanwhile, with the founding of Facebook and Twitter in 2004 and 2006 respectively, social media has become a convenient communication tool and powerful self-media platform, making user generated content the primary production method for network information resources. Information resources generated by users inevitably contain many personal viewpoints and emotions, and these subjective comments mostly involve social hot events or consumption evaluations of products/services. Sentiment analysis has therefore developed rapidly and been widely applied in public opinion monitoring and commercial marketing. IPM has also produced special issues on text mining and sentiment analysis, such as "Managing and mining multilingual documents," "Emotion and sentiment in social and expressive media," "Narrative extraction from texts," and "Mining actionable insights from social networks." Overall, IPM' s research themes are interconnected, and thematic evolution is closely related to network technology development and the rise of social media. Meanwhile, the emergence of emerging technologies such as artificial intelligence has also brought new opportunities to computer and information science.

Through keyword cloud and co-occurrence network mapping, we divided IPM's research themes over the past 20 years into three major categories: information retrieval, text analysis, and user research. In information retrieval, related research comprehensively discussed information retrieval theories and methods from information retrieval model construction to search engines and ranking algorithms, while advancing the semantic development of image retrieval technology. In text analysis, text mining is the main research direction, upon which sentiment analysis in social media has become a recent research hotspot, and knowledge research and analysis based on knowledge graphs have also seen continuous development and application. In user research, after the COVID-19 pandemic, research on health information seeking and rumor identification and propagation has received more attention, while research on personalization of service systems and human-computer interaction highlights the user-centered information service concept.

The thematic analysis reveals three main evolution characteristics: (1) consistently maintaining information retrieval as the core theme, with content retrieval based on information retrieval remaining a key focus of IPM; (2) shifting from text and information analysis to multimedia and knowledge analysis, with research objects expanding from text information to multimedia containing more content, and emerging technologies like artificial intelligence driving the upgrade from information analysis to knowledge analysis; and (3) in-depth analysis and mining of user emotions, with user generated content on social media driving user sentiment analysis and enabling deeper user research.

From the thematic analysis results, IPM papers focus on key issues in computer and information science, employing cutting-edge computer technologies and mathematical statistical methods that can reflect the international academic research and practical status of the field. However, using a single journal to reflect progress across an entire discipline still has limitations, and future research could conduct bibliometric studies and thematic analyses of more high-impact journals to more comprehensively grasp the evolution patterns of the discipline.

References

[1] Chu Jingli. High-end exchange platform construction requires innovation in academic exchange models[J]. Think Tank: Theory & Practice, 2021, 6(1): 7-9.
[2] DEHART F E. Monographic references and information science journal literature[J]. Information processing & management, 1992, 28(5): 629-635.
[3] TSAY M Y. A bibliometric analysis and comparison on three information science journals: JASIST, IPM, JOD, 1998–2008[J]. Scientometrics, 2011, 89(2): 591-606.
[4] Wang Yuefen, Jin Jialin. Comparative analysis of *Modern Library and Information Technology* near 2016(9): 1-16.

