This is a repository copy of *Towards automated analysis of research methods in library and information science*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/172452/

Version: Published Version

Article:

The MIT Press

RESEARCH ARTICLE

Towards automated analysis of research methods in library and information science

Ziqi Zhang[1] (iD), Winnie Tam[2] (iD), and Andrew Cox[1] (iD)

[1]Information School, The University of Sheffield
[2]University of Manchester Library, The University of Manchester

ABSTRACT

Previous studies of research methods in Library and Information Science (LIS) lack consensus in how to define or classify research methods, and there have been no studies on automated recognition of research methods in the scientific literature of this field. This work begins to fill these gaps by studying how the scope of "research methods" in LIS has evolved, and the challenges in automatically identifying the usage of research methods in LIS literature. A total of 2,599 research articles are collected from three LIS journals. Using a combination of content analysis and text mining methods, a sample of this collection is coded into 29 different concepts of research methods and is then used to test a rule-based automated method for identifying research methods reported in the scientific literature. We show that the LIS field is characterized by the use of an increasingly diverse range of methods, many of which originate outside the conventional boundaries of LIS. This implies increasing complexity in research methodology and suggests the need for a new approach towards classifying LIS research methods to capture the complex structure and relationships between different aspects of methods. Our automated method is the first of its kind in LIS, and sets an important reference for future research.

1. INTRODUCTION

Research methods are one of the defining intellectual characteristics of an academic discipline (Whitley, 2000). Paradigmatic fields use a settled range of methods. Softer disciplines are marked by greater variation, more interdisciplinary borrowing, and novelty. In trying to understand our own field of Library and Information Science (LIS) better, a grasp of the changing pattern of methods can tell us much about the character and directions of the subject. LIS employs an increasingly diverse range of research methods as the discipline becomes increasingly entwined with other subjects, such as health informatics (e.g., Lustria, Kazmer et al., 2010), and computer science (e.g., Chen, Liu, & Ho, 2013). As a result of a wish to understand these patterns, a number of studies have been conducted to investigate the usage and evolution of research methods in LIS. Many of these (Bernhard, 1993; Blake, 1994; Chu, 2015; Järvelin & Vakkari, 1990) aim to develop a classification scheme of commonly used research methods in LIS, while some (Hider & Pymm, 2008; VanScoy & Fontana, 2016) focus on comparing the usage of certain methods (e.g., qualitative vs. quantitative), or recent trends in the usage of certain methods (Fidel, 2008; Grankikov, Hong et al., 2020).

However, we identify several gaps in the literature on research methods in LIS. First, there is an increasing need for an updated view of how the scope of "research methods" in LIS has evolved. On the one hand, as we shall learn from the literature review, despite continuous interest in this research area, there remains a lack of consensus in the terminology and the classification of research methods (Ferran-Ferrer, Guallar et al., 2017; Risso, 2016). Some (Hider & Pymm, 2008; Järvelin & Vakkari, 1990) classify methods from different angles that form a hierarchy, while others (Chu, 2015; Park, 2004) define a flat structure of methods. In reporting their methods, scholars also undertake different approaches, such as some that define their work in terms of data collection methods, and others that define themselves through modes of analysis. Therefore, this "lack of consensus" is difficult to resolve, but reflects that LIS is not a paradigmatic discipline where it is agreed how knowledge is built. Rather, the field sustains a number of incommensurable viewpoints about the definition of method.

On the other hand, as our results will show, the growth of artificial intelligence (AI) and Big Data research in the last decade has led to a significant increase of data-driven research published in LIS that extends to these fast-growing disciplines. As a result of this, the conventional scope and definitions of LIS research methods have difficulty in accommodating these new disciplines. For example, many of the articles published around the AI and Big Data topics are difficult to fit into the categories of methods defined in Chu (2015).

The implication of the above situation is that it becomes extremely challenging for researchers (particularly new to LIS) to develop and maintain an informed view of the research methods used in the field. Second, there is an increasing need for automated methods that can help the analysis of research methods in LIS, as the number of publications and research methods both increase rapidly. However, we find no work in this direction in LIS to date. Although such work has already been attempted in other disciplines, such as Computer Science (Augenstein, Das et al., 2017) and Biomedicine (Hirohata, Okazaki et al., 2008) there is nothing comparable in LIS. Studies in those other fields have focused on automatically identifying the use of research methods and their parameters (e.g., data collected, experiment settings) from scientific literature, and have proved to be an important means for the effective archiving and timely summarizing of research. The need for providing structured access to the content of scientific literature is also articulated in Knoth and Herrmannova (2014)'s concept of "semantometics." We see a pressing need for conducting similar research in LIS. However, due to the complexity of defining and agreeing with a classification of LIS research methods, we anticipate the task of automated analysis will face many challenges. Therefore, a first step in this direction would be to gain an in-depth understanding of such technical challenges.

To address these limitations in previous literature, this work combines both content analysis and text mining methods to conduct an analysis of research methods reported in the LIS literature, to answer the following questions:

- How has the scope of "research methods" in LIS evolved, compared to previous definitions of this subject?
- To what extent can we automatically identify the usage of research methods in LIS literature, and what are the challenges?

We review existing definitions and the scope of "research methods" in LIS, and discuss their limitations in the context of the increasingly multi-disciplinary nature and diversification of research methods used in this domain. Following on from this, we propose an updated

classification of LIS research methods based on an analysis of the past 10 years' publications from three primary journals in this field. Although this does not address many of the limitations in the status quo of the definition and classification of LIS research methods, it reflects the significant changes that deviate from the previous findings and highlights issues that need to be addressed in future research in this direction. Second, we conduct the first study of automated methods for identifying research methods from LIS literature. To achieve this, we develop a data set containing human-labeled scientific publications according to our new classification scheme, and a text mining method that automatically recognizes these labels. Our experiments revealed that, compared to other disciplines where automated classification of this kind is well established, the task in LIS is extremely challenging and there remains a significant amount of work to be done and coordinated by different parties to improve the performance of the automated method. We discuss these challenges and potential ways to address them to inform future research taking this direction.

The remainder of this paper is structured as follows. We discuss related work in the next section, followed by a description of our method. We then present and discuss our results and the limitations of this study, with concluding remarks in the final section.

2. RELATED WORK

We discuss related work in two areas. First, we review studies of research methods in LIS. We do not cover research in similar directions within other disciplines, as research methods can differ significantly across different subject fields. Second, we discuss studies of automated methods for information extraction from scholarly data. We will review work conducted in other disciplines, particularly from Computer Science and Biomedicine, because significant progress has been made in these subject fields and we expect to learn from and generalize methods developed in these areas to LIS.

2.1. Studies of Research Methods in LIS

Chu (2015) surveyed pre-2013 studies of research methods in LIS and these have been summarized in Table 1. To avoid repetition, we only present an overview of this survey and refer readers to her work for details. Järvelin and Vakkari (1990) conducted the first study on this topic and proposed a framework that contains "research strategies" (e.g., historical research, survey, qualitative strategy, evaluation, case or action research, and experiment) and "data collection methods" (e.g., questionnaire, interview, observation, thinking aloud, content analysis, and historical source analysis). This framework was widely adopted and revised in later studies. For example, Kumpulainen (1991) showed that 51% of studies belonged to "empirical research" where "interview and questionnaire" (combined) was the most popular data collection method, while 48% were nonempirical research and contained no identifiable methods of data collection. Bernhard (1993) defined 13 research methods in a flat structure. Some of these have a connection to the five research strategies by Järvelin and Vakkari (1990) (e.g., "experimental research" to "empirical research"), while others would have been categorized as "data collection methods" by Järvelin and Vakkari (e.g., "content analysis," "bibliometrics," and "historical research"). Other studies that proposed flat structures of method classification include Blake (1994), who introduced a classification of 13 research methods largely resembling those in Bernhard (1993), and Park (2004), who identified 17 research methods when comparing research methods curricula in Korean and U.S. universities. The author identified new methods such as "focus group," and "field study," possibly indicating the changing scene in

Table 1. A summary of literature on the studies of research methods in LIS

Study	Data sample	Key findings w.r.t. research methods
Järvelin and Vakkari (1990)	833 articles from 37 journals in 1985	A classification scheme consisting of five "research strategies" and seven "data collection methods"
Kumpulainen (1991)	632 articles from 30 LIS journals in 1975	51% "empirical research," 48% "nonapplicable," 13% "historical method," 11% "questionnaire and interview"
Bernhard (1993)	Including journals, theses, textbooks, and reference sources in LIS	13 research methods; some relate to the "research strategies" while others relate to the "data collection methods" in Järvelin and Vakkari (1990)
Blake (1994)	LIS dissertations between 1975 and 1989	13 research methods, most of which are similar to Bernhard (1993)
Park (2004)	71 syllabus of Korean and U.S. universities between 2001 and 2003	17 research methods, some not reported before (e.g., field study, focus group)
Fidel (2008)	465 articles from LIS journals between 2005 and 2006	Only 5% used "mixed methods," while many that claimed to do so actually used "multiple methods" or "two approaches"
Hider and Pymm (2008)	834 articles from 20 LIS journals in 2005	Based on the Järvelin and Vakkari (1990) classification, "survey" remained as the predominant "research strategy" and "experiment" had increased significantly
Chu (2015)	1,162 articles from LIS journals between 2001 and 2010	A classification that extends earlier work in this area; "survey" no long dominating; instead, "content analysis," "experiment," and "theoretical approach" become more popular
VanScoy and Fontana (2016)	1,362 journal articles published between 2000 and 2009	A classification scheme similar to the previous work; majority of research was "quantitative", with "descriptive studies" based on "surveys" most common
Ferran-Ferrer et al. (2017)	580 Spanish LIS journal articles between 2012 and 2014	Proposed nine "research methods" and 13 "techniques." "Descriptive research" was the most used "research method," and "content analysis" was the most used "technique"
Togia and Malliari (2017)	440 LIS journal articles between 2011 and 2016	A similar classification of 12 "research methods" similar to that in Chu (2015). "Survey" remained the dominant method
Grankikov et al. (2020)	386 LIS journal articles between 2015 and 2018	Showed an increase in the use of "mixed methods" in this field

LIS. Hider and Pymm (2008) conducted an analysis that categorized articles from 20 LIS journals into the classification scheme defined by Järvelin and Vakkari (1990). They showed that "survey" remained the predominant research strategy but there had been a notable increase of "experiment." Fidel (2008) examined the use of "mixed methods" in LIS. She proposed a definition of "mixed method" and distinguished it with other concepts that are often misused as "mixed methods" in this field. Overall, only a very small percentage of LIS literature (5%) used "mixed methods" defined in this way. She also highlighted that in LIS, researchers often do not use the term "mixed methods" to describe their work.

Drawing conclusions from the literature, Chu (2015) highlighted several patterns from the studies of research methods in LIS. First, researchers in LIS are increasingly using more

sophisticated methods and techniques instead of the commonly used survey or historical method of the past. Methods such as experiments and modeling were on the rise. Second, there has been an increase in the use of qualitative approaches compared with the past, such as in the field of Information Retrieval. Building on this, Chu (2015) conducted a study of 1,162 research articles published from 2001 to 2010 in three major LIS journals—the largest collection spanning the longest time period in previous studies. She proposed a classification of 17 methods that largely echo those suggested before. However, some new methods included were "research journal/diary" and "webometrics" (e.g., link analysis, altmetrics). The study also showed that "content analysis," "experiment," and "theoretical approach" overtook "survey" and "historical method" to secure the dominant position among popular research methods used in LIS.

