International Journal of Information Science and Management
Vol. 21, No. 3, 2023, 363-382
DOI: https://doi.org/10.22034/ijism.2023.1977962.0 / DOR: https://dorl.net/dor/20.1001.1.20088302.2023.21.3.21.5

Original Research

Heuristical Research on Twelve Decades of Information and Knowledge Creation Utilizing Python and NLP

Farshid Danesh
Assistant Prof., Research Department of Information Management, Islamic World Science & Technology Monitoring and Citation Institute (ISC), Shiraz, Iran.
farshiddanesh@ricest.ac.ir
ORCID iD: https://orcid.org/0000-0001-5481-3988
Scopus Author ID: 36349853500

Meisam Dastani
PhD in Knowledge and Information Science (Information Retrieval), Director of Statistics and Information Technology, Gonabad University of Medical Sciences, Gonbad, Iran.
Corresponding Author: meisam.dastani@gmail.com
ORCID iD: https://orcid.org/0000-0002-5631-539X

Received: 26 September 2022
Accepted: 19 April 2023

Abstract

Publications on knowledge and information creation have grown significantly due to their importance in information and knowledge management. This study aims to discover and analyze the hidden thematic topics of information and knowledge creation publications. The research applied was performed using text mining techniques and an analytical approach. The research population comprises publications on knowledge and information creation from 1900 to 2021, retrieved from the Web of Science Core Collection (WOSCC). The data were analyzed by Latent Dirichlet Allocation (LDA) algorithm and Python Programming Language. Forty-eight thousand two hundred sixty-five documents were retrieved and analyzed. "Data production," "Health seeking behavior," "Human Brain and Information processing," "Decision-making models," "Knowledge production," "Information needs," and "Digital Literacy" are among the essential topics in order of publication rate. The results indicated that the spectrum of the fourteen topics covered a variety of dimensions, including "data and knowledge creation," "information processing," "information needs and behavior," "digital literacy," and "critical thinking." The study's findings have shown the conceptual relationships between textual data and the presentation of the knowledge structure of information and knowledge creation. Based on this, it can be concluded that the creation of knowledge and information includes human mental and behavioral processes concerning knowledge.

Keywords: Knowledge Creation, Information Creation, Data Creation, Data Mining, Text Mining, Topic Modeling, Latent Dirichlet Allocation (LDA).

Introduction

Human beings have sought, organized, and used information for millennia as they evolved and developed patterns of human information behaviors to assist them in resolving human problems and ensuring their continued survival. The first step towards searching, organizing, and using information (or, indeed, foraging for, exchanging, interpreting, or making sense of information) is to create it (Trace, 2007; Kari, 2011). Information-science scholars have recently become more interested in the scientific study and production of information and

knowledge creation (Gorichanaz, 2019; Koh, 2013). Information creation can be considered a part of the information management cycle and an aspect of information behavior that describes how human beings interact with information (Bates, 2010). Koh (2013) also indicated that information-creating behavior is a technique in which people create messages and content to fulfill information requirements (Linde, Snodgrass & Kerne, 2014; Rosenbaum, 2011). Information creation possesses various dimensions so that people use information practically in their lives and use the produced data to advance their everyday lives (Lee, Ocepek & Makri, 2021). In other words, information creation is closely related to daily life activities in the information era (Whittaker, 2011). Information science has begun to focus more on creating information in the past decade rather than just organizing and retrieving it (Kari, 2007; Koh, 2013). Information creation is conceptually part of information behavior, which aims to explain how humans interact with information (Bates, 2010). In the past decade, research on information creation has become increasingly popular in information sciences and beyond. There are several theoretical models and frameworks in addressing information creation from different perspectives.

Information creation has long been acknowledged as part of the information life cycle and the totality of human information behavior (Bates, 2015). For instance, the significance of information creation is represented to differing extents in several classic models of information activities and processes (Gorichanaz, 2019). Moreover, calls for research on information creation in information science and knowledge creation in knowledge management date back to the 1960s (Huvila, Douglas, Gorichanaz, Koh & Suorsa, 2022). Still, there has been little explicit attention to information creation in research and practice (Huvila, 2011; Woxland, Cochran, Davis, & Lundstrom, 2017). In recent years, a new body of work has explored different information creation aspects. At the same time, information creation has been acknowledged as a critical aspect of information literacy. Scholars and practitioners increasingly recognize how the usefulness and relevance of the information being sought and retrieved depend on the conditions and process of its creation.

Over the past decade, information science has become interested in exploring the creation of information rather than just its organization and retrieval (Kari, 2007; Koh, 2013). Conceptually, the information creation area falls within information behavior, which strives to account for how humans relate to information in their lives (Bates, 2010). Research on information creation has become increasingly popular over the past decade, both in and beyond information sciences. Among theoretical models and frameworks that address information creation from different perspectives. As a result, a more comprehensive perspective of information creation can be provided with a correct understanding of literature and scientific publications in this field, which leads researchers to study information phenomena from a new perspective.

Accordingly, Lee et al. (2021) showed the information creation process and its application and impact on everyday life in a conceptual model framework (Figure 1). This model showed the human understanding of the information creation process. Based on this, the emphasis is on information's lively and distinctive character beyond its practical dimension. Information creation also brings entertainment and pleasure to people during the process.

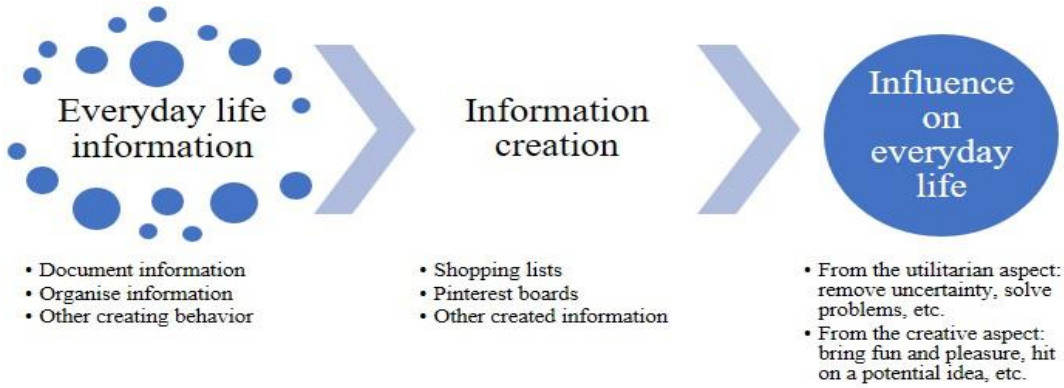

Figure 1: Information Creation in Everyday Life (Lee et al., 2021)

Therefore, it can be noted that knowledge and information creation focus on the entire process of how data is produced, constructed, or produced and covers a wide range of activities such as knowledge creation, knowledge production, records creation, information making, documentation, and document creation, information production and records creation, information-creating behavior, and making and maker spaces. The quantity of international articles published in prestigious journals on knowledge and information creation has significantly increased due to the expanse and diversity of the subject area of knowledge and information creation and the necessity of researching to solve various problems in this field. The dramatic increase in the literature on knowledge and information creation and the strategic nature of this subject area makes it more necessary to extract the hidden knowledge in knowledge creation publications automatically. Among these, natural language processing and textual data mining techniques are the methods of automatic knowledge extraction from a massive amount of text data (Fan, Wallace, Rich & Zhang., 2006). As scientific, reliable, and accepted techniques, text mining techniques automatically extract knowledge from a large amount of textual Data (Halevy, Norvig & Pereira, 2009).