[5] Jiao Li. A review of information retrieval research in China[J]. Information Exploration, 2007(6):
[6] Wang Bin. Introduction to information retrieval[M]. Beijing: Posts & Telecom Press,
[7] SPARCK JONES K, WALKER S, ROBERTSON S E. A probabilistic model of information retrieval: development and comparative experiments Part 1[J]. Information processing & management, 2000, 36(6): 779-808.
[8] SPARCK JONES K, WALKER S, ROBERTSON S E. A probabilistic model of information retrieval: development and comparative experiments Part 2[J]. Information processing & management, 2000, 36(6): 809-840.
[9] ROBERTSON S E, SPARCK JONES K. Relevance weighting of search terms[J]. Journal of the American Society for Information Science, 1976, 27(3): 129-146.
[10] Gong Qingxiong. Research on entity-based information retrieval models[D]. Wuhan: Central China Normal University, 2020.
[11] XU Z B, AKELLA R. Improving probabilistic information retrieval by modeling burstiness of words[J]. Information processing & management, 2010, 46(2): 143-158.
[12] DAHAK F, BOUGHANEM M, BALLA A. A probabilistic model to exploit user expectations in XML information retrieval[J]. Information processing & management, 2017, 53(1): 87-105.
[13] TAI X Y, REN F, KITA K. An information retrieval model based on vector space method by supervised learning[J]. Information processing & management, 2002, 38(6): 749-
[14] GUO J F, FAN Y X, PANG L, et al. A Deep Look into neural ranking models for information retrieval[J]. Information processing & management, 2020, 57(6):
[15] VAUGHAN L. New measurements for search engine evaluation proposed and tested[J]. Information processing & management, 2004, 40(4): 677-691.
[16] CAN F, NURAY R, SEVDIK A B. Automatic performance evaluation of Web search engines[J]. Information processing & management, 2004, 40(3): 495-514.
[17] KIM H, SEO J Y. High-performance FAQ retrieval using an automatic clustering method of query logs[J]. Information processing & management, 2006, 42(3): 650-
[18] JUNG S, HERLOCKER J L, WEBSTER J. Click data as implicit relevance feedback in Web search[J]. Information processing & management, 2007, 43(3): 791-807.
[19] WANG J M, PAN M, HE T T, et al. A Pseudo-relevance feedback framework combining relevance matching and semantic matching for information retrieval[J]. Information processing & management, 2020, 57(6):
[20] BIDOKI A M Z, YAZDANI N. DistanceRank: an intelligent ranking algorithm for Web pages[J]. Information processing & management, 2008, 44(2): 877-
[21] Sun Junding, Yuan Fang. Content-based image retrieval technology[J]. Computer Systems & Applications, 2011, 20(8): 240-244.

[22] HUANG P W, DAI S K. Design of a two-stage content-based image retrieval system using texture similarity[J]. Information processing & management, 2004, 40(1): 81-
[23] LU T C, CHANG C C. Color image retrieval technique based on color features and image bitmap[J]. Information processing & management, 2007, 43(2): 461-472.
[24] Image retrieval: Content-based image retrieval technology[EB/OL].[2021-09-17].https: //www.cnblogs.com/king-lps/p/11407206. html.
[25] PANDEY S, KHANNA P, YOKOTA H. A semantics and image retrieval system for hierarchical image databases[J]. Information processing & management, 2016, 52(4): 571-
[26] Yuan Junpeng, Zhu Donghua, Li Yi, et al. Research progress in text mining technology[J]. Application Research of Computers, 2006(2): 1-4.
[27] Xiao Jianguo. On text mining and its applications[J]. Library Science Research,
[28] ALTINEL B, GANIZ M C. Semantic text classification: a survey of past and recent advances[J]. Information processing & management, 2018, 54(6): 1129-1153.
[29] ELNAGAR A, AL-DEBSI R, EINEA O. Arabic text classification using deep learning models[J]. Information processing & management, 2020, 57(1): 102121.
[30] MOHASSEB A, BADER-EL-DEN M, COCEA M. Question categorization and classification using grammar based approach[J]. Information processing & management, 2018, 54(6): 1228-1243.
[31] KASTRATI Z, IMRAN A S, YAYILGAN S Y. The impact of deep learning on document classification using semantically rich representations[J]. Information processing & management, 2019, 56(5): 1618-1632.
[32] HU G B, ZHOU S G, GUAN J H, et al. Towards effective document clustering: a constrained K-means based approach[J]. Information processing & management, 2008, 44(4): 1397-1409.
[33] CHEN C L, TSENG F S C, LIANG T. Mining fuzzy frequent item sets for hierarchical document clustering[J]. Information processing & management, 2010, 46(2): 193-
[34] ZHU S F, TAKIGAWA I, ZENG J. Field independent probabilistic approach for Web document clustering[J]. Information processing & management, 2009, 45(5): 555-570.
[35] FERSINI E, MESSINA E, ARCHETTI F. A probabilistic relational approach for Web document clustering[J]. Information processing & management, 2010, 46(2): 117-
[36] SOULIER L, TAMINE L, SHAH C. MineRank: leveraging users' latent roles for unsupervised collaborative information retrieval[J]. Information processing & management, 2016, 52(6): 1122-1141.
[37] KUCUKYILMAZ T, CAMBAZOGLU B B, AYKANAT C, et al. Chat mining: predicting user and message attributes in computer-mediated communication[J]. Information processing & management, 2008, 44(4):
[38] TSENG Y H, LIN C J, LIN Y I. Text mining techniques for patent