Since Chu (2015), a number of studies have been conducted on the topic of research methods in LIS, generally using a similar approach. Research articles published from some major LIS journals are sampled and manually coded into a classification scheme that is typically based on those proposed earlier. We summarize a number of studies below. VanScoy and Fontana (2016) focused on reference and information service (RIS) literature, a subfield of LIS. Over 1,300 journal articles were first separated into research articles (i.e., empirical studies) and those that were not research. Research articles were then coded into 13 research methods that can be broadly divided into "qualitative," "quantitative," and "mixed" methods. Again, these are similar to the previous literature, but add new categories such as "narrative analysis" and "phenomenology." Authors showed that most of the RIS research was quantitative, with "descriptive methods" based on survey questionnaires being the most common. Ferran-Ferrer et al. (2017) studied a collection of Spanish LIS journal articles and showed that 68% were empirical research. They developed a classification scheme that defines nine "research methods" and 13 "techniques." Different categories to the previous studies include "log analysis," "text interpretation," etc. However, the exact difference between these concepts was not clearly explained. Togia and Malliari (2017) coded 440 LIS journal articles into a similar classification of 12 "research methods" to that in Chu (2015). However, in contrast to Chu, they showed that "survey" remained in the dominant position. Grankikov et al. (2020) studied the use of "mixed methods" in LIS literature. Different from Fidel (2008), they concluded that the use of "mixed methods" in LIS has been on the rise.

In addition to work within LIS there has been work more widely in the social sciences to produce typologies for methodology (e.g., Luff, Byatt, & Martin, 2015). This update to an earlier seminal work by Durrant (2004) introduces a rather comprehensive typology of methodology, differentiating research design, data collection, data quality, and data analysis, among other categories. While offering a detailed approach for the gamut of social science methods, it does not represent the full range of methods of use in LIS which draws on approaches beyond the social sciences. Thus, while contributing to the development of our own taxonomy, this work could only offer a useful input.

In summary, the literature shows a continued interest in the studies of research methods in LIS in the last two decades. However, there remains significant inconsistency in the interpretation of terminologies used to describe the research methods, and in the different categorizations of research methods. This "lack of consensus" was discussed in Risso (2016) and VanScoy and Fontana (2016). Risso (2016) highlighted that firstly, studies of LIS research methods take different perspectives that can reflect research sub-areas within this field, object of study delimitation, or different ways of considering and approaching it. Secondly, a severe problem is the lack of category definitions in the different research method taxonomies proposed in the literature, and as a result, some were difficult to distinguish from each other.

VanScoy and Fontana (2016) pointed out that existing methodology categorizations in LIS are difficult to use, due to "conflation of research design, data collection, and data analysis methods," "ill-defined categories," and "extremely broad 'other' categories." As examples, while Chu (2015) proposed a classification primarily based on data collection techniques, methods such as "bibliometrics" and "webometrics" are arguably not for data collection, and were seen to be classified as "techniques" or "methods" in Ferran-Ferrer et al. (2017). On the contrary, "survey," "interview," and "observation" are mixed with "content analysis" and "experiment" and all considered as "techniques" by Ferran-Ferrer et al. (2017). In terms of the disagreement on the use of hierarchy, many authors have adopted a simple flat structure (e.g., Bernhard, 1993; Chu, 2015; Hider & Pymm, 2008; Park, 2004), while some introduced simple but inconsistent hierarchies (e.g., "research strategies" vs. "data collection methods" in Järvelin and Vakkari (1990) and "qualitative" vs. "quantitative" in VanScoy and Fontana (2016)). While intuitively we may argue that a sensible approach is to split methods primarily into data collection and analysis methods, apparently the examples shown above suggest that this is not a view that warrants consensus.

We argue that this issue reflects the ambiguity and complexity in research methods used in LIS. As a result of this, the same data can be analyzed in different ways that reflect different conceptual stances. Adding to this is the lack of consistency among authors in reporting their methods. Researchers sometimes define their work in terms of data collection methods, others through modes of analysis. For this reason, we argue that it is intrinsically difficult, if not impossible, to fully address these issues with a single universally agreed LIS research method definition and classification. Nevertheless, it remains imperative for researchers to gain an updated view of the evolution and diversification of research methods in this field, and to appreciate the different viewpoints from which they can be structured.

2.2. Automated Information Extraction from Scholarly Data

Information extraction (IE) is the task of automatically extracting structured information from unstructured or semi-structured documents. There has been increasing research in information extraction from scientific literature (or "scholarly data") in the last decades, due to the rapid growth of literature and the pressing need to effectively index, retrieve, and analyze such data (Nasar, Jaffry, & Malik, 2018). Nasar et al. (2018) reviewed recent studies in this area and classified them into two groups: those that extract metadata about an article, and those that extract key insights from the content. Research in this area has been predominantly conducted in the computer science, medical, and biology domains. We present an overview of these studies below.

Metadata extraction may target "descriptive" metadata that are often used for discovery and indexing, such as title, author, keywords, and references; "structural" metadata that describe how an article is organized, such as the section structures; and "administrative" metadata for resource management, such as file type and size. A significant number of studies in this area focus on extracting information from citations (Alam, Kumar et al., 2017), or header level metadata extraction from articles (Wang & Chai, 2018). The first targets information in individual bibliographic entries, such as the author names (first name, last name, initial), title of the article, journal name, and publisher. The second targets information usually on the title page of an article, such as title, authors, affiliations, emails, publication venue, keywords, and abstract. Thanks to the continuous interest in the computer science, medical, and biology domains, several gold-standard data sets have been curated over the years to be used to benchmark IE methods developed for such tasks. For example, the CORA data set (Seymore,

McCallum, & Rosenfeld, 1999) was developed based on a collection of computer science research articles, and consists of both a set for header metadata extraction (935 records) and a set for citation extraction (500 records). The FLUX-CiM data set (Cortez, da Silva et al., 2007) is a data set for citation extraction, containing over 2,000 bibliography entries for computer science and health science. Th UMASS data set consists of bibliographic information from 5,000 research papers in four major domains that include physics, mathematics, computer science, and quantitative biology.

According to Nasar et al. (2018), key-insights extraction refers to the extraction of information within an article's text content. The types of such information vary significantly. They are often *ad hoc* and there is no consensus on what should be extracted. However, typically, this can include mentions of objectives, hypothesis, method, related work, gaps in research, result, experiment, evaluation criteria, conclusion, limitations of the study, and future work. Augenstein et al. (2017) and QasemiZadeh and Schumann (2016) proposed more fine-grained information units for extraction, such as task (e.g., "machine learning," "data mining"), process (i.e., solutions of a problem, such as algorithms, methods and tools), materials (i.e., resources studied in a paper or used to solve the problem, such as "data set," "corpora"), technology, system, tool, language resources (specific to computational linguistics), model, and data item metadata. The sources of such information are generally considered to be either sentence- or phrase-level, where the first aims to identify sentences that may convey the information either explicitly or implicitly, and the second aims to identify phrases or words that explicitly describe the information (e.g., "CNN model" in "The paper proposes a novel CNN model that works effectively for text classification").

Studies of key-insight extraction are also limited to computer science and medical domains. Due to the lack of consensus over the task definition, which is discussed above, different data sets have been created focusing on different tasks. Hirohata et al. (2008) created a data set of 51,000 abstracts of published biomedical research articles, and classified individual sentences into objective, method, result, conclusion, and none. Teufel and Moens (2002) coded 80 computational linguistics research articles into different textual zones that describe, for example, background, objective, method, and related work. Liakata, Saha et al. (2012) developed a corpus of 256 full biochemistry/chemistry articles which are coded at sentence-level for 11 categories, such as hypothesis, motivation, goal, and method. Dayrell, Candido et al. (2012) created a data set containing abstracts from Physical Sciences and Engineering and Life and Health Sciences (LH). Sentences were classified into categories such as background, method, and purpose. Ronzano and Saggion (2015) coded 40 articles of the computer imaging domain and classified sentences into similar categories. Gupta and Manning (2011) pioneered the study of phrase-level key-insight extraction. They created a data set of 474 abstracts of computational linguistics research papers, and annotated phrases that describe three general levels of concepts: "focus," which describes an article's main contribution; "technique," which mentions a method or a tool used in an article; and "domain," which explains the application domain of a paper, such as speech recognition. Augenstein et al. (2017) created a data set of computational linguistics research articles that focus on phrase-level insights. Phrases indicating a concept of task, process, and material are annotated within 500 article abstracts. QasemiZadeh and Schumann (2016) annotated "terms" in 300 abstracts of computational linguistics papers. The categories of these terms are more fine-grained, but some are generic, such as spatial regions, temporal entities, and numbers. Tateisi, Ohta et al. (2016) annotated a corpus of 400 computer science paper abstracts for relations, such as "apply-to" (e.g., a method applied to achieve certain purpose) and "compare" (e.g., a method is compared to a baseline).

Q2

In terms of techniques, the state of the art has mostly used either rule-based methods or machine learning. With rule-based methods, rules are coded into programs to capture recurring patterns in the data. For example, words like "results," "experiments," and "evaluation" are often used to represent results in a research article, and phrases like "we use," and "our method" are often used to describe methods (Hanyurwimfura, Bo et al., 2012; Houngb & Mercer, 2012). With machine learning methods, a human annotated data set containing a large number of examples is first created, and is used subsequently to "train" and "evaluate" machine learning algorithms (Hirohata et al., 2008; Ronzano & Saggion, 2015). Such algorithms will consume low-level features (e.g., words, word sequences (n-grams), part of speech, word-shape (capitalized, lower case, etc), and word position, which are usually designed by domain experts) to discover patterns that may help capture the type of information that is to be extracted.

In summary, although there have been a plethora of studies on information extraction in the scientific literature, these have been limited to only a handful of disciplines and none has studied the problem in LIS. Existing methods will not be directly applicable to our problems for a number of reasons. First, previous work that extracts "research methods" only aims to identify the sentence or phrase that mentions a method (i.e., sentence- or phrase-level of extraction), but not recognize the actual method used. This is different, because the same research method may be referred to in different ways (e.g., "questionnaire" and "survey" may indicate the same method). Previous work also expects the research methods to be explicitly mentioned, which is not always true in LIS. Studies that use, for example, "content analysis," "ethnography," or "webometrics" may not even use these terms in their work to explain their methods. For example, instead of stating "a content analysis approach is used," many papers may only state "we analyzed and coded the transcripts…." For these reasons, a different approach needs to be taken and a deeper understanding of these challenges as well as to what extent they can be dealt with will add significant value for future research in this area.

3. METHODOLOGY

We uncorrected our method in four parts. First, we explain our approach to data collection. Second, we describe an exploratory study of the data set, with the goal of developing a preliminary view of the possible research methods mentioned in our data set. Third, guided by the literature and informed by the exploratory analysis, we propose an updated research method classification scheme. Instead of attempting to address the intrinsically difficult problem of defining a classification hierarchy, our proposed scheme will adopt a flat structure. Our focus will be the change in the scope of research methods (e.g., where previous classification schemes need a revision). Finally, we describe how we develop the first automated method for the identification of research methods used in LIS studies.

3.1. Data Collection

Our data collection methods are subject to the following criteria. First, we select scientific publications from popular journals that are representative of LIS. Second, we use data that are machine readable, such as those in an XML format that preserves all the structural information of an article, instead of PDFs. This is because we would like to be able to process the text content of each article, while OCR from PDFs is known to create noise in converted text (Nasar et al., 2018). Finally, we select data from the same or similar sources reported from the previous literature such that our findings can be directly compared to early studies. This may allow us to discover trends in LIS research methods.

