Terminological Analysis of Publications as a Method of Research Trends in Science

Mokhnacheva Yu.V., Tsvetkova V.A
Library for Natural Sciences of the Russian Academy of Sciences
119991, Russia, Moscow, Znamenka 11/11
{j-v-m@yandex.ru} {vats08@mail.ru}

ABSTRACT: *This article proposes a terminological approach to understanding the dynamics of the development of scientific topics based on the analysis of the sharing of key terms, which is the basis for the development of an algorithm for the distribution of scientific topics into categories: actively developing topics; topics that are in a stable state; topics that are losing relevance or fading away. A hypothesis is proposed that the more keywords with dynamics, greater than 0% in the topic, the higher the probability that this topic is promising and is actively developing. Conversely, the more essential terms in a topic with negative dynamics, the more likely it will indicate the researcher's interest decreases. The subject area Immunology & Microbiology in SciVal was chosen as a model for studying the dynamics of the development of scientific areas. We assume that if a critical term reduces its usage in one topic, it may appear and grow in other ones. The relevance of various vital terms changes over time for different topics, both in negative and positive dynamics. Using the example of the term "Bibliometric analysis," the dynamics of this crucial term in other topics are shown.*

.

Keywords: Terminological Analysis, Trends In Science, Dynamics of Keywords, *Scival* Topics, Scopus

Received: 18 August 2022, Revised 25 September 2022, Accepted 3 October 2022

DOI: 10.6025/jcl/2022/13/4/83-87

1. Introduction

Research topics are most often based on analyzing knowledge bases (Chen C. et al. 2006). Understanding the evolution of scientific issues is critical to identifying new trends in science (Wang X.G. et al., 2014). Kontostathis A. (Kontostathis A. et al. 2004) suggests the concept of detecting recent trends in scientific fields (or Emerging Trend Detection, ETD) through the identification of new topics and the revealing of their interrelationships based on both citation and common usage of keywords (co-words). Various methods have been developed to identify research trends, with citation-based and keyword-based analysis preferable (Kontostathis A. et al. 2004). To predict trends in the research development of scientific studies, it is essential to track the evolution and life cycle of all topics in the selected area, which is of great importance in identifying new topics and predicting the dynamics of their development (Wang X.G. et al. 2014; Roche I. et al. 2010). The result of scientific areas is reflected in the terms used in the description of the thematic categories (Mokhnacheva Yu. V. et al., 2021). A high degree of semantic similarity between the content of the cited and citing article is revealed by citation. Considering this, the following method was proposed that feels, in addition to sharing keywords, citations: "Keyword—Citation—Keyword" (Cheng Q. et al. 2020).

A successful model for such a classification is the system developed in Scopus, *SciVal* topics. Almost all publications indexed in Scopus are classified as one of the topics 2000-2020. Each case consists of three distinctive vital phrases: the first two are

typically high-frequency, keyword phrases that are chosen to provide a macro-level description of the topic in the field of study, and the third key phrase is a more specific description of the topic (Xia W. et al. 2021; Cui Y. et al. 2018). Almost all publications indexed in Scopus have been classified as one of the topics since 1996.

A Topic is a collection of documents with a shared intellectual interest and can be large or small, new or old, growing or declining. Over time, new Topics will surface and evolve. Many Topics are multidisciplinary and old Topics may be dormant, but they exist. Researchers themselves are mobile and work in various research areas[1]

2. Hypothesis and methodology

The dynamics of the frequency of the same keywords on topics vary greatly: somewhere there is active growth, and somewhere, on the contrary - a decrease (up to 100%). In the publication (Mokhnacheva Yu. V. et al. 2021), we hypothesized that the more keywords with dynamics greater than 0% in the topic, the higher the probability that this topic is promising and is actively developing. Conversely, the more essential terms (KT) in a topic with negative dynamics, the more likely it is that it will indicate the researcher's interest decrease.

The use of key terms in *SciVal* topics varies: in some topics, the most used key terms had a positive trend, while others had a negative trend. According to this, all topics are supposed to be divided into groups. For this purpose, the coefficient of the topic development dynamics (Tlv) is proposed to be used (the proposed calculation formula is applicable only for those cases when at least one key term had a positive growth dynamic during the study period):

Error! Filename not specified.

Where is: KTpos din - number of KTs with positive growth dynamics in the topic; KT 0din - number of KTs in the marginal development zone (i.e., with 0% growth); KTneg din - number of KTs with negative growth dynamics

If indicator Tlv is the smaller relative to others, then the more developing and promising the topic appears to be against the background of the others, and vice versa: the higher the Tlv value is, the more likely it is that the topic is losing its relevance. If the majority of key terms are in a marginal zone, the growth for the studied period is equal to 0%, and the temporal stability of the topic occurs. If most key terms have positive growth, it can be assumed that the topic is actively developing. This approach applies to different periods of development of specific topics. If we trace the dynamics of key terms for additional periods, we can see the period of origin and period of decline in the topic's relevance.