According to the issues raised and the importance of identifying topics in information creation, this study uses text mining techniques and topic modelling algorithms to explore the hidden issues in knowledge and information creation publications. Consequently, the present study responded to the following questions:

RQ1. On what topics are knowledge and information creation publications?

RQ2. What is the percentage of publications on each topic in knowledge and information creation publications?

RQ3. How are the knowledge and information creation topics overlapping and similar?

RQ4. What are the trends in publishing knowledge and information creation topics over time?

Literature Review

In the last decade, the use of text mining techniques to extract knowledge from published literature has increased in various topic areas, such as the analysis of articles in the Journal of Educational Technology (Chen, Zou & Xie, 2020), Asset Management (Jung & Kim, 2021), Coronavirus and COVID-19 (Anderson, 2021; Danesh, Dastani & Ghorbani, 2021; Dastani & Danesh, 2021) and particularly library and information science. In this regard, Park and Song

(2013) identified the topic trend in library and information science in Korea using text mining techniques and LDA topic modeling. This study's results have shown that information science and library services are concentrated among these sub-topics. Also, there is an increasing trend in research topics such as services and evaluation based on library type, Internet, and metadata. However, research topics such as books, classification, and cataloging have shown a downward trend. (Park & Song, 2013).

Yan (2014) provided a topic analysis of the publications of scientific journals on library and information science publications during the 11 years indexed in WOS from 2001 to 2011. The results of this study revealed that the topics "web information retrieval," "citation and bibliometrics," "system and technology," and "health science" has the highest average popularity (Yan, 2014). In a follow-up to a previous study, Yan (2015) identified library and information science publication topics, including "related to online technologies," "informetrics," "information retrieval systems," "health communication and informatics," and "online social networks," on the WOS database during 1955-2013 using Latent Dirichlet Allocation (LDA) topic modelling. Following Yan's research, another study assessed topics published in information retrieval using a topic modelling algorithm. The results of this article identified the topics "User study," "image retrieval," "database querying," "query processing," and "text retrieval" (Chen, Tsutsui, Ding & Ma, 2017). In 2017, another study analyzed 92075 titles and abstracts of library and information articles indexed in LISA from 1978 to 2014. The Latent Dirichlet Allocation (LDA) algorithm was used to cluster the articles. This study also identified nineteen topical clusters divided into four main categories: processes, information technology, library, and information application (Figuerola, García Marco & Pinto, 2017). In line with previous studies, topic modelling was performed in another article using the LDA Library and Information Science (LIS) publications, in which thirty topics were identified and labelled. Conventional techniques and research processes were obtained by drawing a topical map (Kurata, Miyata, Ishita, Yamamoto, Yang & Iwase, 2018). Lamba and Madhusudhan (2019) topically categorized DESIDOC Journal of Library & Information Technology articles with text mining techniques. The statistical population consisted of the full text of 928 articles from 1981 to 2018. The results identified 50 major topics in four decades, of which only twenty-six were unique (Lamba & Madhusudhan, 2019). Consistent with previous literature, in another study, the Library and Information Science topics were identified using LDA in 2000-2002 and 2015-2017, in which 30 topics were identified for the research population (Miyata Ishita, Yang, Yamamoto, Iwase & Kurata. 2020).

In another study, Han (2020) has shown library and information science evolution in 5 periods from 1996 to 2019 using text mining techniques and LDA thematic modeling algorithm. According to the results of this study, library science has become less prevalent over time, as no top topic clusters relevant to library issues have been identified in the period 2000-2005. In bibliometrics, particularly citation analysis, there has been overall stability over the past five decades, as indicated by stable sub-clusters and consistent keywords; and (c) information retrieval has remained the predominant domain, but interests are gradually shifting to model-based text processing. There is also a stable field of information-seeking behavior known for being dispersed across several subjects rather than being presented as a separate discipline. E-commerce has developed a closer relationship with information systems and organizational activities. A technology change has brought about the re-occurrence of topics that occurred only once, such as the Internet, social media, and mobile applications.

In the same year, the topics of Iranian articles indexed in Scopus were identified in another study, in which topics were divided into two groups with high and low popularity. Topics with high popularity were: "library services on social networks," "research models," "social capital," "medical databases," "data mining," "scientific production process," "interdisciplinary topics," "cyberspace algorithms," "knowledge management," "social network studies," "research approaches" and "futures studies." In addition, topics with low popularity were: "electronic resources," "information management systems," "search engines," "lending services," "remote services," "e-learning," "e-government," "indices for journal evaluation," "evaluation of web resources," and "digital libraries" (Baghmohammad, Mansouri & CheshmehSohrabi, 2021). The topics of articles in Medical Librarianship and Information Science (MLIS) in Iran were analyzed in another study, the results of which identified scientometrics, information literacy, health information, knowledge management, webometrics, and the quality of the website and hospital information systems for these publications (Dastani, Mousavi Chelak, Ziaei & Delghandi, 2020).

Due to the thematic overlap of information creation with LIS and MLIS, a literature review was done in these areas. The literature review indicated that most of the research conducted using text-mining techniques has studied the subject area of librarianship and information. Also, text mining analyses have been performed on scientific texts published in specific and limited periods. In most of the studies conducted with LDA, the most important topics and their time trends are shown. Despite all the searches, no study was observed directly addressing the topic of knowledge and information creation and all its related aspects over more than a century (1900-2021). Therefore, this article is innovative and original in its findings, and this research is required more than ever.

Materials and Methods

Data Collection

This study is applied research using text mining techniques, Natural Language Processing, and an LDA topic modeling algorithm with a heuristical approach. The research data was obtained using the advanced search of the Web of Science Core Collection (WOSCC), and 48265 documents related to the subject area of knowledge and information creation were retrieved from 1900 to 15 December 2021. Then, the titles and abstracts of the articles were extracted into an Excel file to perform the text-mining process.

The documents indexed in WOSCC in the form of articles are original articles taken from original research methods (experiments, observations, surveys, interviews, and questionnaires) and show the results of these researches (Arrom, Huguet, Errando, Breda & Palou, 2018). The most critical concern of the methodology section was the search strategy, which can provide the most relevant information creation articles with a maximum retrieval rate for further analysis. In this part of the research, we asked twenty faculty members with the academic rank of associate professor and professor experts in knowledge creation and information and knowledge management to prepare the key and essential words and phrases they wanted and send them to us via email. In the second stage, after performing the necessary preprocessing on the words. We prepared a list and sent it to three expert faculty members with the academic rank of professor. Finally, subject experts approved thirty-four keywords and critical phrases for advanced search in WOSCC (Figure 2). The search strategy used advanced search, the "TI" tag, and the "OR" operator to explore and retrieve knowledge and information creation

publications.

