Educational Administration: Theory and Practice
2023, 29(2) 658-664
ISSN: 2148-2403
https://kuey.net/

Research Article

A Study Of The Impact Of Big Data On Information Science Research Methods

Dr. Sandip Banerjee[1*], Anant Mittal[2]

[1*]Assistant Programmer, Department of Election, Govt. of Arunachal Pradesh
[2]IPS (2015), Superintendent of Police, Special Investigation Cell (Vigilance), Govt of Arunachal Pradesh

Citation: Dr. Sandip Banerjee, et.al (2023) A Study Of The Impact Of Big Data On Information Science Research Methods, *Educational Administration: Theory and Practice*, 29(2) *658-664*
DOI: 10.53555/kuey.v29i2.7827

ARTICLE INFO

ABSTRACT

This abstract presents a comprehensive analysis of the impact of big data on research methodologies in information science. The study investigates how the defining characteristics of big data, namely its volume, variety, and velocity, have introduced novel challenges and opportunities for researchers in the field.
The key aspects addressed include:

- The conceptualization and significance of big data in information science research.
- The complexities associated with unstructured big data and the necessity for data refinement and unit definition.
- The ongoing relevance of inferential statistical analysis methods, encompassing sampling techniques and considerations related to p-value manipulation.
- The application of exploratory and statistical methodologies to big data, including correlation and regression analysis, natural language processing, sentiment analysis, temporal analysis, and visualization techniques.
- The utilization of link analysis and clustering methods to identify patterns and relationships within large-scale datasets.

The research elucidates how these methodologies have been adapted and implemented to analyse extensive datasets derived from sources such as social media platforms, search engine logs, and digital libraries. It underscores the importance of comprehending and effectively employing these research methods to address the challenges and capitalize on the opportunities presented by big data in information science research.

Keywords: Big data, Information science, Data analysis, P-value hacking, Correlation analysis, Regression analysis, Data visualization, Clustering methods, Social media analytics, Search engine logs, Scientometrics

1. Brief Introduction

The digital omnipresence of big data is remarkable, as it captures a wide array of human activities and natural occurrences in digital form. The pursuit of big data is propelled by the epistemological belief that extensive datasets provide superior intelligence and knowledge (Mills, 2018). Big data is defined by the renowned five Vs: Volume, Velocity, Variety, Veracity, and Value. These refer to the enormous quantity of data generated and stored; the rapid speed at which data grows; the diverse formats and types of data; the quality of the captured information; and the utility derived from the data, respectively. These characteristics of data have introduced unprecedented challenges and possibilities for data utilization and analysis methods in information science. Big data necessitates advanced research methodologies (boyd & Crawford, 2012).

Every empirical study comprises key components: research goals/objectives, research methods, data, and their analysis. These elements are interconnected. Research methods serve as guidelines, techniques, and procedures used in a study to gather, process, and analyze data, produce findings, and draw conclusions to achieve research aims. The availability, nature, and size of a dataset can significantly influence the choice of research methods and even research topics. A large dataset is not self-explanatory and requires appropriate

methodologies for analysis and processing. A dataset that is rich in content, diverse in type, and large in size would not only expand the range of potential research topics but also provide researchers with considerable flexibility to apply a wide range of research methods to their studies. The emerging big data trend has influenced data-driven empirical research in information science. It is crucial to conduct a study that systematically examines and discusses the impact of this trend on research methods in information science. The importance of such a study is multifaceted. It stimulates healthy discussion and reflection on the challenges and opportunities presented by the big data trend in the information science research community, aids researchers in the field in developing robust research designs for big data-oriented studies, and supports information practitioners in addressing challenges they encounter in the big data era.

2. Analytics And Big Data in Information Science

The definition of information science has evolved over time (Borko, 1968; Williams, 1987/1988; Saracevic, 2009; Stock & Stock, 2013), yet its fundamental components remain consistent. The field concentrates on information selection, organization, retrieval, dissemination, and utilization. A significant focus of research in this area is the interplay between information systems and their users.

Information technology, particularly the Internet, has enabled users to access diverse online information resources. Search engines, social media platforms, and Web portals have become essential information sources for the general public. Users employ search engines to find relevant information on the Internet, with Google processing over 63,000 searches per second on average (SEM Knowledge Base, 2018). Web portals, which are specially designed websites offering synthesized information from various sources on specific topics, help users navigate integrated information resources with features like roadmaps, subject directories, and internal search engines. For example, the CDC Website, a public health portal, recorded more than 3 billion page views in 2021 (CDC Digital Media Metrics, 2022).