The subject area Immunology & Microbiology in *SciVal* was chosen as a model for studying the dynamics of the development of scientific areas.

According to the study, data on the 500 most actively developing topics in this category according to the *SciVal* version were uploaded to carry out the analysis. The 100 most actively used key terms in 2015-2019 were taken from Scopus publication records in each of the 500 topics uploaded into separate Excel tables: 50000 key terms.

Analysis of the dynamics of the use of key terms for 2015–2019 in Scopus showed that while the central part of key terms had positive dynamics in some topics, the situation was the opposite in other topics: most key terms had negative dynamics, that is, the use of terms decreased in most cases.

3. Results and Discussion

As a result, we divided topics into 4 groups:

1. Tlv = 0.01–0.99: 141 topics (258 860 publications). In this group in 97 topics, 250 key terms had an increase of 1000% or more, and the average Tlv = 0.57;

[1] What is *SciVal*'s Topic Prominence? URL: https://service.elsevier.com/app/answers/detail/a_id/27947/supporthub/scopus/

2. Tlv = 1–1.5: 110 topics (158318 publications). In this group in 35 topics, 46 key terms had an increase of 1000% or more, and the average Tlv = 1.24;3. Tlv = 1.56–2.5: 139 topics (168724 publications). In this group in 27 topics, 32 key terms had an increase of 1000% or more, and the average Tlv = 1.97;

4. Tlv = 2.56 or more: 110 topics (122959 publications). In this group in 11 topics, 15 key terms had an increase of 1000% or more, and the average Tlv = 4.65.

Since the analysis covered the most actively developing topics in the world with an average level of relevance (relevance percentile) of at least 92%, in this case none of them can be said to be fading.

SciVal Topic percentile	**Prominence din**	**KT pos**	**KTneg din+KT 0 din**	**Tlv**
Intestine Flora; Ruminococcaceae; Dysbiosis	99.992	98	2	0.02
Antibiotic Resistome; Tetracycline Resistance; Integron	99.947	94	6	0.06
Photobioreactor; Nannochloropsi; Chlorellum Sorokiniana	99.904	90	10	0.11
1-Aminocyclopropane-1-Carboxylate Deaminase; Plant Growth-Promoting Rhizobacterium; Rhizosphere Bacteria	98.545	87	13	0.15
Macrophage Colony-Stimulating Factor; Tumor Microenvironment; Myeloid Cell	99.728	86	14	0.16
Mycobiome; RNA Gene; Oti	99.828	86	14	0.16
Acidobacteria; Chloroflexus; Nitrospirae	99.391	86	14	0.16
Extracellular Trap; Peptidylarginine Deiminase Type IV; Neutrophile	99.591	85	15	0.18
Endophyte; Plant Growth-Promoting Rhizobacterium; Phytoaccumulation	98.503	84	16	0.19
Microglia; Macrophage Colony-Stimulating Factor; Synapse	99.818	82	18	0.22

Table 1. Top 10 most developing topics in Immunology and Microbiology by value Tlv

However, it should be clarified that the obtained Tlv values apply only to an array of publications from the 500 most actively developing topics according to the SciVal version in the Immunology & Microbiology category. The values will be different for other arrays and should be calculated depending on the available information about the actual document arrays.

Thus, we assume that if a key term reduces its usage in one topic, it may appear and grow in other ones. The relevance of various key terms changes over time for different topics, both in negative and positive dynamics.

This process may indicate both the broad interdisciplinarity of research and the application of common methods to solve various problems and objects of research in different topical contexts. Terms are moving toward those topics where their more intense positive dynamics are an indicator of a shift in scientific fields to the changing research environment.

This approach applies to different periods of development of specific topics. Thus, if we trace the dynamics of key terms for different periods, we can see the period of origin and period of decline in the topic's relevance. In addition, by applying the analysis of the dynamics of KTs in combination with the study of co-citation (Marshakova I.V. 1981; Small H. 1973; Small H. 1997; Small H. et al. 1985; Boyack K.W. et al. 2010), we can see how and to what extent different topics are related to each other, as well as how the research emphasis is redistributed between them.

Using the example of the term "Bibliometric Analysis", we will show the dynamics of the same key term in different topics. The active use of the term "Bibliometric analysis" was revealed in 83 topics in July 2021. SciVal Topics in which we can see the highest increase in the use of "Bibliometric analysis":

- "Scholarly Publishing; Scientometrics; COVID-19": growth of 2200 %

Table 1 shows the top 10 most developing topics in the field of immunology and microbiology by indicator Tlv. The Tlv index here does not exceed 0.22, which indicates that all these topics are very actively developing.

- "Information Literacy; Library Instruction; Librarians": growth of 700%.

- "Knowledge Organization; Paul Otlet; Library Science": growth of 600%.

SciVal Topics in which the term "Bibliometric analysis" was actively used can be divided into 2 categories:
1. Topics of the Library and Information Science.