Q3

Thus, building on Chu (2015), we selected research articles published between January 1, 2008 and December 31, 2018 and from *Journal of Documentation* (JDoc), *Journal of the American Society for Information Science & Technology* (JASIS&T; now *Journal of the Association for Information Science and Technology*), and *Library & Information Science Research* (LISR). These are among the core journals in LIS and were also used in Chu (2015), thus allowing us to make a direct comparison against earlier findings. We used the CrossRef API[1] to fetch the XML copies of these articles, and only kept articles that describe empirical research. This is identified with a category label assigned to each article by a journal. However, we notice a significant degree of inter- and intra-journal inconsistency in terms of how their articles are labeled. Briefly, each journal used between 14 and 19 categories to label their articles. There appear to be repetitions in these categories within each journal, and a lack of consensus on how each journal categorizes its articles. We show details of this later in our results section. For JDoc, we included a total of 381 (out of a total of 508 articles published in this period) articles labeled as "research article" and "case study". For JASIS&T, we included 1,837 "research articles" (out of 2,150). For LISR, we included 382 "research articles" and "full length articles (FLA)". This created a data set of 2,599 research articles, twice more than that in Chu (2015).

The XML versions of research articles allow programmatic access to the structured content of the articles, such as the title, authors, abstract, sections of main text, subsections, and paragraphs. We extract this structured content from each article for automated analysis later. However, it is worth noting that different publishers have adopted different XML templates to encode their data, which created obstacles during data processing.

3.2. Exploratory Analysis

To support our development of the classification scheme, we begin by undertaking an exploratory analysis of our data set to gain a preliminary understanding of the scope of methods potentially in use. For this, we use a combination of clustering and terminology extraction methods. VOSviewer (Van Eck & Waltman, 2010), a bibliometric software tool, is used to identify keywords from the publication data sets and their co-occurrence network within the three journals. Our approach consisted of three steps detailed below.

First, for each article, we extract the text content that most likely contains descriptions of its methodology (i.e., the "methodology text"). For this, we combine text content from title, keywords, abstracts, and also the methodology section (if available) of each article. To extract the methodology section from an article, we use a rule-based method to automatically identify the section that describes the research methods (i.e., the "methodology section"). This is done by extracting all level 1 sections in an article together with their section titles, and then using a list of keywords to match against these section titles. If a section title contains any one of these keywords, we consider that section to be the methodology section. The keywords include[2]: "methodology, development, method, procedure, design, study description, data analysis/ study, the model." Note that although these keywords are frequently seen in methodology section titles, we do not expect them to identify all variations of such section titles, nor can we expect every article to have a methodology section. However, we did not need to fully recover them as long as we have a sufficiently large sample that can inform our development of the classification scheme later on. This method identified methodology sections from 290 (out of 381), 1,283 (out of 1,837), and 346 (out of 383) of JDoc, JASIS&T, and LISR articles

---

[1] https://www.crossref.org/services/metadata-delivery/, last retrieved in March 2020.
[2] Their plural forms are also considered.

respectively. Still, there remains significant variation in terms of how researchers name their methodology section. We show this later in the results section. When the methodology section cannot be identified by our method, we use the title, keywords, and abstract of the article only. We apply this process to each article in each journal, creating three corpora.

Second, we import each corpus to VOSviewer[3] (version 1.614) and use its text-mining function to extract important terms and create clusters based on co-occurrences of the terms. VOSviewer uses natural language processing algorithms in the process of identifying terms. It involves steps such as copyright statement removal, sentence detection, part-of-speech tagging, noun phrase identification, and noun phrase unification. The extracted noun phrases are then treated as term candidates. Next, the number of articles in which a term occurs is counted (i.e., document frequency, or DF). Binary counting is chosen to avoid the analysis being skewed by terms that are very frequent within single articles. Then we select the top 60% relevant terms ranked by document frequency, and exclude those with a DF less than 10. These terms are used to support the development of the classification scheme.

To facilitate our coders in their task, the terms are further clustered into groups using the clustering function in VOSviewer. Briefly, the algorithm starts by creating a keyword network based on the co-occurrence frequencies within the title, abstract, keyword list, and methodology section. It then uses a technique that is a variant of the modularity function by Newman and Girvan (2004) and Newman (2004) for clustering the nodes in a network. Details of this algorithm can be found in Van Eck and Waltman (2014). We expect terms related to the same or similar research methods to form distinct clusters. Thus, by creating these clusters, we seek to gain some insight into the methods they may represent.

The term lists and their cluster memberships for the three journals are presented to the coders, who are asked to manually inspect them and consider them in their development of the classification scheme below.

3.3. Classification Scheme

Our development of the classification of research methods is based on a deductive approach informed by the previous literature and our exploratory analysis. A sample of around 110 articles ("shared sample") were randomly selected from each of the three journals to be coded by three domain experts. To define "research methods," we asked all coders to create a flat classification of methods primarily following the flat scheme proposed by Chu (2015) for reference. They could identify multiple methods for an article, and when this was the case, they were asked to identify the "main" (i.e., "first" as in Chu) method and other "secondary" methods (i.e., second, third, etc. in Chu). While Chu (2015) took a view focusing on data collection methods, we asked coders to consider both modes of analysis and data collection methods as valid candidates, as in Kim (1996). We did not ask coders to explicitly separate analysis from data collection, because (as reflected in our literature review) there is disagreement in how different methods are classified from these angles.

Coders were asked to reuse the methods in Chu's classification where possible. They were also asked to refer to the term lists extracted before, to look for terms that may support existing theory, or terms that may indicate new methods that were not present in Chu's classification. When no codes from Chu's model can be used, they were asked to discuss and create new codes that are appropriate, particularly informed by the term lists. Once the codes were finalized, the coders split the remaining data equally for coding. An Inter-Annotator-Agreement

---

[3] https://www.vosviewer.com/. Last accessed May 2020.

(Kappa statistics) of 86.7 was obtained on the shared sample when only considering the main method identified.

One issue at the beginning of the coding process is the notable duplicative and overlapping nature in the methods reported in the existing literature, as well as those proposed by the coders. Using Chu's scheme as an example, *ethnography* often involves participant *observation*, while *bibliometrics* may use methods such as link analysis (as part of *webometrics*). Another issue is the confusion of "topic" and "method." For example, an article could clearly discuss a bibliometrics study, but it was debatable whether it uses a "bibliometrics" method. To resolve these issues, coders were asked to follow the following principles. The first was to distinguish the goal of an article and the means implemented to achieve it. The second was to treat the main method as the one that generally takes the larger part of the text. Examples will be provided later in the results section.

During the coding process, coders were also asked to document the keywords that they found to be often indicative of each research method. For example, "content analysis" and "inter coder/rater reliability" are often seen in articles that use the "content analysis" method, while "survey," "Likert," "sampling," and "response rate" are often seen in articles that use "questionnaire." Note however, that it is not possible to create an exhaustive vocabulary for all research methods. Many keywords could also be ambiguous, and some research methods may only have a very limited set of keywords. However, these keywords form an important resource for our automated methods to be proposed below. Our proposed method classification contains a total of 29 methods. These, together with their associated keywords, are shown and discussed later in the results section.

3.4. Information Extraction of Research Methods

In this section, our goal is to develop automated IE methods that are able to determine the type of research method(s) that are used by a research article. As discussed before, this is different from the large number of studies on key-insights extraction that are already conducted in other disciplines. First, previous studies aim to classify text segments (e.g., sentences, phrases) within a research article into broad categories including "methods," without identifying what the methods are. As we have argued, these are two different tasks. Second, compared to the types of key insights for extraction, our study tackles a significantly larger number of fine-grained tasks—29 research methods. This implies that our task is much more challenging and that previous methods will not be directly transferable.

As our study is the first to tackle this task in LIS, we opt for a rule-based method for two reasons. First, compared to machine learning methods, rule-based methods were found to have better interpretability and flexibility when requirements are unclear (Chiticariu, Li, & Reiss, 2013). This is particularly important for studies in new domains. Second, despite increasing interest in machine learning-based methods, Nasar et al. (2018) showed that they do not have a clear advantage over rule-based methods. In addition, we also focus on a rather narrow target: identifying a single main method used. Note that this does not imply an assumption that each article will use only one method. It is rather a built-in limitation of our IE method. The reasons, as we shall discuss in more detail later, are two-fold. On the one hand, almost every article will mention multiple methods, but it is extremely difficult to determine automatically which are actually used for conducting the research and which are not. On the other hand, as per Chu (2015), articles that report using multiple methods remain a small fraction (e.g., 23% for JDoc, 13% for JASIS&T, and 18% for LISR in 2009–2010). With these in mind, it is extremely easy for automated methods to make false positive extractions of multiple
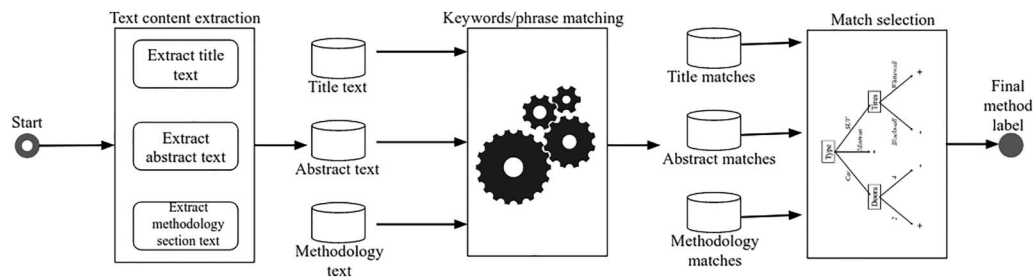
Q4

Figure 1. Overview of the IE method for research method extraction.

methods. Therefore, our aim here is exploring the feasibility and understanding the challenges in achieving our goal, rather than maximizing the potential performance of the automated methods.

We used a smaller sample of 30 coded articles to develop the rule-based method, with the remaining 300 for evaluation later on. Generally, our method searches the keywords (as explained before) associated with each research method within the restricted sections of an article. The method receiving the highest frequency will be considered to be the main research method used in that study. As we have discussed previously, many of these keywords can be ambiguous, but we hypothesize that by restricting our search within specific contexts, such as the abstract or the methodology section, there will be a higher possibility of recovering true positives. Figure 1 shows the overall workflow of our method, which will be explained in detail below.

3.4.1. Text content extraction

In this step, we aim to extract the text content from the parts of an article that are most likely to mention the research methods used. We focus on three parts: the title of an article, its abstract, and the methodology section, if available. Titles and abstracts can be directly extracted from our data set following the XML structures. For methodology sections, we use the same method introduced before for identifying them.

3.4.2. Keywords/keyphrase matching

In this step, we aim to look up the keywords/keyphrases (to be referred to uniformly as "keywords" below) associated with each research method within the text elements identified above. For each research method, and for each associated keyword, we count its frequency within each of the identified text elements. Note that the inflectional forms of these keywords (e.g., plural forms) are also searched. Then we sum the frequencies of all matched keywords for each research method within each text element to obtain a score for that research method within that text element. We denote this as $freq(m, text_i)$, where m denotes one of the research methods, $text_i$ denotes the text extracted from the part i of the article, with $i \in \{title, abstract, methodsection\}$.