TI=(Digital Literacy OR Seeking behavior OR Research data creation OR Record making OR Record production OR Knowledge making OR Record creation OR knowledge production OR Knowledge creation Theory OR Knowledge creation process OR Knowledge creation OR Knowledge creation model OR Information searching OR Information production OR Information process OR information need OR information making OR Information literacy OR Information dissemination OR Information creation Theory OR Information creation research OR Information creation process OR Information creation model OR Information creation implication OR information creation OR information behavior OR document theory OR Document making OR Document creation OR data production OR data making OR Critical thinking OR Creation of Information OR Data creation Critical thinking)

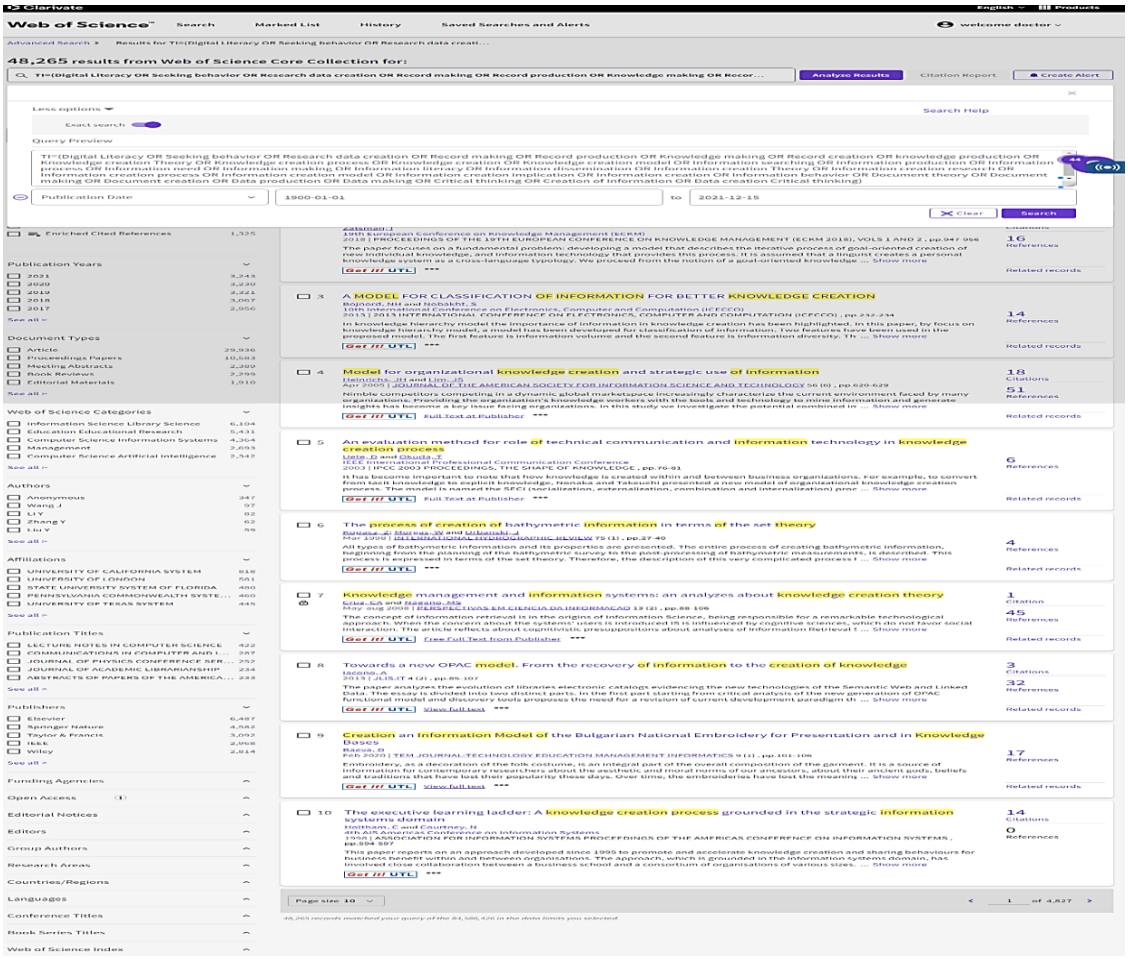

Figure 2: Findings of an Advanced Search Using the Keywords in the Search Query (Screenshot taken from WOSCC on 20 July 2022)

Text Mining and Topic Modeling

In the following, the title and abstract fields of the articles were merged to perform the text mining process, and preprocessing operations were applied to these texts. Preprocessing and data cleaning was applied to the study data to increase the data quality, the validity of the patterns, and the extracted relations. It should be noted that data cleansing retains only the required relevant textual data and removes irrelevant data to continue working (Indurkhya & Damerau, 2010). These preprocessing texts include word tokenization, lowercase conversion

for text homogenization, synchronization of synonymous words, lemmatization (conversion of words into their basic form), and elimination of stop words.

It should be noted that the preprocessing operation was performed automatically and using commands in the Python programming language on the texts of scientific articles in the information and knowledge creation field. After the preprocessing operation, texts were analyzed and clustered using the LDA algorithm.

Topic modeling is one of the most widely used text mining techniques, identifying the topics from documents (Kobayashi, Mol, Berkers, Kismihók & Den Hartog, 2018). Topic modeling algorithms are statistical methods that analyze words in a text and thus extract topics from textual data. They also determine how these topics relate to each other and how they alter over time. Topic modeling aims to analyze documents with different topics and obtain probable topics from the distribution of words in documents (Alghamdi & Alfalqi, 2015). This technique can significantly identify the conceptual relations underlying textual data and present their knowledge structure (Hannigan, et al, 2019; Schmiedel, Müller & Vom Brocke, 2019).

LDA, the most common algorithm for topic modeling (Blei, Ng & Jordan, 2003; Jelodar et al., 2019), was used for topic modeling in this paper. This algorithm assumes an unstructured set of documents comprising a limited number of topics with different ratios. Each topic is a probability distribution on a fixed set of words. Accordingly, the LDA algorithm analyzes the words in the set of studied texts to determine the salient topics and their distribution in each document (Blei, 2012). The LDA is one of the most extensively used algorithms and effectively identifies related semantic topics in literature (ibid).

The limitation in clustering using LDA is not determining the number of optimal topics for clustering articles. Therefore, solutions have been provided in past studies to overcome this limitation. "CV Coherence" is a suitable solution to calculate and determine the optimal number of topics. "CV Coherence" essentially measures the co-occurrence of the words extracted by the topic model. If those words from the same topic co-occur very often (i.e., the CV Coherence is high), the model performs well (Wang & Mengoni, 2020). According to the literature, the CV Coherence measure has shown good performance in determining the number of topics and is closely related to human judgments about topic interpretation (Röder, Both & Hinneburg, 2015; Syed & Spruit, 2017). Figure 3 shows the CV Coherence value to select the number of topics.

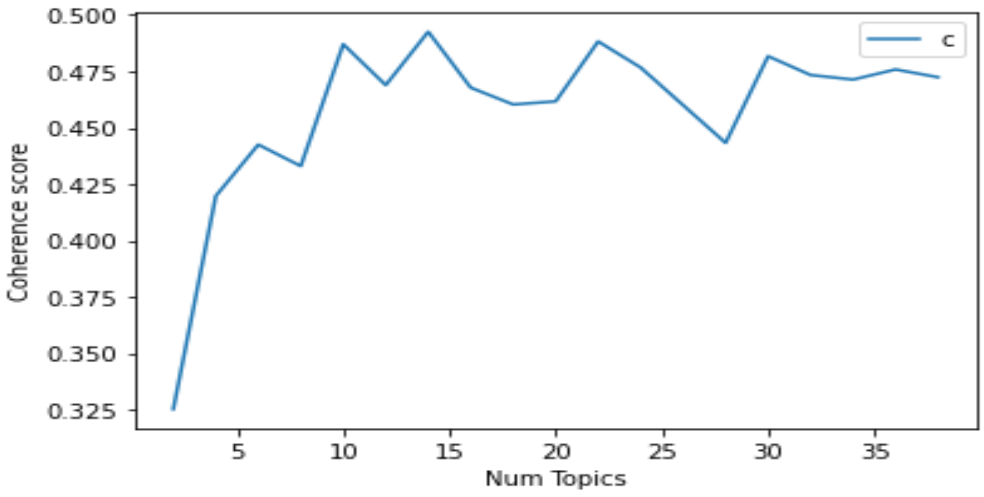

Figure 3: Value of CV Coherence in Determining the Number of Different Subjects

As indicated in Figure 3, the *CV* Coherence value is between two and forty topics, and the highest *CV* Coherence value is related to fourteen topics. Therefore, the present study identified fourteen topics as appropriate for topic modeling. After determining the rank of different subjects, we selected the highest subject rank and carried out the subject modeling process. Then, we categorized the articles manually and examined each subject in a specialized and technical manner. Finally, the selected model was approved for illustration and interpretation. It is worth noting that the topics were labeled after checking the keywords and articles assigned to each topic.