Social media platforms, powered by Web 2.0 technologies, allow users to create, share, and exchange information in virtual communities. These interactive platforms can be categorized into social networks, blogs, microblogs, social news, social bookmarking, media sharing, wikis, question-and-answer sites, and review sites (Barbier & Liu, 2011). The openness and interactivity of social media generate vast amounts of data, with user-generated content and network entity relationships serving as key information sources (Gandomi & Haider, 2015). Facebook reported over 2.7 billion monthly active users in the second quarter of 2020 (Statista, 2020). As of May 2019, YouTube saw more than 500 hours of video uploaded every minute and over one billion hours of video watched daily (Tankovska, 2021).

Search engines store all submitted queries in transaction logs. Web portals continuously capture and save users' navigation footprints. Social media platforms record, store, and aggregate user interactions and activities, creating vast data repositories. These logs and repositories constitute user-generated data, contributing to the growth of big data and serving as crucial sources for information science researchers. These extensive datasets are invaluable to researchers in the field, as they accurately document users' detailed information-seeking behaviors and expand the range of research topics. The available data encompasses search queries, web portal browsing paths, topic-related conversations, user reactions to posts (such as comments, likes, and dislikes), event responses, user-generated hashtags, and various other user activities.

Beyond data from social media and general search sites, big data methodologies are applicable to the analysis of scientometric datasets, which have also experienced significant growth. These include conventional bibliographic and citation databases, as well as an increasing number of data sources providing access to bibliographic records, citation connections, and altmetrics data (e.g., Google Scholar, Crossref, Microsoft Academic, Dimensions Database, Semantic Scholar, Altmetric). With potential access to hundreds of millions of bibliographic records, billions of citation links, and extensive full-text corpora of scientific research covering all areas of human knowledge, opportunities have emerged to examine interdisciplinary relationships at both broad and detailed levels.

Undoubtedly, the scale of emerging big data, the diversity of data types, the wide range of topics, the heterogeneity of user groups, the variety of data sources, and the complexity of users' information needs in new contexts have created both challenges and opportunities for research methods in information science. Research findings derived from big datasets extend beyond the datasets themselves, facilitating knowledge accumulation in information science in unprecedented ways. Researchers in the field must reevaluate and adapt traditional methods of data collection, processing, and analysis to address these new challenges. Understanding these challenges and opportunities, and adopting appropriate research methods, is crucial for researchers to effectively conduct studies in the new big data environment. It is not surprising that researchers in the field are leveraging this data abundance to conduct studies aimed at understanding users and their information use behaviors across many existing and emerging areas of information science, thanks to big data.

3.The Unstructured Nature of Big Data

In most cases, the majority of information within a large dataset, excluding bibliographic or citation-related data, is unrefined, lacks structure (or is partially structured), and is disorganized (or partially organized). When

collecting data from sources like social media platforms, the information is raw and unprocessed. Not all elements in the gathered dataset are valuable, and the useful information is not well-arranged. The raw data requires cleansing before further analysis can occur. The relevance of data varies depending on the specific research study. Data that is suitable for one project may not be applicable to another. Processes such as data partitioning, separation, extraction, and tokenization are employed to refine and select relevant information for a particular research topic. It's clear that the refined and chosen dataset is a subset of the original large dataset. This process is guided by researchers' perspectives on the data, which are influenced by the research objectives and goals. Consequently, different research projects may yield distinct refined datasets. For instance, if researchers are focused on queries in a transaction log, they extract all query-related information. When studying consumer search behavior for a specific disease in a public Q&A forum, only questions, answers, responses, and related information about that disease are selected. Similarly, when examining a trending topic like COVID-19 on a social media platform, all posts, reposts, likes/dislikes, comments, and shares related to the topic are extracted for analysis.

Establishing a logical data unit is a crucial initial step. This unit represents a comprehensive record of an item, encompassing all potentially useful attributes or characteristics. Items are the subjects of interest in a study, and their attributes are often multidimensional and concealed. The identification of significant and valuable attributes is challenging due to their unstructured and hidden nature. Formulating a meaningful unit has proven to be a complex task.