2. Other topics on a variety of scientific categories outside Library and Information Science.

There are studies on the analysis of document flows in various scientific fields. Such topics are characterized by the appearance of this term among the most used ones and its complete disappearance from this number. As a result, we can conclude that bibliometric analysis is widely used in various subject categories.

Next, we will show the dynamics of using the keyterm "Bibliometric analysis" in various topics.

Thus, in May 2021, the activity of using the "Bibliometric Analysis" in the topic "Academic Libraries; Bibliometric Analysis; "Research Evaluation" showed an increase of 200%, but in July 2021, it already showed a decrease of 100%;

In the Topic: "Economic Journals; Citation Index»; Bibliometric Indicators" in May 2021, "Bibliometric Analysis" showed an increase of 100%, but in July 2021, it already showed a decrease of 100 %;

In the Topic: "Bibliometric Analysis; Citation Index; Document Type" "Bibliometric Analysis" in May 2021 showed an increase of 48.7%, while in July 2021 decreased by 35.2%.

5. Conclusion

Thus, we can conclude that understanding the evolution of research topics at the level of studying the dynamics of keywords (terms) usage frequency is crucial for identifying new trends in science. Answers to the questions of what topics are well-established and stable; which issues are going through the active stage of development; which topics are losing their popularity and fading away, or spill over into other similar but newer topics; and how are the topics related to each other - will allow one to understand the trends in the development of science, the landscape of scientific research as well as to predict the outcome of scientific areas (Mokhnacheva Yu. V. et al. 2022 in press).

Until recently, the terminological approach to the scientometric analysis of the development of scientific research topics was very laborious and, therefore, difficult to implement. However, due to the capabilities of global thematic and bibliographic systems such as Scopus and *SciVal*, this approach has become available for widespread use. It has several significant capabilities: from analyzing the distribution of publication arrays on various scientific topics and studying the interdisciplinarity of scientific areas to predicting the development of scientific areas.

Understanding the evolution of research topics at the level of studying the dynamics of keywords (terms) frequency is crucial for identifying new trends in science. Answers to the questions of what topics are well-established and stable; which topics are going through the active stage of development; which topics are losing their popularity and fading away, or spill over into other similar but newer topics; and how are the topics related to each other - will allow one to understand the trends in the development of science, the landscape of scientific research, as well as to predict the outcome of scientific areas.

References

[1] Boyack, K. W., & Klavans, R. (2010). Co-Citation Analysis, Bibliographic Coupling, and Direct Citation: Which Citation

Approach Represents the Research Front Most Accurately? *Journal of the American Society for Information Science and Technology*, 61(12) 2389-2404.

[2] Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature // *Journal of the American Society for Information Science and Technology*. – 2006. – Vol. 57, 13. – P. 359–377.

[3] Cheng, Q. K., Wang, J. M., Lu, W., Huang, Y., & Bu, Y. (2020). Keyword-citation-keyword network: a new perspective of discipline knowledge structure analysis. *Scientometrics*, 124 (3) 1923-1943.

[4] Cui, Y., Wang, S., Gao, X., Yang, H., Cao, X. (2018). Detecting and Characterizing Research Fronts Topics Based on Global-Micro Model. Library and Information Service. 62 (15), 75-82.

[5] Kontostathis, A., Galitsky, L., Pottenger, W. M., Roy, S., Phelps, D. J. (2004). A survey of emerging trend detection in textual data mining. A comprehensive survey of text mining / M. Berry (Ed.). - Heidelberg: Springer, 2004, 185–224.

[6] Marshakova, I. V. (1981). Citation networks in information-science. *Scientometrics*, 3 (1) 13-25.

[7] Mokhnacheva, Y. V., & Tsvetkova, V. A. (2021). Development of Research Topics Based on the Terminological Approach (for Example, Immunology and Microbiology According to Scopus-SciVal Data). Scientific and Technical Information Processing, 48 (2) 139-145.

[8] Mokhnacheva, Yu. V., Tsvetkova, V. A. (2022). Russian publications on Library and Information Sciences in Scopus. Nauchnye i tekhnicheskie biblioteki - Scientific and technical libraries. (in press).

[9] Roche, I., Besagni, D., Francois, C., Horlesberger, M., Schiebel, E. (2010). Identification and characterisation of technological topics in the field of Molecular Biology. *Scientometrics*. 82 (3) 663-676.

[10] Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*. 24 (4) 265–269.

[11] Small, H. (1997). Update on science mapping: Creating large document spaces. Scientometrics, 38 (2) 75-293.

[12] Small, H., & Garfield, E. (1985). The geography of science - disciplinary and national mappings. *Journal of Information Science,* 11(4) 147-159.

[13] Wang, X. G., Cheng, Q. K., & Lu, W. (2014). Analyzing evolution of research topics with NEViewer: a new method based on dynamic co-word networks. *Scientometrics*, 101 (2) 1253-1271.

[14] Xia, W. J., Jiang, Y. P., Zhu, W. F., Zhang, S., & Li, T. R. (2021). Research Fronts of Computer Science: A Scientometric Analysis. *Journal of Scientometric Research*, 10 (1) 18-26.