Title Length Variation Trends in Papers on Algorithms in IEEE Xplore

Mohammad Reza Besharati (✉ besharati@ce.sharif.edu)

Sharif University of Technology <https://orcid.org/0000-0003-3190-411X>

Nafiseh Jafari

Malek-Ashtar University of Technology

Mohammad Izadi

Sharif University of Technology

Research Article

Keywords: Algorithm, Paper title, IEEE Xplore, Scientometrics

Posted Date: April 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-316558/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title Length Variation Trends in Papers on Algorithms in IEEE Xplore

Mohammad Reza Besharati¹, PhD Candidate, Sharif University of Technology, Tehran, Iran, besharati@ce.sharif.edu

Nafiseh Jafari, Graduate Student, Malek Ashtar University of Technology, Tehran, Iran

Dr. Mohammad Izadi, Assistant Professor, Sharif University of Technology, Tehran, Iran

Abstract

The title of any paper showcases its identity and the scientific orientation of its authors. The title length is one of the scientometrics indicators used in the statistical analysis of papers. The researchers have already found a relationship between the number of authors of a paper and its title length. In this study, we have analyzed the metadata of 19,469 papers on algorithms indexed in the IEEE Xplore database to conduct a more extensive study on this indicator and its trends, with three major findings: 1) Over time, the title lengths of the algorithm-related papers have increased, going from 70 printed letters in 1995 to 80 in 2009; 2) Papers with more authors also have longer titles; and 3) With time, as more scholars delve into the subfields of a scientific discipline, the title lengths of papers would increase. So that the peaks of the frequency charts of this variable exhibit a wave-like behavior and shift to the right.

Keywords: Algorithm; Paper title; IEEE Xplore, Scientometrics

1. Introduction

With the rapid growth of the number of scientific papers, the study of the nature of the development of human knowledge is both appealing and imperative. It is precisely what scientometrics is concerned with. Analyzing the plethora of published papers using statistical methods is one of the theoretical and yet, at the same time, practical aspects of this science [1].

The title of a paper is an indicator of its key orientation and showcases its content. It is the essence of the efforts and focus of the author(s). Therefore, statistical analysis of the trends observed in the titles of papers allows us to identify the behavioral trends of researchers and to study how knowledge is developed in various disciplines.

The title length of the paper is one of the metrics that is used in many studies on various corpora of papers. Researchers have already found that there is a relationship between the number of authors and the title length of a paper [2]. Furthermore, Semantic Complexity of a text is related to its length [3]. So the title length is a proxy of title Semantic Complexity.

---

¹ Responsible Author, the Lead Contact, Corresponding Author

We could consider understanding of a title as a cognitive process. In the formal computation semantics, the length of a process is a key computational parameter [4].

This work is limited to the study of the title length of algorithm-related papers indexed in the IEEE Xplore database. In this paper, we will see that over time the title length of algorithm-related papers has increased. We will also explain how a larger number of authors contribute to a longer title.

Lastly, we study the frequency of papers based on their title length to see how over time the local and global peaks of these charts shift, like a wave traveling through a rope. This finding can also illustrate how knowledge is developed in the fields related to the studied papers.

In this paper, we first introduce several statistical indicators that are used in the study of the trends of variations in the title length of papers. In section 2, we describe the three key findings of this study about the trends of the variations of paper title lengths. Section 3 provides our observations and the results of the study, while section 4 concludes the paper. Section 5 is devoted to the suggested ideas for future researches in this field.

1.1. The paper title length indicator

The paper title length indicator is equal to the number of printed letters used in the paper's title.

1.2. Author count indicator

This indicator represents the number of authors listed in the authors section of the paper.

1.3. Content Index (CI) indicator

This indicator is used to compare the content of the titles of two sets of papers. Assume that the first set of papers are called A, while the second set is denoted with B. We find all the words used in the titles of the papers in set A and put them in a sorted list called $list(A)$, which is sorted according to the frequency of their usage in all of the papers. It means that the first element of $list(A)$ is the word with the highest frequency of usage in the titles of the papers of set A. Then, we use the same method to generate $list(B)$. We set the initial value of the CI indicator to zero. Then, for each member X of the $list(A)$ we repeat the following steps:

If the position of word X in $list(A)$ and $list(B)$ are denoted by $P(X)_A$ and $P(X)_B$, respectively, we have:

$$\forall X \in list(A),$$

$$CI = CI + |P(X)_A - P(X)_B| \quad (1)$$

If X is not in $list(B)$, $P(X)_B$ is equal to the index of the last member of $list(B)$. Finally, the CI will be divided by the number of the elements in $list(A)$, so we can use it as an intensity indicator [5]. This indicator shows the strength of the relationship between the content of the titles in two sets of papers. The larger the CI of the sets A and B is, the more relevant the content of the titles in the two sets would be.