Visualization Topics

Next, the topical map of the obtained model based on LDA was drawn using the LDA visualization package (pyLDAvis python library) to show the relationships and overlap between the topics (Fiandrino & Tonelli, 2021). The LDAvis tool shows the coverage of an issue in the overall transcripts and the identified topic as part of comparative visual analytics. After successfully tuning the model, we display the inferred topic results in an Intertopic Distance Map (Katre, 2019; Sievert & Shirley, 2014). Python programming language and Pandas, Gensim, NLTK, and Spacy libraries performed the above steps. Python is an open-source programming language with a simple syntax and is compact and versatile. It is also easy to develop and provides various libraries to users for working with texts (Řehůřek & Sojka, 2010). Figure 4 shows the steps of the study methodology:

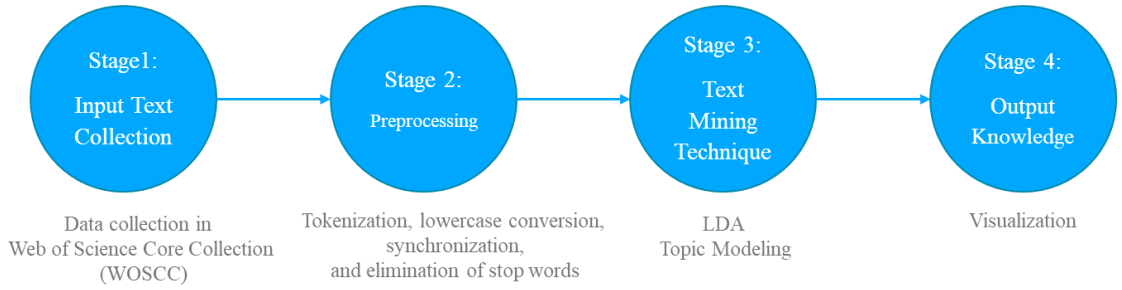

Figure 4: Steps of the Study Methodology

Results

Topics of knowledge and information creation's publications

Forty-eight thousand two hundred sixty-five documents on knowledge and information creation were retrieved from 1900 to 15 December 2021 from the WOSCC using the advanced search strategy listed in the methodology section. In general, the words cloud obtained from the reviewed articles is shown in Figure 5.

Figure 5: Information and knowledge creation word cloud

Figure 5 shows the frequently repeated words in the information and knowledge creation articles in a word cloud format. The frequently repeated words are offered in a larger size.

The results of the topic modeling of knowledge and information creation publications are shown in Table 1. The characteristics of each topic, including the title selected for each topic, and the most important and most extensively used words in each topic, are presented.

Table 1

The Topics Extracted from the Implementation of the Topic Modeling Algorithm

No.	Topic	the highest important words
1	Data production	production, datum, chemical, solution, analysis, make, article, animal, cost, high
2	Health seeking behavior	record, health, patient, report, medical, need, make, regard, care, clinical
3	The human brain and Information processing	information, processing, process, effect, visual, task, human, stimulus, response, memory
4	Decision-making models	information, process, system, make, decision, problem, management, regard, model, concern
5	knowledge production	knowledge, production, acid, scientific, creation, development, theory, recording, science, new
6	Information needs	information, search, need, behavior, computer, system, source, user, analysis, electronic
7	Digital Literacy	technical, library, school, program, student, digital, university, read, course, education
8	Critical thinking	critical, thinking, learn, think, student, skill, test, teach, thought, ability
9	Consumer behavior	behavior, effect, influence, choice, theory, consumer, relationship, factor, attitude, market
10	Information seeking	behavior, seek, drug, effect, rat, utilization, behavior, help,

No.	Topic	the highest important words
	behavior	subject, light
11	computer information processing	model, processing, process, control, system, application, optical, time, machine, spatial
12	social networks	dissemination, social, network, medium, determinant, attribute, mass, cognition, evolution, diffusion
13	Information supply	child, industry, estimate, plan, patent, agricultural, economic, age, technological, regional
14	Information production	language, risk, man, laboratory, cue, woman, reduction, film, financial, investment

Based on the data in Table 1, there are fourteen topics, which are: "Data production," "Health seeking behavior," "Human brain and Information processing," "Decision-making models," "knowledge production," "Information needs," "Digital Literacy," "Critical thinking," "Consumer behavior," "Information seeking behavior," "computer information processing," "social networks," "Information supply" and "Information production."

Publication percentage of knowledge and information creation topics

As shown in Figure 6, the most significant contributions of publications on knowledge and information creation were related to "Data production" with 18.55%, "Health seeking behavior" with 14.61%, and "Human brain and Information processing" with 12.67%. Moreover, the topics "social networks" with 0.17%, Information supply with 0.08%, and "Information production" with 0.04% have the lowest contribution to knowledge and information creation publications.

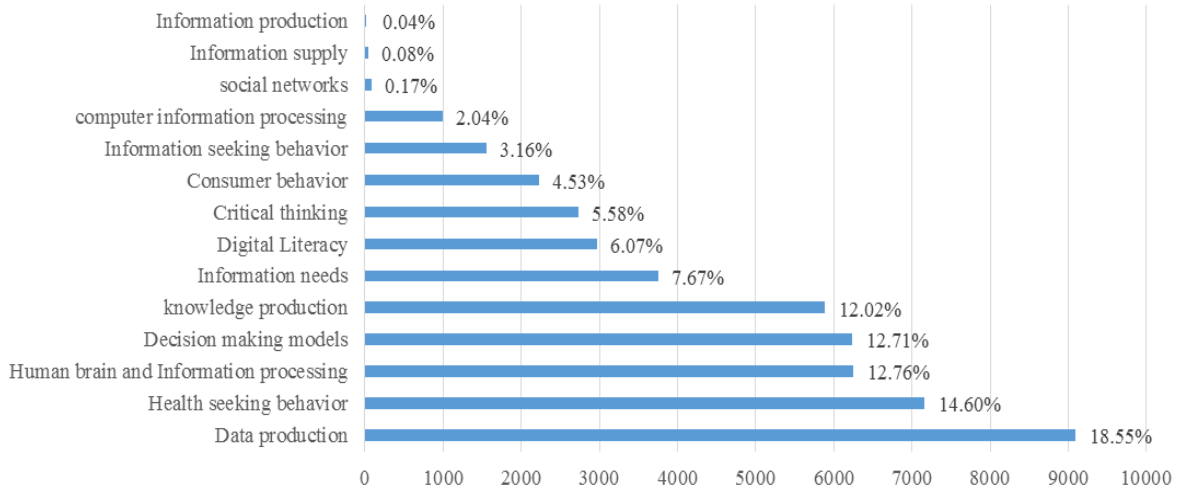

Figure 6: Publication Percentage of Knowledge and Information Creation Topics (WOSCC, 1900-2021)

Knowledge and Information creation topics in terms of overlapping and similarity

The topics of data production or the topical distance map are shown in Figure 7, where each topic is indicated in a circle. The size of each of these circles indicates the publication rate of the relevant topic, among other topics. Also, the circles that all or part of is inside other circles show the degree of similarity of that topic with other topics. The following topics have

a high overlap:

3. Decision-making models; 5. The human brain and Information processing, and 6. Information needs; 2. Health-seeking behavior, and 4. Knowledge production; 8. Information-seeking behavior, 10. Digital Literacy challenges and opportunities and 11. Critical thinking; 12. Information supply, 13. Social networks, and 14. Information production.