3.4.3. Match selection

In this step, we aim to determine the main research method used in an article based on the matches found before. Given the set of matched research methods for a particular type of text element, that is, for a set of $\{freq(m_1, text_i), freq(m_2, text_i)..., freq(m_k, text_i)\}$, where i is fixed, we simply choose the method with the highest frequency. As an example, if "content analysis" and "interview" have frequencies of 5 and 3, respectively, in the abstract of an article, we

select "content analysis" to be the method detected from the abstract of that paper. Next, we select the research method based on the following priority: title > abstract > methodology section. In other words, if a research method is found in the title, abstract, and methodology section of an article, we choose only the one found in the title. Following the example above, if "content analysis" is the most frequent method based on the abstract of an article, and "questionnaire" is the one selected for its methodology section, we choose "content analysis" to be the research method used by the study. If none of the research methods are found in any of the three text elements, we consider the article to be "theoretical." If multiple methods are found to tie based on our method, then the one appearing earlier in the text will be chosen to be the main method.

3.4.4. Evaluation

Typically, automated methods cannot obtain perfect results as judged by humans and their performance needs to be formally evaluated. Thus, to understand to what extent we can correctly identify the research method used by a study, we propose to use the standard Precision, Recall, and F1 measures used for classification tasks. Specifically, these are defined in Eqs. 1, 2, and 3.

$$\text{Precision} = \frac{\#true\ positives}{\#total\ predicted\ positives} \tag{1}$$

$$\text{Recall} = \frac{\#true\ positives}{\#total\ actual\ positives} \tag{2}$$

$$\text{F1} = 2 \times \frac{precision \times recall}{precision + recall} \tag{3}$$

Given a particular type of research method in the data set, the number of research articles that reported using that method is "total actual positives," while the number predicted by the IE method is "total predicted positives." The intersection of the two is "true positives." Because the problem is cast as a classification task, and in line with the work in this direction but in other disciplines, we treat Precision and Recall with equal weights in computing F1. Also, we compute the "micro" average of Precision, Recall, and F1 over the entire data set across all research methods, where the "true positives," "total predicted positives," and "total actual positives" will simply be the sum of the corresponding values for each research method in the data set.

4. FINDINGS

4.1. Data Collection

As mentioned previously, we notice a significant degree of inter- and intra-journal inconsistency in how different journals categorize their articles. We show the details in Table 2.

First, there is a lack of definition of these categorization labels from the official sources, and many of the labels are not self-explanatory. For example, it is unclear why fine-grained JASIS&T labels like "advances in information science" and "AIS review" deserve to be separate categories, or what "technical paper" and "secondary article" entail in JDoc. For LISR, which uses mostly acronym codes to label its articles, we were unable to find a definition of these codes[4].

---

[4] All available codes are defined at: https://www.elsevier.com/__data/assets/text_file/0005/275666/ja5_art550_dtd.txt. However, no explanation of these codes can be found. A search on certain Q&A platforms found "FLA" to be "Full Length Article."

Table 2. Different categorizations of published articles by the three different journals

JASIS&T		JDoc		LISR	
Research article	1837	Research paper	370	FLA	350
Brief communication	115	Conceptual paper	121	EDI	40
Letter to the editor	65	Review	75	research-article	32
Editorial	31	Secondary article	52	e-review	23
Advances in information science	31	Literature review	16	ANN	12
Erratum	17	Viewpoint	14	BRV	11
In this issue	13	Editorial	11	e-non-article	11
Perspectives on design: information technologies and creative practice	12	Case study	11	IND	7
Opinion paper	10	Article	8	e-conceptual-paper	5
Opinion	7	General view	8	SCO	5
AIS review	7	Book review	5	EDB	3
Review	2	Technical paper	3	review-article	2
Opinion piece	1	Guest editorial	1	E-literature review	2
Depth review	1	List of referees 2013	1	REV	2
Guest editorial	1			ERR	2
				PRP	1
				COR	1
				DIS	1
				PUB	1

Second, different journals have used a different set of labels to categorize their articles. While the three journals appear to include some types that are the same, some of these are named in different ways (e.g., "opinion paper" in JASIS&T and "viewpoint" in JDoc). More noticeable is the lack of consensus in their categorization labels. For example, only JASIS&T has "brief communication," only JDoc has "secondary article," and only LISR has "non-article."

A more troubling issue is the intra-journal inconsistency. Each journal has used a large set of labels, many of which appear to be redundant. For example, in JASIS&T, "opinion paper," "opinion," and "opinion piece" seem to refer to the same type. "Depth review" and "AIS review" seem to be a part of "review." In JDoc, "general review" and "book review" seem to be a part of "review." And "article" seems to be too broad a category. In LISR, it is unclear why "e-review" is needed in addition to "review-article." Also, note that for many categories, there are only a handful of articles, an indication that those labels may be no longer used, or were even created in error.

4.2. Exploratory Analysis

Figures 2–4 visualize the clusters of methodologyrelated keywords found in the articles from each of the three journals. All three journals show a clear pattern of three separated large clusters. For LISR, three clusters emerge as follows: One (green) centers on "interview," with keywords like "interviewee," "theme," and "transcript"; one (red) centers on "questionnaire," with keywords like "survey," "respondent," and "scale"; and one (blue) with miscellaneous keywords, many of which seem to correlate weakly with studies of scientific literature (e.g., keywords like "author," "discipline," and "article") or bibliometrics generally.

For JDoc, the two clusters around "interview" (green) and "questionnaire" (blue) are clearly visible. In contrast to LISR, the third cluster (red) features keywords that are often indicative of statistical methods, algorithms, and use of experiments. Overall, the split of the clusters seems to indicate the separation of methods that are typically qualitative (green and blue) and quantitative (red).

The clusters from JASIS&T appear to be more different from LISR and JDoc and also have clearer boundaries. One cluster (red) appears to represent methods based on "interview" and "survey"; one (green) features keywords indicative of bibliometrics studies; and one (blue) has keywords often seen in studies using statistical methods, experiments, or algorithms. Comparing the three journals, we see a similar focus of methodologies between LISR and JDoc, but quite different patterns in JASIS&T. The latter appears to be more open to quantitative and data science research.

Figure 2. Cluster of terms extracted from the LISR corpus (top 454 terms ranked by frequency extracted from the entire corpus of 382 articles). Size of font indicates frequency of the keyword.

Figure 3. Cluster of terms extracted from the JDoc corpus (top 451 terms ranked by frequency extracted from the entire corpus of 381 articles). Size of font indicates frequency of the keyword.

Figure 4. Cluster of terms extracted from the JASIS&T corpus (top 2,027 terms ranked by frequency extracted from the entire corpus of 1,837 articles). Font size indicates frequency of the keyword.

4.3. Classification Scheme

Table 3 displays our proposed method classification scheme, together with references to previous work where appropriate, and keywords that were indicative of the methods. Notice that some of the keywords are selected based on the clusters derived from the exploratory studies. Also, the keywords are by no means a comprehensive representation of the methods, but only serve as a starting point for this type of study. In the following we define some of the methods in detail and explain their connection to the literature.

Our study was able to reuse most of the codes from Chu (2015). We revised Chu's "ethnography/field study" to two categories: "ethnography/field study," which refers to traditional ethnographic research (e.g. using participant observation in real world settings), and "digital ethnography," referring to the use of ethnographic methods in the digital world, including work following Kozinets' (2010) suggestions for "netnography" as an influential branch of this work.

The major change we have introduced concerns the "experiment" category. Chu (2015) argued for a renewed perspective on "experiment," in the sense that this refers to a broad range of studies where "new procedures (e.g., key-phrase extraction), algorithms (e.g., search result ranking), or systems (e.g., digital libraries)" are created and subsequently evaluated. This differs from the classic "experimental design" as per Campbell and Stanley (1966). However, we argue that this is an "over-generalization", as Chu showed that more than half of the articles from JASIS&T have used this method. Such a broad category is less useful as it hides the complex multidisciplinary nature in LIS. Therefore, in our classification, we use "experiment" to refer to the classic "experimental design" method and introduce a more fine-grained list of methods that would have been classified as "experiment" by Chu. These include "agent based modeling/simulation," "classification," "clustering," "information extraction," "IR related indexing/ranking/query methods," and "topic modeling," all of which focus on developing procedures or algorithms (rather than simple application of such techniques for a different purpose) that are often subject to systematic evaluation; and "comparative evaluation," which focuses on following scientific experimental protocols to systematically compare and evaluate a set of methods.

Further, we added methods that do not necessarily overlap with Chu's classification. For example, "annotation" refers to studies that involve users annotating or coding certain content, with the coding frame or the coded content being the primary output of a study. "Document analysis" refers to studies that analyze a collection of documents (e.g., government policy papers) or media items (e.g., audio or video data) to discover patterns and insights. "Mixed methods" is added, as studies such as Grankikov et al. (2020) revealed an upward trend in the usage of this research method in LIS. Note that in this context, "mixed methods" refers to Fidel's (2008) definition, which refers to research that combines data collection in a particular sequence for some reason, rather than any research that happens to involve multiple forms of data. "Statistical methods" has a narrow scope encompassing studies of correlation between variables or hypothesis testing, as well as those that propose metrics to quantify certain problems. This excludes metrics specifically targeting the bibliometrics domain (e.g., h-index), as the level of complexity and the extent of effort devoted to that area justifies it being an independent umbrella term that encompasses various statistical metrics. Statistical methods also exclude generic comparison based on descriptive statistics, which is very common (and thus can be over-generalizing) in quantitative research; also, the majority of computational methods for classification, clustering, or regression are statistical-based in a more general sense. Finally, "user task based studies" refers to systematic methods that involve human users undertaking

Table 3. The proposed research method classification scheme

Method	Relation to previous work	Indicative key-words/phrases[5]
***From* Chu (2015)**		
bibliometrics	Same as Chu	impact factor, scientometric, bibliometric, citation analysis, h-index…
content analysis	Same as Chu	content analysis, inter coder reliability, inter annotator agreement, krippendorff
delphi study	Same as Chu	delphi study
ethnography/ field study	Traditional ethnographic studies excluding those done in a digital context (see "digital ethnography" below).	Hammersley, participant observation, ethnography, ethnographic, ethnographer…
experiment	Classic experimental studies, not the generalized concept as per Chu.	dependent variable, independent variable, experiment
focus group	Same as Chu	focus group
historical method	Same as Chu	historical method
interview	Same as Chu	interview, interviewed, interviewer, interviewee, interviewing
observation	Same as Chu	observation
questionnaire	Same as Chu	respondent, questionnaire, survey, Likert, surveyed…
research diary/ journal	Same as Chu	diary study, cultural probe
think aloud protocol	Same as Chu	think aloud
transaction log analysis	Same as Chu	log analysis/technique
theoretical studies	Same as Chu	Studies that cannot be classified into any of the other method categories
webometrics	Same as Chu	webometrics, cybermetrics, link analysis
New		
agent based modeling/ simulation	Studies that use computational modeling methods for the purpose of simulation	agent model/modeling, multi-agents

Table 3.   (continued)

Method	Relation to previous work	Indicative key-words/phrases[5]
annotation	Studies that focus on using human users to create coded data	annotation, tagging
classification	Studies that focus on developing computational classification techniques	classification, classify
clustering	Studies that focus on developing computational clustering techniques	cluster, clustering
comparative evaluation	Studies that follow systematic evaluation procedures to compare different methods	comparative evaluation, evaluative studies
document analysis	Studies that analyze secondary document collections (e.g., historical policy documents, transcripts) with a critical close reading	document/textual analysis, document review
information extraction	Studies that develop computational methods for the purpose of extracting structured information from texts	named entity recognition, NER, relation extraction
IR related indexing/ ranking/query methods	Studies that develop methods with a goal to improve search results	Learning to rank, term weighting, indexing method, query expansion, question answering
mixed method	Fidel (2008)	cresswell and plano clark, mixed method
digital ethnography	Studies applying ethnography to the digital context	digital ethnography, netnography, netnographic
network analysis	Studies that apply network theories with a focus to understand the properties of social networks	network analysis/study
statistical methods	Studies of correlations between variables, hypothesis testing; proposing new statistical metrics that quantify certain problems other than bibliometrics. This category excludes comparisons based on simple descriptive statistics	correlation, logistic test, t-test, chi-square, hypothesis test…
topic modeling	Studies that develop computational topic modeling methods	topic model, topic modeling, LDA

Table 3. (*continued*)

Method	Relation to previous work	Indicative key-words/phrases[5]
user task based study	Studies that require human users to carry out certain tasks (sometimes using a system) to produce data for further analysis	user study, user analysis

certain tasks following certain (often different) processes, with a goal to compare their behaviors or evaluate the processes.