For example, when defining a visiting user as a unit in a transaction log, pinpointing a unique user can be problematic. In public spaces like libraries, multiple individuals may use the same computer, sharing an IP address. Consequently, an appropriate session method must be employed to distinguish different users sharing the same IP address. A session encompasses a set of user interactions occurring within a specific timeframe, including multiple page views, searches, link jumps, and other activities. It can be defined by time, campaign changes, or other meaningful criteria.

Another illustration involves a study on network analysis in a social media platform. Here, a user is defined as a record, with their connections to others serving as attributes. If node centrality is used to describe these connections, the corresponding centrality values are hidden and cannot be directly extracted from raw data. Instead, they must be calculated based on the available information. The identification of attributes is largely determined by the research questions or problems at hand. These attributes may function as independent or dependent variables in the proposed research, allowing for the exploration of various relationships, patterns, and trends.

Big data has created a remarkable opportunity for serendipitous discoveries. Researchers may experience the unexpected unearthing of relevant and useful information as they identify and define units and attributes from unstructured datasets.

4.Analyses Based on Inferential Statistics

Scientific research studies primarily seek to elucidate and anticipate natural and social phenomena through the application of scientific methodologies. Researchers employ inferential statistical analysis to accomplish these goals of explanation and prediction. This analytical approach draws conclusions about a larger population based on sample data. A sample, which represents a smaller subset of a broader population, is used to illustrate the characteristics of the entire group. When a population is too vast for comprehensive testing, samples are utilized in inferential analysis. Sampling plays a crucial role in this type of analysis. Despite the advent of big data, the epistemological issues associated with certain traditional inferential analysis techniques remain significant in contemporary research.

4.1.The Sampling Process

Some argue that extensive datasets derived from sources like social media and search engines eliminate the need for sampling and inferential statistics due to their completeness. However, sampling and inferential analysis remain valuable and necessary. Firstly, for massive datasets, utilizing a sample and conducting inferential analysis can be computationally efficient and resource-saving. When sampling methods are appropriate, inferential analysis results can be reliable. Two sampling techniques specifically designed for large datasets have been proposed (Kim & Wang, 2019). One incorporates external auxiliary information, while the other combines big data samples with independent probability samples.

Secondly, a collected large dataset may represent only a subset of an expanding data source. For example, data gathered from social media platforms or search engine logs typically covers a specific timeframe, ranging from a month to several years. Although this dataset may include all raw data for that period, it remains a subset of an ongoing, growing data source. New content and emerging topics can be added to social media or search logs, which may not be reflected in older datasets. Consequently, it's impossible to treat the collected dataset as an entire population due to the nature of expanding data. Instead, it can be viewed as a special sample of the ongoing dataset, representing a section of the entire population for a limited time. This approach can be termed the "section sampling method."

Sampling methods are generally categorized as probability and non-probability sampling (Etikan & Bala, 2017). Probability sampling includes stratified, simple random, cluster, and systematic sampling, while non-probability sampling encompasses convenience and judgmental sampling. The key distinction between these categories is whether items in a sample are randomly selected. The section sampling method falls under non-probability sampling since items are not randomly chosen but come from a data selection. Random selection of sample items ensures results are representative of the population. Without random selection, there's a risk of systematic bias, potentially leading to over- or under-representation of a sample.
The section sampling method is limited by its non-random selection process. Additionally, research has demonstrated that sample size influences the outcomes of inferential analysis (Ajiferuke et al., 2006). When dealing with large datasets, utilizing a comparatively substantial sample size results in more accurate, less biased, and more robust inferential analysis findings.

4.2.Issues Associated With P-Value Hacking in Inferential Analysis

The p-value plays a crucial role in inferential statistical analysis. It denotes the likelihood of obtaining a set of observations assuming the null hypothesis is correct. Researchers use the p-value from analytical results to determine whether to reject the null hypothesis at a given significance level. As the null hypothesis is derived from a research question or problem, the p-value directly influences the study's outcome. In information science research, a significance level of 0.05 is commonly employed.
P-value hacking, also known as data-dredging or significance chasing, occurs when researchers manipulate their methods to achieve a desirable p-value. This practice can introduce bias into inferential studies. Examples of p-value hacking include altering samples, switching analytical methods, removing unsuitable data points, modifying measurements, changing data types, or adjusting experimental conditions. Research has shown that p-value hacking is closely linked to publication bias and can significantly increase false positive rates (Friese & Frankenbach, 2020). This issue is prevalent across various scientific disciplines (Head et al., 2015).
The root of p-value hacking lies in the inferential analysis process itself. However, studies utilizing comprehensive big datasets that encompass entire populations eliminate the need for sampling and inferential analysis. Consequently, employing a complete big dataset in research can effectively prevent p-value hacking issues.