Also, the Computer information processing topic, i.e., 9 is more distant from the other topics in Figure 7; it does not overlap or resemble other subjects.

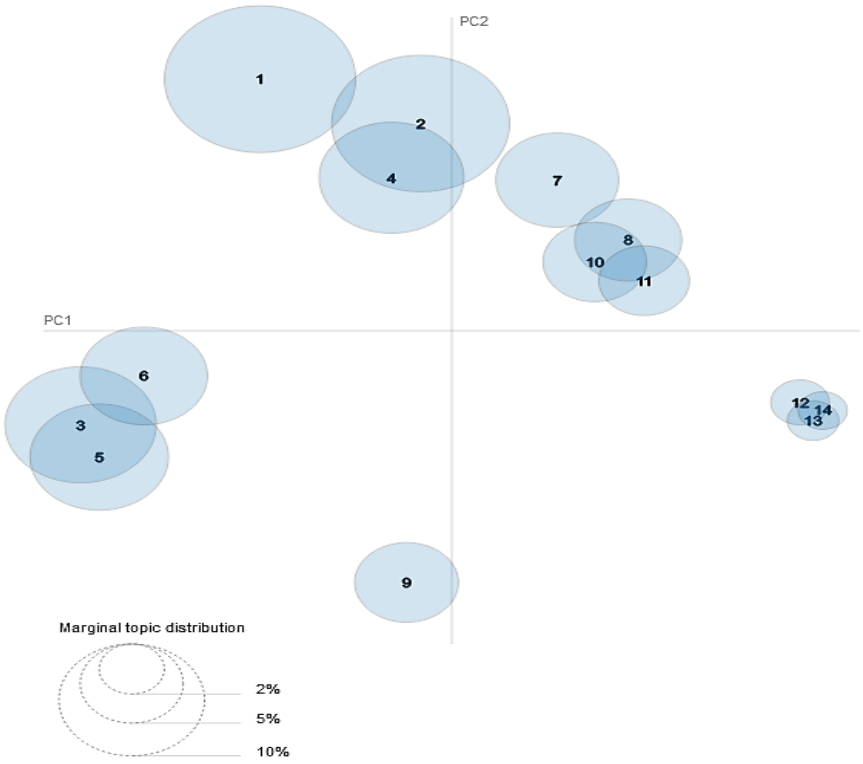

Figure 7: Knowledge and Information Creation Topic Modeling Map

Trends in the publishing of knowledge and information creation topics

Figure 8 illustrates the process of knowledge and information creation publications in each of the fourteen topics from 1900 to 2021. The horizontal axis of this figure shows the year of publication, while the vertical axis shows the topic's name. In addition, the color of each cell indicates the number of publications on each topic, in which greenish and yellowish colors indicate more and fewer publications per year, respectively. Therefore, the relationship between each topic and its publishing trend each year has been determined.

According to the data in Figure 8, the publications of Topics "Data production" and "Health seeking behavior" have grown significantly in recent years, particularly since 2005, indicating a significant growth of these topics compared to other topics. Furthermore, the topics "Human brain and Information processing" and "knowledge production" from 1965 onwards have had a mild and uniform green color, suggesting a steady trend in the publication of these topics, and also had no significant increase or decrease in the publication rate since 1965. In general, other

topics of information creation have been growing in recent years; however, the publications of Topics "Critical thinking," "social networks," "Information supply," and "Information production" have been accompanied by a slight growth but have not grown significantly compared to other topics.

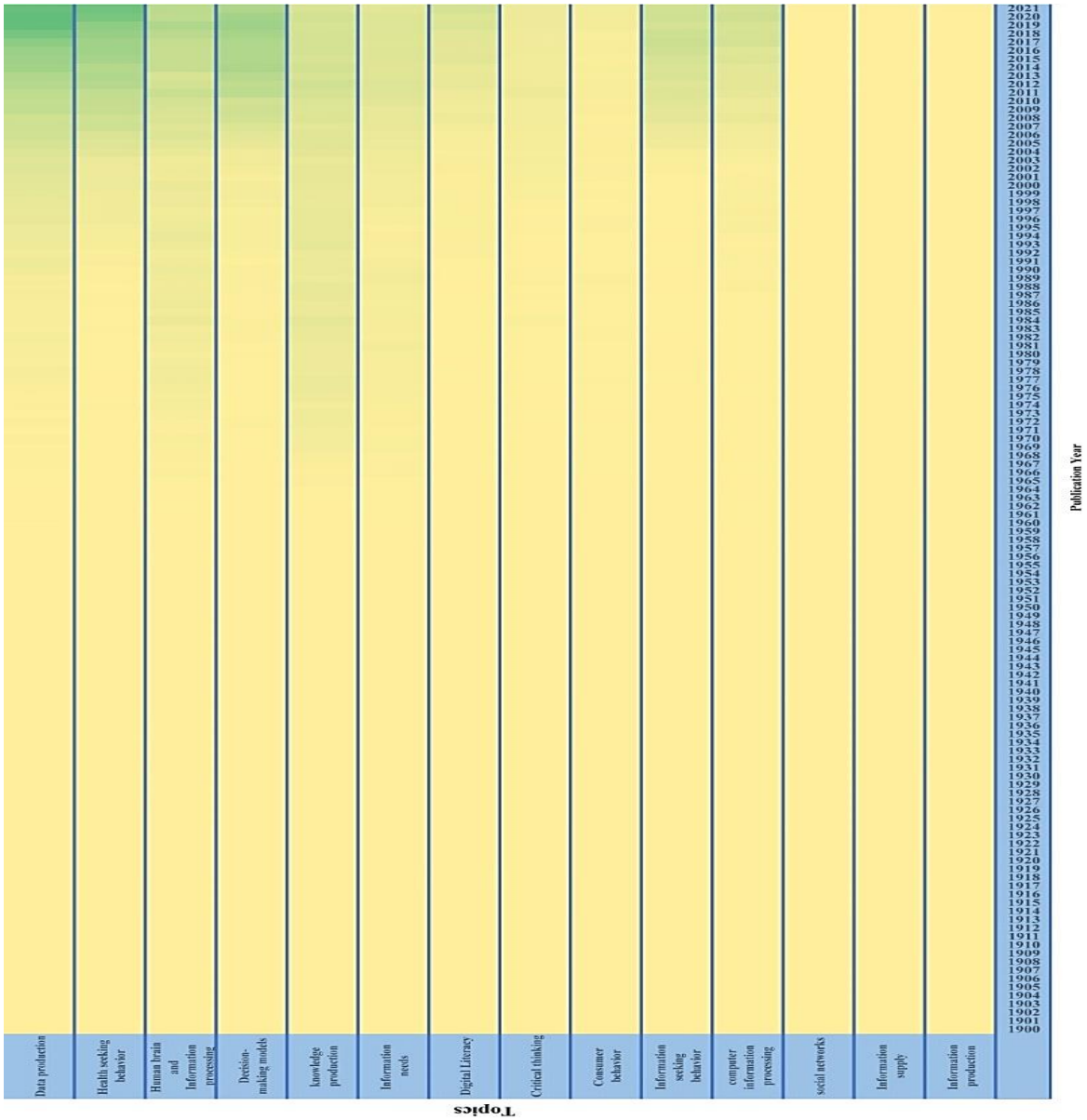

Figure 8: Publishing Trends of Knowledge and Information Creation Topics from 1900 to 2021

Discussion

The present study showed an overview of knowledge and information creation topics, providing fourteen topics for knowledge and information creation articles from 1900 to 2021. Accordingly, the topic "Data production" has had the highest quantity of knowledge and information creation publications with significant publishing growth in recent years. The term "data production" refers to the production of data with the intention of its use by others, so "data production" can involve several activities from data generation to data sharing, including its description, management, packaging, archive, and access (Baker & Mayernik, 2020). The

growth and development of modern information technology tools can be one of the reasons for the growth of this topic in recent years, which has led to the production, sharing, and easy accessibility of data. The data volume and diversity increase, and access to their platforms, computers, software, and applications have resulted in new data-related issues.

