Ontology-based Approach for Research Activity Mapping

Miroslav Kubelskiy
Graduate School of Management
St. Petersburg State University
St. Petersburg, Russian Federation
mirqube@gmail.com

Anna Kuznetsova
Lyceum No. 408
St. Petersburg, Russian Federation
anna.romantseva@gmail.com

Irina Leshcheva
Graduate School of Management
St. Petersburg State University
St. Petersburg, Russian Federation
leshcheva@gsom.spbu.ru

Tatiana Gavrilova
Graduate School of Management
St. Petersburg State University
St. Petersburg, Russian Federation
gavrilova@gsom.spbu.ru

Vladislav Shvankin
JSC "SalesAI"
Moscow, Russian Federation
v.shvankin@salesai.ru

***Abstract*—Visual knowledge mapping simplifies the process of knowledge search and processing that helps to find potential partners for joint projects, and organization management to facilitate the processes of decision making more effectively. The article proposes a new approach to university teachers' knowledge mapping. The approach is based on ontology engineering and extensive use of bibliometric analysis. As part of the proposed framework, a set of ontologies was created to describe the scientific activity of a faculty member in terms of his competencies and performance indicators. Then the BIB-METR visualization system was created to collect and visualize information about faculty members and researchers. The proposed approach was justified on the sample of one department of the university-based business school. The developed ontology works as a skeleton of a multidimensional portrait of university faculty member in the part of their research activity. And the set of such portraits generates the information research landscape of the university.**

Keywords— ontology, bibliometric data analysis, bibliometrics, knowledge maps, visualization system

INTRODUCTION

Insufficient transparency of the professional profile and experience of teachers and researchers in scientific institutes and universities makes it difficult to find partners and contractors for joint projects. Classic academic resumes do not allow you to quickly assess the expertise of specialists. Internal university portals do not always provide complete information about the competencies of teachers.

All this leads to a decrease in the effectiveness of scientific activities, inaccessibility of experts to potential partners and difficulties in the prompt search for counterparties for joint projects.

The report proposes using visual knowledge mapping techniques [1-3] to provide educators and researchers with the tools of the visual representation of their expertise. This will also help the administration make more objective decisions, and will promote interaction between students and teachers.

Researchers note the importance of visualization for understanding the presented information by different groups of people [4,5]. Moreover, visual models help to comprehend, compress and visually present large amounts of quantitative and verbal information, explicit and implicit knowledge. In the field of modern management and information technology, an innovative branch of "knowledge management" has been developing for more than twenty years [6], in which visualization acts as the most powerful means of codifying, preserving and disseminating knowledge [7-9].

The paper discusses the issue of visualizing information resources as part of the Russian Science Foundation project "Methodology and technology for developing digital knowledge maps for educational and scientific teams (METAKARTA)". The METAKARTA project is dedicated to analyzing the state of research in the field of digital knowledge maps of specialists and developing a methodology for designing such maps based on ontologies. Ontologies as conceptual models of a knowledge domain are one of the most perspective approaches to the development of knowledge bases and knowledge graphs.

This paper is devoted to the co-design of ontologies of university professors' knowledge in research with the subsequent development of knowledge maps to simplify the processes of searching and processing knowledge and further better communication and co-operation in research. Studies show that bibliometric analysis is an important tool for managing science, shaping research policy and developing a strategy for the publication activity of scientific and educational organizations. Bibliometric analysis has gained immense popularity in business research in recent years [10]. Special bibliometric software makes bibliometric analysis easy and accessible (VOSviewer, GOOGLE SCHOLAR, RSCI and ELIBRARY). The popularity of bibliometric analysis for universities is a reflection of its utility for leveraging and handling large amounts of scientific data, and producing high research impact.

As part of the implementation of the METAKARTA project, a pilot study was carried out on the development of bibliometric knowledge maps of the university departments. The data was used from open sources, then the dataset was analyzed, and the results were presented in a visual form. A software prototype of the BIB-METR visualization system was developed for a number of the bibliometric indicators.

The work was carried out by Gavrilova T.A., Kubelsky M. as part of the Russian Science Foundation grant No. 23-21-00168, https://rscf.ru/project/23-21-00168/

979-8-3315-3202-4/24/$31.00 ©2024 IEEE

The software prototype of the system is focused on the automated construction of knowledge maps based on the open data sources.

