Mapping retrieval method for academic publications in the field of aerospace technology safety

A L Serdechnyi, A A Goncharov, A G Ostapenko and I L Bataronov
Voronezh State Technical University, 14, Moscow ave., Voronezh, 394026, Russian Federation

E-mail: csit@bk.ru

Abstract. The article presents result of the development and application of the Mapping retrieval method. The method is based on the construction, location in two-dimensional space and the analysis of the citation graph, which found in bibliographic databases. The method allows increase the speed and quality of the scientific research, as well as to review publications on a given topic. It is possible to identify the most authoritative co-authors communities in the field of topics of interest, as well as identify trends and predict the development of the topic. In addition, the Mapping retrieval method allows finding academic publications on related interdisciplinary research without known keywords that are needed for traditional text retrieval methods. The method is demonstrated by the example of constructing and analyzing the retrieval map in the field of safety of aerospace technologies, which made it possible to identify and track the development of thematic clusters. As a result of a mapping retrieval, a number of the most interesting academic publications were identified. Their brief review is presented in the article

1. Introduction
Information retrieval is the basic stage of any scientific research. Its speed and depth significantly affect the quality of research. At the same time, with the development of information technologies, the volume of published scientific literature grows in geometric progression and it becomes quite difficult to monitor only annotations of daily published scientific articles, not to mention the search and analysis of the full texts of all publications on the subject under study. The development of automated search engines solves the indicated problem only partially, because the algorithms for determining the relevance of publications implemented by such systems are often closed and do not guarantee the location of the most important publications on the first pages of search results. In addition, the ambiguity of terms, as well as the use of different terms to denote the same concepts, forces the expert to more carefully approach the choice of keywords for a search query. Moreover, the expert has to process several search results containing multiple repetitions at once. Therefore, the development of new approaches is required to increase the effectiveness of scientific research. One such approach is the graphical presentation and analysis of search results. As part of this approach, a method has been developed called the Mapping retrieval method for academic publications.

The developed method is demonstrated on the example of the security search academic publications aerospace technology, the issue of which is particularly acute in connection with two cases of the crash of Boeing aircraft due to software vulnerabilities (SW) [1-3].
The presence of such tragic cases is caused by the magnitude of the complexity of aircraft automation systems, and therefore there is a need to increase security at all stages of software development. This requires a generalization of the existing experience in testing and identifying vulnerabilities in aviation software, which is described in the relevant academic publications.

In this regard, the article presents the results of the development and application of the Mapping retrieval method using academic publications in the field of software development security in the aircraft industry as an example.

2. The algorithm of cartographic search

The proposed method of cartographic search for academic publications is a development of the ideas presented in [4-10]. The proposed method is based on a combination of techniques for constructing, analyzing, and visualizing a citation graph for academic publications. Such a graph is stacked in two-dimensional space using the "gravitational" algorithm. Algorithms of gravitational laying involve the calculation of forces of "attraction", "repulsion" for nodes that act as "physical bodies" with which "other bodies" interact. The coordinates of the nodes are calculated in such a way as to minimize the integral estimate of the totality of the relationships of all pairs of nodes in the graph. With this arrangement of the graph, the most interconnected nodes will be closer to each other than the less connected ones. This makes it possible to consider the graph as an analogue of a geographical map in which the metric principle of distance operates. Moreover, the "landscape" is not a combination of geographical objects, but an informational semantic space formed by academic publications related to citation relations.

A packed and colored graph forms the basis of the search map, which enables the expert to conduct interactive research of individual areas, similar to how geographical maps are studied in order to solve the problems of searching for objects or routes. The map for constructing a search map is shown in figure 1.

![Figure 1](image)

**Figure 1.** The process of building a search map s (A - Search the map for an interactive research domain and B - thermal semantic map domain for presenting search results in the thematic terrain).

3. The results of constructing a map of academic publications in the field of aerospace technology safety

As a result of the application of the developed method it was well built and search map in the field of safety of aerospace technology. Source data for it was carried out using the search engine Google Scholar based on the execution results of such search requests as a «Wireless aircraft security», «Boeing software vulnerability» et al. The information obtained as a result of performing search queries, as well as publications citing the found ones, made up the search map shown in figure 2.

And coming from the analysis of this graph, we can draw a number of conclusions. The first conclusion is as follows: on the periphery there are separate nodes that do not have any quoting links. This feature of the map may be due to several factors: 1) sufficient initial data were not obtained to determine the relationship of peripheral nodes with the core of the map; 2) topics of publications expressed by peripheral nodes are not related to the subject.

Secondly, in the presented column e it is easy to recognize two clusters (figure 3), and both clusters (cluster “Security of data transmission systems in aviation” - figure 3.a, and cluster “Patents”, figure...
3.b) have several vertices with a predominant number of citation links. On a larger cluster (figure 3.b), the central nodes (in the circle of which other publications are grouped) are four patents dating to 1998 [11], 2001, [12], 2002 [13] and 2013 [14] for years. Patents describe systems for observing and registering aircraft in mid-air. In the context of the safety of aerospace technologies, this information may make it possible to identify technological features that affect safety.

![Figure 2. Aerospace Safety Publications Search Map.](image)

Also in the clusters one can distinguish “tails” (separate publications on related topics, for example, a general description of a flight control system, such publications have much less connections with the cluster than those inside it).

In addition to two large clusters, covering most of the publications in the subject area, there are small clusters, for example, a cluster of publications related to Boeing incidents [2-4]. These publications collect publicly available technical information related to the causes and consequences of a plane crash that occurred due to software vulnerabilities. These incidents revealed a number of problems in the certification system of aviation software, which is also mentioned in the texts of the relevant articles [15, 16].