5.Statistical Analysis and Exploratory Research

The analysis of large datasets heavily depends on various quantitative methods applicable to numerical, textual, or media data. While many library and information science researchers come from diverse academic backgrounds, those without extensive training in computing or statistics may find some of these techniques unfamiliar. Data mining encompasses a wide array of exploratory and statistical approaches aimed at uncovering patterns within datasets. Although this brief overview cannot cover all techniques, those relevant to information science research include correlation and regression analysis, natural language processing, sentiment analysis, temporal analysis, and visualization methods. These language-based techniques are particularly useful for extracting insights from textual data.

5.1. Analyzing Correlations and Regressions

Correlation analysis is a statistical technique used to assess the relationship strength between quantitative variables. This method is particularly effective in the realm of big data, with predictions derived from correlation analysis forming the core of big data applications (Mayer-Scho€n- berger & Cukier, 2013). The extensive and complex nature of big datasets enables researchers to identify and define various quantitative variables and perform correlation analyses on them. Variable predictions are widely utilized across numerous fields and serve as the driving force behind many information science studies. For example, understanding health consumers' information needs during a viral outbreak like Covid-19 on social media can help raise public awareness about pandemics and mitigate their societal impact. Similarly, discovering patterns in people's information-seeking behavior regarding asthma across seasons can assist health consumers in preventing potential asthma-related risks.
Regression analysis is a method for modeling variable relationships. A key feature of regression analysis is its ability to predict future outcomes for one variable based on another. While correlation analysis reveals relationships between variables, regression analysis shows how one variable influences another. For instance, a regression analysis could examine the relationship between user preferences and various factors in diabetes-related YouTube videos, such as presenter characteristics, gender, settings, or topics (Wang & Zhang, 2020). This analysis can uncover viewer attitudes towards specific video attributes, benefiting health consumers, professionals, and content creators. Correlation and regression analyses are closely related and can be applied to the same dataset. It's important to note that various regression methods, including linear, logistic, ridge, lasso, polynomial, Bayesian, negative binomial, and others, can be employed on big datasets due to the rich data relationships present in big data.

5.2. Technologies for Natural Language Processing and Text Mining

The accessibility of extensive text collections for research creates opportunities to employ Natural Language Processing (NLP) techniques for text and data mining. Large corpora, such as the HathiTrust Digital Library, accessible through the HathiTrust Research Center, provide researchers with over 17 million volumes, exceeding 6 billion pages and 2.9 trillion tokens (https://analytics.hathitrust.org/datasets). While NLP methods are also utilized on smaller datasets, analyzing linguistic patterns and mining text on a grand scale can yield insights into language usage variations across time and geography, as demonstrated by tools like Google Trends (https://trends.google.com/trends/) and Google Ngram Viewer (https://books.google.com/ngrams#).

Topic modeling, which combines NLP and machine learning techniques to condense textual relationships within corpora into a smaller set of "topics," has become widespread in information science studies. Researchers have applied it to examine the development of library and information science (Figuerola et al., 2017) and investigate topical areas on social media platforms like Twitter (Hong & Davison, 2010).

Latent Dirichlet Allocation (LDA) has been utilized to uncover autism-related themes on Facebook (Zhao et al., 2019). Additionally, it serves as an alternative method for exploring author relationships based on their work's content, complementing traditional approaches that rely on citations or term co-occurrence (Lu & Wolfram, 2012).

5.3. Methodology for Sentiment Analysis

Sentiment analysis is a technique that integrates text, computational linguistic, and statistical analyses to methodically detect, measure, and expose the emotional states and intensity within a given text. This text can range from a single word to an entire section. The outcome of sentiment analysis is typically expressed as a sentiment score, indicating whether the text conveys a positive, negative, or neutral emotion. This score is usually normalized on a scale from -1 to 1, or similar scales, where -1 represents negative emotion, 0 denotes neutral emotion, and 1 signifies positive emotion. This scoring system provides researchers with a logical metric for quantitative sentiment evaluation.

