ZIN

Studia Informacyjne
Information Studies

VOL. 57 2019 NO. 2A(114A)
e-ISSN 2392-2648

WDIiB UW

SBP
STOWARZYSZENIE
BIBLIOTEKARZY
POLSKICH

Wersja elektroniczna jest wersją pierwotną czasopisma.
Zawartość czasopisma jest indeksowana w CEJSH, CSA, LISTA, Knowledge Organization Literature, PBN, PBB.

The electronic version is the original version of the journal.
The content of the journal is indexed in CEJSH, CSA, LISTA, Knowledge Organization Literature, PBN, PBB.

Pełne teksty roczników dostępne są w Archiwum SBP: www.sbp.pl/archiwumcyfrowe
oraz (do 2017 r.) w Mediotece WDIB UW: http://mediateka.uw.edu.pl/dlibra/

ZAGADNIENIA INFORMACJI NAUKOWEJ

Studia Informacyjne

ISSUES IN INFORMATION SCIENCE

Information Studies

VOL. 57 2019 NO. 2A(114A)
e-ISSN 2392–2648

Ministerstwo Nauki i Szkolnictwa Wyższego

Publikacja sfinansowana z grantu DUN „Wydawanie czasopisma *Zagadnienia Informacji Naukowej* – zwiększenie zasięgu czasopisma w obiegu krajowym i międzynarodowym" – zadanie finansowane w ramach umowy 921/P-DUN/2019 ze środków Ministra Nauki i Szkolnictwa Wyższego przeznaczonych na działalność upowszechniającą naukę.

Ministry of Science and Higher Education
Republic of Poland

The publication of this issue was funded by the grant "Publishing of *ZIN. Issues in Information Science. Information Studies* journal – increasing the national and international circulation" – an action financed under agreement 921/P-DUN/2019, supported by the Ministry of Science and Higher Education program "Actions to Promote Science: Publishing" (DUN).

WDIiB UW

SBP
STOWARZYSZENIE BIBLIOTEKARZY POLSKICH

WYDAWNICTWO SBP

Warszawa 2019

ISSUES IN INFORMATION SCIENCE – INFORMATION STUDIES

The core purpose of *Issues in Information Science – Information Studies (Zagadnienia Informacji Naukowej – Studia Informacyjne,* ZIN – *Studia Informacyjne)* is to provide a forum for the dissemination of scientific papers and research results in the field of information science and other disciplines which analyze social and technological aspects of various information-related activities performed by contemporary communities. Moreover, the journal is to disseminate critical reviews and summaries of new publications in the field of information science and reports from important conferences discussing contemporary information problems.

We publish papers in Polish or English. For each paper a set of metadata is provided: an abstract and keywords in both languages) as well as author's bio and contact information.

The subtitle of the journal – *Information Studies* – emphasizes the interdisciplinary nature of its subject profile covering a broad spectrum of issues studied by various academic disciplines and professional activity domains related to access to resources of recorded information and knowledge and the use of these resources by contemporary man and society. Other subjects to be covered by ZIN – *Information Studies* involve: (1) theoretical ponderings on the practice of information-related activities performed by various communities, (2) the results of research on the conditions influencing those activities and ways of improving methods and tools employed for the activities in question, (3) the methodology of information science research, information science history and education concerning the information science. The subject profile of ZIN – *Information Studies* covers, among else, the issues of:

- information services in institutions of science, culture, business, education and administration,
- information and knowledge management,
- traditional and online scholarly communication,
- information and knowledge organization,
- metadata theory and practice,
- Web 2.0,
- Semantic Web,
- information architecture,
- information websites usability,
- digital humanities,
- human-computer interaction,
- natural language processing,
- information retrieval,
- use of information and behavior of the information users,
- social response to modern information technologies,
- culture of information,
- information, digital and media skills,
- information policy,
- information ethics.

ZIN – *Information Studies* is addressed to: (1) information science teachers and lecturers, researchers and students, (2) practitioners of information-related activities who analyze methods and tools used to implement those activities in various domains and organizational environments, (3) politicians and donators related to information activities in various domains. The journal content may also be of some interest to teachers, students and researchers in other disciplines of science which deal with various aspects of information existence and use in the contemporary world.

