How do Author-Selected Keywords Function Semantically in Scientific Manuscripts?†

Wei Lu*, Xin Li**, Zhifeng Liu***, Qikai Cheng****

Wuhan University, School of Information Management,
Information Retrieval and Knowledge Mining Laboratory, Wuhan 430072, China,
*<weilu@whu.edu.cn>, **<lucian@whu.edu.cn>, ***<zfliu17@163.com>, ****<chengqikai@whu.edu.cn>

Wei Lu is Professor of information science, Director of the Information Retrieval and Knowledge Mining Center, and Vice Dean of the School of Information Management, Wuhan University. He was also a visiting scholar at the City University of London, UK (2005-2006) and The Royal School of Library and Information Science, Denmark (2011-2012). His research interests are information retrieval, knowledge mining and visualization, knowledge organization and knowledge management. He is also an editorial advisory board member of the *Journal of Data and Information Science* and the *Journal of the China Society for Scientific and Technical Information*.

Xin Li is a doctoral student at the School of Information Management, Wuhan University. He is in the last year of a masters-doctorate combined program now. He received his bachelor's degree in health information management from Tongji Medical College, Huazhong University of Science and Technology (HUST) and in English literature from the School of Foreign Language, HUST. His research interests include health knowledge organization, bioinformatics, data mining and science of science.

Zhifeng Liu is a graduate student at the School of Information Management, Wuhan University. He received his bachelor's degree in information management and systems from Huazhong University of Science and Technology. His research interests include informetrics, text mining and medical informatics.

Qikai Cheng is a lecturer at the School of Information Management, Wuhan University. He received his doctoral degree in information science at Wuhan University in 2015. He is also a visiting scholar at the University of Pittsburgh and a researcher at the Information Retrieval and Knowledge Mining Center of Wuhan University. His research interests are natural language processing, information retrieval and text mining.

Lu, Wei, Xin Li, Zhifeng Liu and Qikai Cheng. 2019. "How do Author-Selected Keywords Function Semantically in Scientific Manuscripts?" *Knowledge Organization* 46(6): 403-418. 57 references. DOI:10.5771/0943-7444-2019-6-403.

Abstract: Author-selected keywords have been widely utilized for indexing, information retrieval, bibliometrics and knowledge organization in previous studies. However, few studies exist concerning how author-selected keywords function semantically in scientific manuscripts. In this paper, we investigated this problem from the perspective of term function (TF) by devising indicators of the diversity and symmetry of keyword term functions in papers, as well as the intensity of individual term functions in papers. The data obtained from the whole *Journal of Informetrics* (*JOI*) were manually processed by an annotation scheme of keyword term functions, including "research topic," "research method," "research object," "research area," "data" and "others," based on empirical work in content analysis. The results show, quantitatively, that the diversity of keyword term function decreases, and the irregularity increases with the number of author-selected keywords in a paper. Moreover, the distribution of the intensity of individual keyword term function indicated that no significant difference exists between the ranking of the five term functions with the increase of the number of author-selected keywords (i.e., "research topic" > "research method" > "research object" > "research area" > "data"). The findings indicate that precise keyword related research must take into account the distinct types of author-selected keywords.

Received: 11 April 2019; Revised: 21 June 2019; Accepted: 27 June 2019

Keywords: term functions, author-selected keywords, research topic

† This study was supported by the Major Project of the National Social Science Foundation of China (17&ZDA292) and the National Natural Science Foundation of China (71473183). The support provided by China Scholarship Council (CSC) during a visit of Xin Li to Indiana University Bloomington is also acknowledged. The authors would like to express special gratitude to Yi Bu and Yong Huang for their valuable comments and editorial assistance. The authors are also grateful to the anonymous referees and editors for their invaluable and insightful comments.

1.0 Introduction

Author-selected keywords are considered as a significant conduit of scientific concepts, ideas and knowledge (Cobo, López-Herrera, et al. 2011; Ding, Chowdhury, and Foo 2001; Névéol, Doğan, and Lu 2010; Van Raan and Tijssen 1993) and have been widely utilized in indexing, knowledge management, bibliometrics and information retrieval. For instance, a keyword co-occurrence network was constructed to map the knowledge structure of technology foresight research by (Su and Lee 2010). Khan and Wood (2015) conducted a co-keywords clustering to detect emerging themes in the information technology management domain. More recently, Wu (2016) adopted a keyword-based patent network approach to identify technological trends and evolution in the field of green energy. All of these studies can be summarized as "keyword analysis," whose general workflow entails data retrieval and collection, keywords identification and preprocessing, frequency counting, network generation, analysis and visualization, and interpretation and conclusion.

However, the undiscriminating use of keyword analysis remains controversial given the existence of certain problems such as the lack of an authoritative criterion for the selection of keywords (e.g., Chen and Xiao 2016; Milojević et al. 2011; Smiraglia 2013), the presence of possible bias due to the "indexer effect" (Michel Callon, Rip, and Law 1986; He 1999), ignoring semantic roles and their relationships between keywords (Wang et al. 2012) and the discipline attributes of keywords (Chen and Xiao 2016; J. Choi, Yi, and Lee 2011).

Actually, each author-selected keyword plays a specific semantic role or function, which can be called a "term function" (TF) in a scientific paper. Specifically, a keyword could be the topic discussed or the method adopted or it also could play another semantic role in a scientific paper. In most extant studies of keyword analysis, keywords that play different semantic roles that should have been weighted differently are treated as equally important by simple counting and aggregation for different tasks (Ferrara and Salini 2012). However, "topic," "domain," "method" and "application" keywords should have been assigned unequal weights for generating accurate research topic networks. Hence, to overcome these problems, the semantic function of author-selected keywords played in scientific manuscripts should be elucidated. In addition, understanding how the author-keywords function semantically in scientific manuscripts is also beneficial to the organization and indexing of scientific papers in databases and to determine papers' accessibility and citations in scientific communities.

The overall aim of this paper is to reveal the patterns of author-selected keywords in scientific papers from the perspective of term function, whose results will substantially contribute to the improvement of keyword indexing and keyword analysis. To realize this goal, the following research questions are posed:

1) What is the distribution of author-selected keyword term functions in scientific papers?
2) What is the regularity of the diversity and symmetry of author-selected keyword term functions in scientific papers?
3) What is the distribution of the intensity of individual keyword term functions in scientific papers?
4) What is the relationship between the author-selected keyword ranking and its term functions in scientific papers?

In this study, we first annotated term functions for all author-selected keywords in our dataset, for which an annotation scheme based on empirical work in content analysis is presented. Then, we introduced a framework to compute the diversity and symmetry of keyword term functions in a single paper, as well as the distribution of the intensity of individual keyword term functions, using concepts from network science and "true diversity," which can be understood as a normalization for the Shannon entropy. We also analyze the relationships between keyword rankings and keyword term functions.

The remainder of this paper is organized as follows. Section 2.0 reviews studies regarding author-selected keywords and term function (TF). Section 3.0 presents the dataset and the annotation scheme for keyword term function, as well as the framework to represent and evaluate the diversity, intensity and symmetry of author keyword term functions in papers. In Section 4.0, the main results of this study are described in detail. Finally, in Section 5.0, conclusions and directions for future work are presented.

2.0 Literature review

2.1 Author-selected keywords

Author keywords have been generally regarded as one of the most important forms of bibliographic metadata in bibliometrics and scientometrics, as well as being a significant conduit of scientific concepts, ideas and knowledge (Cobo, López-Herrera, et al. 2011; Ding, Chowdhury and Foo 2001; Névéol, Doğan and Lu 2010; Van Raan and Tijssen 1993). Therefore, author-selected keyword analysis has a long tradition of widespread application in hotspot detection, trend analysis and mapping the knowledge structures in both natural and social sciences, e.g., in environmental acidification (Law et al. 1988), polymer chemistry (M. Callon, Courtial and Laville 1991), chemical engineering (Peters and van Raan 1993), software engineering (Coulter, Monarch and Konda 1998), knowledge discovery (He 1999), in-

formation retrieval (Ding, Chowdhury and Foo 2001), ethics and dementia (Baldwin et al. 2003), geographic information system (GIS) (Tian, Wen and Hong 2008), biomedical science (Névéol, Doğan and Lu 2010), technology foresight (Su and Lee 2010), fuzzy sets theory (Cobo, López-Herrera et al. 2011b), tourism (B. Wu et al. 2012), strategic management (Keupp, Palmié and Gassmann 2012), information technology management (Khan and Wood 2015) and biofuels (Wu 2016). However, with the wide-ranging applications of author-selected keyword analysis, problems with the method have become increasingly evident and have begun to be actively discussed by researchers. For example, Callon, Rip and Law (1986) and He (1999) pointed out the "indexer effect" of author-selected keywords at a theoretical and technical level. More recently, Wang et al. (2012) suggested that experts' knowledge be integrated into the process of co-word analysis to improve precision; Chen and Xiao (2016) put forward methods for keyword selection that take keyword discrimination into account by considering their frequency both in and out of the domain. In this paper, we will analyze author-selected keywords of different term functions, which should have been weighted unequally in different bibliometric tasks.