Revisiting the issue of duplication and overlap often seen in the scope of LIS research methods discussed before, we use examples to illustrate how our classification should be used to avoid such an issue. In Table 4, articles by Zuccala, van Someren, and van Bellen (2014), Wallace, Gingras, and Duhon (2008), Denning, Soledad, and Ng (2015), and Solomon and Björk (2012) all study bibilometrics problems, but their main research method is classified differently under our scheme. Zuccala et al. (2014) focuses on developing a classifier to automatically categorize sentences in reviews by their scholarly credibility and writing style. The article studied a problem of bibilometrics nature, and used human coders to annotate training data. However, its ultimate goal is to develop and evaluate a classifier, as is the focus of the majority of the text. Therefore, the main research method is considered to be "classification," while "annotation" may be considered a secondary research method and "bibliometrics" is more appropriate as a topic of the study. Wallace et al. (2008) has a similar pattern, where the content is dominated by technical details of how the "network analysis" method is constructed and applied to bibliometrics problems. Denning et al. (2015) describes a tool whose core method is formulating a statistical indicator, which the authors propose to measure book readability. Thus its main method qualifies under "statistical methods". Solomon and Björk (2012) uses descriptive statistics to compare open access journals. By definition, we do not classify such an approach as "statistical methods." But it can be argued that the authors used certain metrics to quantify a specific bibliometrics problem and therefore, we label its main method as "bibliometrics." In terms of our very own article, arguably, we consider both "content analysis" and "classification" as our main methods, and "annotation" as a secondary method because it serves a purpose for content analysis and creating training data for classification. "Bibliometrics" is more appropriate as the topic rather than the method we use, because our work actually adapts generic methods to bibliometric problems.

Figure 5 compares the distribution of different research methods found in the samples of the three journals. We notice several patterns. First, compared to JDoc and LISR, work published at JASIS&T has a clear emphasis on using a wider range of computational methods. This is consistent with findings from Chu (2015). Second, JASIS&T also has a substantial focus on bibliometrics research, which lacks representation in JDoc or LISR. Instead (the third pattern) for JDoc and LISR, questionnaire and interview remain the most dominant research methods. These findings resonate with those from our exploratory analysis. Fourth, for all three journals, a noticeable fraction of published work (between 10% and 18%) is of a theoretical nature, where no data collection or analysis methods are documented. Finally, we could not identify studies using "webometrics" as methods, but many may qualify under such a topic. However,

---

[5] Only up to five examples are shown. For the full list of keywords, see supplementary material in the appendix.

Table 4. Example articles and how their main research method will be coded under our scheme

Article	Reference	Main method
A machine-learning approach to coding book reviews as quality indicators: Toward a theory of megacitation	Zuccala et al. (2014)	Classification
A new approach for detecting scientific specialties from raw cocitation networks	Wallace et al. (2008)	Network analysis
A readability level prediction tool for K-12 books	Denning et al. (2015)	Statistical methods
A study of open access journals using article processing charges	Solomon and Björk (2012)	Bibliometrics

Q6

they often use other methods (e.g., content analysis of web collections, annotation of web content) to study a webometrics problem.

4.4. Information Extraction of Research Methods

We evaluate our IE method using 300 articles from the coded sample data[6] (disjoint with the smaller set for developing the method), and present the Precision, Recall and F1 scores below. As mentioned before, we only evaluate the main method extracted by the IE process using Eqs. 1–3. We then show the common errors made by our method.

4.4.1. Overview of Precision, Recall, and F1

Table 5 shows the Precision, Recall and F1 of our IE method obtained on the annotated samples from the three journals. Overall, the results show that the task is a very challenging one, as our method has obtained rather poor results on most of the research methods. Across the different journals and considering the size of the sample, our method has generally performed consistently on "interview," "questionnaire," and "bibliometrics." Based on the nature of our method (i.e., keywords lookup), this suggests that terminologies related to these research methods may be used more often in nonambiguous contexts. The average performance of our IE method achieves a micro-average F1 of 0.783 on JDoc, 0.811 on LISR, and 0.61 on JASIS&T. State-of-the-art methods on key-insights extraction generally achieve an F1 of between 0.03 (Lin, Ng et al., 2010) and 0.53 (Kovačević, Konjović et al., 2012) on tasks related to "research methods" at either sentence or phrase levels. Notice that the figures should not be compared directly as-is, because the task we deal with is different: We aim to identify specific methods, while all the previous studies only aim to determine whether a specific piece of text describes a research method or not.

4.4.2. Impact of the article of abstract

We conducted further analysis to investigate the quality of abstracts and its impact on our IE method. This includes three types of analysis. To begin with, we disabled the "methodology section" extraction component in our method, and retested our method on the same data set, but excluded articles where methods can only be identified from the methodology section. The results are shown in Table 6. On average, we obtained noticeable improvement on the

---

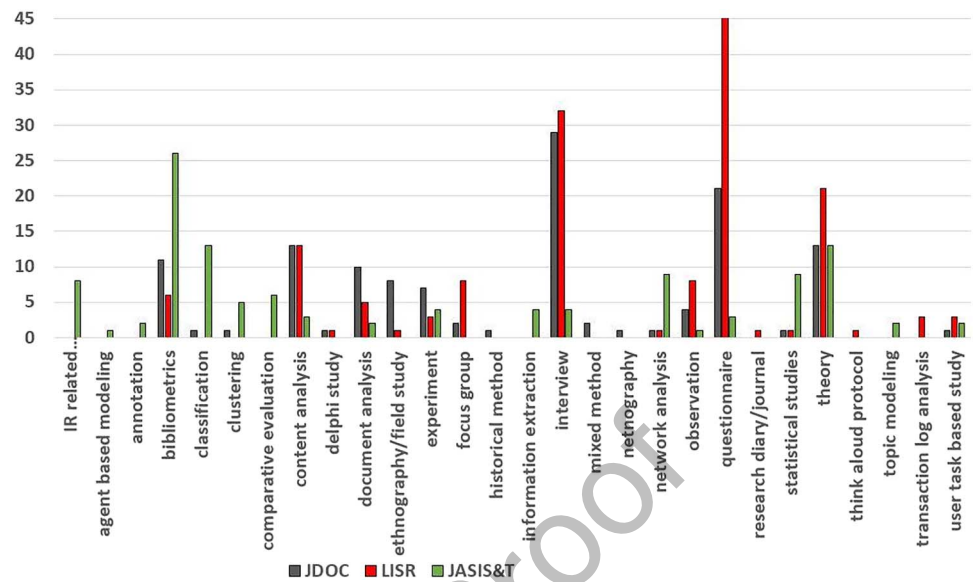[6] Data can be downloaded at https://doi.org/10.5281/zenodo.4486156.

Figure 5. Distribution of research methods found in the samples of the three journals. The y-axis indicates percentages represented by a method within a specific journal collection.

JDoc data set, but not on LISR or JASIS&T. Among the three journals, JDoc is the only one that enforces a structured abstract. Arguably, this ensures consistency and quality in writing the abstracts, from which our IE methods may have potentially benefited.

To verify this, we conducted the second type of analysis. We asked coders to revisit the articles they coded and identify the percentage of articles for which they were unable to identify its main method confidently without going to the full texts. This provides an alternative but more direct view of the quality of abstracts from the three journals, without the bias from the IE method. The figures are 5%, 6%, and 12% for JASIS&T, JDoc, and LISR respectively. This shows that to a human reader, comparatively, both JDoc and JASIS&T abstracts are more explicit than LISR when it comes to explaining their methods. This may be an indication of better-quality abstracts. To some extent, this is consistent with the pattern we observed from the previous analysis. The quality in JASIS&T abstracts does not translate to better performance of our IE method when focusing on only the abstracts. This could be partially attributed to the wider diversity of methods noted in JASIS&T articles (Figure 5) as well as the implicitness in the description of many of those methods that deviate from LISR and JDoc. For example, none of the articles using "comparative evaluation" used the keywords shown in Table 3. Instead, they used generic words that, if included, could have significantly increased false positives (e.g., "compare" and "evaluate" are typically used but will be nondiscriminative to identify studies that solely focus on comparative evaluations). Similarly, only one article using "user based task studies" used our proposed keywords. We will cover this issue again in the later sections.

Our third type of analysis involves studying the association between the length of an abstract and its quality, and subsequently (and potentially) its impact on our IE method. We notice that the three journals have different requirements on the length of abstract: 150 for LISR, 250 for JDoc, and 200 for JASIS&T. We do not make hypothesis a correlation between an abstract's length and its clarity (hence affecting its quality), as this can be argued from contradictory angles. On the one hand, one may argue that a shorter length can force authors to be

Table 5. Precision (P), Recall (R) and F1 on the three journals. "–" indicates that no articles are classified under method by the coders; neither does our method predict that method for any articles. For the absolute number of instances for each method, see Figure 5

Method	JDoc			LISR			JASIS&T		
	P	R	F1	P	R	F1	P	R	F1
***Methods from* Chu (2015)**									
bibliometrics	0.917	1.00	0.957	1.00	0.833	0.909	0.846	0.846	0.846
content analysis	0.857	0.462	0.600	0.462	0.462	0.462	0.500	0.667	0.571
delphi study	0	0	0	1.00	1.00	1.00	—	—	—
ethnography/field study	0.875	0.875	0.875	0.250	1.00	0.400	—	—	—
experiment	1.00	0.714	0.833	0.400	0.667	0.500	0.071	0.250	0.111
focus group	0.500	0.500	0.500	0.667	0.750	0.706	—	—	—
historical method	1.00	1.00	1.00	—	—	—	—	—	—
interview	0.793	0.821	0.807	0.735	0.781	0.758	1.00	0.500	0.667
observation	0.375	1.00	0.545	0.444	0.500	0.471	0	0	0
questionnaire	0.600	0.750	0.667	0.827	0.915	0.869	0.333	0.667	0.444
research diary/journal	–	–	–	0	0	0	–	–	–
think aloud protocol	–	–	–	0.250	1.00	0.400	–	–	–
transaction log analysis	–	–	–	0	0	0	–	–	–
webometric	–	–	–	–	–	–	–	–	–
theoretical studies	0.310	0.923	0.462	0.474	0.857	0.610	0.476	0.769	0.588
New									
agent based modeling	–	–	–	–	–	–	1.00	1.00	1.00
annotation	–	–	–	–	–	–	0	0	0
classification	0.143	1.00	0.250	–	–	–	0.467	0.583	0.519
clustering	0	0	0	–	–	–	0.833	1.00	0.909
comparative evaluation	–	–	–	–	–	–	0	0	0
document analysis	1.00	0.200	0.333	1.00	0.600	0.750	0	0	0
information extraction	–	–	–	–	–	–	1.00	0.250	0.400
IR related indexing/ranking/ query methods	–	–	–	–	–	–	0.67	0.25	0.363
mixed method	0.667	1.000	0.800	–	–	–	–	–	–
digital ethnography	1.00	1.00	1.00	–	–	–	–	–	–
network analysis	0	0	0	1.00	1.00	1.00	0.750	0.333	0.462

Table 5. (*continued*)

Method	JDoc			LISR			JASIS&T		
	P	R	F1	P	R	F1	P	R	F1
statistical methods	0	0	0	0.167	1.00	0.286	0.500	0.444	0.471
topic modeling	–	–	–	–	–	–	1.00	0.500	0.667
user task based study	0	0	0	0	0	0	1.00	0.500	0.667
AVERAGE[7]	0.783	0.783	0.783	0.811	0.811	0.811	0.610	0.610	0.610

more explicit about their methodology; on the one hand, one could also argue that a shorter length may result in more ambiguity, as authors have little space to explain their approach clearly. Instead, we started with analyzing the distribution of abstract length in our data sets across the three journals. We wrote a program that counts the number of words in each abstract, where words are delimited by white space characters only. We made surprising findings, as shown in Figure 6: a very large proportion of articles did not comply with the limit of the abstract length.