ZIN – *Information Studies* is included in the list of journals scored by Polish Ministry of Science and Higher Education and indexed by: Central European Journal in Social Sciences and Humanities (CEJSH), Cambridge Scientific Abstracts (CSA), Library and Information Science and Technology Abstracts (LISTA), Polish Bibliography of Book Studies (PBB), Knowledge Organization Literature, Worldcat and Polish Scholarly Bibliography (PBN). The journal is registered in the European Reference Index for the Humanities (ERIH Plus).

ZAGADNIENIA INFORMACJI NAUKOWEJ – STUDIA INFORMACYJNE

Głównym celem *Zagadnień Informacji Naukowej – Studiów Informacyjnych* (ZIN – *Studia Informacyjne*) jest zapewnienie forum dla rozpowszechniania artykułów naukowych i wyników badań z zakresu nauki o informacji (informatologii) oraz innych dyscyplin, w których podejmowane są analizy społecznych i technologicznych aspektów działalności informacyjnej prowadzonej w różnych sferach współczesnego życia społecznego. Czasopismo służyć ma również rozpowszechnianiu krytycznych recenzji i omówień publikacji z tego zakresu oraz problemowych sprawozdań z ważnych konferencji poświęconych współczesnym problemom informacyjnym.

Publikujemy artykuły w językach polskim i angielskim. Każdy artykuł posiada zestaw metadanych: abstrakt i słowa kluczowe (w obu językach) oraz nota biograficzna autora i dane do kontaktu z nim.

Czasopismo adresowane jest zarówno do czytelnika polskiego jak i zagranicznego, publikujemy artykuły zarówno w języku polskim jak i angielskim. Podtytuł czasopisma – *Studia Informacyjne* – podkreśla interdyscyplinarny charakter jego profilu tematycznego, który obejmuje szeroki zakres problemów podejmowanych przez dyscypliny akademickie i dziedziny działalności zawodowej związane z zapewnianiem dostępu do utrwalonych zasobów informacji i wiedzy oraz ich wykorzystywaniem przez współczesnego człowieka i współczesne społeczeństwo. Czasopismo publikuje też artykuły prezentujące teoretyczną refleksję o praktycznej działalności informacyjnej prowadzonej w różnych dziedzinach i obszarach życia społecznego, a także wyniki badań służących poznaniu różnych uwarunkowań tej działalności oraz doskonaleniu jej metod i narzędzi. Na łamach ZIN publikowane są także artykuły poświęcone metodologii badań informatologicznych, historii nauki o informacji oraz edukacji w zakresie nauki o informacji. Profil tematyczny półrocznika ZIN – *Studia Informacyjne* obejmuje m.in. problematykę:

- usług informacyjnych w instytucjach nauki, kultury, biznesu, edukacji i administracji,
- zarządzania informacją i wiedzą,
- komunikacji naukowej i cyfrowej komunikacji naukowej,
- organizacji informacji i wiedzy,
- teorii i praktyki metadanych,
- zagadnień Web 2.0,
- zagadnień Sieci Semantycznej,
- architektury informacji,
- projektowania użytecznych serwisów informacyjnych,
- humanistyki cyfrowej,
- interakcji człowiek – komputer,
- przetwarzania języka naturalnego,
- wyszukiwania informacji,
- wykorzystywania informacji i zachowań informacyjnych użytkowników,
- społecznej recepcji nowoczesnych technologii informacyjnych,
- kultura informacji,
- kompetencji informacyjnych i cyfrowych,
- polityki informacyjnej,
- etyki informacyjnej.

Zagadnienia Informacji Naukowej – Studia Informacyjne adresowane są do wykładowców, badaczy i studentów nauki o informacji, a także praktyków działalności informacyjnej, krytycznie analizujących metody i narzędzia jej realizacji w różnych środowiskach dziedzinowych i organizacyjnych oraz polityków i donatorów działalności informacyjnej w różnych dziedzinach. Lektura czasopisma może też zainteresować wykładowców, studentów i badaczy innych dyscyplin, które zajmują się różnymi aspektami funkcjonowania informacji we współczesnym świecie.