The application of sentiment analysis is extensive, encompassing the examination of opinions, attitudes, feedback, and reactions to events or individuals. It is inherently subjective in nature. For example, social media platforms, which are open to the public and facilitate user interaction, cover a broad range of topics. Certain subjects, such as public health, politics, and news, can be particularly emotionally charged. Consequently, it is common for users to share posts, comments, and other forms of expression on these topics with strong emotional content.

Traditional methods of processing informational text concentrate on subject terms, eliminating stop words and other insignificant words. During this process, non-subject terms, including sentiment words, are often excluded due to their lack of subject-related meaning. These terms cannot serve as indexing terms to represent the text's themes and are therefore not utilized for subject analysis or information retrieval.

Sentiment analysis introduces a unique dimension to the analysis of social media-related data. It enhances understanding of the analyzed text's context and assists researchers in uncovering users' information-seeking behavior on social media from a novel perspective. It adds a "colorful sentiment layer" to users' information behavior patterns. For instance, a study (Zhang et al., 2020) discovered that negative sentiment scores associated with terms related to panic (e.g., danger, worried, concern, sad, scared, epidemic, and kill) on a social media platform suggested that the Zika virus outbreak induced public panic.

5.4. Methodology for Temporal Analysis

Examining and modeling the change patterns of a specific variable in a dataset over time is known as temporal analysis. A distinctive feature of large datasets generated by computer systems is their temporal nature. When data are produced in a computer-based system, time is invariably linked to the generated information. Most data transaction logs employ a time stamp to automatically record the exact moment of data item creation resulting from an online action. This time stamp includes the year, month, date, hour, minute, and second. The temporal aspect enables researchers to analyze datasets from a time-based perspective. It can show the precise duration of a significant activity based on time stamp data, assisting researchers in connecting the elements of an event of interest from start to finish. This provides researchers with a new temporal dimension to examine, evaluate, and model the behavior change of specific variables over time, observing their change trends and patterns. The longitudinal or temporal analysis method serves as a natural and effective approach. For example, the parallel coordinate analysis technique allows researchers to effectively observe how multiple variables evolve over time (Stevenson & Zhang, 2015). Using temporal analysis, researchers uncovered theme changes during Zika virus outbreaks on a public question and answer forum (Zhang et al., 2020). The time information associated with activities in a transaction log allows researchers to specify and define variables for studies related to users' browsing behaviors, such as the duration a user spends on a meaningful object like a webpage or node of a subject directory, and the path a user navigates in a Web portal.

5.5. Visualization Of Data and Information

Visualization techniques for data and information possess a remarkable ability to manage extensive datasets and complex relationships within them. These methods not only alleviate the cognitive load associated with comprehending large volumes of data but also serve as potent tools for analysis (Zhang, 2008). When appropriately applied to substantial datasets, visualization approaches reduce data dimensionality, simplify intricate relationships, highlight key features, connect relevant points, and project refined data onto visual spaces. This process unveils hidden hierarchies, networks, and other relationships in an intuitive manner, facilitating exploration and analysis. It presents comprehensive overviews of datasets at various levels and from multiple perspectives. For example, search terms can be grouped and displayed based on their topical connections, revealing patterns and trends for analysis. Both visualization and clustering methods are extensively employed in subject analysis. Visualization techniques encompass, but are not limited to, multidimensional scaling (MDS), self-organized maps (SOM), and social network analysis (SNA). To illustrate, MDS was employed to uncover subject themes from malaria-related YouTube clip transcripts (Omwando & Zhang, 2021), SOM was utilized to examine data mining topics on Wikipedia (Wang & Zhang, 2017), and SOM was also used to optimize subject directory node relationships on a public health portal (Zhang et al., 2016). Nevertheless, researchers continue to face challenges in effectively visualizing vast amounts of information within limited display spaces, interpreting and explaining patterns generated by various visualization models, projecting high-dimensional objects onto low-dimensional visual spaces, and evaluating visual environments.

5.6. Methods of Link Analysis

Link analysis is a technique for examining connections between nodes in network environments. These nodes can represent various entities, such as websites, individuals, organizations, events, citations, or other objects of interest. The relationships between nodes are identified and defined based on the objects themselves, and can take various forms. This analytical method enables researchers to quantitatively assess relationships, determine node importance, and uncover network patterns.

In social media contexts, for example, the following and follower relationships among users create a virtual network. Link analysis can effectively pinpoint influential figures within this social structure. The method's diffusion analysis encompasses several aspects: breadth, time, speed, acceleration, and delay. This allows researchers to track events, such as the spread of misinformation, from start to finish, assess their scope, and monitor their propagation speed within a community.