Thus, the topological features of the search map allow you to visually find areas of interest for closer examination. Unlike text search, you do not need to scroll through several tabs for various search queries or trust the search engine ranking system. All publications are grouped by density of links, expressing, as a rule, thematic similarities.

The most interesting from the point of view of information security technologies of the aerospace field are located in cluster 3.a, so we will consider it in more detail.

First, one should pay attention to the publication in 2016 «the On perception and reality in wireless air traffic communication security», marked with the arrow in figure 3.a This publication refers to a large (compared to other publications) number of sources located on the search map. This feature may
be evidence acce that publication is a review or a book, in which several topics discussed. This article is devoted to an overview of attacks on wireless technologies used in the aviation industry and serve as a good source for finding information about attacks on protocols such wireless communication as the GPS, the ACARS, the ADS-B, the ILS and others.

![Figure 3](image)

**Figure 3.** The main components of the search card in the field of safety of aerospace technologies (A - cluster "Transmission systems data security in aviation", B - to cluster "Patents")

The search map also allows you to analyze the development of the subject area under study by "filtering" the nodes of the graph that compose it according to specified criteria (for example, by author or year of publication). To illustrate an example of such an analysis, figure 4 shows the growth dynamics of the cluster “Security of data transmission systems in aviation”.

![Figure 4](image)

**Figure 4.** The growth dynamics of the cluster “Security of data transmission systems in aviation”.

Beginning in 2016 the number of publications has increased dramatically, and it is connected with the appearance of a review article [17], which is sponsored by the Martin Strohmeier, Vincent Landers and Matthew Smith. These authors are also the most published researchers in the examined cluster. On the topic of security of data transmission systems in aviation Martin Strohmeier in the cluster contains 18 publications, from Vincent Landers - 12 pieces, from Matthew Smith - 11 pieces, all of whom are located in the most active growth (figure 4).

4. Conclusion

Thus, the article is a description of the method and the search and scientific publications. Some of the capabilities of this method are demonstrated by the example of a search and analysis of the results of
scientific research in the field of safety of aerospace technologies. With the use of this technique and built a search engine map, allowing due to the visual analysis of data to increase the speed and quality of search and identify the authors of activity and determine the development of specific topics within a given search area.

Using this method, key publications were identified within the security topics of data transmission systems in aviation. Further research will be aimed at expanding the search map due to additional search results regarding the security of the space industry, as well as methods for identifying avionics vulnerabilities. A quantitative assessment of the effectiveness of the method and the development of the most effective search strategies using appropriate cartographic tools requires further research. The developed method can also be applied in other areas of scientific research.

References
[1] Sgobba T 2019 B-737 MAX and the crash of the regulatory system Journal of Space Safety Engineering 6(4) 299-303
[2] Bergstra J A and Burgess M 2019 A Promise Theoretic Account of the Boeing 737 Max MCAS Algorithm Affair arXiv preprint arXiv:2001.01543
[3] Bergstra J A and Burgess M 2020 Candidate Software Process Flaws for the Boeing 737 Max MCAS Algorithm and Risks for a Proposed Upgrade arXiv preprint arXiv:2001.05690
[4] Shneiderman B 1996 The eyes have it: A task by data type taxonomy for information visualizations Proceedings of the IEEE symposium on visual languages 336-43
[5] Klouche K, Ruotsalo T, Micallef L, Andolina S and Jacucci G 2017 Visual re-ranking for multi-aspect information retrieval Proceedings of the 2017 conference on human information interaction and retrieval 57-66
[6] Liu S, Wang X, Chen J, Zhu J and Guo B 2014 Topicpanorama: A full picture of relevant topics. IEEE Conference on Visual Analytics Science and Technology (VAST) 183-92
[7] Fried D and Kobourov S G 2014 Maps of computer science IEEE Pacific Visualization Symposium 113-20
[8] Von Landesberger T, Kuijper A, Schreck T, Kohlhammer J, van Wijk J J, Fekete J D and Fellner D W 2011 Visual analysis of large graphs: state-of-the-art and future research challenges Computer graphics forum 30(6) 1719-49
[9] Azimzadeh F and Norouzi F 2018 Incorporating structural information in scientific document retrieval 4th International Conference on Web Research (ICWR) 103-10
[10] Tsaregorodtsev A V, Zelenina A N and Savelev V A 2017 Two-stage procedure of quantitative assessment of information security risk of cloud computing Modeling, optimization and information technology 4(19) https://moit.vivt.ru/wp-content/uploads/2017/10/ZaregorodzevSoavtori_4_1_17.pdf
[11] Lee F A 1998 U.S. Patent No. 5,742,336
[12] Wright T H and Delpak R 2001 U.S. Patent No. 6,173,159
[13] Wright T H and Ziarno J J 2002 U.S. Patent No. 6,353,734
[14] Monroe D 2006 U.S. Patent No. 7,131,136
[15] Ozkaya I 2019 What Should a Software Engineer Know? IEEE Software 37(1) 3-6
[16] Johnston P and Harris R 2019 The Boeing 737 MAX saga: lessons for software organizations Software Quality Professional 21(3) 4-12
[17] Strohmeier M, Schäfer M, Pinheiro R, Lenders V and Martinovic I 2016 On perception and reality in wireless air traffic communication security IEEE transactions on intelligent transportation systems 18(6) 1338-57