Zagadnienia Informacji Naukowej znajdują się na liście czasopism punktowanych Ministerstwa Nauki i Szkolnictwa Wyższego. Czasopismo jest indeksowane w bazach: Central European Journal in Social Sciences and Humanities (CEJSH), Cambridge Scientific Abstracts (CSA), Library and Information Science and Technology Abstracts (LISTA), Polska Bibliografia Bibliologiczna (PBB), Knowledge Organization Literature, Worldcat, Polska Bibliografia Naukowa (PBN). Czasopismo jest zarejestrowane w European Reference Index for the Humanities (ERIH Plus).

QUANTITATIVE INFORMATION RESEARCH: METHODS, APPLICATIONS, PROBLEMS

ILOŚCIOWE BADANIA INFORMACJI: METODY, ZASTOSOWANIA, PROBLEMY

Contents | Spis treści

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.

Nonlinear Estimation of Similarity Among Scientists' Disciplinary Profiles. A Case Study

Veslava Osińska
ORCID 0000-0002-1306-7832
Institute of Information and Communication Research, Nicolas Copernicus University in Toruń, Poland

Oleksandr Sokolov
ORCID 0000-0002-6531-2203
Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University in Toruń, Poland

Aleksandra Mreła
ORCID 0000-0002-2059-864X
Institute of Informatics, Kazimierz Wielki University, Poland

Abstract

Purpose/Thesis: Authors estimate the disciplinary similarity of researchers according to selected academic units with a different cross-section of specializations. The paper presents the model for studying disciplinary diversity of scientific units. The premise of the article is that knowledge of disciplinary profiles of researchers can be applied to create interdisciplinary teams, or one disciplinary team with a focused specializations.
Approach/Methods: The approach is based on the visualization and comparison of disciplinary space and space of co-authorship. Fuzzy logic and aggregation norm were used to calculate disciplinary weights of each journal listed in the database. For visualization, new, dimension reduction algorithm t-SNE was applied. Achieved results were verified by using the expert's knowledge.
Results and conclusions: In the evaluation of scientific collaboration, a co-authorship relationship can be complemented by researchers' disciplinary profiles represented by aggregation norm. Thanks to the continuity of researchers' publishing activity, the proposed measure based on the disciplinary profile is stable.
Research limitations: The sample of both selected teams and journals database is limited. The journals from WoS/Scopus list were considered because analyzed researchers publish articles there. Additionally, during linking these two databases, problems of matching journals titles appeared.
Practical implications: The authors proposed a model of evaluating scientists' disciplinary similarity and further, to estimate the potential of their collaboration.
Originality/Value: This approach applies fuzzy logic algorithms to quantifying scientific interests and is another rare instance of practical application of artificial intelligence algorithms (fuzzy logic) in scientometric studies.

Keywords

Aggregation norm. Authorship. Disciplinary profiles. Fuzzy logic. Scientometrics. t-SNE plot.

Received: 25 August 2019. Reviewed: 28 October 2019. Accepted: 30 November 2019.

1. Introduction

In scientific communication, the disciplinary orientation of an individual researcher carries essential information about them, their groups, cooperation areas, fields in which they conduct their research. One can say that it constitutes the principal identifier of individuals in terms of specialization, but it can also indicate a shared research framework within a particular scientific community. If fields of research were to be represented graphically, such a representation would show where they activities overlap, and where they lie far away from each other, which could be used for organizing and designing the multidisciplinary teams. This insight would be especially valuable in the context of contemporary science policy in Poland, where the Ministry of Science emphasizes conducting joint research in all scientific domains, team fusion and multi-authorship publishing.

From the researcher's disciplinary profile, their publications output narrowed to the journal data set, and their association to disciplines might be inferred. Scientific journals are usually assigned a specific database, such as Web of Science (WoS) or Scopus collections by the professionals such as editors, domain classifiers, librarians, information specialists, and so on (Leydesdorff et al., 2019). A table of journal titles, mapping their disciplinary allegiance according to last Science Classification (MNSW, 2018) changes, introduced by the Ministry of Science and Higher Education has been recently published online.