Additionally, author-selected keywords have also been widely utilized for the classification and clustering of scientific documents (Jones and Mahoui 2000), the "gold-standard" for automatic keyword indexing and extraction (Matsuo and Ishizuka 2004; Ren 2014; Gil-Leiva 2017), automatic thesaurus development (Gil-Leiva and Alonso-Arroyo 2007; Tseng 2002; J. Wang 2006), the retrieval and recommendation of scientific papers in digital libraries (Lu and Kipp 2014; Schaffner 2009), citation counts prediction (Sohrabi and Iraj 2017; Uddin and Khan 2016) and the comparison with social tags (Y. Choi and Syn 2016; Lu and Kipp 2014).

2.2 Term function in scientific texts

Term function (TF) refers to the specific semantic role that a word, a term or a phrase plays in scientific texts (Xin, Qikai and Wei 2017), including "topic," "method," "technology," etc. For instance, in the paper entitled "Knowledge discovery through co-word analysis" (He 1999), the TF of the term "knowledge discovery" is a "topic"; whereas, for the term "co-word analysis," it is a "method." Notably, the TF of the same term can differ in different contexts, for example, the TF of the term "knowledge discovery" is a "method" in the article entitled "Intelligent query answering by knowledge discovery techniques" (Han et al. 1996). In addition, academic terms have numerous other functions according to different classifications, such as "goal," data and "application," which are also quite common in scientific contexts.

With the dramatic growth in the number of scientific publications, it has become a challenge to understand a scientific community by identifying important topics, methods, applications and the relations between them. In the extant literature, this question has been mainly addressed using bibliometric methods, for example, considering citation networks and topic models (Ding 2011; Song et al. 2014) and generating crude topic clustering based on contextual cues. However, several researchers concluded that these methods could not answer certain key questions, such as "what methods were used for a particular topic?" and pointed out that the need to identify the semantic roles of scientific terms by analyzing the text itself, i.e., the identification of the term function (TF) in scientific texts (Kondo et al. 2009; Tsai, Kundu and Roth 2013).

Identification of term functions has received increasing interest with the rapid development of natural language processing and machine learning. Key terms that play different semantic roles have been identified, such as the identification of "head," "goal" and "method" in research papers' titles based on a rule extracted from the structure of titles (Kondo et al. 2009), the recognition of "technology" and "effect" from research papers and patents based on machine learning (Nanba, Kondo and Takezawa 2010), the identification of "focus," "techniques" and "domain" from article abstracts by using semantic extraction patterns (Gupta and Manning 2011), the recognition of "techniques" and "application" from scientific literature using an unsupervised bootstrapping algorithm (Tsai, Kundu and Roth 2013) and the identification of "method" and "task" from scientific papers based on the Markov Logic Network (Huang and Wan 2013).

More recently, a comprehensive framework for term function in academic texts was presented by Xin, Qikai and Wei (2017). In his study, Cheng categorized term functions into "domain-independent term function" (including "topic" and "method" in three levels) and "domain-related term function" (different sub categories in different domains). Based on this classification, approaches have been used, including conditional random fields with word2vec and machine learning to rank, for automatic recognition of domain-independent term functions in scientific papers in computer science. In addition, Heffernan and Teufel (2018) presented an automatic classifier for identifying problems and solutions in scientific texts. It remains unknown, however, precisely how author-selected keywords function semantically in scientific manuscripts. Understanding qualitatively and quantitatively the patterns of author-selected keywords from term function perspectives, in our view, is of great benefit for improving keyword indexing and keyword analysis in bibliometric tasks.

3.0 Methodology

This section examines the overall process of the methodology, as illustrated in Figure 1. To investigate the author-selected keyword patterns, the approach is designed to be executed in four discrete steps: 1) data collection and pre-processing; 2) term function annotation; 3) indicator computing; and, 4) patterns analysis.

3.1 Step 1: data collection and processing

In this step, we collected the publication records from the *Journal of Informetrics* (*JOI*). To probe the author-selected keyword patterns from the term function (TF) perspective in scientific manuscripts, all 842 articles published between 2007 to 2017 from *JOI* were manually collected from the Web of Science. A total of 149 articles were excluded, because they were not articles but, for example, brief communications, book reviews, editorial statements, errata or critical remarks. Finally, 693 articles were selected as the dataset in this study. For each of these articles, we have not only obtained the author-selected keyword lists, but have also extracted the title, abstract and the hyperlink to its detailed information page for term function annotation in the subsequent step. To investigate the relationship between term functions and the ranking of keywords, we also recovered the position of each author-selected keyword in the keyword lists.

The distribution of the number of author-selected keywords per paper is shown in Figure 2. There are a total of

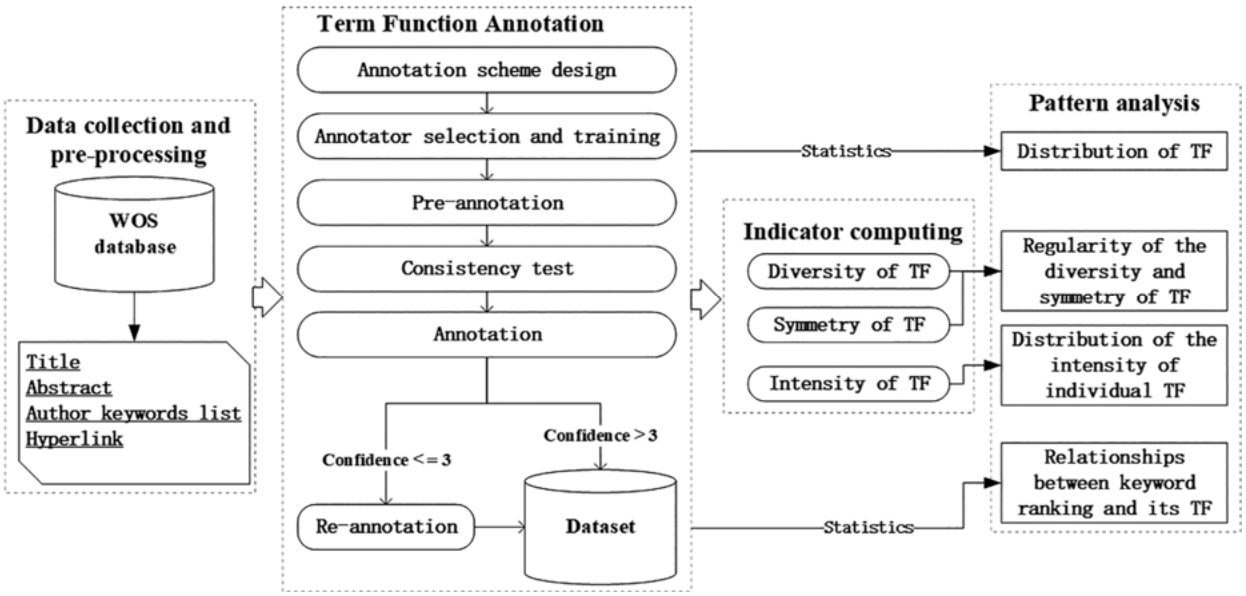

Figure 1. Framework of author keyword pattern analysis from the term function (TF) perspective.

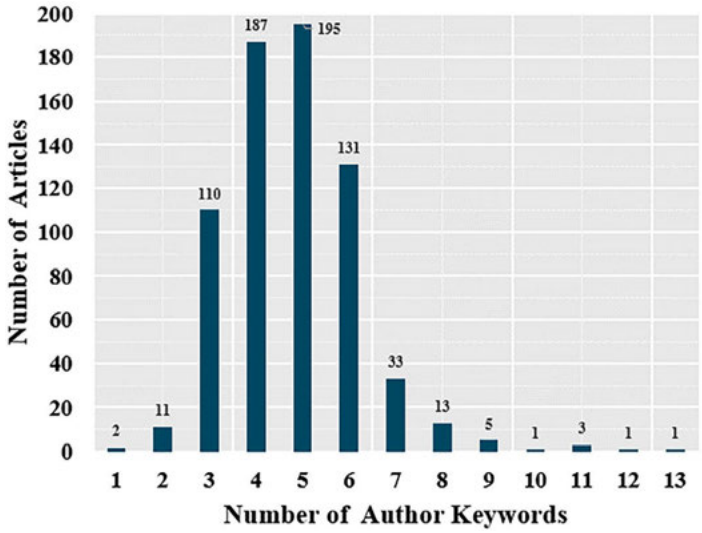

Figure 2. Histogram of the number of keywords in the *Journal of Informetrics* (*JOI*). An irregular distribution is found, in which most of the papers include three to six keywords.

3,311 author-selected keywords in all 693 articles, and the average number of author-selected keywords per article is found to be 4.78. It is also found that the range of author-selected keywords for each paper varied from one to thirteen. A few papers contained fewer than two keywords or more than eight keywords (approximately 1.9%), while most papers contained three to six keywords (approximately 89.9%).