1.4. The studied dataset

This study has worked on 19,469 papers indexed in IEEE Xplore that had the word “algorithm” or “algorithms” in their titles. To do this, the metadata of the papers was collected from the database. Then we used a program to calculate the indicators introduced in the previous section.

2. Hypotheses

The following statements are the key hypotheses discussed in this study, and the author(s) are trying to prove them.

- 1) Over time, the title length of algorithm-related papers increases;
- 2) Papers with more authors have longer titles;
- 3) With time, local and global peaks of the frequency charts of the papers based on the title length would shift to the right, like a propagating wave.

3. Results and Findings

Our findings are based on the graphs obtained by processing the metadata of the papers. In all Figures, “title length of 80” refers to papers whose titles contain 70–80 printed letters. To compare the graphs presented in Figure 1, the numbers of each graph were scaled to 1000. As if in each period, the number of papers was equal to each other and equal to the number 1000.

Figure 1 shows the frequency chart of the title length of papers for different periods. As can be seen, the peak, as well as the rest of the frequency chart, shifts to the right, indicating the increased title length of papers over time. Figure 1, which is one of the key findings of this study, supports Hypotheses 1—that is, over time, the title length of algorithm-related papers has increased.

Figure 2 also shows the frequency chart of title length. But in this figure, different colors are used to compare the frequencies of papers with different numbers of authors. These plots are obtained using all of the papers published in the 1990–2010 period, which were included in the study dataset. Figure 2 shows that with more authors, the plots have shifted to the right, indicating that papers with more authors have longer titles.

Figure 3 compares the frequency of papers in the 1998–2001 and 2002–2005 periods, based on their title length. In this figure, the dotted line (the purple plot), we have a peak for the title length of 90 and a pseudo peak for the title length of 100. It seems that these two, are the results of a right-shift in the two peaks of the dashed green plot.

Figure 1. The frequency chart of title length of papers; Comparing the effect publication year on title length.

Figure 2. The frequency chart of title length of the papers; Comparing the effects of the number of authors on title length. Note that this graph is provided for all the papers in the 1990–2010 period.

Figure 3. Frequency versus title length of the papers.

If the content of the papers with a title length of 100 that are published in the 2002–2005 period is similar to that of the papers with a title length of 90, published in the 1998–2001 period, we can draw such a conclusion. It means that in the next four years, the researchers whose papers had a title length of near 90, have published more detailed papers and therefore, used longer titles to express the additional details.

Therefore, as the scholars added further detail, part of the papers with a title length of 90 was shifted near the papers with a title length of 100 after four years, which has created a pseudo-peak near 100 in the purple plot in Figure 3.

However, this would be true only if the similarity between the contents of these two sets of papers is proven, which is why we refer you to Figure 4. In this Figure, the CI indicator is used to show that the titles of papers published in 2002–2005 with a title length of 100 are similar to the titles of the papers published in 1998–2001 with a title length of 90.

4. Conclusion and Discussion

In this study, we proved three hypotheses listed in Section 2:

- 1) Over time, the title length of algorithm-related papers increases;
- 2) Papers with more authors have longer titles;
- 3) Over time, the local and global peaks of the frequency chart of papers based on the title length have shifted to the right, similar to the peaks of a moving wave.

We obtained these results through statistical analysis of 19,469 papers indexed in the IEEE Xplore database.

Figure 4. The similarity of the papers published in 2002–2005 featuring a title length of 100 with the papers published in 1998–2001 featuring a title length of 90.

5. Future works

We believe that a similar analysis on the length of the abstract section of papers would yield similar findings.

References

- [1] A. Ambihler and B. Mooler, *Scientometrics*. John Wiley & Sons, pp. 10-20, 2002.
- [2] W. Koehler, et Al., "A profile in statistics of journal articles", *CYBERmetrics*, Vol. 4, No. 2, pp. 120-180, 2000.
- [3] Besharati, MohammadReza, and Mohammad Izadi. "DAST Model: Deciding About Semantic Complexity of a Text." *arXiv preprint arXiv:1908.09080* (2019).
- [4] Izadi, Mohammad, and Ali Movaghar Rahimabadi. "An equivalence based method for compositional verification of the linear temporal logic of constraint automata." *Electronic Notes in Theoretical Computer Science* 159 (2006): 171-186.
- [5] Gholizadeh, B., *Numerical Analysis Methods*, Sharif University Press, 2012.

Figures

Figure 1

The frequency chart of title length of papers; Comparing the effect publication year on title length.

Figure 2

The frequency chart of title length of the papers; Comparing the effects of the number of authors on title length. Note that this graph is provided for all the papers in the 1990–2010 period.

Figure 3

Frequency versus title length of the papers.

Figure 4

The similarity of the papers published in 2002–2005 featuring a title length of 100 with the papers published in 1998-2001 featuring a title length of 90.