analysis[J]. Information processing & management, 2007, 43(5): 1216-1247.
[39] PONS-PORRATA A, BERLANGA-LLAVORI R, RUIZ-SHULCLOPER J. Topic discovery based on text mining techniques[J]. Information processing & management, 2007, 43(3): 752-768.
[40] Hong Jiangtao, Chen Liuyin, Huang Pei. Research on the impact of third-party review websites on catering enterprise brand image and consumer behavior −A case study of Dianping.com[J]. Finance & Trade Economics, 2013(10): 108-117.
[41] LIU B. Sentiment analysis and opinion mining[M]. San Rafael: Morgan & Claypool Publishers, 2012.
[42] Ma Li, Gong Yulong. A review of text sentiment analysis research[J]. Electronic Science and Technology, 2014, 27(11): 180-184.
[43] MOHAMMAD S M, ZHU X D, KIRITCHENKO S, et al. Sentiment, emotion, purpose, and style in electoral tweets[J]. Information processing & management, 2015, 51(4): 480-499.
[44] BALAHUR A, PEREA-ORTEGA J M. Sentiment analysis system adaptation for multilingual processing: the case of tweets[J]. Information processing & management, 2015, 51(4): 547-556.
[45] SEVERYN A, MOSCHITTI A, URYUPINA O, et al. Multi-lingual opinion mining on YouTube[J]. Information processing & management, 2016, 52(1): 46-60.
[46] AL-SMADI M, AL-AYYOUB M, JARARWEH Y, et al. Enhancing aspect-based sentiment analysis of Arabic Hotels' reviews using morphological, syntactic and semantic features[J]. Information processing & management, 2019, 56(2): 308-319.
[47] KUMAR A, SRINIVASAN K, CHENG W H. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data[J]. Information processing & management, 2020, 57(1): 102141.
[48] MAHMOOD Z, SAFDER I, NAWAB R M A, et al. Deep sentiments in Roman Urdu text using recurrent convolutional neural network model[J]. Information processing & management, 2020, 57(4): 102233.
[49] Qi Guilin, Gao Huan, Wu Tianxing. Research progress on knowledge graphs[J]. Intelligence Engineering, 2017, 3(1): 4-25.
[50] CHO H C, OKAZAKI N, MIWA M, et al. Named entity recognition with multiple segment representations[J]. Information processing & management, 2013, 49(4): 954-
[51] DERZYNSKI L, MAYNARD D, RIZZO G, et al. Analysis of named entity recognition and linking for tweets[J]. Information processing & management, 2015, 51(2): 32-49.
[52] TANG X, CHEN L, CUI J, et al. Knowledge representation learning with entity descriptions, hierarchical types, and textual relations[J]. Information processing & management, 2019, 56(3): 809-822.
[53] BOUNHAS I, SOUDANI N, SLIMANI Y. Building a morpho-semantic knowledge graph for Arabic information retrieval[J]. Information processing & management, 2020, 57(6): 102124.