The topics "health-seeking behavior" and "knowledge production" are other important topics that have been identified, and the results also show that these topics mostly overlap. Their substantial quantitative growth has begun since 2005. Health-seeking behavior is a set of individual activities to address information requirements for physical problems or diseases (Broussard & Doty, 2016). It also includes medical and health information retrieval measures for people at risk for disease or symptoms (Huang et al., 2019). Acquisition of such information modifies the individual's knowledge level and effectively makes the right decisions related to the individual's health (Cutrona, Mazor, Vieux, Luger, Volkman & Finney Rutten, 2015; Wilson, 2009). In this regard, the World Health Organization (WHO) during the COVID-19 pandemic indicated that searching behavior for the correct health information helps reduce anxiety and distress among people (World Health Organization, 2021). As the use of new information technologies and the web has become an inevitable part of people's everyday lives, this can be the main reason for the growth of this topic in recent years. In general, using reliable sources is a principle in knowledge production. Acquiring, recording, and storing information is a secondary function in creating knowledge (Goldman & Scardamalia, 2013).

The information-searching process is a planned and purposeful knowledge acquisition process primarily performed by a person with insufficient required knowledge (Griffin, Dunwoody & Neuwirth, 1999). These information technologies also provide many opportunities to acquire information and knowledge to meet individual requirements (Zhang *et al.*, 2019), which can be one of the reasons for the relationship and overlap between the topics of "Health seeking behavior" and "knowledge production." "Human brain," "Information processing," and "Decision-making models" are topics in the publications on information creation. These two topics significantly overlap and are considered old topics in this field, so a steady trend has accompanied the growth of these two topics from 1965 onwards. The main capabilities of the human brain are the storage of large volumes of information and its high processing power (Rousselet, Thorpe & Fabre-Thorpe, 2004; Von Neumann, 2012). After gaining information, people process and evaluate it, then use it to meet their needs. In other words, searching, processing, and using data are rotational behaviors because patients always find, evaluate, and employ data to alleviate their pains and recover (Preum, Clark, Davis, Khutsishvilli & Valdez, 2019).

Information processing has a fundamental role in making decisions regarding the use of information. The information processing involved in visual search culminates in a target detection decision, which sometimes may be as important as the search operations that preceded it (Wickens & Carswell, 2021). Accordingly, Cao, Duan and Li (2015) indicated that the ability to process information is to capture, integrate, and analyze information and utilize it in decision-making. Moreover, some studies reported that information processing capability influences decision-making effectiveness (Cao, Duan & Cadden, 2019; Cao *et al.*, 2015; Ransbotham, Kiron & Prentice, 2016).

"Information need" is another topic identified in knowledge and information creation publications, which means the need to utilize information services or resources. It is also a subjective process, i.e., the question in the mind that makes the person think and strive to find

the answer, and the person moves in the process of finding the answer to this inner need (Schmiedel et al., 2019). "Information need" also overlaps with "Decision-making models," "Human brain," and "Information processing," and there are similarities among them, which can be due to the subjective process of this topic in recognizing, selecting, and the use of information.

It would be a timely addition that the information field updates the notion of information literacy to include creating information and recognizing needs, seeking, and using information. In an era of participatory culture, a truly information-literate individual should be able to participate in the information world meaningfully by creating and sharing information that meets one's needs (Koh, 2013).

"Digital Literacy," "Information seeking behavior," and "Critical thinking" are other topics found in information creation articles. These topics are also related and overlap with each other. "Digital literacy" is an individual's knowledge, attitude, and ability to make appropriate utilization of digital tools and facilities to identify, access, manage, integrate, evaluate, analyze, and combine digital resources, build new knowledge, and communicate with others (Martin & Madigan, 2006). The issue of digital literacy has become increasingly important in recent times due to the increasing use of social media for information interaction.

Bawden (2008) also described the four main skills of digital literacy: internet searching, hypertext navigation, knowledge assembly, and content evaluation. "Information-seeking behavior" is also a set of individuals" activities to fulfill their information needs. "Information-seeking behavior" includes information search goals, search methods, search problems, and factors affecting it (Broussard & Doty, 2016), as well as, in general, human behavior concerning information sources and channels, including the search for active and inactive information, and the utilization of information (Wilson, 2000).

Information behavior models also refer to creating information tangentially or implicitly, such as Wilson's Information-Seeking Behavior Model, Robson and Robinson's (2015) information-seeking and communication model, and the Document theory of Lund and Skare (2009). Koh (2013) has also shown that information creation is one of the most innovative and complex behaviors in which intensive information search and use coincide (Koh, 2013). Furthermore, critical thinking is the decision-making about the actions that should be taken or believed, performed using meditative thinking and reasoning (Ennis, 2015). The skilled individual in critical thinking uses rational standards that guide the thinking process (Paul, 1995). Due to the definitions of these three topics and the fact that processes are associated with searching and selecting information, these topics may overlap. Kari (2007, 2011) also states that creating knowledge includes judging information and interacting with information and its results. Information results can be cognitive, affective, social, functional, and/or autonomous (ibid).

"Consumer behavior" is another topic identified in scientific publications on information creation. In this regard, Clemons (2008) suggested that information accessibility increases consumer awareness to manage and optimize their selections. This awareness also modifies the behavior of consumers and the market. Consumers also make decisions based on the received information, so if little information is available, consumers are more likely to adapt their behaviors and decisions based on incomplete information (Allcott & Taubinsky, 2015). In addition, the topics "computer information processing," "social networks," "information supply," and "information production" had the lowest quantity of publications in information

creation articles, suggesting information production and sharing.

In social media, the information generated is remarkably multimodal and based on a combination of different types of information, including experiential, embodied, and scientific information. Navigating social media environments—and using, evaluating, and creating information in these environments—requires a set of competencies. In addition, in social media, young people are often placed in situations where they can influence others and be recognized as potential cognitive authorities (Multas & Hirvone, 2022). "Computer information processing" is less similar to other topics in knowledge creation, which process information by computer systems.

Conclusion

A new insight has been gained about the subject of information and knowledge creation as a result of the work of this study. The results of this study have identified fourteen main topics in information and knowledge creation. According to this study, knowledge creation encompasses a vast realm, including various topics such as information production, presentation of information and knowledge, information-seeking behaviors, decision-making models, and critical thinking. Information-seeking behaviors also include information needs, search, evaluation, and selection of the most extensive topics in the scientific field of knowledge creation. The topic "computer information processing" also has a small contribution to the field of knowledge creation. According to the results, information creation focuses more on a subjective human-dependent process than machine-dependent processes.