3.2 Step 2: term function annotation

3.2.1 Annotation scheme design

In prior studies regarding term function recognition (TFR), words in scientific papers that have been recognized include "topic," "method," "problem," "solution," "goal," "technology," "focus," "domain," etc. (Heffernan and Teufel 2018; Xin, Qikai and Wei 2017; Tsai, Kundu and Roth 2013; Kondo et al. 2009; Gupta and Manning 2011; Huang and Wan 2013). Concerning the term function of each author-selected keyword in each article, we present an annotation scheme for author-selected keywords, based on empirical work in content analysis. In the first place, we captured all possible term functions of author-selected keywords. Then, to simplify our analysis, these term functions were integrated and reduced to a smaller set comprising only the most frequent term functions. This set, i.e., the annotation scheme for term functions of author-selected keywords includes the following categories: 1) research topic; 2) research method; 3) research object; 4) research area; 5) data; and, 6) others. The detailed description and source for each category of term function is shown in Table 1.

In order to guarantee the precision of term function annotation, the method of human annotation is selected. The term function of author-selected keywords is difficult to annotate, because, in principle, it requires interpretation of the author's intentions and the content of the entire paper. Consequently, in most cases, it is impossible to know exact term function without understanding academic context, because the same keyword can have a totally different term function in different conditions.

3.2.2 Annotators selection and training

Before term function annotating, four PhD students were selected from the School of Information management, Wuhan University. Four criteria were used in the selection of annotators. Specifically, the annotators had to: 1) be very familiar with informetrics and bibliometrics; 2) have good English reading and writing skills; 3) have published more than two academic articles in peer-reviewed journals in the field of informetrics; and, 4) be in or beyond their second year in the PhD program. Then, the selected annotators were trained and asked to point to textual evidence for assigning a particular term function.

3.2.3 Pre-annotation and consistency test

To guarantee annotation consistency, prior to starting the annotating, we randomly chose sixty-nine articles (9.96%) comprising of 337 author-selected keywords from the JOI dataset and arranged for four annotators to annotate term functions in two parallel groups. Then, the kappa coefficient (Carletta 1996), which is a statistic measuring pairwise

No.	**Categories**	**Description**	**Source**
1	Research Topic (T)	Problems or topics discussed in research articles.	Hoey 2013; Kondo et al. 2009; Heffernan and Teufel 2018; Xin, Qikai and Wei 2017)
2	Research Method (M)	Methods or solutions used in research articles, including theories, bibliometric indicators, algorithms, math formulas, models, etc. For examples, "Bradford's law," "h-index," "PageRank algorithm," "Hall's model."	Augenstein et al. 2017; Heffernan and Teufel 2018; Xin, Qikai and Wei 2017; Mesbah et al. 2017; Tsai, Kundu and Roth 2013; Sahragard and Meihami 2016
3	Research Object (O)	The object that the research studied, including people, group, organization, materials or objects.	Xin, Qikai and Wei 2017; Tsai, Kundu, and Roth 2013
4	Research Area (A)	The academic area or background of the article, for instance, "bibliometrics," "physics," "science of science," and "library and information science (LIS)."	Hoey 2013; Carletta 1996; Sahragard and Meihami 2016
5	Data (D)	The dataset used in the study or the data created by the study, for examples, "APS dataset," "X corpus," or "Web of Science," etc.	Kondo et al. 2009; Mesbah et al. 2017; Sahragard and Meihami 2016
6	Others (OT)	Cannot be included in the former categories.	Kondo et al. 2009; Xin, Qikai and Wei 2017

Table 1. The detailed description for each category of term function of author-selected keywords.

agreements among a set of coders' category judgements, was used for quantifying the consistency. Finally, the coefficients were 0.843 and 0.817 respectively (average 0.830 > 0.75), which was considered sufficiently high for annotating to proceed separately, particular given the conservative nature of the kappa coefficient.

3.2.4 Annotation

In the process of annotating, annotators were asked to carefully read the title and abstract for a comprehensive understanding of the academic context of each keyword in the original dataset and were encouraged to click the hyperlink for its full text to make a further confirmation. Moreover, annotators were asked to record the Annotation Confidence (ac) of each article. The value of $ac \in [1,2,3,4,5]$, in which a higher value of ac represents that the annotator is more confident in his or her work. If an article's value of ac is below four, the article will be annotated again by all annotators together.

3.3 Step 3: indicator computing

To quantify the intensity of individual term functions in a paper, as well as the diversity and symmetry of term functions of author-selected keywords in each article, the information provided in each article of our dataset is treated as a bipartite network (Newman 2010), which is a network with links established only among nodes and belonging to distinct groups. As shown in Figure 3, the bipartite network derived from each paper establishes links between author-selected keywords and their possible term functions. As can be seen from Figure 3, each author-selected keyword is annotated to one term function, while one term function can have multiple author-selected keywords assigned, which can represent the regularity of term functions of author-selected keywords in a paper.

3.3.1 Term function intensity

The term function intensity measure was used to calculate the strength of an individual term function in a scientific paper's author-selected keyword list. In this paper, we first define f as the matrix storing the relationship between author-selected keywords and their term functions in the bipartite network. The following equation was used:

$$f_{ij} = \begin{cases} 1, & \text{if the term function of author keyword } j \text{ is } i \\ 0, & \text{otherwise} \end{cases}$$

In the example provided in Figure 3, $f_{1j} = 1$ only for j = "1st keyword" and j = "4th keyword." Then the intensity of a given term function is given by the following equation:

$$TF\ Intensity\ (I_i) = \frac{\sum_j \omega_j f_{ij}}{\sum_i \sum_j \omega_j f_{ij}}$$

where ω_j is the weight associated to the j-th author-selected keyword. Differently from Edilson et al. (2017), we weighted the importance of each author-selected keyword to the research according to its rank in the keyword list, as defined by the following equation:

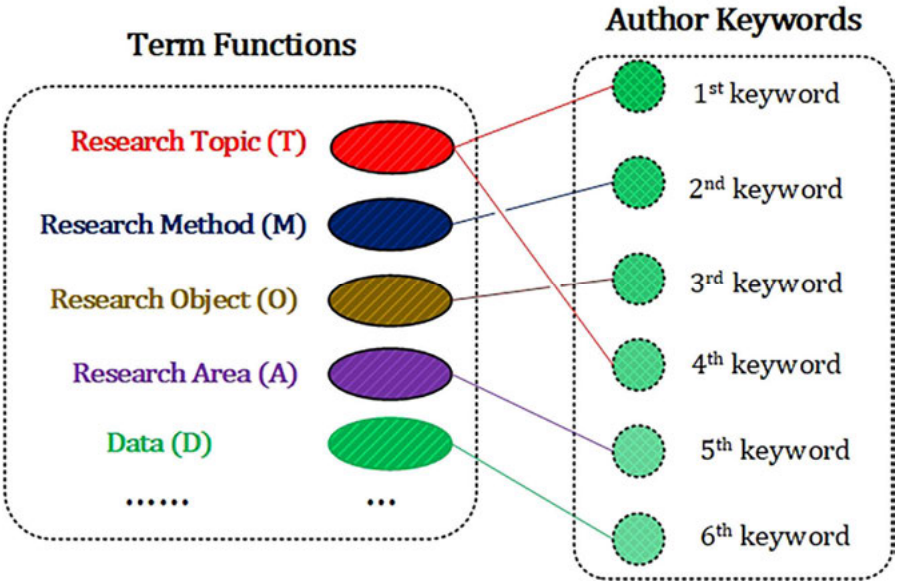

Figure 3. Example of a bipartite network representing the relationship between author-selected keywords and their term functions. Note that the total amount of keywords and particular term functions vary according to article.

$$\omega_j = \begin{cases} -0.1R + 0.6, & R < 5 \\ 0.1, & R \geq 5 \end{cases}$$

3.3.2 Term function diversity

For an article, the "term function diversity" measure calculates the level of variety of the term functions for author keyword lists. Drawing on the accessibility concept, a centrality measurement that can be understood as a normalization for the Shannon entropy was employed in this study. This measurement was originally proposed by Travençolo and Costa (2008) to compute the effective number of access nodes when an agent walks randomly on a network from a starting node. Compared to the traditional measurements, network features are used that go beyond the simple static network topology and can be utilized to quantify the effective number of neighbors (Amancio, Oliveira jr, and da F. Costa 2015). In this paper, a simple interpretation of the diversity measure in terms of network quantities was used to compute term function diversity, which has been extensively done in several studies (Corrêa Jr et al. 2017; Silva et al. 2016; Travençolo and Costa 2008). Notably, the "term function intensity" of each term function ranges in the interval [0,1], and thus we can measure its distribution of it using the entropy concept for all elements in the set of term functions. The following equation was then used to calculate the "term function diversity" of an article:

$$TF\ Diversity\ (\varphi) = \exp\left(-\sum_{i \in I} I_i log I_i\right)$$

3.3.3 Term function symmetry

The measure of "term function symmetry" examines the distributions of the "term function intensity" of each term function in a scientific paper. Thus, this measure represents how intensity varies across different term functions in a paper using a normalization of "term function diversity." The normalized TF diversity, referred to as a symmetry of the intensity of individual term function in a paper, takes a range of values restricted in the interval [0,1]. Therefore, the term function symmetry was represented by the following equation:

$$TF\ Sysmmetry\ (\sigma) = \frac{\varphi}{n_I}$$

where $n_I \in [1,6]$ is the total number of term functions in the paper. Note that σ is a symmetry measure, because it reaches its maximum value $(\sigma = 1)$ when all term functions are assigned equally to the paper.