On web portals, user visit log data can generate networks. When a user navigates from one website to another within the portal, it establishes a link between those sites. The accumulation of visit data produces a user visit network. By applying link analysis to this network, significant webpages within the portal can be identified.

5.7. Techniques for Clustering

A crucial aspect of analyzing large datasets involves identifying similar traits within the data to establish groupings of observations or records. These groupings can be utilized to simplify data relationships, thereby reducing the computational burden associated with data processing. Statistical methods employed for this purpose encompass cluster analysis, factor analysis, principal component analysis, and multidimensional scaling. Clustering techniques have found widespread application in information science research, including information retrieval, where they can be employed to identify related documents and perform topic modeling in IR systems, thus simplifying vocabulary relationships between documents. These techniques can also be applied to search log analysis, enabling the identification of session pattern groups (Chen & Cooper, 2001; Wolfram et al., 2009) or the visualization of co-occurring post relationships through multidimensional scaling (Zhang et al., 2020). Various forms of clustering are commonly used in scientometric research to identify groups of entities of interest, such as authors, publications, journals, or other units of measurement. Notably, the connections established between entities based on citations, co-authorship, or term co-occurrence allow for the use of link-based methods to identify clusters (Zhao & Strotmann, 2014). Language-based groupings can incorporate topic modeling, as mentioned earlier, to cluster scientific documents (Yau et al., 2014).

6. Summary

In conclusion, the advent of big data has significantly impacted research methods in information science, presenting both challenges and opportunities. The vast scale, diversity, and complexity of big datasets have necessitated the adaptation of traditional research methodologies and the development of new approaches.

Key findings from this review include:

- ✓ The importance of data refinement and structuring in dealing with unstructured big data.
- ✓ The continued relevance of sampling and inferential statistical analysis, even in the era of big data.
- ✓ The need for awareness and mitigation of p-value hacking issues in inferential analysis.
- ✓ The application of various exploratory and statistical methods, including correlation and regression analysis, natural language processing, sentiment analysis, temporal analysis, and visualization techniques.

- ✓ The effectiveness of link analysis and clustering methods in uncovering patterns and relationships within large datasets.

These methodological adaptations have enabled researchers to extract meaningful insights from big data, enhancing our understanding of information-seeking behaviors, user interactions, and content patterns across various platforms and contexts.

Moving forward, it is crucial for information science researchers to continue developing and refining methodologies that can effectively harness the power of big data while maintaining scientific rigor and ethical standards. This includes addressing challenges such as data quality, privacy concerns, and the need for interdisciplinary collaboration.

As big data continues to evolve and grow, so too must the research methods employed to study it. By embracing these methodological innovations and challenges, the field of information science can continue to advance its understanding of information phenomena in the digital age.

References

1. boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society, 15(5), 662–679.
2. Ajiferuke, I., Wolfram, D., & Famoye, F. (2006). Sample size and informetric model goodness-of-fit outcomes: A search engine log case study. Journal of Information Science, 32(3), 212–222.
3. Barbier, G., & Liu, H. (2011). Data mining in social media. In C. C. Aggarwal (Ed.), Social network data analytics (pp. 327–352). Springer.
4. Borko, H. (1968). Information science: What is it? American Documentation, 19(1), 3–5.
5. CDC Digital Media Metrics. (2022). CDC.gov 2022 monthly page views. Available at htt ps://www.cdc.gov/digital-social-media-tools/metrics/cdcgov/index.htm. (Accessed 28 January 2023).
6. Chen, H. M., & Cooper, M. D. (2001). Using clustering techniques to detect usage patterns in a Web-based information system. Journal of the American Society for Information Science and Technology, 52(11), 888–904.
7. Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design & analysis issues in field settings. Boston, MA: Houghton Mifflin.
8. Etikan, I., & Bala, K. (2017). Sampling and sampling methods. Biometrics & Biostatistics International Journal, 5(6), Article 00149.
9. Figuerola, C. G., Marco, F. J. G., & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. Scientometrics, 112(3), 1507–1535.
10. Frick´e, M. (2015). Big data and its epistemology. Journal of the Association for Information Science and Technology, 66(4), 651–661.
11. Friese, M., & Frankenbach, J. (2020). P-hacking and publication bias interact to distort meta-analytic effect size estimates. Psychological Methods, 25(4), 456–471.