Based on this, it can be concluded that the creation of knowledge and information includes human mental and behavioral processes concerning knowledge. In other words, cognitive processes in knowledge and information creation include information needs, information processing, critical thinking, and decision-making. Behavioral processes regarding knowledge creation include data and knowledge production, information supply, information retrieval behavior, and consumer behavior. All the mentioned items are under the knowledge and information creation category and can be seen in 14 topics discovered and introduced in this article. In other words, knowledge is created by humans in a complex process. After converting implicit knowledge into explicit and personalizing it based on the information needs of people, it is available to them. In the information society, the created explicit knowledge is available to people with information literacy skills and creative and critical thinking and is used for personal and social development. In other words, the created knowledge is used to make effective and correct decisions.

Therefore, based on the findings, the process of creating interconnected social, material, and embodied information literacy practices, including planning, information-seeking, organization, editing, and presentation, has been formed. Finally, it is recommended that scholars analyze each topic in this study individually for future studies using text-mining techniques. Furthermore, publications on information creation can be topically analyzed with other text-mining techniques, such as supervised classification algorithms. Estimating the topic quantity before topical modeling operations is one of the methodological limitations of this study. Based on the number of topics, a manual review and interpretation of topical models were also required, even though some algorithms were used in this study. Another limitation of the present study was the manual labeling of topics, so some topics may not be easily labeled, requiring a more detailed study of the topic and consultation with experts.

Endnote

1. Data production, 2: Health-seeking behavior, 3: Decision-making models, 4: Knowledge production, 5: Human brain and Information processing, 6: Information needs, 7: Consumer behavior, 8: Information-seeking behavior, 9: Computer information processing 10: Digital Literacy, 11: Critical thinking, 12: Information supply, 13: Social networks, 14: Information production

References

Alghamdi, R. & Alfalqi, K. (2015). A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications* (IJACSA), 6(1), 147-153. Retrieved from https://thesai.org/Downloads/Volume6No1/Paper_21-A_Survey_of_Topic_Modeling_in_Text_Mining.pdf

Allcott, H. &Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review*, 105(8), 2501-2538. https://doi.org/10.1257/aer.20131564

Anderson, B.S. (2021). Using text mining to glean insights from COVID-19 literature. *Journal of Information Science*, 49(2), 373-381. https://doi.org/10.1177/01655515211001661

Arrom, L.M., Huguet, J., Errando, C., Breda, A.& Palou, J. (2018). How to write an original article. *Actas Urológicas Españolas (English Edition)*, 42(9), 545-550.

Baghmohammad, M., Mansouri, A. & CheshmehSohrabi, M. (2021). Identification of topic development process of knowledge and information science field based on the topic modelling (LDA). *Iranian Journal of Information Processing and Management*, 36(2), 297-326. https://doi.org/10.35050/JIPM010.2020.001 [in Persian]

Baker, K.S. & Mayernik, M.S. (2020). Disentangling knowledge production and data production. *Ecosphere*, 11(7), e03191. https://doi.org/10.1002/ecs2.3191

Bates, M.J. (2010). Information behavior. *Encyclopedia of library and information sciences*, 3, 2381-2391.

Blei, D.M., Ng, A.Y. & Jordan, M.I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022. Retrieved from https://dl.acm.org/doi/pdf/10.5555/944919.944937

Blei, D.M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. https://doi.org/10.1145/2133806.2133826

Broussard, R. & Doty, P. (2016). Toward an understanding of fiction and information behavior. *Proceedings of the Association for Information Science and Technology*, 53(1), 1-10. https://doi.org/10.1002/pra2.2016.14505301066

Cao, G., Duan, Y. & Cadden, T. (2019). The link between information processing capability and competitive advantage mediated through decision-making effectiveness. *International Journal of Information Management*, 44, 121-131. https://doi.org/10.1016/j.ijinfomgt.2018.10.003

Cao, G., Duan, Y. & Li, G. (2015). Linking business analytics to decision-making effectiveness: A path model analysis. *IEEE Transactions on Engineering Management*, 62(3), 384-395. https://doi.org/10.1109/TEM.2015.2441875

Chen, B., Tsutsui, S., Ding, Y. & Ma, F., 2017. Understanding the topic evolution in a scientific domain: An exploratory study for the field of information retrieval. *Journal of Informetrics*, 11(4), 1175-1189. https://doi.org/10.1016/j.joi.2017.10.003

Chen, X., Zou, D. & Xie, H. (2020). Fifty years of British Journal of Educational Technology: A topic modelling based bibliometric perspective. *British Journal of Educational Technology*, 51(3), 692-708. https://doi.org/10.1111/bjet.12907

Clemons, E.K. (2008). How information changes consumer behavior and how consumer behavior determines corporate strategy. *Journal of management information systems*, 25(2),13-40. https://doi.org/10.2753/MIS0742-1222250202

Cutrona, S.L., Mazor, K.M., Vieux, S.N., Luger, T.M., Volkman, J.E. & Finney Rutten, L.J. (2015). Health information-seeking on behalf of others: characteristics of surrogate seekers. *Journal of Cancer Education*, 30(1), 12-19. https://doi.org/10.1007/s13187-014-0701-3

Danesh, F., Dastani, M. & Ghorbani, M. (2021). Retrospective and prospective approaches of coronavirus publications in the last half-century: A Latent Dirichlet allocation analysis. *Library Hi Tech*, 39(3), 855-872. https://doi.org/10.1108/LHT-09-2020-0216

Dastani, M. & Danesh, F. (2021). Iranian COVID-19 publications in LitCovid: Text mining and topic modeling. *Scientific Programming*, 1-12. Article ID 3315695, https://doi.org/10.1155/2021/3315695

Dastani, M., Mousavi chelak, A., Ziaei, S. & Delghandi, F. (2020). Topic analysis of published articles in medical librarianship and information science in iran using text mining techniques. *Depiction of Health*, 11(4),355-367. https://doi.org/10.34172/doh.2020.43 [in Persian]

Ennis, R.H. (2015). Critical thinking: A streamlined conception. In *The Palgrave handbook of critical thinking in higher education* (pp. 31-47). Palgrave Macmillan, New York.

Fan, W., Wallace, L., Rich, S. & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9),76-82.

Fiandrino, S. & Tonelli, A. (2021). A text-mining analysis on the review of the non-financial reporting directive: bringing value creation for stakeholders into accounting. *Sustainability*, 13(2),763. https://doi.org/10.3390/su13020763

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

Goldman, S.R. & Scardamalia, M. (2013). Managing, understanding, applying, and creating knowledge in the information age: Next-generation challenges and opportunities. *Cognition and Instruction*, 31(2),255-269. https://doi.org/10.1080/10824669.2013.773217

Gorichanaz, T. (2019). Information creation and models of information behavior: Grounding synthesis and further research. *Journal of Librarianship and Information Science*, 51(4), 998-1006. https://doi.org/10.1177/0961000618769968

Griffin, R.J., Dunwoody, S. & Neuwirth, K. (1999). Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. *Environmental Research*, 80(2),S230-S245. https://doi.org/10.1006/enrs.1998.3940

Halevy, A., Norvig, P. & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2), 8-12. https://doi.org/10.1109/MIS.2009.36

Han, X. (2020). Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model. *Scientometrics*, 125(3), 2561-2595. https://doi.org/10.1007/s11192-020-03721-0

Hannigan, T.R., Haans, R.F., Vakili, K., Tchalian, H., Glaser, V.L., Wang, M.S., Kaplan, S. and Jennings, P.D. (2019). Topic modelling in management research: Rendering new theory from textual data. *Academy of Management Annals*, 13(2), 586-632. https://doi.org/10.5465/annals.2017.0099