3.4 Step 4: patterns analysis

In this paper, we reveal the patterns of author-selected keywords from four aspects. First, we described the distribution of author-selected keyword term functions using a statistical method. Second, the results of indicators including "term function diversity" and "term function symmetry" were employed to represent the regularity of author-selected keyword term functions in a scientific manuscript. Third, we also used the indicator "term function intensity" to depict the distribution of the strength of individual term functions in the dataset. Finally, the relationships between author-selected keyword ranking in the article's keyword list and their term functions were identified to analyze the author's potential indexing patterns.

Term Function	Percentage
Research Topic (T)	40.75%
Research Method (M)	37.79%
Research Object (O)	7.66%
Research Area (A)	9.55%
Data (D)	1.05%
Others (OT)	3.19%

Table 2. Frequency of appearance for each type of keyword term function, considering all of the papers in the dataset. Each author-selected keyword was counted as a distinct occurrence, even if it appeared in more than one paper in the dataset.

4.0 Results

4.1 The distribution of author-selected keyword term functions

The overall count for the author-selected keyword term functions in the dataset are shown in Table 2. The most common was "research topic," accounting for 40.75% of the total. "Research method" was a clear second, comprising more than a third of the total (37.79%). The other term functions scored between 7% and 10%, except for "data," which had very low frequency. In addition, the average number of "research topic[s]" per paper was 2.19, which is the highest among the five term functions. The average number of "research method[s]" per paper is 1.90, ranking second. The other term functions' average number per paper scored around 0.50, except for "data," whose average number was very low (0.18).

The distribution of the article numbers of different term functions in the dataset are presented in Figure 4 from which it can also be seen that "research topic" and "research method" are the top two term functions. We also find that the range of the number of "research topic" or "research method" for a paper varies from one to eight. A

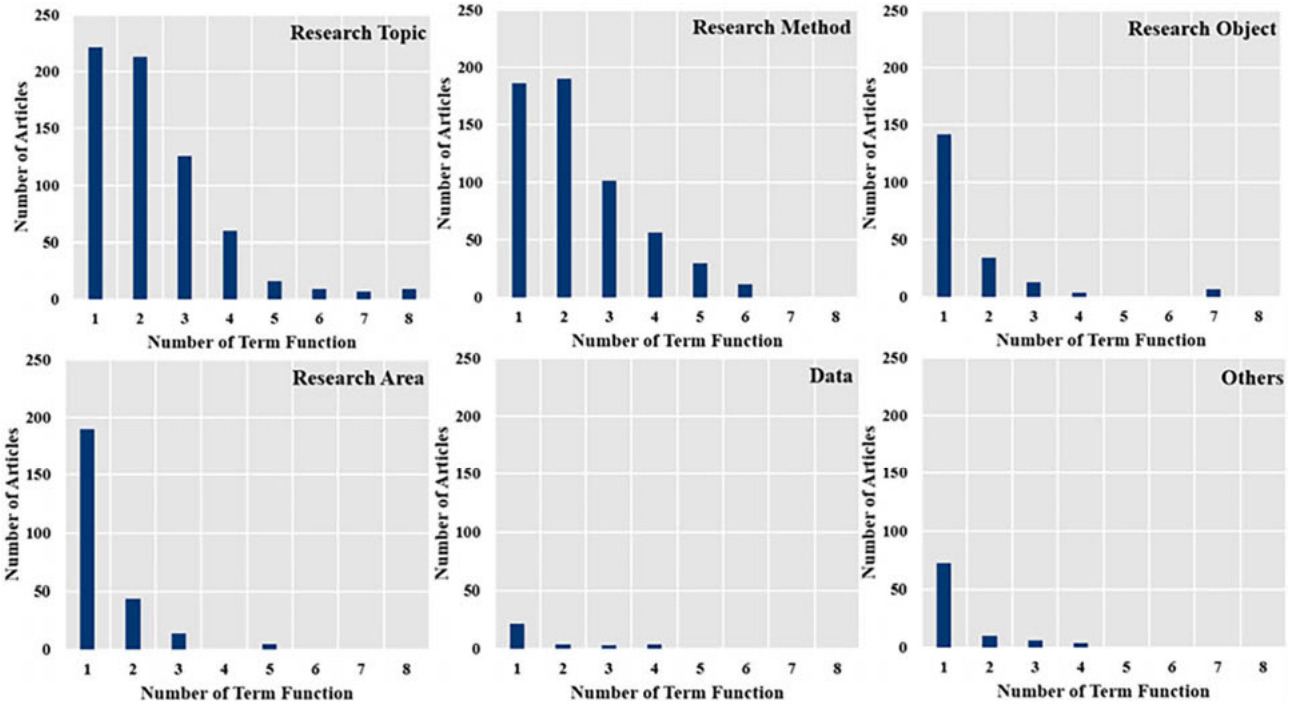

Figure 4. The distribution of the article numbers of different term functions in the dataset.

few papers (less than 15%) contain more than three "research topic" or "research method" keywords, while the most common scenario covers papers that contain one or two individual term functions (more than 50%). Moreover, the range of the number of the other three term functions for a paper is between one and four, while very few papers contain five "research area" or seven "research object" term functions. Most individual term functions have only occurred once in a paper, 70.6% for "research object," 75.1% for "research area" and 66.7% for "data."

4.2 The regularity of author-selected keyword term functions in papers

4.2.1 The diversity of author-selected keyword term functions

To investigate how keyword term functions vary in scientific papers, we used the "diversity of term functions in a paper" (φ) as a measure of the variability, as defined in Section 3.3.2. Considering that the value of $\omega_j f_{ij}$ varies according to the number of author-selected keywords (n_K) and the ranking of author-selected keywords (R), we decided to separately compute the values of φ for each n_K. As shown in Figure 5, the red line is the reference curve when the number of keywords assigned to each term function is equal; and if the keyword ranking (i.e. $\omega_j = 1$) is ignored, the reference will fit to the curve $\varphi = n_K$. The other curve denotes the points observed in our *JOI* dataset, from which one can find that, when n_K increases, the diversity of term functions of author-selected keywords in a paper also increases, thus confirming a relatively strong correlation between these quantities. Moreover, it reaches its highest point (φ is approximately 2.5) when the number of author-selected keywords is six. When $n_K = 1, \varphi = 1$, as one should anticipate from the equation above. One can also find that the largest deviations between these quantities ($i.e., n_k - \varphi$) were found for the papers tagged by many author-selected keywords. Note that, in general, the number of "research topic" keywords in papers tagged by more than eight author-selected keywords was usually more than five, which makes the diversity of term functions quite irregular. Considering that this set of articles has eight author-selected keywords, the paper with the most irregular distribution of keyword term functions has a total diversity of term functions of only approximately two. Despite these discrepancies, we can conclude that, in a typical paper tagged by three to six author-selected keywords, the diversity of term functions is relatively high and the difference between n_K and φ is relatively small, as the differences in the number of keyword term functions tagged in these studies is insignificant.

4.2.2 The symmetry of author-selected keyword term functions

The irregularity of author-selected keyword term functions was also investigated in terms of "symmetry of term

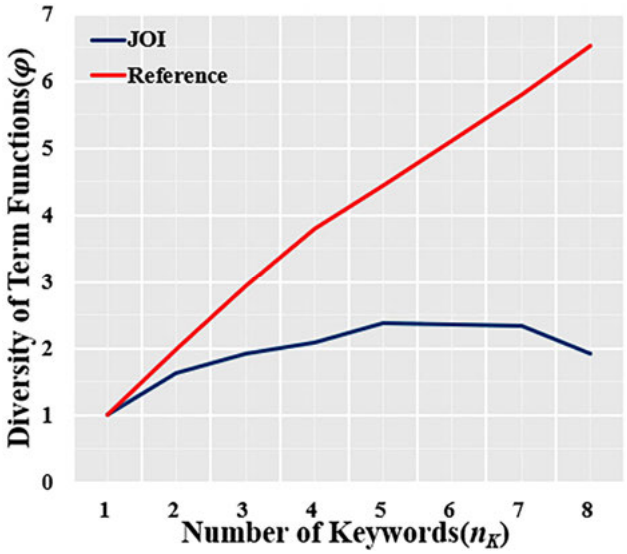

Figure 5. The diversity of author-selected keyword term functions (φ) as a function of the number of author-selected keywords (n_K). Because, in some cases, some term functions are tagged by more than others, their diversity of them is lower than the reference when each term function is tagged equally. The largest deviations occur for the papers tagged by many author-selected keywords.