[54] LI D F, MADDEN A. Cascade embedding model for knowledge graph inference and retrieval[J]. Information processing & management, 2019, 56(6): 102093.
[55] SHIN S, JIN X, JUNG J, et al. Predicate constraints based question answering over knowledge graph[J]. Information processing & management, 2019, 56(3): 445-462.
[56] JANSSENS F, LETA J, GLANZEL W, et al. Towards mapping library and information science[J]. Information processing & management, 2006, 42(6): 1614-1642.
[57] Zhang Yihan, Yuan Qinjian. Research progress on user information behavior in China[J]. Journal of the National Library of China, 2014, 23(6): 91-98.
[58] Jin Ronglin. Exploring influencing factors of information seeking behavior in college students' autonomous learning context[D]. Baoding: Hebei University, 2019.
[59] MAKRI S, BLANDFORD A, COX A L. Investigating the information-seeking behaviour of academic lawyers: from Ellis' s model to design[J]. Information processing & management, 2008, 44(2): 613-634.
[60] JAMALI H R, NICHOLAS D. Interdisciplinarity and the information-seeking behavior of scientists[J]. Information processing & management, 2010, 46(2): 233-243.
[61] LYKKE M, PRICE S, DELCAMBRE L. How doctors search: a study of query behavior and the impact on search results[J]. Information processing & management, 2012, 48(6): 1151-1170.
[62] PIAN W J, SONG S J, ZHANG Y. Consumer health information needs: a systematic review of measures[J]. Information processing & management, 2020, 57(2):
[63] ZHANG X F, GUO F, XU T X, et al. What motivates physicians to share free health information on online health platforms?[J]. Information processing & management, 2020, 57(2): 102166.
[64] HUVILA I, ENWALD H, ERIKSSON-BACKA K, et al. Anticipating ageing: older adults reading their medical records[J]. Information processing & management, 2018, 54(3): 394-407.
[65] JOHNSON J D. Health-related information seeking: is it worth it?[J]. Information processing & management, 2014, 50(5): 708-717.
[66] Zhao Yuxiang, Fan Zhe, Zhu Qinghua. Conceptual analysis and research progress of user generated content (UGC)[J]. Journal of Library Science in China, 2012, 38(5): 68-81.
[67] GE Y D, QIU J N, LIU Z Y, et al. Beyond negative and positive: exploring the effects of emotions in social media during the stock market crash[J]. Information processing & management, 2020, 57(4): 102218.
[68] LI L F, WANG Z Q, ZHANG Q P, et al. Effect of anger, anxiety, and sadness on the propagation scale of social media posts after natural disasters[J]. Information processing & management, 2020, 57(6): 102313.
[69] LIU Y H, JIN X L, SHEN H W. Towards early identification of online rumors based on long short-term memory networks[J]. Information processing

& management, 2019, 56(4): 1457-1467.
[70] ALKHODAIR S A, DING S H H, FUNG B C M, et al. Detecting breaking news rumors of emerging topics in social media[J]. Information processing & management, 2020, 57(2): 102018.
[71] WANG J, CLEMENTS M, YANG J, et al. Personalization of tagging systems[J]. Information processing & management, 2010, 46(1): 58-70.
[72] BELEM F M, MARTINS E F, ALMEIDA J M, et al. Personalized and object-centered tag recommendation methods for Web 2.0 applications[J]. Information processing & management, 2014, 50(4): 524-553.
[73] RENJITH S, SREEKUMAR A, JATHAVEDAN M. An extensive study on the evolution of context-aware personalized travel recommender systems[J]. Information processing & management, 2020, 57(1): 102078.
[74] WANG P L, HAWK W B, TENOPIR C. Users' interaction with World Wide Web resources: an exploratory study using a holistic approach[J]. Information processing & management, 2000, 36(2): 229-251.
[75] KUMARAN G, ALLAN J. Adapting information retrieval systems to user queries[J]. Information processing & management, 2008, 44(6): 1838-1862.
[76] LOPATOVSKA I, ARAPAKIS I. Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction[J]. Information processing & management, 2011, 47(4): 575-592.
[77] COLE M J, GWIZDKA J, LIU C, et al. Inferring user knowledge level from eye movement patterns[J]. Information processing & management, 2013, 49(5):
[78] CLARK M, RUTHVEN I, HOLT P O, et al. You have e-mail, what happens next? tracking the eyes for genre[J]. Information processing & management, 2014, 50(1): 175-198.
[79] HILBERINK-SCHULPEN B, NEDERSTIGT U, VAN MEURS F, et al. Does the use of a foreign language influence attention and genre-specific viewing patterns for job advertisements? an eye-tracking study[J]. Information processing & management, 2016, 52(6): 1018-1030.
[80] State Council of the People' s Republic of China. Notice of the State Council on Issuing the New Generation Artificial Intelligence Development Plan[EB/OL].[2021-10-26]. http://www.gov.cn/zhengce/content/2017-07/20/content_{5211996}.htm.
[81] SPINK A, QIN J. Introduction to the special issue on Web-based information retrieval research[J]. Information processing & management, 2000, 36(2): 205-206.
[82] Guo Fei. Discussion and application of text mining methods[D]. Chengdu: Chengdu University of Technology, 2006.

Author Contributions:
Li Hanxiao: Data collection, initial draft writing
Du Xingye: Topic proposal, paper revision and finalization

Note: Figure translations are in progress. See original paper for figures.