The classification of sciences in Poland consists of scientific disciplines (previously) and eight branches of science, resembling WoS research areas: medicine and health sciences, technical sciences and engineering, exact and natural sciences, humanities, social sciences, theological sciences, agricultural sciences and arts.

While creating scientific profiles of researchers, bibliometrics should take into consideration that some scholar databases are oriented towards a particular scientific domain (Kulczycki & Rozkosz, 2017). For example, WoS covers mostly natural sciences and engineering (Kokowski, 2015); Scopus also specializes in hard sciences, although to a lesser degree, and PubMed database is dedicated to medical sciences.

2. The objectives

By using mapping algorithms, a visual representation of researcher network can be constructed. Ties will represent collaboration between people in co-authorship relationships (Fields, 2015; Small & Garfield, 1986). Based on the citations database, it is possible to see which research fields are close, thanks to techniques such as direct citations, co-citations, or bibliographic coupling (see Boyack et al., 2005; Huang, 2015; Jarneving, 2005; Shibata et al., 2009). In the bibliometric study, it is a common practice to use the linear counting of records after aggregation processes according to selected units, for example, author, keyword, discipline, or topic. However, the calculation of impact weights can be based on fuzzy logic algorithms, rather than on direct operations. Scientometricians turn to fuzzy logic rarely, which is confirmed by only nine items retrieved from WoS (the result of a query "fuzzy logic" AND "scientometric*"). Most papers refer to bibliometric study on fuzzy sets and algorithms used in other scientific domains. The query also yielded analyses of citation networks presenting the ties between information sciences and fuzzy systems (Merigóa et al., 2108; Yu & Shi, 2015).

The main difficulty which the authors met in the context of citations study is that there is no unified citation database concerning Polish scientists; however the authors proposed to analyze the similarity between the researchers established by the consideration of the characteristics of journals where they publish. Thus, taking into account both co-authorships and disciplinary embedding should compose a comprehensive view of the similarity and collaboration between the members of the scientific community. Final evaluation of these characteristics was made using visual maps generated by t-SNE algorithm.

Two sets of scientists were taken into consideration, the first team from the Department of Informatics and the second one from the Interdisciplinary Centre for Modern Interdisciplinary Technologies at the Nicolaus Copernicus University in Torun (NCU), Poland. These teams were chosen because two of the authors know the staff of these groups of scientists and can interpret the achieved results. Thus, the authors can include their experts' assignment of the spatial arrangement of data on the visualization thanks to authors' acquaintance with the members. Moreover, the first team is rather homogeneous (it is called GROUP_H) because the members represent one knowledge domain – computer science with the relation to technology and engineering domain. The other team (GROUP_M) is more multidisciplinary, gathering researchers whose research encompasses natural sciences (chemistry, physics, biology), social sciences (psychology, cognitive sciences) and health sciences (neurology, physiotherapy).

Mapping results are represented in t-SNE space and co-authorships graphs, which are compared and interpreted by experts.

Research questions:

- How do two approaches, co-authorship and disciplinary similarity, characterize researchers, and do they complement each other?
- Do co-authorship and disciplinary similarity characterize researchers, and do they complement each other?
- Does disciplinary similarity studies foster collaboration potential?
- Is fuzzy logic appropriate for the representation of relationships between disciplines and journals, and disciplines and authors, in the multidisciplinary world of science?
- Is fuzzy logic more fit to calculate the contribution to the researchers' disciplines (numbers from the interval) which constitutes their scientific profiles, than the linear sum of published papers?

Hypothesis: In the disciplinary representation of scientists, co-disciplinary relationships based on fuzzy logic should be compared with the classically calculated co-authorships. These combinations give the best results of a real disciplinary profile of the selected group; the expert's assignment can confirm that.

3. Nonlinear aggregation of scientific contribution based on fuzzy logic

The paper aims to find a method to calculate the scientists' contribution to discipline/science. The basis of calculating scientific achievements is the number of articles published by researchers considered. Next, aggregation functions are applied, of which one is the total number of papers. However, this method has weaknesses; the result belongs to the interval $[0,+\infty)$, so it is difficult to compare the contribution to science between two

different scientific units (with a varied number of their members). The authors propose the nonlinear function of aggregation, whose result belongs to the interval [0,1], makes it much easier to compare the contribution to science of a few different scientific units.