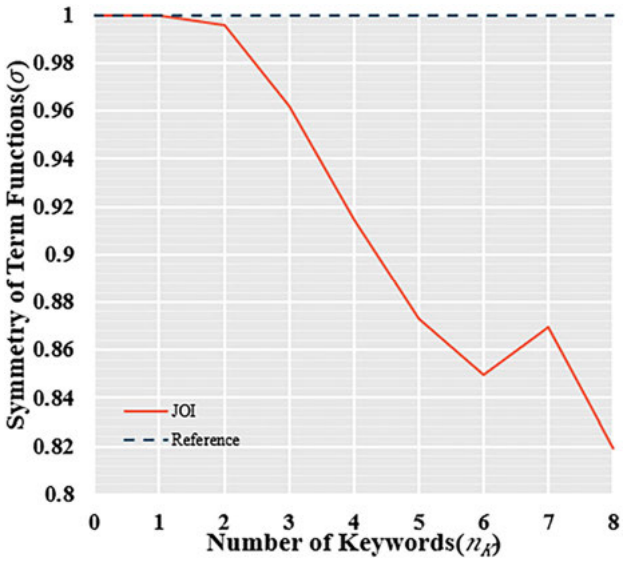

Figure 6. Symmetry of author-selected keyword term functions in papers (σ) as a function of the number of author-selected keywords (n_K).

functions" (σ), as defined in Section 3.3.3. As illustrated in Figure 6, for each value of n_K, we can obtain the corresponding value of TF symmetry. The blue dotted line is the reference line $\sigma_{\text{maximum}} = 1$, and the other line represents the curve obtained by linking the points representing the average symmetry obtained for each n_K, when $n_K = 1, \sigma = 1$. Overall, one can find that the average symmetry of author-selected keyword term functions monotonically decreases when the number of author-selected keywords increases from $n_K = 1$ to $n_K = 6$. However, when the number of author-selected keywords is more than five, the falling rate of the symmetry decreases significantly. This indicates that the distribution of keyword term functions becomes more irregular when the number of author-selected keywords increases. However, the average value of symmetry is always above 0.80. So, we

further count the number of papers whose symmetry is below 0.8 and find that most of them are in the $n_K = 4$ or $n_K = 5$ group. The reason for this phenomenon might be that due to the large number of papers tagged by four to five keywords, outliers are more common in this subset of papers. In addition, it is evident that values of $\sigma < 0.8$ are not frequent in the dataset with more than six or fewer than four keywords.

4.3 The distribution of the individual term function's intensity

In this section, we will investigate which term functions tend to be tagged more frequently by an author when indexing keywords for a scientific paper. Although no straightforward studies currently exist regarding this issue, the consensus among scientists is that the nature of a research process can be viewed as a problem-solving activity (Heffernan and Teufel 2018; Jordan 1980). When indexing keywords for a paper, authors are asked to use phrases that constitute an adequate description of the paper's content (Ding, Chowdhury and Foo 2001; Gil-Leiva 2017). A pertinent question is then which keywords are indexed more by authors, "research topic" or "research method"? "Data" is also of major significance to scientific research, especially in the field of information science, in which data constitute the essential materials. In addition, "research object" and "research area" are also essential for a rigorous design of scientific activity. This analysis illustrates the frequent occurrence of all five of these term functions of author-selected keywords. However, what are the differences among the five individual term functions according to the indexing behavior of authors?

To answer the question above, we described the distribution of the "intensity of individual term function" (I_i). We also analyzed the term function as a function of rankings to identify whether there is an implicit factor leading to the organization of rankings according to term functions.

In Figure 7, the distribution of the intensity of individual term functions of the *JOI* dataset is shown. The results are organized by the total number of author-selected keywords considered in Figure 7, with papers tagged by: a) 2; b) 3; c) 4; d) 5; e) 6; and, (f) all author-selected keywords in the *JOI* dataset. In Figure 7, as expected (Heffernan and Teufel 2018; Ding, Chowdhury and Foo 2001; Jordan 1980), it is evident that "research topic" and "research method," in general, obtain higher intensity than the others. Nonetheless, the values of TF intensity are not very different, since, on average, "research topic" and "research method" comprise approximately 40% and 30% of the intensity in paper level, respectively. When more author-selected keywords are included, one can observe a very similar pattern: while "research topic" obtain most of the in-

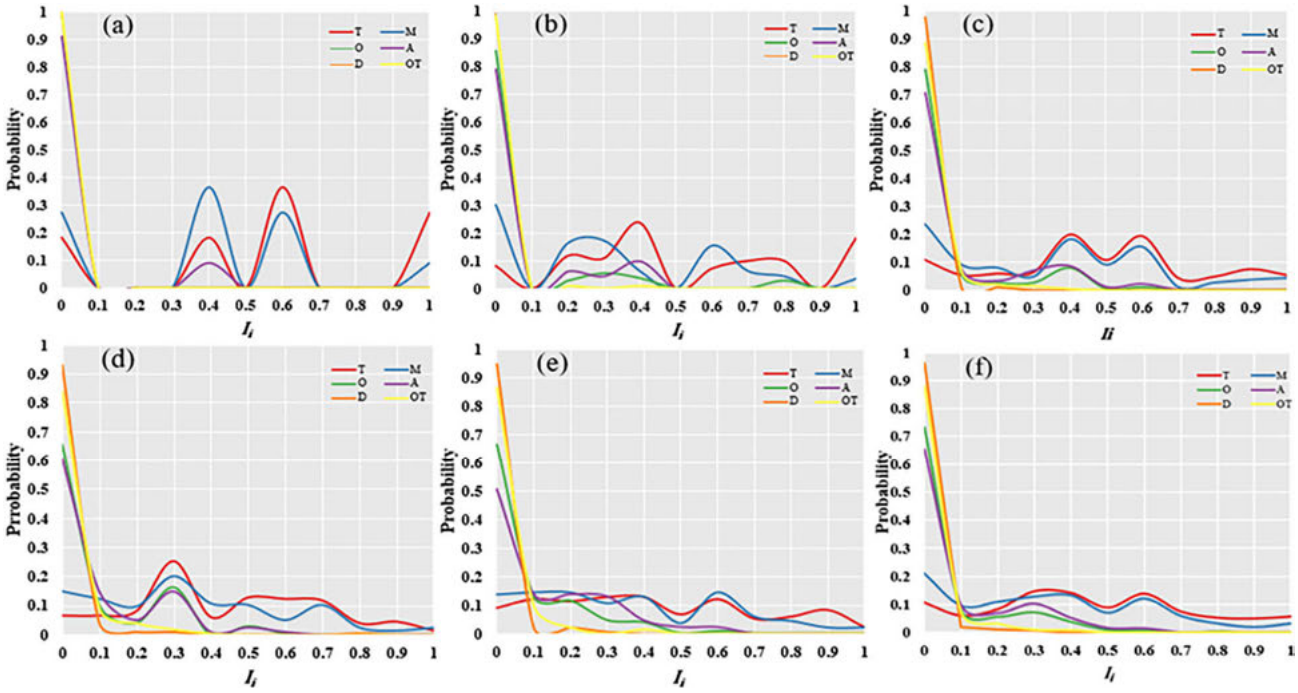

Figure 7. The distribution of individual term functions' intensity in the dataset. The results are shown considering the following number of author-selected keywords: a) 2; b) 3; c) 4; d) 5; e) 6; and, f) all. "Research topic" and "research method" are the first and second term functions, respectively, with a relative larger intensity value.

tensity, "research method" is usually ranked as the second most common term function; and "research object" (about 15%) and "research area" (10%) are third and fourth, respectively. "Data" has the least value of TF intensity in all conditions (less than 5%).

These patterns can also be observed in in Figure 8, which summarizes the average intensity of individual term function (I_i) in terms of the number of keywords. "Research topic" (upper red curve) always obtains most of the intensity, while "research method" usually appears in the second position in the ranking of average intensity. As the number of author-selected keywords increases, however, there is not a larger difference between the ranking of the five term functions on the value of TF intensity (i.e., "research topic" > "research method" > "research object" > "research area" > "data").