Ontology Based Knowledge Mapping for Research

Research as such largely depends on the subject area, and it is obvious the types and forms of research activity will differ significantly. However, the general forms of research knowledge are similar and always imply creative activity to increase the amount of research knowledge about the research domain of one's research interest. From a pragmatic point of view, we can talk about various types of fundamental and applied research, as well as the development of new theories, technologies, methods and recommendations.

Thus, the question arises about creating an ontology for research. There are various ontologies for research, for example, Core Ontology for Scientific Research Activities (COSRA) [11], EMPIRION [12], DINGO - ontology for scientific projects [13]. However, they all describe the nature and structure of empirical data from research.

First, we set a narrower and more pragmatic goal - to describe research activities of teachers of the university-based business school, therefore we rely on traditional forms of research (for example, projects, presentations at conferences and preparation of publications).

Methods of ontology engineering have been widely developed for almost 30 years [14, 15].

The number of steps of such methodologies used for building knowledge maps varies from 3 to 11, depending on the degree of detail of each development phase. Based on the analysis, summarizing it and our experience, we have proposed the novel methodology with 4 steps called it ONE-VID [16, 17]:

I - Set of Ontologies Design (ON) (Ontology);

II - Knowledge elicitation for ontology population (E) (Elicitation);

III - Visual knowledge mapping (VI) (Visualization);

IV - Knowledge map justification and dissemination (D) (Dissemination).

The first stage of the methodology (ON) is focused on the development of a set of ontologies defining the conceptual structure of the future knowledge map. Our task was to create an ontology providing a big picture view of the teacher's activities at the university in the domain of management studies. Ontology sketches were developed using mind mapping and conceptual modeling tools, and then transferred to the PROTÉGÉ ontology editor for further software implementation and use.

Each faculty member's responsibilities include teaching, research, practical consulting work, and administrative work. Fig. 1 presents a simplified top-level ontology describing the research work of a teacher (developed by Irina Leshcheva, described in [17].

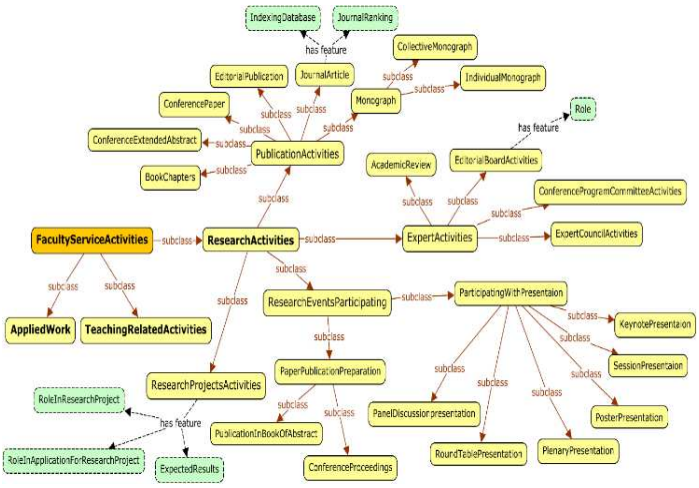

Fig. 1. Ontology of research activities (upper levels).

The ontology of the teacher's research includes

1) research projects, the role of the researcher;

2) preparation of publications, presentations at conferences;

3) expert work.

Then we developed the ontology further and included bibliometric data of the author as presented in Fig. 2.

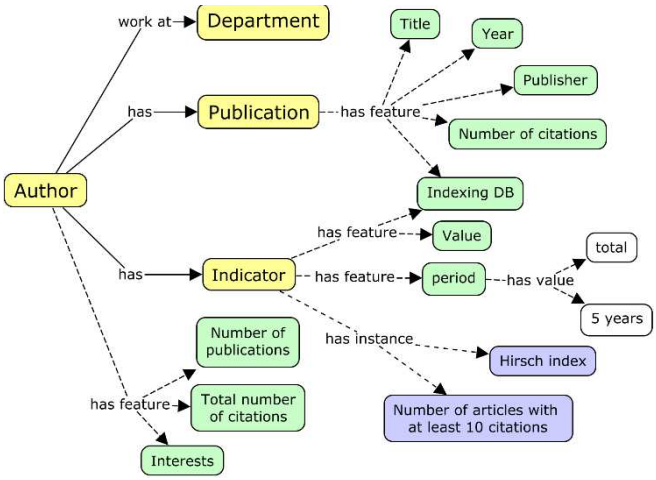

Fig. 2. Ontology of research activities (including bibliometric data).

