Quantitative Information Research: Introduction

In the first thematic issue of Zagadnienia Informacji Naukowej – Studia Informacyjne (*Issues in Information Science – Information Studies, ZIN*) for this year, designated as 1A and published in September 2019 on the open access website of Polish Librarians Association (sbp.pl), I declared that, thanks to the financial support which the Ministry of Science and Higher Education provided in 2019 under the program "Actions to Promote Science: Publishing" to increase the national and international circulation of *ZIN*, the editorial committee would be releasing two additional thematic issues devoted to the newest topics in information science, all in English. Each of these issues focused on a specific research area of information science, which is currently attracting research interest. Our goal is to make the contents of *ZIN* more accessible and more attractive to a wider, international audience.

The first thematic issue focused on the question of open science and open access to scientific data and materials. It also presented the challenges, which this new approach to conducting research and publishing its results posed before information science. The second thematic issue, which we are sharing with our readers as the year comes to an end, focuses on the issues in quantitative information research – on its methodology, its application, and first and foremost, on the problems which emerge in the process and interpretation of the results of quantitative analysis.

Quantitative information research is one of the most dynamically developing research areas in the contemporary information science. If we refer to Marcia J. Bates's model of the intellectual structure of information science, based on three big questions of information science[1], we may frame information metrics as an attempt to answer the first of these questions: the physical question of the features and laws of the recorded information universe. Alongside thus described object of study, the research in this area is distinguished by the application of statistical methods to establish the features and laws ruling the recorded information universe, their conditions, time variability, and the relations between the studied entities and their properties. Tefko Saracevic labelled these studies as "metrics", or "metric studies in information science"[2], emphasizing that the label should encompass a number of specific subdisciplines: bibliometrics, scientometrics, informetrics, webometrics, altmetrics.

---

[1] Information science's big questions: (1) the physical question: What are the features and laws of the recorded information universe? (2) The social question: How people relate to, seek, and use information? (3) The design question: How can access to recorded information be made most rapid and effective? (M.J. Bates (2016). The Invisible Substrate of Information Science. In: M.J. Bates (ed.). *Information and the Information Professions: Volume I of the Selected Works.* Berkeley, CA: Ketchikan Press, 119. First published as M.J. Bates (1999). The Invisible Substrate of Information Science. *Journal of the American Association for Information Science*, 50(12), 1043–1050.

[2] T. Saracevic (2010). Information Science. In: M.J. Bates, M.N. Maack (eds.). *Encyclopedia of Library and Information Sciences. Third Edition* (2570–2586). New York: Taylor & Francis, 2580, http://doi.org/10.1081/E-ELIS3-120043704

They differ according to the types of objects they study; they emerged at different points in time; they realize different goals, and enter different interdisciplinary relations. The oldest of these is bibliometrics, i.e., the quantitative study of written communication – the properties of publications and processes related to them. Bibliometrics emerged as a part of information science in the 1950s, preceded by statistical studies of written communication already conducted in the first half of the 20th century. These studies revealed the existence of repetitive patterns in the distribution of productivity of authors of scientific publications (Alfred Lotka, 1926) and in the distribution of relevant articles on a certain issue in scientific journals (Samuel Bradford, 1934). Scientometrics, which emerged in the 1960s, drawing on the work of Derek De Soll Price, is concerned with statistical analysis of scientific research and its results, basing on different sources of information, including but not limited to bibliometric investigation of written scholarly communication. In 1980s, informetrics began to develop as a field of quantitative research studying information units of all types, including both bibliometric studies of literature as well as other forms of communication and other aspects of information processes. Webometrics, which emerged in the late 1990s, is concerned with quantitative analysis of the resources of the World Wide Web (WWW), the changes occurring in its communication environment, its structure and functioning[3]. Altmetrics, the latest subdiscipline of information metrics, which emerged in the second decade of the 21st century, measures the impact of scientific research on the development of science, as well as on the society, basing on the statistical analysis of the activities registered online, which relate to citation, viewing and downloading articles, posters, chapters, and research datasets published online, as well as to sharing information regarding these via social media[4].

Quantitative information research is the central subject of two international scientific journals, whose impact indicators are one of the highest in information science: *Scientometrics*, a journal published monthly since 1978, whose thematic scope encompasses scientometrics, and *Journal of Informetrics*, published quarterly since 2007, whose thematic scope encompasses all types of information metrics. High impact indicators testify not only to the quality of the research whose results are published therein, but also to the appeal of these studies, and the growth of the community concerned with them. The issues in quantitative information research are often discussed in other journals devoted to information science. In recent years, it became a subject of many articles published in *ZIN*.