4.4 The relationship between the keyword's rank and its term function

It is conjectured that, in general, the first keywords are more frequently tagged as "research topic" or "research method," which are considered as the core part of a paper, while the last keywords have the least significance, such as "others." However, guidelines for ranking author-selected keywords are not always strictly followed, and thus there is no widespread evidence that exists relating ranking of author-selected keywords and specific term functions. To highlight the potential patterns in ranking keywords according to the type of their term functions, Figure 9 and Table 3 show the total amount of keywords in a particular ranking that made specific term functions. In Figure 9(a), it can be seen that, in papers tagged by only two author-selected keywords, both

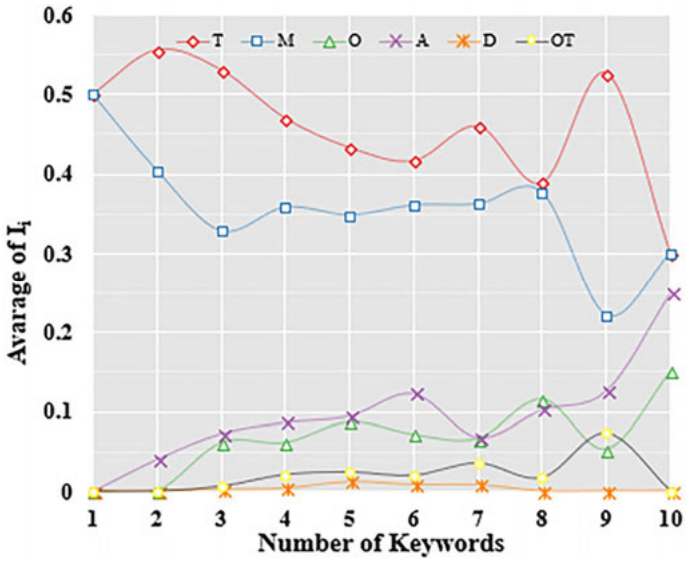

Figure 8. Average intensity of individual term function (I_i) as a function of the number of author-selected keywords (n_K) in the dataset. In general, "research topic" > "research method" > "research area" > "research object" > "data."

n_K	Term Function (TF)				
	Research Topic (T)	Research Method (M)	Research Object (O)	Research Area (A)	Data (D)
$n_K = 2$	1st>2nd	1st>2nd	1st>2nd	1st>2nd	1st>2nd
$n_K = 3$	3rd>2nd>1st	3rd>2nd>1st	1st>2nd>3rd	1st>2nd>3rd	1st>3rd>2nd
$n_K = 4$	1st>2nd>3rd>4th	3rd>4th>1st>2nd	1st>2nd>3rd>4th	1st>2nd>4th>3rd	2nd>4th>3rd>1st
$n_K = 5$	2nd>1st>3rd>4th>5th	4th>5th>3rd>2nd>1st	1st>2nd>3rd>4th>5th	1st>5th>2nd>4th>3rd	5th>4th>2nd>3rd>1st
$n_K = 6$	1st>2nd>3rd>4th>6th>5th	5th>6th>4th>3rd>2nd>1st	1st>2nd>3rd>4th>5th>6th	1st>2nd>6th>3rd>5th>4th	3rd>6th>1st>2nd>5th>4th

Table 3. The relationship between the number of author-selected keywords tagged as specific term functions and their rankings in author-selected keyword lists.

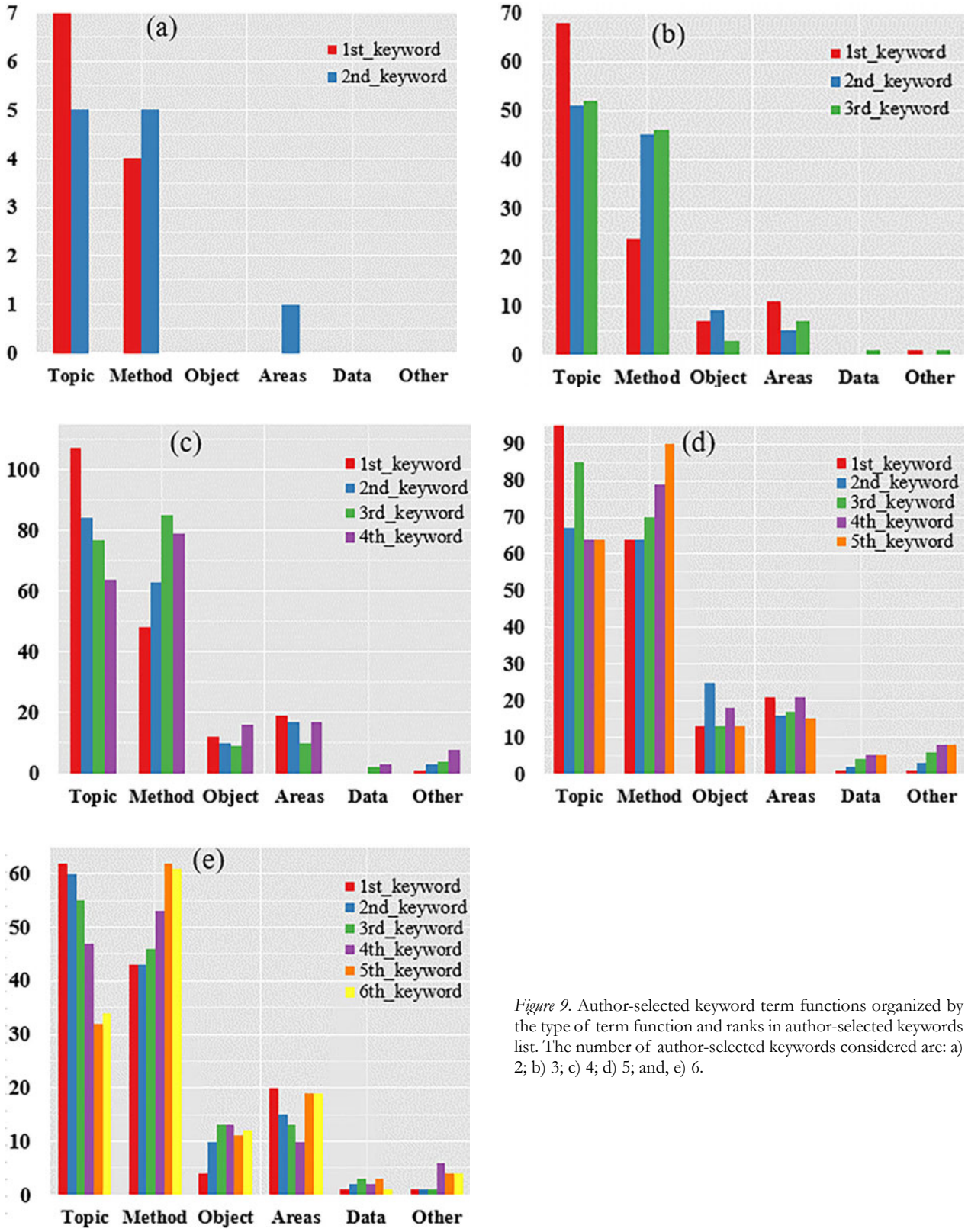

Figure 9. Author-selected keyword term functions organized by the type of term function and ranks in author-selected keywords list. The number of author-selected keywords considered are: a) 2; b) 3; c) 4; d) 5; and, e) 6.

keywords are usually tagged as "research topic," "research method," and "research area." However, in most cases, the first keywords are tagged as "research topic," as could be anticipated. Moreover, all of the "research areas" are tagged by the second author-selected keywords.

Specific term functions tagged by author-selected keywords in papers with three keywords are shown in Figure 9(b). Note that, when comparing the number of "research topic" and "research method," the proportions are very similar. However, when considering the number according to the ranking of author-selected keywords, the first keywords obtain the largest number (1st keywords > 3rd keywords > 2nd keywords in "research topic," 1st keywords > 2nd keywords > 3rd keywords in "research method," which is the same as "research area"), which is different with "research object" (2nd keywords > 1st keywords > 3rd keywords) and "data" (3rd keywords >1st keywords = 2nd keywords).

Regarding papers tagged by four author-selected keywords, as shown in Figure 9(c) it can be observed that, the least number of "research method" are tagged by first keywords. Interestingly, the second-to-last keywords take the largest number of "research method" (i.e., 3rd keywords > 4th keywords > 2nd keywords > 1st keywords). Similar patterns of contributions have also been found for papers tagged by five keywords (see Figure 9(d)) and six keywords (see Figure 9(e)). However, the first keywords are always the keywords that take the largest number of "research topic."

According to Figure 9 and Table 3, we can summarize the several patterns relating to author-selected keyword rankings and their term functions as follows:

1) Pattern I: Overall, the total amount of "research topic" and "research method" keywords possesses an absolute advantage over keywords of other term functions. More specifically, when the number of author-selected keywords is less than four, the total amount of "research topic" is predominant. Meanwhile, the total amount of "research method" increases rapidly from four to more keywords, and "research topic" and "research method" are almost equal. This pattern reveals the significance of topics and methods to a scientific research in the author's cognition, which is also in accordance with previous studies that interpret scientific research as a problem-solving activity (Heffernan and Teufel 2018; Jordan 1980). Interestingly, several studies maintain that the semantic role of all domain-independent terms in a scientific paper can be divided into topics or methods (Xin, Qikai and Wei 2017).
2) Pattern II: Different keyword term functions have their own preferential positions in author-selected keyword lists, although all of these keyword term functions can appear at every position. Specifically, "research topic" tends more to be tagged by keywords at the first three positions (i.e., 1st, 2nd, and 3rd keyword in the list, see Figure 9). Conversely, "research method" keywords are more likely to appear at the last two keywords in the list. Moreover, the first two and the last two positions are where "research area" keywords always occur, which exhibits a symmetric behavior as a function of keyword ranking.
3) Pattern III: The number of "research topic" keywords approximately decreases with keyword ranking, while the number of "research method" keywords increases with keyword ranking. This indicates that it is easier for authors to think of the topic of the research than the methods used in the study when they index keywords.