The presented ontologies were used as the basis for the development of a knowledge map of research portfolio. A knowledge map is a representation in a diagram of the main components of an organization's knowledge. Knowledge maps are one of the tools used in knowledge engineering to organize and represent knowledge, forming a graphical landscape in tasks of visualizing complex concepts, decision support, and knowledge sharing [18]. In a knowledge map, the location of the represented knowledge, their relationships are especially important; the carriers (knowledge holders) and sources are indicated [19, 20]. In general, a knowledge map is a means of bringing together experts, sources of tacit and explicit knowledge of the organization, the relationship between elements and knowledge flows in the organization.

979-8-3315-3202-4/24/$31.00 ©2024 IEEE

BIB-METR VISUALIZATION SYSTEM FRAMEWORK

Bibliometrics, the quantitative analysis of scholarly publications, has emerged as a valuable tool for assessing research output, impact, and trends [10]. Visual diagrams play a crucial role to communicate effectively the complex data and relationships revealed through bibliometrics. This short literature review examines the use of bibliometric visual diagrams, their types, benefits, and challenges.

Bibliometric visual diagrams are essential tools for communicating the insights gained from quantitative analysis of scholarly literature [21]. They enhance data visualization, facilitate pattern identification, simplify communication, and support decision-making. However, it is important to consider the challenges associated with their use, including data completeness, interpretation, software limitations, and potential subjectivity [22]. Researchers can effectively leverage bibliometric visual diagrams to gain valuable insights from scholarly literature by addressing these challenges and employing appropriate visualization techniques.

The main types of bibliometric visual diagrams are [23, 24]:

- Subject Category Maps: Organize scholarly publications into subject categories, providing a hierarchical view of research areas and their relationships.
- Citation Networks: Visualize the connections between scholarly publications based on their citations and co-citations. Nodes represent publications, and edges represent citations.
- Keyword Co-occurrence Networks: Maps of the co-occurrence of keywords in scholarly literature reveal research themes and relationships.
- Author Collaboration Maps: Illustrate the collaborative relationships between researchers based on their co-authorship patterns.
- Citation Impact Visualizations: Present the citation counts and impact factors of scholarly publications, allowing for comparisons and identification of highly cited works.

Many authors discuss the challenges of bibliometric visual diagrams. They consider the data completeness, which can impact the reliability of visualizations [25]. Also data interpretation is considered as visual diagrams can be complex and require careful review to avoid misunderstanding. All the researchers face the software limitations that can regulate their effectiveness [26]. Also there is a group of publications that study the subjectivity of the interpretations.

A pilot study on the development of bibliometric maps required solving a number of problems within the framework of the concept (proof-of-concept). That is, our aim was to prove the viability of the chosen approach.

The following steps are performed to achieve this research goal:

a) Determination of requirements for the bibliometric data

b) Analysis and selection of open data sources

c) Collection of open data

d) BIB-METR software prototype architecture design

e) Data analysis and visualization

Let us focus on each of these 5 stages.

A. Determination of requirements for the bibliometric data

To collect bibliometric data, it was necessary to create access to the register of collaborating authors in connection with the departments for which information would be collected and searched.

The provided register contains links to employee profiles on the main open research bibliometric resources. This fact made it possible to significantly reduce the labor intensity while implementing this part of the project, because there was no need to manually search for the links to employee profiles. The register includes reference information about employees and contains a set of fields, for example

- Name, family name / Department / Position / Degree
- Elibrary_SPIN / Scopus_AuthorID / WoS_ResearcherID / WoS_link
- Orcid / Google Scholar / Pure_link

B. Analysis and selection of open data sources

The main open databases of publications available on the Internet were considered as open sources of data on the scientific activities of university department employees:

- Elibrary (https://elibrary.ru/)
- Scopus (https://www.scopus.com/home.uri)
- Web of Science (https://webofknowledge.com/)
- Google Scholar (https://scholar.google.com/)
- ResearchGate (https://www.researchgate.net/)

The following criteria were formed to select a data source: (a) Possibility of automatic data download / (b) Availability from Russia / (c) Free access / (d) Data completeness.