Quantitative information research has both theoretical and practical dimensions. Its fundamental epistemological goal is to determine the qualities of the information universe, its features (the size of information resources, their growth in time, their distribution in relation to various criteria – disciplinary, geographical, institutional, and so on) and its laws (concentration, scattering, aging). This research makes it possible to establish the current state of, and the changes in, communication of knowledge occurring in various areas, and to predict the direction of its future development. The analysis of distribution of, and relations between, publications yields insight regarding the models of authorship, national

---

[3] M. Skalska-Zlat (2017). Webometria. W: *Encyklopedia Książki*. Wrocław: Wydaw. Uniwersytetu Wrocławskiego, T. 2, 605.

[4] A. Tattersall, ed. (2016). *Altmetrics. A Practical Guide for Librarians, Researchers and Academics*. London: Facet Publishing.

and institutional output, the development of collaboration, the intellectual structure of research areas, the shaping of the so-called frontier of research, the emergence of new directions and disciplines. The diagnostic and predictive functions of information metrics constitute its practical dimension. Quantitative research of scientific publications is applied to planning publishing activity, optimizing the collections of the libraries for specific users, directing the science policy, identifying research priorities, drawing comparisons between various scientific disciplines, research institutions, and individual researchers. The emergence of citation indexes gave rise to the application of quantitative analysis of scientific publications in evaluation, which intends to objectively assess the research's influence on the development of science, or more generally, the influence of scientific research on social and cultural development. The application of bibliometrics and scientometrics in research evaluation has caused some controversy; their role in the assessment of research institutions and specific researchers' activity inspires particularly strong reactions. In Poland, criticisms of such a mode of evaluation grew in force after the government introduced a new classification of disciplines and new rules for the evaluation of scientific institutions in 2018, to which Polish scientific institutions will become subject in 2021. The critics of quantitative information research draw attention to the questionable quality of research reliant on the increasingly numerous and accessible digital data resources, which are not always adequate to the purposes for which they are employed.

The sources of data subject to statistical analysis are the foundation of all quantitative studies; thus they determine the results, their quality and reliability. Bibliometrics studies the sets of bibliographic data, currently, these are bibliographic databases and citation indexes. Scientometrics uses also other sources recording information regarding various aspects of the development of science: research project and financing thereof, academic conferences, international collaboration, and so on. Digital libraries and institutional digital repositories are used as sources of data with increasing frequency. Webometrics relies on the resources available on the WWW, and on the structure of hypertext links. The data for altmetric research is drawn from social media. The crucial issues for quantitative analysis are the quality of the sources of data used and the researcher's knowledge of its contents, its organization, the rules according to which the data was gathered and their limitations, and the appropriate selection of these sources for the particular research purposes. The results of the research are dependent on the quality and the representativeness of the data analyzed, and on the quality of the analysis itself. We should bear in mind that a reliable interpretation of results requires a sound knowledge of studied phenomena and a thorough assessment of the chosen sources' influence on the results they yielded. Therefore, although information metrics are based on quantitative methods, the selection and preparation of research material, as well as an in-depth interpretation of the results require an application of suitable qualitative methods to ensure that the researcher will make appropriate choices and draw only thoroughly justified conclusions. With the development of digital technology, the number of information sources, and of tools allowing automatic filtering of resources according to various criteria, conducting statistical analysis, and creating attractive visual models of the results, is rapidly growing[5]. They facilitates quantitative information research, which makes it increasingly popular. However, it does not follow

---

[5] Zob. M. Thelwall (2008). Bibliometrics to Webometrics. *Journal of Information Science*, 34 (4), 605–621.

that the researcher will always be aware of the studied problems, and the sources and methods employed. Superficial bibliometric studies of datasets which are not sufficiently representative, and whose analysis is limited by the capacities of the available automatic tools, are subject to justified criticism. Nevertheless, we should not undermine the value of thorough quantitative information research, and the actual achievements and potential of bibliometrics and scientometrics.

Information metrics is therefore a dynamically developing research area of information science, concerned with a wide set of research problems, whose achievements increasingly often have practical applications. Various aspects of the information universe are subject to intensive exploration, while new research methods are being developed. These studies engage researchers from various disciplines, as well as interdisciplinary research teams. The present issue of ZIN publishes six articles which offer different approaches to various issues in quantitative information research.