On the whole, it can be concluded that the keyword ranking, and its term function are strongly related by evidence of the aforementioned patterns. These patterns also confirm that there is no obvious relationship between the intensity and ranking of keyword term functions, although the rank of keywords is weighted in this study, as shown in Section 3.3.1. For example, "research topic" ranked in the first positions and has the maximum intensity, on average; whereas, "research method" obtains the second largest intensity and is always tagged by last two keywords in the list. Meanwhile, from pattern I, one can find that the key factor that affects the value of intensity of individual term functions is the number of specific term function keywords in author-selected keywords lists. In addition, we note here that, since the scale of "data" keywords is very small, no obvious regularity is found.

5.0 Conclusion and future work

Although author-selected keywords have long been utilized in knowledge organization, information retrieval, social tags, keyword extraction, indexing and thesaurus development, few studies have investigated the patterns of author-selected keywords in scientific papers. However, for a more fine-grained indexing and retrieval of scientific papers, for example, retrieving studies in which co-word analysis comprises the "research topic" but not "research method," it is necessary to identify the term functions of keywords in scientific papers. Additionally, analyzing the patterns of author-selected keywords from the term function perspective also constitutes the basis for the construction of a semantic network of keywords, which will be of great significance for knowledge organization and traditional bibliometric tasks, such as hot spot identification, trends analysis and mapping the knowledge structure of hard sciences and social sciences. Therefore, in this paper, we have mainly analyzed the potential patterns of author-selected keywords from the perspective of term function (TF).

The main contributions of this study are threefold. First, in order to investigate the patterns of author-selected keywords in scientific manuscripts, this paper, by treating the relationship between author-selected keywords and term functions as a bipartite network, proposes a new method based on the concept of accessibility and true diversity to quantify the diversity and symmetry of keyword term functions (φ and σ) at the paper level and the intensity of individual term function (I_i) at the function level. These measures can effectively describe the irregularity of author-selected keywords from the term function perspective. Second, this study also found that a strong relationship exists between a keyword's ranking and its term function. We confirmed that "research topic" and "research method" keywords are the most frequent in scientific papers. Despite this well-known pattern, three patterns of author-selected keywords are also found, depending on the relationship between the amount of specific term function keywords and their rankings. For instance, "research topic" tended to be tagged more by keywords at the first three positions. Interestingly, "research method" keywords were more likely to appear at the last two keywords in the list, which indicates that there is no obvious relationship between the intensity and ranking of keyword term functions. Third, we also designed an annotation scheme for author-selected keyword term functions, with which a corpus comprising 3,311 author-selected keywords from 693 scientific papers (all original research papers published between 2007 and 2017 in the *Journal of Informetrics*) are obtained with rigorous human annotation. Great care was taken in constructing this corpus by professionals to ensure the quality. Hence, this corpus could be valuable for the tasks of term function recognition, keyword extraction and more fine-grained co-word network analysis in the further study.

The results of this study should be interpreted in the context of its limitations. The main defect is that we analyzed the author-selected keywords only from the field of informetrics and bibliometrics. The reason for this is that the annotation of term functions manually for keywords is difficult due to its huge workload to interpret author intentions and the content of the whole article. In the future, we will perform studies that analyze and compare patterns of author-selected keywords among different natural sciences and social sciences. Furthermore, we will also investigate the patterns of other kinds of keywords from the perspective of term function, for example, KeyWords Plus in the Web of Science or *MeSH* (*Medical Subject Headings*) terms in PubMed. Finally, we raise an open-ended question of whether the diversity of keyword term functions (φ), the symmetry of keyword term functions (σ) and the intensity of individual term function (I_i) can affect scientific papers' citations. We believe that much room still exists for further research, and we anticipate interesting results in consequent work.

References

Amancio, Diego R., Osvaldo N. Oliveira jr and Luciano da F. Costa. 2015. "Topological-Collaborative Approach for Disambiguating Authors' Names in Collaborative Networks." *Scientometrics* 102: 465-85. doi:10.1007/s11192-014-1381-9

Augenstein, Isabelle, Mrinal Das, Sebastian Riedel, Lakshmi Vikraman and Andrew McCallum. 2017. In *Proceedings of the 11th International Workshop on Semantic Evaluation SemEval-2017*, ed. Steven Bethard, Marine Carpuat, Marianna Apidianaki, Saif M. Mohammad, Daniel Cer and David Jurgens. Vancouver, BC: Association for Computational Linguistics, 546–55. doi:10.18653/v1/S17-2091

Baldwin, Clive, Julian Hughes, Tony Hope, Robin Jacoby and Sue Ziebland. 2003. "Ethics and Dementia: Mapping the Literature by Bibliometric Analysis." *International Journal of Geriatric Psychiatry* 18: 41-54. doi:10.1002/gps.770

Callon, M., J. P. Courtial and F. Laville. 1991. "Co-Word Analysis as a Tool for Describing the Network of Interactions between Basic and Technological Research: The Case of Polymer Chemsitry." *Scientometrics* 22: 155-205. doi:10.1007/BF02019280

Callon, Michel, Arie Rip and John Law. 1986. *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. Cham: Springer.

Carletta, Jean. 1996. "Assessing Agreement on Classification Tasks: The Kappa Statistic." *Computational Linguistics* 22: 249-54.

Chen, Guo and Lu Xiao. 2016. "Selecting Publication Keywords for Domain Analysis in Bibliometrics: A Comparison of Three Methods." *Journal of Informetrics* 10: 212-23.

Choi, Jinho, Sangyoon Yi and Kun Chang Lee. 2011. "Analysis of Keyword Networks in MIS Research and Implications for Predicting Knowledge Evolution." *Information & Management* 48: 371-81.

Choi Youngok and Sue Yeon Syn. 2016. "Characteristics of Tagging Behavior in Digitized Humanities Online Collections." *Journal of the Association for Information Science and Technology* 67: 1089-104.

Cobo, Manolo J, Antonio Gabriel López-Herrera, Enrique Herrera-Viedma and Francisco Herrera. 2011. "An Approach for Detecting, Quantifying and Visualizing the Evolution of a Research Field: A Practical Application to the Fuzzy Sets Theory Field." *Journal of Informetrics* 5: 146-66.

Cobo, Manolo J, Antonio Gabriel López-Herrera, Enrique Herrera-Viedma and Francisco Herrera. 2011. "Science Mapping Software Tools: Review, Analysis and Cooperative Study among Tools." *Journal of the American Society for Information Science and Technology* 62: 1382-402.

Corrêa Jr, Edilson A, Filipi N Silva, Luciano da F Costa and Diego R Amancio. 2017. "Patterns of Authors

Contribution in Scientific Manuscripts." *Journal of Informetrics* 11: 498-510.

Coulter, Neal, Ira Monarch and Suresh Konda. 1998. "Software Engineering as Seen through Its Research Literature: A Study in Co-word Analysis." *Journal of the American Society for Information Science* 49: 1206-23.

Ding, Ying. 2011. "Topic-based PageRank on Author Cocitation Networks." *Journal of the American Society for Information Science and Technology* 62: 449-66.

Ding, Ying, Gobinda G Chowdhury and Schubert Foo. 2001. "Bibliometric Cartography of Information Retrieval Research by Using Co-Word Analysis." *Information Processing & Management* 37: 817-42.

Ferrara, Alfio and Silvia Salini. 2012. "Ten Challenges in Modeling Bibliographic Data for Bibliometric Analysis." *Scientometrics* 93: 765-85.

Gil-Leiva, Isidoro and Adolfo Alonso-Arroyo. 2007. "Keywords given by Authors of Scientific Articles in Database Descriptors." *Journal of the American Society for Information Science and Technology* 58: 1175-87.

Gil-Leiva, Isidoro. 2017. "SISA-Automatic Indexing System for Scientific Articles: Experiments with Loca- tion Heuristics Rules Versus TF-IDF Rules." *Knowledge Organization* 44: 139-62.

Gupta, Sonal and Christopher Manning. 2011. "Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers." In *Proceedings of Fifth International Joint Conference on Natural Language Processing 8-13 Noverber 2011 Chiang Mai, Thailand.* Asian Federation of Natural Language Processing, 1-9. https://www.aclweb.org/anthology/I11-1

Han, Jiawei, Yue Huang, Nick Cercone and Yongjian Fu. 1996. "Intelligent Query Answering by Knowledge Discovery Techniques." *IEEE Transactions on Knowledge and Data Engineering* 8: 373-90.