Figure 6 suggests that at least 50% of articles in our JASIS&T and LISR data sets have exceeded the abstract word limits. The situation of JDoc is not very much better. Across all three journals, there are also very long abstracts that almost doubled the word limit[8]; and there are noticeable articles with very short abstracts, such as those containing fewer than 100 words: 1 for JDoc, 34 for LISR, and 14 for JASIS&T. Overall, we do not see significantly different patterns in the distributions across the three journals. We further manually inspected a sample of 20 articles from each journal to investigate whether there were any patterns in terms of the publication year of those articles that exceeded the word limit. This is because we were uncertain whether during the abstract word limit changed during the history of each journal. Again, we could not find any consistent patterns. For JDoc, the distributions are: 2010 (3), 2011 (3), 2013 (4), 2014 (1), 2015 (2), 2016 (2), 2017 (2), and 2018 (3). For LISR, the distributions are: 2010 (5), 2011 (1), 2012 (4), 2013 (2), 2014 (1), 2015 (4), 2016 (2), and 2018 (1). For JASIS&T, the distributions are: 2010 (3), 2011 (4), 2012 (2), 2013 (1), 2014 (4), 2015 (2), 2016 (1), 2017 (2), and 2018 (1). Articles exceeding the abstract length limit can be found in any year in all three journals. For these reasons, we argue that there is no strong evidence indicating any association between the abstract length and its impact on our IE method. However, the lack of compliance with the journal requirement is rather concerning. While the quality of abstracts may be a factor that affects our method, it is worth noting that our method for detecting the methodology section has its limitations. Some articles do not have an explicit "methodology" section. Instead, they may describe different parts of their method in several top-level sections (e.g., see Saarikoski, Laurikkala et al., 2009). Some may have a "methodology" section that is a subsection of the top-level sections (e.g., the method section is within the "Case Study" section in Freeburg, 2017). A manual inspection of 50 annotated samples revealed that there were 10% of articles on which this method failed to identify the methodology section. In other words, the method has a 10% error rate. Thus arguably, with a more reliable method for finding methodology sections or generally content sections that describe methodology, our IE method could perform better.

---

[8] Examples: 10.1108/JD-10-2012-0138, 10.1108/00220410810912415, 10.1002/asi.21694.

Table 6. Precision (P), Recall (R) and F1 on the three journals when the text from the methodology section (if available) is ignored. "–" indicates that no articles are classified under method by the coders, and neither does our method predict that method for any articles. **Bold** indicates better results while <u>underline</u> indicates worse results compared to Table 4. For the absolute number of instances for each method, see Figure 5

Method	JDoc			LISR			JASIS&T		
	P	R	F1	P	R	F1	P	R	F1
Methods from Chu (2015)									
bibliometrics	0.917	1.00	0.957	1.00	0.833	0.909	0.846	0.846	0.846
content analysis	**1.00**	0.462	**0.632**	<u>0.444</u>	<u>0.363</u>	<u>0.400</u>	0.500	0.667	0.571
delphi study	0	0	0	1.00	1.00	1.00	–	–	–
ethnography/field study	0.875	0.875	0.875	0.250	1.00	0.400	–	–	–
experiment	1.00	0.714	0.833	**1.00**	<u>0.500</u>	**0.667**	0.071	0.250	0.111
focus group	0.500	0.500	0.500	**0.833**	**0.833**	**0.833**	–	–	–
historical method	1.00	1.00	1.00	–	–	–	–	–	–
interview	**0.821**	0.821	**0.821**	**0.885**	<u>0.766</u>	**0.821**	1.00	0.500	0.667
observation	**0.429**	1.00	**0.600**	**0.600**	<u>0.429</u>	**0.500**	0	0	0
questionnaire	0.600	0.750	0.667	**0.891**	<u>0.891</u>	**0.891**	0.333	0.667	0.444
research diary/journal	–	–	–	0	0	0	–	–	–
think aloud protocol	–	–	–	**0.500**	1.00	**0.667**	–	–	–
transaction log analysis	–	–	–	0	0	0	–	–	–
webometric	–	–	–	–	–	–	–	–	–
theoretical studies	**0.324**	0.923	**0.480**	<u>0.457</u>	**0.889**	<u>0.604</u>	0.476	0.769	0.588
New									
agent based modeling	–	–	–	–	–	–	1.00	1.00	1.00
annotation	–	–	–	–	–	–	0	0	0
classification	**0.167**	1.00	**0.286**	–	–	–	0.467	0.583	0.519
clustering	0	0	0	–	–	–	0.833	1.00	0.909
comparative evaluation	–	–	–	–	–	–	0	0	0
document analysis	1.00	**0.286**	**0.444**	1.00	<u>0.500</u>	<u>0.667</u>	0	0	0
information extraction	–	–	–	–	–	–	1.00	0.250	0.400
IR related indexing/ranking /query methods	–	–	–	–	–	–	0.67	0.25	0.363
mixed method	0.667	1.000	0.800	–	–	–	–	–	–
digital ethnography	1.00	1.00	1.00	–	–	–	–	–	–

Table 6. (*continued*)

Method	JDoc			LISR			JASIS&T		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
network analysis	**1.00**	**1.00**	**1.00**	1.00	1.00	1.00	0.750	0.333	0.462
statistical methods	0	0	0	0.167	1.00	0.286	0.500	0.444	0.471
topic modeling	–	–	–	–	–	–	1.00	0.500	0.667
user task based study	0	0	0	0	0	0	1.00	0.500	0.667
AVERAGE	**0.804**	**0.804**	**0.804**	**0.812**	**0.812**	**0.812**	0.610	0.610	0.610
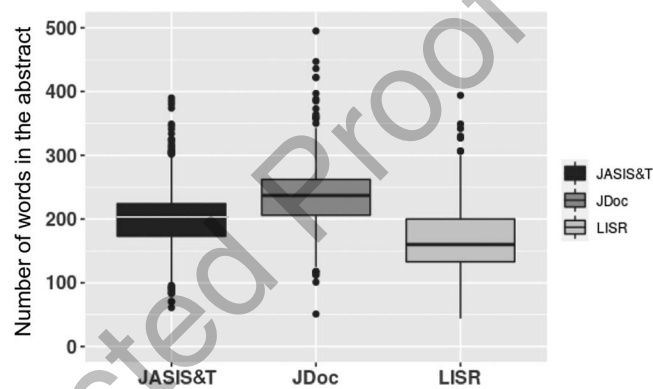
Figure 6. Distribution of abstract length across the three different journals.

4.4.3. *Error analysis*

To further understand the challenges of this task, we analyzed all errors made by our IE method and explain these below. Of the errors, 67%[9] are due to keywords used in different contexts than expected. For example, we define "classification" to be methods that use computational approaches for classifying data. However, the keywords "classify" or "classification" are also used frequently in work that may use, for example, content analysis or document analysis, to study library classification systems. A frequent error of this type is when a method is mentioned as future or previous work, such as in "In future studies, e.g., families' focus-group interviews could bring new insights". Some 10% of errors are due to ambiguity of the keywords themselves. For example, "bibliometrics" was identified as the wrong research method from the sentence "This paper combines practices emerging in the arts and humanities with research evaluation from a scientometric perspective…". A further 33% of errors are due to the lack of keywords, or when a method is mentioned implicitly and can only be inferred from reading the context. As examples, we discussed "comparative evaluation" and "user based task studies" before. More examples include "information extraction", which is a very broad topic and can be difficult to include all possible keywords; and "document analysis", which is particularly difficult to capture because researchers rarely use distinctive keywords to describe their study. In all these cases, a lot of inference with background knowledge is required.

---

[9] More than one error category can be associated with each article.

5. DISCUSSION

We discuss the lessons learned from this work with respect to our research questions, as well as limitations of our work.

5.1. Research Method Classification

Our first research question concerns the evolution of "research methods" in LIS. We summarize three key points below.

First, following a deductive coding process informed by literature as well as our data analysis, we developed a classification scheme that largely extends that of Chu (2015). In particular, we refined Chu's "experiment" category to include a range of methods that are based on computational approaches, used in the creation of procedures, algorithms, or systems. These are often found in work belonging to the "new frontier" of LIS (i.e., those that often cross boundaries with other disciplines, such as information retrieval, data mining, human computer interaction, and information systems). We also added new categories that were not included in the existing classification schemes by earlier studies. Overall, we believe that our significantly wider classification scheme indicates the increasing trend of diversification and interdisciplinary research in LIS. This could be seen as a strength in terms of LIS drawing fruitfully on a wide range of fields and influences, from humanities, social science, and science. It does not suggest a field moving towards the mature position of paradigmatic consensus, but it could be seen to reflect a healthy dynamism. More troubling might be considered the extent to which novelty comes largely from computational methods, suggesting a discipline without a long history of development and whose direction is subordinate to that of another.

Second, coming with this widening scope is the increasing complexity in defining "research methods." While our proposed classification scheme remains a flat structure, as is the case for the majority of studies in this area, we acknowledge that the LIS community may benefit from a hierarchical classification that reflects different perspectives of research methodology. However, as we have discussed in extended depth earlier on, it has been difficult to achieve consensus, simply because researchers in different traditions view methodology differently and use terminology differently. While not an aim of this very study, we anticipate that this can be partially addressed by developing a framework for defining and classifying LIS research methods from multiple, complementary perspectives. For example, a study should have a topic (e.g., "bibliometrics" could be both a method and topic), could use certain modes of analysis and data collection methods (resonating with the "research strategy" and "data collection method" model by Järvelin and Vakkari (1990)), and adopt a certain methodological stance (e.g., mixed-methods, multi-methods, quantitative) based on the mode of analysis (resonating with that by Hider and Pymm (2008)).