The first article, written by Veslava Osińska (Institute of Information and Communication Research at Nicolas Copernicus University in Torun), Oleksander Sokolov (Department of Informatics, Faculty of Physics, Astronomy and Informatics at Nicolaus Copernicus University in Torun) and Aleksandra Mreła (Technical Department at Kazimierz Wielki University), *Nonlinear Estimation of Similarity Between Scientists' Disciplinary Profiles. Case Study*, focuses on the development of information metrics methodology. The authors seek to establish new methods of automatic analysis of scientometric data, employing the technology of artificial intelligence to generate disciplinary profiles of researchers alongside an in-depth analysis of the disciplinary structure of research units and teams.

The following two articles by Zbigniew Osiński (Department of Digital Humanities, Faculty of Humanities, Maria Curie-Skłodowska University in Lublin), are devoted to bibliometric analysis of publications on library and information science (LIS). First, *Analysis of the Thematic Overlap Between Library and Information Science and Other Sub-disciplines of the Social Communication and Media Sciences in Poland* employs the citation analysis of the contents of journals devoted to LIS, media studies, and cognition and social communication science, to determine their thematic proximity, which would justify classifying them as one discipline, social communication and media sciences, as Polish government's classification of scientific disciplines did in 2018. The subject of Osiński's second article, *The Usefulness of Data from Web of Science and Scopus Databases for Analyzing the State of a Scientific Discipline. The Case of Library and Information Science* is the capacity of the two most important multidisciplinary international databases of bibliographic citations to represent the state of scholarship in a given discipline. The article focuses on the peculiar case of LIS, which is a highly interdisciplinary research area, and for which it is difficult to identify the representative corpus of canon journals publishing the results of its research. In these circumstances, it would be disingenuous to rely on the simplified disciplinary classification systems of journals employed by the databases and arbitrarily limit the corpus only to those journals which publish the articles classified as belonging to the said discipline only. The tools for automatic filtering and analyzing large datasets available in these databases might be easy to use, but they do not produce reliable results.

The article of Viviane Couzinet (LERASS, University of Toulouse III Paul Sabatier – IUT, France), Regina Marteleto (Brazilian Institute of Information in Science and Technology, Rio de Janerio, Brazil) and Icléia Thiesen (Federal University of the State of Rio de Janeiro,

Brazil), *The Evolution of the Researchers' Bibliography: From Systematic Organization to Citation* highlighted a different problem of quantitative analysis based on digital sources of information processed by digital tools. The authors evaluated the resources indexed by the popular Google Scholar and discussed the quality (completeness, reliability, usefulness) of the bibliographic lists it generates, which additionally provide the number of citations for each listed text. Nowadays such lists are often intended to replace bibliographies of given researchers. The article focused on two outstanding scholars, Jean Meyriat from France, and Edson Nera da Fonesca from Brazil, whose research had a definitive influence on the development of information science in their respective countries. The analysis of those two cases vividly shows the limitations of the view offered by the Google Scholar.

The aim of Adam Jachimczyk's article, *Patent Applications for Electronic Publishing Market (2014–2018). Selected Issues* was to establish the state and pace of development of the e-publishing market basing on the analysis of patent applications from years 2014–2018 registered in the universally accessible database lens.org. The author analyzed the data according to various criteria: quantitative, geographical, the type of the innovations to be patented.

The issue ends with the article of Marcin Roszkowski, *Citation Type Analysis for Zagadnienia Informacji Naukowej – Studia Informacyjne (2016–2017).* The research, whose results are discussed in the article, was based on the articles published in ZIN and employed the method of citation analysis and a qualitative analysis of the types of the citations found; the author compared this results with the classification of bibliographic citations established by Bluma C. Peritz for the journals from the discipline of social sciences. The simplified understanding of bibliographic citations as an exponent of the quality of cited works, assumed by Robert Merton and others[6], is founded on the premise that authors of scientific articles refer to important earlier works, which influenced their own understanding of their subject of study. Thus, it is generally assumed that a citation number of a given scientific publication indicates its influence on the development of science, and its quality. The factors used to measure such influence generally do not account for the variation in citations: different motivations of the authors citing and varied use of the cited content. This approach has been criticized for a long time, and some studies in the discipline of bibliometrics and scientometrics seek to establish methods to identify different types of citations occurring in scientific writing, and their roles in solving further research problems. Automatic identification would allow a more specific view of the real impact of various ideas on the development of science. The analysis of citations in the articles published in *ZIN* is such a study.

The second thematic issue of *ZIN* thus presents varied studies of information metrics, highlighting the related methodological problems. As we share it with our readers, we hope that it will help them to better understand the role of quantitative information research, and its significance for the development of information science.

Barbara Sosińska-Kalata
ORCID 0000-0002-4511-4701

Warsaw, November 28, 2019

---

[6] R.K. Merton (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: Chicago University Press.