He, Qin. 1999. "Knowledge Discovery through Co-Word Analysis." *Library Trends* 48, no. 1: 133-59.

Heffernan, Kevin and Simone Teufel. 2018. "Identifying Problems and Solutions in Scientific Text." *Scientometrics* 116: 1367-82. doi:10.1007/s11192-018-2718-6

Hoey, Michael. 2013. *Textual Interaction: An Introduction to Written Discourse Analysis*. Abingdon: Routledge.

Huang, Shanshan and Xiaojun Wan. 2013. "AKMiner: Domain-Specific Knowledge Graph Mining from Academic Literatures." In *Proceedings of 14th International Conference on Web Information Systems Engineering October 2013, Nanjing China.* Cham: Springer, 241-55.

Jones, Steve and Malika Mahoui. 2000. "Hierarchical Document Clustering Using Automatically Extracted Keyphrases." In *Proceedings of the Third International Asian Conference on Digital Libraries, Seoul, Korea.* Berkeley, CA: ACM Press, 113-20.

Jordan, Michael P. 1980. "Short Texts to Explain Problem-Solution Structures and Vice Versa." *Instructional Science* 9: 221-52.

Keupp, Marcus Matthias, Maximilian Palmié and Oliver Gassmann. 2012. "The Strategic Management of Innovation: A Systematic Review and Paths for Future Research." *International Journal of Management Reviews* 14: 367-90

Khan, Gohar Feroz and Jacob Wood. 2015. "Information Technology Management Domain: Emerging Themes and Keyword Analysis." *Scientometrics* 105: 959-72.

Kondo, Tomoki, Hidetsugu Nanba, Toshiyuki Takezawa and Manabu Okumura. 2009. "Technical Trend Analysis by Analyzing Research Papers' Titles." In *Human Language Technology. Challenges for Computer Science and Linguistics: 4th Language and Technology Conference, LTC 2009, Poznan, Poland, November 6-8, 2009, Revised Selected Papers.* Lecture Notes in Computer Science 6562. Lecture Notes in Artificial Intelligence 6562. Berlin: Springer, 512-21. doi:10.1007/978-3-642-20095-3_47

Law, John, Serge Bauin, J Courtial and John Whittaker. 1988. "Policy and the Mapping of Scientific Change: A Co-Word Analysis of Research into Environmental Acidification." *Scientometrics* 14: 251-64.

Lu, Kun and Margaret E. I. Kipp. 2014. "Understanding the Retrieval Effectiveness of Collaborative Tags and Author Keywords in Different Retrieval Environments: An Experimental Study on Medical Collections." *Journal of the Association for Information Science and Technology* 65: 483-500.

Matsuo, Yutaka and Mitsuru Ishizuka. 2004. "Keyword Extraction from a Single Document Using Word Co-Occurrence Statistical Information." *International Journal on Artificial Intelligence Tools* 13: 157-69.

Mesbah, Sepideh, Kyriakos Fragkeskos, Christoph Lofi, Alessandro Bozzon and Geert-Jan Houben. 2017. "Facet Embeddings for Explorative Analytics in Digital Libraries." In *Research and Advanced Technology for Digital Libraries: 21st International Conference on Theory and Practice of Digital Libraries, TPDL 2017, Thessaloniki, Greece, September 18-21, 2017, Proceedings,* ed. Jaap Kamps, Giannis Tsakonas, Yannis Manolopoulos, Lazaros Iliadis and Ioannis Karydis. Lecture Notes in Computer Science 10450. Information Systems and Applications 10450. Cham: Springer, 86-99.

Milojević, Staša, Cassidy R Sugimoto, Erjia Yan and Ying Ding. 2011. "The Cognitive Structure of Library and Information Science: Analysis of Article Title Words." *Journal of the American Society for Information Science and Technology* 62: 1933-53.

Nanba, Hidetsugu, Tomoki Kondo and Toshiyuki Takezawa. 2010. "Automatic Creation of a Technical Trend Map from Research Papers and Patents." In *Proceedings of the 3rd International Workshop on Patent Information Retrieval*

26 October 2010 Toronto, Onatario, Canada. New York: ACM, 11-16. doi:10.1145/1871888.1871891

Névéol, Aurélie, Rezarta Islamaj Doğan and Zhiyong Lu. 2010. "Author Keywords in Biomedical Journal Articles." In *AMIA Annual Symposium Proceedings 2010*. Bethesda, MD: AMIA, 537-41.

Newman, Mark. 2010. *Networks: An Introduction*. Oxford: Oxford University Press.

Peters, H. P. F. and Anthony F. J. van Raan. 1993. "Co-Word-Based Science Maps of Chemical Engineering. Part I: Representations by Direct Multidimensional Scaling." *Research Policy* 22: 23-45.

Raan, Anthony F. J. van and Robert J. W. Tijssen. 1993. "The Neural Net of Neural Network Research: An Exercise in Bibliometric Mapping." *Scientometrics* 26: 169-92. doi:10.1007/bf02016799

Ren, Feiliang. 2014. "An Unsupervised Cascade Learning Scheme for 'Cluster-Theme Keywords' Structure Extraction from Scientific Papers." *Journal of Information Science* 40: 167-79.

Sahragard, Rahman and Hussein Meihami. 2016. "A Diachronic Study on the Information Provided by the Research Titles of Applied Linguistics Journals." *Scientometrics* 108: 1315-31.

Schaffner, Jennifer. 2009. *The Metadata Is the Interface: Better Description for Better Discovery of Archives and Special Collections: Synthesized from User Studies*. Dublin, OH: OCLC Programs and Research.

Silva, Filipi N., Diego R. Amancio, Maria Bardosova, Luciano da F. Costa and Osvaldo N. Oliveira Jr. 2016. "Using Network Science and Text Analytics to Produce Surveys in a Scientific Topic." *Journal of Informetrics* 10: 487-502.

Smiraglia, Richard P. 2013. "Keywords, Indexing, Text Analysis: An Editorial." *Knowledge Organization* 40: 155-9.

Sohrabi, Babak and Hamideh Iraj. 2017. "The Effect of Keyword Repetition in Abstract and Keyword Frequency per Journal in Predicting Citation Counts." *Scientometrics* 110: 243-2.51.

Song, Min, SuYeon Kim, Guo Zhang, Ying Ding and Tamy Chambers. 2014. "Productivity and Influence in Bioinformatics: A Bibliometric Analysis Using PubMed Central." *Journal of the Association for Information Science and Technology* 65: 352-71.

Su, Hsin-Ning and Pei-Chun Lee. 2010. "Mapping Knowledge Structure by Keyword Co-Occurrence: A First Look at Journal Papers in Technology Foresight." *Scientometrics* 85: 65-79.

Tian, Yangge, Cheng Wen and Song Hong. 2008. "Global Scientific Production on GIS Research by Bibliometric Analysis from 1997 to 2006." *Journal of Informetrics* 2: 65-74.

Travençolo, Bruno Augusto Nassif and L da F Costa. 2008. "Accessibility in Complex Networks." *Physics Letters A* 373: 89-95.

Tsai, Chen-Tse, Gourab Kundu and Dan Roth. 2013. "Concept-Based Analysis of Scientific Literature." In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. New York: ACM, 1733-8. doi:10.1145/2505515.2505613

Tseng, Yuen-Hsien. 2002. "Automatic Thesaurus Generation for Chinese Documents." *Journal of the American Society for Information Science and Technology* 53: 1130-8.

Uddin, Shahadat and Arif Khan. 2016. "The Impact of Author-Selected Keywords on Citation Counts." *Journal of Informetrics* 10: 1166-77.

Wang, Jun. 2006. "Automatic Thesaurus Development: Term Extraction from Title Metadata." *Journal of the American Society for Information Science and Technology* 57: 907-20.

Wang, Zhong-Yi, Gang Li, Chun-Ya Li and Ang Li. 2012. "Research on the Semantic-Based Co-Word Analysis." *Scientometrics* 90: 855-75.

Wu, Bihu, Honggen Xiao, Xiaoli Dong, Mu Wang and Lan Xue. 2012. "Tourism Knowledge Domains: A Keyword Analysis." *Asia Pacific Journal of Tourism Research* 17: 355-80.

Wu, Chao-Chan. 2016. "Constructing a Weighted Keyword-Based Patent Network Approach to Identify Technological Trends and Evolution in a Field of Green Energy: A Case of Biofuels." *Quality & Quantity* 50: 213-35.

Xin, L., Qikai C. and Wei L. 2017. "CS-LAS: A Scientific Literature Retrieval and Analysis System Based on Term Function Recognition (TFR)." In *Proceedings of 16th International Conference of the International Society for Scientometrics and Informetrics 16-17 March 2017, Wuhan, China*. Wuhan: Wuhan University Press, 1346-56.