Fuzzy logic is one of the most efficient tools for describing uncertainty characterizing artificial intelligence. The research area analysis, the characterization of the researchers in terms of their contribution to the discipline, level of collaboration, the similarity of discipline, etc. can be more efficient described based on fuzzy logic. All of these concepts are fuzzy for two reasons.

The first reason is the lack of objective measuring instruments. It is difficult to find a unit of measurement of the scientist's contribution to the discipline. Of course, it may be measured by a number of articles in a journal; however, a linear dependence of the sum of papers may lead to the need for constant scaling during the comparison, for instance, of two scientists. The limitation of the linear calculation of contribution values became more significant in the case of more than two scientists with the high diversity of their output. If we have teams of researchers publishing very differently, then normalization needs to be updated with every new article. Linear normalization by the maximum number of articles leads to a huge divergence of contribution value, which does not coincide with reality. The authors propose a fuzzy value for describing a contribution unit for the scientist.

The second reason is the subjective nature of discipline/disciplines assigned to a scientific journal, which is established by experts/editorial boards. Instead of a binary relation, it is more natural to use fuzzy one, which describes the degree of such a relation more realistically or accurately. Moreover, in the current version of the scientific journals' classification, measures assigned to the journals, for example, Impact Factor, are defined. They also need to be taken into account. The following example will demonstrate this approach.

Let the fuzzy value A be the contribution of one scientist to one discipline after publishing one paper. Let D, J, S denote spaces of disciplines (as said above, according to the new classification of scientific domains and disciplines used by Polish Ministry of Science and Higher Education (MNSW, 2018), there currently are disciplines in Poland), journals, and scientists, respectively. Moreover, let

$$D = \{D_i, i = 1,2,...,I\}, J = \{J_m, m = 1,2,...,M\}, S = \{S_n, n = 1,2,...,N\}.$$

Let $R \subseteq J \times D$ be a fuzzy relation between journals and disciplines for one scientist, where $J \times D$ denotes the Cartesian product of spaces J and D. Then $R(J_m, D_i)$ is equal to A if discipline D_i is assigned to journal J_m for each i and m and it is equal to 0 otherwise. Hence, if according to the Classification Table journal J_i is assigned only to discipline D_3, then $R(J_1, D_1) = R(J_1, D_2) = 0$ and $R(J_1, D_3) = A$. If the next paper is published in journal J_2 which is assigned to disciplines D_2 and D_3, then $R(J_2, D_1) = 0$ and $R(J_1, D_2) = R(J_1, D_3) = A$. Table 1 presents the total achievement of this scientist when the both journals are taken into account.

Tab. 1. Values of relation for the discussed scientist

Disciplines / **Journals**	D_1	D_2	D_3
J_1	0	0	A
J_2	0	A	A

To calculate the contribution $R(D_i)$, where i = 1,2,3, of this scientist to disciplines D_i, the following function is applied:

$$R(D_i) = R(D_i,J_1) + R(D_i,J_2) - R(D_i,J_1) \cdot R(D_i,J_2). \quad (1)$$

Hence,

$$R(D_1) = R(D_1,J_1) + R(D_1,J_2) - R(D_1,J_1) \cdot R(D_1,J_2) = 0 + 0 - 0 \cdot 0 = 0,$$
$$R(D_2) = R(D_2,J_1) + R(D_2,J_2) - R(D_2,J_1) \cdot R(D_2,J_2) = 0 + A - 0 \cdot A = A,$$
$$R(D_3) = R(D_3,J_1) + R(D_3,J_2) - R(D_3,J_1) \cdot R(D_3,J_2) = A + A - A \cdot A = 2A - A^2.$$

This formula is a special case of an optimistic fuzzy aggregation norm (Sokolov et al., 2018).