However, there exist significant hurdles to achieve this goal. As suggested by Risso (2016), LIS needs to disambiguate and clearly define different categories of "methods" (e.g., to address issues such as "citation analysis" being treated as both research strategy and data collection method in Järvelin and Vakkari (1990)). Further, there is a need to regularly update the framework to accommodate the evolution of the LIS discipline (Ferran-Ferrer et al., 2017). For this, automated IE methods may be useful in coping with the growing amount of literature. Also, significant effort needs to be devoted to encourage the adoption of such standards. Last, but not least, researchers should be encouraged to share their coding frame and the data they coded as examples for future reference. Data sharing has been an obvious gap in LIS research on research methods, compared to other disciplines such as Computer Science and Biomedicine.

Third, there is a clear pattern of different methodological emphasis in the articles published by the three different journals. While JDoc and LISR appear to publish more work that uses "conventional" LIS research methods, JASIS&T appears to be more open to accepting work that uses a diverse range of methods that have an experimental nature and seen more common in other disciplines. This pattern may reflect the different scope of focus of these journals. For example, LISR explicitly states that it "does not normally publish technical information science studies … or most bibliometric studies," while JASIS&T "focuses on the production, …, use, and evaluation of information and on the tools and techniques associated with these processes." However, JDoc's scope description is less indicative of the methodological emphasis, as it states "… welcome submissions exploring topics where concepts and models in the library and information sciences overlap with those in cognate disciplines." This difference in terms of their scope and aims had an impact on our exploratory analysis and, therefore, our resulting classification scheme. However, this should not be considered a limitation of our approach. If an LIS journal expands its scope to cover such a diverse range of fields, then we argue there is a need to develop a more fine-grained classification that better reflects this trend.

5.2. Automated Extraction of Research Methods

Our IE method for detecting the research methods used in a study is the first in LIS. Similar to earlier studies on key-insight extraction from scientific literature, we found this task particularly challenging. Although our method is based on simple rules, we believe it is still representative of the state of the art. This is because, on the one hand, its average performance over all methods is comparable to figures previously reported in similar tasks, even if our task is arguably more difficult. On the other hand, research so far cannot show a clear advantage of more complex methods such as machine learning over rule-based ones. The typical errors we found from our method will be equally challenging for typical machine learning-based methods.

Overall, our method achieved reasonable performance on only a few methods (i.e., "interview," "questionnaire," and "bibliometrics"), while its performance on most methods is rather unsatisfactory. Compared to work in a similar direction from other disciplines, we argue that research on IE of research methods from the LIS literature will need to consider unique challenges. The first is the unique requirement of the task. As we discussed before, existing IE methods in this area only aim to identify the sentence or phrase that mentions a method (i.e., sentence- or phrase-level of extraction), but not to recognize the actual method used. This is not very useful when our goal is to understand the actual method adopted by a study, which may mention other methods for the purposes of comparison, discussion, and references. This implies a formulation of the task beyond the "syntactic" level to the "semantic" level, where the automated IE method needs not only to identify mentions of methods in text, but also to understand the context in which they appear to derive their meanings (e.g., recall the examples we have shown in the error analysis section).

Adding to the above (i.e., the second challenge) is the complexity in defining and classifying LIS "research methods," as we have discussed in the previous section. The need for taking a multi-perspective view and identifying not only the main but also secondary methods only escalates the level of difficulty for IE. Also, there is the lack of standard terminology to describe LIS methods. For example, from our own process of eliciting research methods, we discovered methods that are difficult to identify by keywords, such as "mixed methods" and "document analysis."

Finally, researchers may need to cope with varying degrees of quality in research article abstracts. This is particularly important because, as we have shown, our method can benefit

from well-structured abstracts. In Computer Science for example, IE of research methods has mostly focused on abstracts (Augenstein et al., 2017) because they are generally deemed to be of high quality and information-rich. In the LIS domain, however, we have noticed issues such as how journal publishers differ in terms of enforcing structured abstracts, and that not every study would clearly describe their method in the abstracts (Ferran-Ferrer et al., 2017).

All these challenges mean that feature engineering—a crucial step for IE of research methods from texts—will be very challenging in the LIS discipline. We discuss some possibilities that may partially address this in the following section.

5.3. Other Issues

During our data collection and analysis, we discovered issues with how journal publishers categorize their articles. We have shown an extensive degree of intra- and inter-journal inconsistency, as well as a lack of guidance on how to interpret these categories. This undoubtedly created difficulties for our data collection process and potential uncertainties in the quality of our data set, and will remain an obstacle for future research in this area. We therefore urge the journal publishers to be more transparent about their article categorization system, and to work on improving the quality of their categorization. It might also be useful for publishers to offer common guidelines on describing methods in abstracts and to prompt peer reviewers to examine keywords and abstracts with this in mind.

Our further analysis of the abstract lengths showed a significant extent of non-compliance, as many articles (around, or even exceeding, 50%) are published with an abstract exceeding the word limit, and a small number of articles had a very short abstract. While we were unable to confirm the association between the length of the abstracts and the performance of our IE method, such inconsistency could arguably be considered as a quality issue for the journal.

5.4. Limitations of this Study

First, our proposed classification scheme remains a flat structure, and as we discussed above, it may need to be further developed into a hierarchy to better reflect different perspectives on research methods. Some may also argue that our classification diverges from the core research methods used in LIS. Due to the multidisciplinary nature of LIS, do we really need to integrate method classifications that conventionally belong to other disciplines? Would it be better to simply use the classification schemes from those disciplines when a study crosses those disciplines? These are the questions that we do not have answers to but deserve a debate given the multidisciplinary trend in LIS.

Second, our automated IE method for extracting research methods has large room for improvement. Similar to the previous work on key-insight extraction, we have taken a classification-based approach. Our method is based on keyword lookup, which is prone to ambiguity due to both context and terminology, as we have discussed. As a result, its performance is still unsatisfactory. We envisage an alternative approach to be sentence- or paragraph-level classification that focuses on sentences or paragraphs from certain areas of a paper only, such as abstracts or the methodology section, when available. The idea is that sentences or paragraphs from such content may describe the method used and, compared to simple keywords lookup, provide additional context for interpretation. However, this creates a significant challenge for data annotation, because machine learning methods require a large amount of examples (training data) to learn from, and for this particular task there will be a very large number of

categories that need examples. We therefore urge researchers in LIS to make a collective effort towards data annotation, sharing, and reuse.

Also, our IE method only targets a single, main research method from each article. Detecting multiple research methods may be necessary but will be even more challenging, as features that are usually effective for detecting single methods (e.g., frequency) will be unreliable, and it requires a more advanced level of "comprehension" by the automated method. In addition, existing IE methods only identify the research methods themselves but overlook other parameters of the methods that may also be very interesting. For example, new researchers to LIS may want to know what a reasonable sample size is when a questionnaire is used, whether the sample size has an impact on citation statistics, or what methods are often "mixed" in a mixed method research. Addressing these issues will be beneficial to the LIS research community, but remains a significant challenge to be tackled in the future.

Finally, our work has focused on the LIS discipline. Although this offers unique value compared to the existing work on IE of research methods predominantly covering Computer Science and Biomedicine, the question remains as to how the method can generalize to other social science disciplines or humanities. For example, our study shows that among the three journals, between 13% and 21% of articles are theoretical studies (Figure 5). However, methods commonly used in the humanities (e.g., hermeneutics) would not be described in a manner like empirical studies in LIS. This means that our IE method, if applied to this discipline, can miss-classify some studies that use traditional humanities methods as nonempirical, even though their authors might consider them to be empirical. Nevertheless, LIS is marked by considerable innovation in methods. This reflects wider pressures for more interdisciplinary studies to address complex social problems as well as individual researchers' motives to innovate in methods to achieve novelty. These factors are by no means confined to LIS. We can anticipate that these factors will make the classification of methods in soft and applied disciplines equally challenging. Therefore, something may be learned from this study by those working in other fields.

6. CONCLUSION

The field of Library and Information Science is becoming increasingly interdisciplinary as we see a growing number of publications that draw on theory and methods from other subject areas. This leads to increasingly diverse research methods reported in this field. A deep understanding of these methods would be of crucial interest to researchers, especially those who are new to this field. While there have been studies of research methods in LIS in the past, there is a lack of consensus in the classification and definition of research methods in LIS, and nonexistence of studies of automated analysis of research methods reported in the literature. The latter has been recognized as of paramount importance and has attracted significant effort in fields that have witnessed significant growth of scientific literature, a situation that LIS is also undergoing.

Set in this context, this work analyzed a large collection of LIS literature published in three representative journals to develop a renewed perspective of research method classification in LIS, and to carry out an exploratory study into automated methods—to the best of our knowledge, the first of this nature in LIS—for analyzing the research methods reported in scientific publications. We discovered critical insights that are likely to impact the future studies of research methods in this field.

In terms of research method classification, we showed a widening scope of research methodology in LIS, as we see a substantial number of studies that cross disciplines such as

information retrieval, data mining, human computer interaction, and information systems. The implications are twofold. First, conventional methodology classifications defined by the previous work can be too broad, as certain methodological categories (e.g., "experiment") would include a significant number of studies and are too generic to differentiate them. Second, there is the increasing complexity of defining "research method," which necessitates a hierarchically structured classification scheme that reflects different perspectives of research methodology (e.g., data collection method, analysis method, and methodological stance). Additionally, we also showed that different journals appear to have a different methodological focus, with JASIS&T being the most open to studies that are more quantitative, or algorithm and experiment based.

In terms of the automated method for method analysis, we tackled the task of identifying specific research methods used in a study, one that is novel compared to the previous work in other fields. Our method is based on simple rule-based keyword lookup, and worked well for a small number of research methods. However, overall, the task remains extremely challenging for recognizing the majority of research methods. The reasons are mainly due to language ambiguity, which results in challenges in feature engineering. Our data are publicly available and will encourage further studies in this direction.

Further, our data collection process revealed data quality issues reflecting an extensive degree of intra- and inter-journal inconsistency with regards to how journal publishers organize their articles when making their data available for research. This data quality issue can discourage interest and effort in studies of research methods in the LIS field. We therefore urge journal publishers to address these issues by making their article categorization system more transparent and consistent among themselves.

Our future work will focus on a number of directions. First, we aim to progress towards developing a hierarchical, structured method classification scheme reflecting different perspectives in LIS. This will address the limitations of our current, flat method classification scheme proposed in this work. Second, as discussed before, we aim to further develop our automated method by incorporating more complex features that may improve its accuracy and enabling it to capture other aspects of research methods, such as the data sets involved and their quantity.

Q7

DATA AVAILABILITY

The data are available at Zenodo (https://doi.org/10.5281/zenodo.4486156).

REFERENCES

Q8 Alam, H., Kumar, A., Werner, T., & Vyas, M. (2017). Are cited references meaningful? Measuring semantic relatedness in citation analysis. In *BIRNDL@SIGIR* (vol. 1888, pp. 113–118).

Augenstein, I., Das, M., Riedel, S., Vikraman, L., & McCallum, A. (2017). SemEval 2017 Task 10: ScienceIE—Extracting Keyphrases and Relations from Scientific Publications. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (pp. 546–555). **DOI:** https://doi.org/10.18653/v1/S17-2091

Bernhard, P. (1993). In search of research methods used in information science. *Canadian Journal of Information and Library Science*, *18*(3), 1–35.