As part of the work on the prototype using the proof-of-concept, it was possible to relax the data requirements, for example, not require 100% completeness, i.e. there is no need to collect information about all employees and all publications.

The Elibrary system does not meet criteria (a) and (c), because a subscription payment and an appropriate agreement between the provider and a scientific and educational institution are required to gain access to the system's API. Scopus and Web of Science systems do not meet criteria (b) and (c), because access to them is paid, and interaction with Russian universities is suspended. The ResearchGate system does not meet criterion (d), because it contains verified profiles of a smaller number of employees/authors.

Thus, Google Scholar was chosen as the main source of data on the scientific activities of employees, which meets all the formulated criteria:

- It is possible to automatically upload both via the API and using freely distributed libraries;
- Available from Russia and employee articles continue to be indexed;
- Access is free;

979-8-3315-3202-4/24/$31.00 ©2024 IEEE

- Most employees have profiles in this system, links to which are presented in the register.

C. Collection of open data

Minimum requirements for data from open sources were formulated to ensure the possibility of constructing various versions of knowledge maps from a data set. They are presented in Table 1 and include the information about the author, bibliometric indicators, publication title and abstract as semantically meaningful information.

TABLE I. METADATA

Author/contributor metadata	**Bibliometric indicators for each author/employee**	**Semantically meaningful information to identify the author/contributor's areas of expertise**
a. Surname b. Name c. Unique identifier	a. List of publications, including publication title b. Number of citations for each publication c. H-index and other available aggregation indicators	a. Publication titles b. Publication abstracts

D. BIB-METR software prototype architecture design

BIB-METR software module was designed to automate the collection of data from Google Scholar. It was partly prototyped in the Python programming language, including the following libraries:

- scholarly - the main library, it is used to search for and connect to the Google Scholar proxy;
- json, re - libraries for parsing results;
- google.oauth2, gspread - libraries for automatic record of the results in Google spreadsheets;
- time, csv - libraries to monitor the speed of work and the quality of results.
- The developed software module made it possible to generate a set of bibliometric data (dataset) Bibset. Bibset dataset has got the following structure:
- author_info - information about authors, including the following fields with the number of publications and citations, H-index and areas of interest.
- publication_by_author_info - brief information about publications, contains 3291 records with fields of information about publications.
- more_publication_info - extended information about publications, contains 3152 records with annotations, publishers and links fields.

E. Data analysis and visualization

As part of the current stage of the pilot study project several visual diagrams were tested. For example, a number of publications of the faculty, average number of publications per department, number of employee citations in relation to departments, etc.

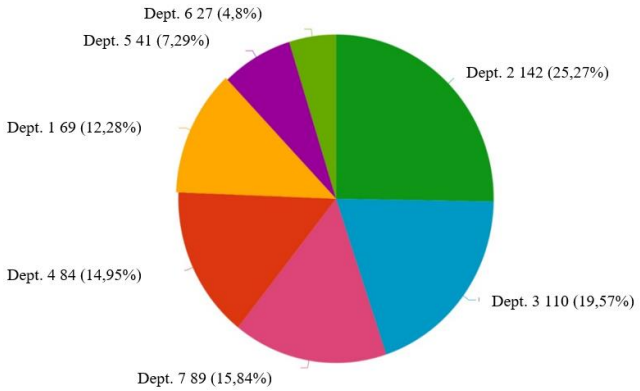

Fig.3. Distribution of all the publications among the school departments.

Figure 3 shows the distribution of all the publications listed in the database among the university school departments. Here the information on the percentage of the total amount of publications of the university school departments is stated. That helps to evaluate the more and less active departments in terms of publications.

Figure 4 shows a portrait of the faculty from department X and gives the information on the number of publications for each of the faculty members, their H-index, the number of citations of each of the teacher.