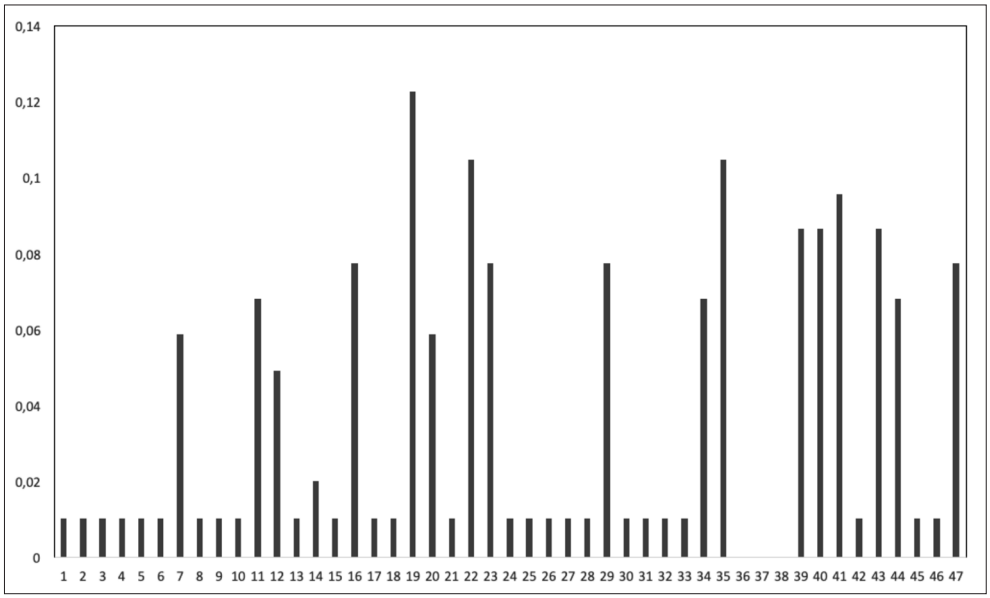

Fig. 1. The graphs of scientific contribution to disciplines of one scientist

Thus, the graph above (Fig. 1) shows the scientific contribution to disciplines of one exemplary scientist. The x-axis presents the numbered disciplines, and the y-axis shows the scientific contribution to these disciplines of this scientist calculated by formula (1). Moreover, by the application of the fuzzy optimistic aggregation norm, the total contribution to science by one scientist can be calculated, and for this scientist, it is equal to 0.818873.

4. Data and methods

4.1. Data

As authors found in previous research (Osińska et al., 2018), the bibliographic database Expertus (n.d.) is the most complementary and updated source of scientific output on the local level, i.e., in a selected university. This Web platform is included in the science-policy

of NCU, whose authorities force the staff to update their bibliographic records of published papers indicating that it is the primary condition to get promoted. Expertus platform has the same interface in most universities in Poland; it allows the user to filter records by researchers' names and organizational units. Bibliographical metadata in both basic and extended forms of selected records can be downloaded in text format. An article's metadata such as author(s), title, source, and publication year have been processed for further analysis.

Following the expectations of authorities to publish in selected domains, the members of these two organizational units (Department of Informatics and CMIT), publish journal articles rather than books and book chapters because the primary Web filter was predefined to set on formal publication type to be equal "002" (journal articles) and the authors assumed that Web of Science and Scopus databases fully cover downloaded data. The current conditions do not require to apply the set Polish journals, but the authors take into account the need for extending their approach to the broader scope of analysis units in the future. Then the database such as Arianta (http://arianta.pl) a continuously updated Polish journals reference list, will be useful (Kulczycki & Rozkosz, 2017).