Blake, V. (1994). Since Shaughnessy: Research methods in library and information science dissertation, 1975–1989. *Collection Development*, *19*(1/2), 1–42. **DOI:** https://doi.org/10.1300/J105v19n01_01

Campbell, D., & Stanley, J. (1966). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Chen, Y., Liu, Y., & Ho, W. (2013). A text mining approach to assist the general public in the retrieval of legal documents. *Journal of the American Society for Information Science and Technology*, *64*(2), 280–290. **DOI:** https://doi.org/10.1002/asi.22767

Chiticariu, L., Li, Y., & Reiss, F. (2013). Rule-based information extraction is dead! Long live rule-based information extraction systems! In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 827–832). Association for Computational Linguistics.

Chu, H. (2015). Research methods in library and information science: A content analysis. *Library and Information Science Research*, *37*, 36–41. **DOI:** https://doi.org/10.1016/j.lisr.2014.09.003

Cortez, E., da Silva, A., Goncalves, M., Mesquita, F., & de Moura, E. (2007). FLUX-CiM: Fexible unsupervised extraction of citation metadata. In *Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries* (pp. 215–224). **DOI:** https://doi.org/10.1145/1255175.1255219

Dayrell, C., Candido, A., Lima, G., Machado, D., Copestake, A., Feltrim, V., & Aluísio, S. (2012). Rhetorical move detection in English abstracts: Multi-label sentence classifiers and their annotated corpora. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (pp. 1604–1609).

Denning, J., Soledad, M., & Ng, P. (2015). A readability level prediction tool for K-12 books. *Journal of the American Society for Information Science and Technology*, *67*(3), 550–565. **DOI:** https://doi.org/10.1002/asi.23417

Durrant, G. (2004): A typology of research methods within the social sciences. *NCRM Working Paper Series*, pp. 1–22.

Ferran-Ferrer, N., Guallar, J., Abadal, E., & Server, A. (2017). Research methods and techniques in Spanish library and information science journals (2012–2014). *Information Research*, *22*(1), 741.

Fidel, R. (2008). Are we there yet?: Mixed methods research in library and information science. *Library & Information Science Research*, *30*, 265–272. **DOI:** https://doi.org/10.1016/j.lisr.2008.04.001

Freeburg, D. (2017). A knowledge lens for information literacy: Conceptual framework and case study. *Journal of Documentation*, *73*(5), 974–991. **DOI:** https://doi.org/10.1108/JD-04-2017-0058

Grankikov, V., Hong, Q., Crist, E., & Pluye, P. (2020). Mixed methods research in library and information science: A methodological review. *Library & Information Science Research*, *42*(1), 101003. **DOI:** https://doi.org/10.1016/j.lisr.2020.101003

Gupta, S., & Manning, C. (2011). Analyzing the dynamics of research by extracting key aspects of scientific papers. In *Proceedings of the 5th International Joint Conference on Natural Language Processing* (pp. 1–9). Asian Federation of Natural Language Processing.

Hanyurwimfura, D., Bo, L., Njogu, H., & Ndatinya, E. (2012). An automated cue word based text extraction. *Journal of Convergence Information Technology*, *7*(10), 421–429. **DOI:** https://doi.org/10.4156/jcit.vol7.issue10.50

Hider, P., & Pymm, B. (2008). Empirical research methods reported in high-profile LIS journal literature. *Library & Information Science Research*, *30*, 108–114. **DOI:** https://doi.org/10.1016/j.lisr.2007.11.007

Hirohata, K., Okazaki, N., Ananiadou, S., & Ishizuka, M. (2008). Identifying sections in scientific abstracts using conditional random fields. In *Proceedings of the Third International Joint Conference on Natural Language Processing* (pp. 381–388).

Houngb, H., & Mercer, R. (2012). Method mention extraction from scientific research paper. In *Proceedings of COLING: Technical paper*, 1211–1222.

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

Kim, M. (1996). Research record. *Journal of Education for Library and Information Science*, *37*(4), 376–383. **DOI:** https://doi.org/10.2307/40324247

Knoth, P., & Herrmannova, D. (2014). Towards semantometrics: A new semantic similarity based measure for assessing a research publications contribution. *D-Lib Magazine*, *20*(11), 8. **DOI:** https://doi.org/10.1045/november2014-knoth

Kovačević, A., Konjović, Z., Milosavljević, B., & Nenadic, G. (2012). Mining methodologies from NLP publications: A case study in automatic terminology recognition. *Computer Speech & Language*, *26*(2), 105–126. **DOI:** https://doi.org/10.1016/j.csl.2011.09.001

Kozinets, R. (2010). *Netnography: Doing ethnographic research online*. Newbury Park, CA: Sage.

Kumpulainen, K. (1991). Library and information science research in 1975. *Libri*, *41*(1), 59–76. **DOI:** https://doi.org/10.1515/libr.1991.41.1.59

Liakata, M., Saha, S., Dobnik, S., Batchelor, C., & Rebholz-Schuhmann, D. (2012). Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics*, *28*(7), 991–1000. **DOI:** https://doi.org/10.1093/bioinformatics/bts071, **PMID:** 22321698, **PMCID:** PMC3315721

Lin, S., Ng, J., Pradhan, S., Shah, J., Pietrobon, R., & Kan, M. (2010). Extracting formulaic and free text clinical research articles

metadata using conditional random fields. In *Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents* (pp. 90–95). Association for Computational Linguistics.

Luff, R., Byatt, D., & Martin, D. (2015). Review of the typology of research methods within the social sciences. *National Centre For Research Methods Report*.

Lustria, M., Kazmer. M., Glueckauf, R., Hawkins, R., Randeree, E., Rosario, I., McLaughlin, C., & Redmond, S. (2010). Participatory design of a health informatics system for rural health practitioners and disadvantaged women. *Journal of the American Society for Information Science and Technology*, 61(11), 2243–2255. **DOI:** https://doi.org/10.1002/asi.21390

Nasar, Z., Jaffry, S., & Malik, M. (2018). Information extraction from scientific articles: A survey. *Scientometrics*, 117, 1931–1990. **DOI:** https://doi.org/10.1007/s11192-018-2921-5

Newman, M. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6), 066133. **DOI:** https://doi.org/10.1103/PhysRevE.69.066133, **PMID:** 15244693

Newman, M., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113. **DOI:** https://doi.org/10.1103/PhysRevE.69.026113, **PMID:** 14995526

Park, S. (2004). The study of research methods in LIS education: Issues in Korean and U.S. universities. *Library & Information Science Research*, 26, 501–510. **DOI:** https://doi.org/10.1016/j.lisr.2004.04.009

QasemiZadeh, B., & Schumann, A. (2016). The ACL RD-TEC 2.0: A language resource for evaluating term extraction and entity recognition methods. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (pp. 1862–1868).

Risso, V. (2016). Research methods used in library and information science during the 1970–2010, *New Library World*, 117(1/2), 74–93. **DOI:** https://doi.org/10.1108/NLW-08-2015-0055

Ronzano, F., & Saggion, H. (2015). Dr. Inventor framework: Extracting structured information from scientific publications. In N. Japkowicz & S. Matwin (Eds.), *Discovery Science* (pp. 209–220). Cham: Springer. **DOI:** https://doi.org/10.1007/978-3-319-24282-8_18

Saarikoski, J., Laurikkala, J., Järvelin, K., & Juhola, M. (2009). A study of the use of self-organising maps in information retrieval. *Journal of Documentation*, 65(2), 304–322. **DOI:** https://doi.org/10.1108/00220410910937633

Seymore, K., McCallum, A., & Rosenfeld, R. (1999). Learning hidden Markov model structure for information extraction. In *Proceedings of the AAAI'99 Workshop Machine Learning for Information Extraction* (pp. 37–42).

Solomon, D., & Björk, B. (2012). A study of open access journals using article processing charges. *Journal of the American Society for Information Science and Technology*, 63(8), 1485–1495. **DOI:** https://doi.org/10.1002/asi.22673

Tahamtan, I., Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, 107, 1195–1225. **DOI:** https://doi.org/10.1007/s11192-016-1889-2

Tateisi, Y., Ohta, T., Pyysalo, S., Miyao, Y., & Aizawa, A. (2016). Typed entity and relation annotation computer science papers. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (pp. 3836–3843).

Teufel, S., & Moens, M. (2002). Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4), 409–445. **DOI:** https://doi.org/10.1162/089120102762671936

Togia, A., & Malliari, A. (2017). Research methods in library and information science. In S. Oflazoglu (Ed.) *Qualitative versus Quantitative Research* (pp. 43–64). London: IntechOpen. **DOI:** https://doi.org/10.5772/intechopen.68749

Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. **DOI:** https://doi.org/10.1007/s11192-009-0146-3, **PMID:** 20585380, **PMCID:** PMC2883932

Van Eck, N., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285–320). Cham: Springer.

VanScoy, A., & Fontana, O. (2016). How reference and information service is studied: Research approaches and methods. *Library & Information Science Research*, 38, 94–100. **DOI:** https://doi.org/10.1016/j.lisr.2016.04.002

Wallace, M., Gingras, Y., & Duhon, R. (2008). A new approach for detecting scientific specialties from raw cocitation networks. *Journal of the American Society for Information Science and Technology*, 60(2), 240–246. **DOI:** https://doi.org/10.1002/asi.20987

Wang, M., & Chai, L. (2018). Three new bibliometric indicators /approaches derived from keyword analysis. *Scientometrics*, 116(2), 721–750. **DOI:** https://doi.org/10.1007/s11192-018-2768-9

Whitley, R. (2000). *The intellectual and social organization of the sciences*. New York: Oxford University Press.

Zuccala, A., van Someren, M., & van Bellen, M. (2014). A machine-learning approach to coding book reviews as quality indicators: Toward a theory of megacitation. *Journal of the American Society for Information Science and Technology*, 65(11), 2248–2260. **DOI:** https://doi.org/10.1002/asi.23104

Q9

APPENDIX A

Keywords associated with each research method

Method	Indicative keywords/phrases
***From* Chu (2015)**	
bibliometrics	impact factor, scientometric, bibliometric, citation analysis/impact/importance/counts/index/report/window/rate/pattern/distributions/score/network, citation-based index, h-index, hindex, citers, citees, bibliometric indicator, leydesdorff, altmetrics
content analysis	content analysis, inter coder reliability, inter annotator agreement, krippendorff
delphi study	delphi study
ethnography/field study	Hammersley, participant observation, ethnography, ethnographic, ethnographer, field note, rich description
experiment	dependent variable, independent variable, experiment, experimental
focus group	focus group
historical method	historical method
interview	interview, interviewed, interviewer, interviewee, interviewing, transcript
observation	observation
questionnaire	respondent, questionnaire, survey, Likert, surveyed, sampling, response rate
research diary/journal	diary study, cultural probe
think aloud protocol	think aloud
transaction log analysis	log analysis/technique
webometrics	webometrics, cybermetrics, link analysis
New	
agent based modeling/simulation	agent model/modeling, multi-agents,
annotation	annotation, tagging
classification	classification, classify, classifier
clustering	cluster, clustering
comparative evaluation	comparative evaluation, evaluative studies
document analysis	document/textual analysis, document review

(*continued*)

Method	Indicative keywords/phrases
information extraction	named entity recognition, NER, relation extraction
IR related indexing/ ranking/query methods	Learning to rank, term weighting, indexing method, query expansion, question answering
mixed method	cresswell and plano clark, mixed method
digital ethnography	digital ethnography, netnography, netnographic,
network analysis	network analysis/study
statistical methods	correlation, logistic test, t-test, chi-square, hypothesis test, null hypothesis, dependence test
topic modeling	topic model, topic modeling, LDA
user task based study	user study, user analysis