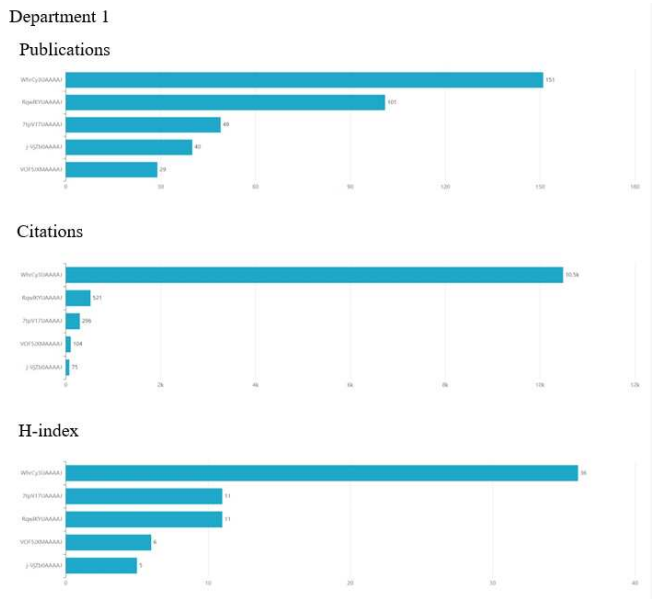

Fig.3. Portrait of department X.

CONLCUSION

Information resources are the main source of data and knowledge that constitute the intellectual capital of an educational organization. The innovative method of visual knowledge mapping is a powerful tool and the core of special knowledge management systems (KMS) [27, 28]. The use of knowledge maps helps make knowledge sharing and finding colleagues with similar scientific interests and/or expertise more accessible to different stakeholder groups.

Based on the proposed methodology and collected data we developed a research ontology that includes bibliometric indicators, in addition to project descriptions, preparation of publications, presentations at conferences, as well as scientific expert work. We proposed a framework to summarize

979-8-3315-3202-4/24/$31.00 ©2024 IEEE

analytical information on the publication activity of the faculty of the selected university unit.

Further fields for development of the proposed approach are:

- The prepared data set can serve as the basis for the development of a variety of visual representations that reflect other analytical sections that are valued by stakeholders.
- A proven approach to the automated construction of knowledge maps based on open data sources can be used in the implementation of an IT solution aimed at supporting management decisions based on current bibliometric data.
- Data presented in natural language, for example, titles and abstracts of articles, can be studied by automated semantic analysis, which will provide an objective understanding of the intellectual assets that a research and educational institution has got.

ACKNOWLEDGMENT

Authors thank Dr. Dmitry Kudryavtsev, who was the initiator of this project and did a lot in terms of research design.

REFERENCES

[1] E. R. Tufte, The visual display of quantitative information. Cheshire, CT: Graphics press, Vol. 2, No. 9, 2001.

[2] X. Liang, L. Luo, S. Hu, and Y. Li, "Mapping the knowledge frontiers and evolution of decision making based on agent-based modeling," Knowledge-Based Systems, 250: 108982, 2022.

[3] O. Borisova, "Infographics as an independent genre. News of higher educational institutions," Problems of printing and publishing (5): 186–1948, 2012. (In Russian)

[4] K. Henderson, On line and on paper: Visual representations, visual culture, and computer graphics in design engineering. MIT press, 1998.

[5] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman, "The science of visual data communication: What works," Psychological Science in the public interest,Vol. 22, No. 3. – P. 110–161, 2021.

[6] M. Mårtensson, "A critical review of knowledge management as a management tool," Journal of knowledge management, Vol. 4. No. 3. – P. 204–216, 2000.

[7] M. J. Eppler, and R. A. Burkhard, Knowledge visualization: towards a new discipline and its fields of application. Università della Svizzera italiana, 2004.

[8] X. Ding, and Z. Yang, "Knowledge mapping of platform research: a visual analysis using VOSviewer and CiteSpace," Electronic Commerce Research, No. 1–23, 2020.

[9] M. Cordeiro, F. Puig, and L. Ruiz-Fernández, "Realizing dynamic capabilities and organizational knowledge in effective innovations: the capabilities typological map," Journal of Knowledge Management, 2022.

[10] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," Journal of business research. 133, 285-296, 2021.

[11] P. M. C. Campos, C. C. Reginato, and J. P. Almeida, "Towards a core ontology for scientific research activities," Lecture Notes in Computer Science 11787: 3–12, 2019.