Huang, M., Zhang, H., Gu, Y., Wei, J., Gu, S., Zhen, X., Hu, X., Sun, X. & Dong, H. (2019). Outpatient health-seeking behavior of residents in Zhejiang and Qinghai Province, China. *BMC Public Health*, 19(1),1-8. Retrieved from https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-7305-0

Huvila, I. (2011). The complete information literacy? Unforgetting creation and organization of information. *Journal of Librarianship and Information Science*, 43(4), 237-245. https://doi.org/10.1177/0961000611418812

Huvila, I., Douglas, J., Gorichanaz, T., Koh, K. & Suorsa, A. (2022). Guest editorial: Advances in research on information creation. *Library & Information Science Research*, 44(3), 101178. https://doi.org/10.1016/j.lisr.2022.101178

Indurkhya, N. & Damerau, F.J., 2010. Handbook of natural language processing. *Chapman and Hall/CRC*.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y. & Zhao, L. (2019). Latent Dirichlet Allocation (LDA) and topic modelling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11),15169-15211. https://doi.org/10.1007/s11042-018-6894-4

Jung, H. & Kim, B. (2021). Identifying research topics and trends in asset management for sustainable use: A topic modeling approach. *Sustainability*, 13(9),4792. https://doi.org/10.3390/su13094792

Kari, J., 2007. Conceptualizing the personal outcomes of information. *Information Research*, 12(2). Retrieved from https://informationr.net/ir/12-2/paper292.html

Kari, J., (2011). Outcomes of information: An analysis of spiritual messages. *The Open Information Science Journal*, 3, 63-75. http://dx.doi.org/10.2174/1874947X01103010063

Katre, P. D. (2019, December). Text mining and comparative visual analytics on large collection of speeches to trace socio-political issues. In *2019 IEEE 9th International Conference on Advanced Computing (IACC)* (pp. 108-114). IEEE.

Kobayashi, V.B., Mol, S.T., Berkers, H.A., Kismihók, G. & Den Hartog, D.N. (2018). Text classification for organizational researchers: A tutorial. *Organizational Research Methods*, 21(3),766-799. https://doi.org/10.1177/1094428117719322

Koh, K. (2013). Adolescents' information-creating behavior embedded in digital Media practice using scratch. *Journal of the American Society for Information Science and Technology*, 64(9),1826-1841. https://doi.org/10.1002/asi.22878

Kurata, K., Miyata, Y., Ishita, E., Yamamoto, M., Yang, F. & Iwase, A. (2018). Analyzing library and information science full-text articles using a topic modelling approach. *Proceedings of the Association for Information Science and Technology*, 55(1),847-848. https://doi.org/10.1002/pra2.2018.14505501143

Lamba, M. & Madhusudhan, M. (2019). Mapping of topics in DESIDOC Journal of Library and Information Technology, India: A study. *Scientometrics*, 120(2), 477-505. https://doi.org/10.1007/s11192-019-03137-5

Lee, L., Ocepek, M.G. & Makri, S. (2021). Creating by me, and for me: investigating the use of information creation in everyday life. *Information Research*, 26(1). Retrieved from https://informationr.net/ir/26-1/paper891.html

Linder, R., Snodgrass, C. & Kerne, A. (2014, April). Everyday ideation: All of my ideas are on Pinterest. *In Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2411-2420).

Lund, N.W. & Skare, R. (2009). Document theory. *Annual Review of Information Science and Technology*, 43(1), 1-55. https://doi.org/10.1002/aris.2009.1440430116

Martin, A. & Madigan, D. eds., (2006). *Digital literacies for learning*. Facet Publishing.

Miyata, Y., Ishita, E., Yang, F., Yamamoto, M., Iwase, A. & Kurata, K. (2020). Knowledge structure transition in library and information science: Topic modelling and visualization. *Scientometrics*, 125(1), 665-687. https://doi.org/10.1007/s11192-020-03657-5

Multas, A.M. & Hirvonen, N. (2021). "Let's keep this video as real as possible": young video bloggers constructing cognitive authority through a health-related information creation process. *Journal of Documentation*, 78(7), 42-64. https://doi.org/10.1108/JD-02-2021-0027

Park, J.H. & Song, M. (2013). A study on the research trends in library & information science in Korea using topic modeling. *Journal of the Korean society for information management*, 30(1),7-32. https://doi.org/10.3743/KOSIM.2013.30.1.007 [in korean]

Paul, R. (1995). *Critical thinking: How to prepare students for a rapidly changing world.* Foundation for Critical Thinking: Santa Rosa, CA

Preum, S.M., Clark, K., Davis, A., Khutsishvilli, K. and Valdez, R.S. (2019). Information seeking and information processing behaviors among Type 2 diabetics. *arXiv preprint arXiv*:1910.12444

Ransbotham, S., Kiron, D. and Prentice, P.K. (2016). Beyond the hype: The hard work behind analytics success. *MIT Sloan Management Review*, 57(3).

Řehůřek, R. & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks*.

Robson, A. & Robinson, L. (2013). Building on models of information behaviour: linking information seeking and communication. *Journal of documentation*, 69(2), 169-193. https://doi.org/10.1108/00220411311300039

Röder, M., Both, A. & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399-408).

Rosenbaum, S. (2011). *Curation nation: How to win in a world where consumers are creators*. *McGraw Hill Professional*.

Rousselet, G.A., Thorpe, S.J. & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in cognitive Sciences*, 8(8),363-370. https://doi.org/10.1016/j.tics.2004.06.003

Schmiedel, T., Müller, O. & vom Brocke, J. (2019). Topic modelling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, 22(4), 941-968. https://doi.org/10.1177/1094428118773858

Sievert, C. & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63-70).

Syed, S. & Spruit, M. (2017, October). Full-text or abstract? examining topic coherence scores using latent dirichlet allocation. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 165-174). IEEE.

Trace, C. B. (2007). Information creation and the notion of membership. *Journal of Documentation*, 63(1), 142-163. https://doi.org/10.1108/00220410710723920

Von Neumann, J. (2012). *The computer and the brain*. Yale University Press.

Wang, M. & Mengoni, P. (2020, December). How pandemic spread in news: Text analysis using topic model. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 764-770). IEEE.

Whittaker, S. (2011). Personal information management: from information consumption to curation. *Annual Review of Information Science and Technology*, 45(1). https://doi.org/10.1002/aris.2011.1440450108

Wickens, C.D. & Carswell, C.M. (2021). Information processing. In *Handbook of human factors and ergonomics*,117-151. John Wiley & Sons, INC. retrived from https://ftp.idu.ac.id/wp-content/uploads/ebook/ip/BUKU%20ERGONOMI/BUKU%20INGGRIS/handbook-of-human-factors-and-ergonomics-fourth-edition-2012.pdf

Wilson, T.D. (2000). Human information behavior. *Informing Science*, 3(2), 49-56. Retrived from http://www.inform.nu/Articles/Vol3/v3n2p49-56.pdf

Wilson, V.L. (2009). Behavioral change in type 1 diabetes self-management: Why and how? *Health Education Journal*, 68(4), 320-327. https://doi.org/10.1177/0017896909339530

World Health Organization (2021). Infodemic management: a key component of the COVID-19 global response. *Weekly Epidemiological Record*, 145-148.

Woxland, C. M., Cochran, D., Davis, E. L. & Lundstrom, K. (2017). Communal and student-centered: Teaching information creation as a process with mobile technologies. *Reference Services Review*, 45(1), 79-99. https://doi.org/10.1108/RSR-03-2016-0023

Yan, E. (2014). Research dynamics: Measuring the continuity and popularity of research topics. *Journal of Informetrics*, 8(1), 98-110. https://doi.org/10.1016/j.joi.2013.10.010

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