[12] A.M. Begler, D.V. Kudryavtsev, and T.A. Gavrilova, "On the use of ontologies for the integration of scientific research data," Proceedings of the 18th National Conference on Artificial Intelligence with International Participation KII-2020 (as part of the I National Congress on Cognitive Research, Artificial Intelligence and Neuroinformatics), Moscow, 3-12, 2020.

[13] D. Chialva, and A. M. Mugabushaka, "Dingo: an ontology for projects and grants linked data," ADBIS, TPDL and EDA 2020 Common Workshops and Doctoral Consortium: International Workshops: DOING, MADEISD, SKG, BBIGAP, SIMPDA, AIMinScience 2020 and Doctoral Consortium, Lyon, France, August 25–27, 2020, Proceedings. 24, 183-194. Springer International Publishing, 2020.

[14] T. Gruber, "A translation approach to portable ontology specification," Knowledge acquisition, 5, 2, 199–220, 1993.

[15] B. Krieg-Brückner, T. Mossakowski, and M. Codescu, "Generic Ontology Design Patterns: Roles and Change Over Time," Advances in Pattern-Based Ontology Engineering, 51, 25, 2021.

[16] T. Gavrilova, O. Alkanova, and A. Kuznetsova, "Ontology-Based Methodology for Knowledge Maps Design," Kovalev, S., Kotenko, I., Sukhanov, A. (eds) Proceedings of the Seventh International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'23). IITI 2023. Lecture Notes in Networks and Systems, vol 776, chapter 23, Springer, Cham, pp. 250–259, 2023.

[17] T. Gavrilova, A. Kuznetsova, and I. Leshcheva, "On the issue of the development of ontologies of scientific and educational work," Twenty-first National Conference on Artificial Intelligence with International Participation, CAI-2023. Proceedings of the Conference. In 2 volumes, vol.1, Smolensk, pp. 60–67,2023. (in Russian)

[18] G. B. Zagorulko, and Yu. A. Zagorulko, Artificial intelligence. Knowledge Engineering, 2019. (In Russian)

[19] T. H. Davenport, and L. Prusak, Working knowledge: How organizations manage what they know. Harvard Business Press, 1998.

[20] R. Haghnazar Koochaksaraei, F. Gadelha Guimarães, B. Hamidzadeh, and S. Hashemkhani Zolfani, "Visualization method for decision-making: a case study in bibliometric analysis," Mathematics, 9(9), 940, 2021.

[21] J. T. McAllister, L. Lennertz, and Z. Atencio Mojica, "Mapping a discipline: a guide to using VOSviewer for bibliometric and visual analysis," Science & Technology Libraries, 41(3), 319-348, 2022. .

[22] S. L. Parfenova, E. G. Grishakina, D. V. Zolotarev, and V. V. Bogatov, "Publication landscape of Russian science," Science Management and Scientometrics, (1 (23)), 53-79, 2017.

[23] K. Subramanyam, "Bibliometric studies of research collaboration: A review," Journal of information Science, 6(1), 33-38, 1983.

[24] M. A. Akoev, V. A. Markusova, O. V. Moskaleva, and V. V. Pislyakov, Guide to scientometrics: indicators of the development of science and technology: [monograph, 2nd edition]. Guide to Scientometrics: Indicators of Science and Technology Development, 2021.

[25] M. E. Bales, D. N. Wright, P. R. Oxley, and T. R. Wheeler, Bibliometric visualization and analysis software: State of the art, workflows, and best practices, 2020.

[26] J. A. Moral-Muñoz, E. Herrera-Viedma, A. Santisteban-Espejo, and M. J. Cobo, "Software tools for conducting bibliometric analysis in science: An up-to-date review," Profesional de la información/Information Professional, 29(1), 2020.

[27] T. Gavrilova, L. Grigoriev, D. Kudryavtsev, "Knowledge management: from words to deeds," Enterprise Systems 12–13(101): 43–48, 2004. (In Russian)

[28] A. Di Vaio, R. Palladino, A. Pezzi, and D. E. Kalisz, "The role of digital innovation in knowledge management systems: A systematic literature review," Journal of business research, No. 123. – P. 220–231, 2021.

979-8-3315-3202-4/24/$31.00 ©2024 IEEE