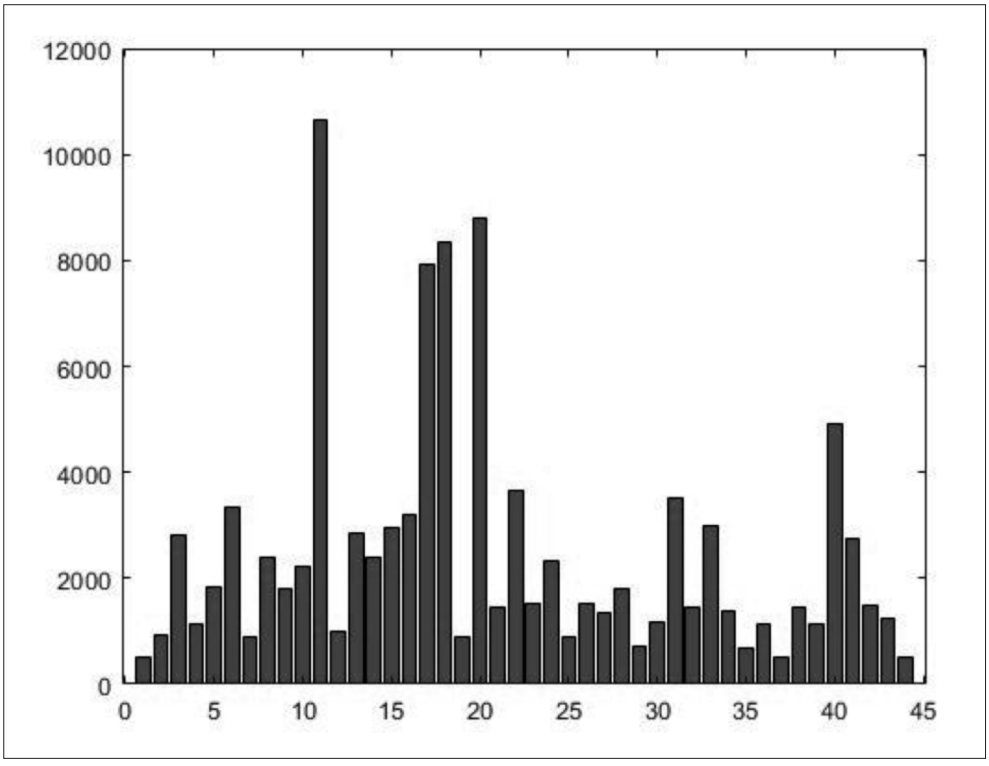

Fig. 2. The histogram of journal disciplinary distribution in the dataset of journals title list (N=29037) in both WoS and Scopus databases in 2019 (x-axis denotes the numbered disciplines and y-axis denotes the frequency of journals connected with these disciplines)

The ministerial table of disciplinary appurtenance of analyzed journals (Komunikat MNSW, 2019) was used to construct the mapping space which would account for all the researchers who belong to selected units. It consists of a list of journals (N=29037) from both WoS and Scopus databases as well as of the sequence of assigned disciplines, in some cases amounting to as many as 20. The disciplinary distribution of journals dataset is presented in Figure 2, where the x-axis denotes the numbered disciplines, and y-axis – the frequency of journals connected with these disciplines. As it can be seen, the majority of journals are described by one, two, or three categories, but often multidisciplinary titles can be found in the right tail of the distribution. Outstanding discipline in Figure 2 numbered as 11, describing more than 10,000 journals, is biomedical engineering.

The Expertus database is constructed with coding external and internal co-authors (i.e. University employees). This information allows filtering the data about collaboration within the University. Finally, by selecting certain units, the co-authorship records are obtained, and the symmetric matrixes are constructed. The rows and columns consist of all members of analyzed Departments, the cells – of the numbers of their joint publications. The Departments GROUP_H and GROUP_M numbered 18 and 16 members, respectively. This representation of data follows mapping by graph layouts algorithms.

4.2. Methods

Figure 3 presents the stages of the research. The data regarding the particular researcher's output was collected, taking into account journals where the articles were published. Journals' datasets were matched with the WoS/Scopus classification table.

In Table 2, the rows represent scientists and columns – their extracted disciplines, as gathered from the journals' list. The table was filled out with weights calculated by using the optimistic fuzzy aggregation norm. The data from the table was visualized in the t-SNE space representing the disciplinary similarity between scientists.

Tab. 2. A part of the fuzzy relation between scientists and disciplines

Disciplines / Scientist	Archeology	Architecture and urban studies	Astronomy	Automatics, electronics and electrotechnics	Economy and Finances	Philosophy
S1	0	0	0	0	0.01	0
S2	0	0	0	0	0	0
S3	0	0	0	0	0	0
S4	0	0	0	0	0	0
S5	0	0.0199	0.01	0.05852	0	0
S6	0	0	0	0	0	0
S7	0.01	0	0.095618	0.05852	0.039404	0.029701

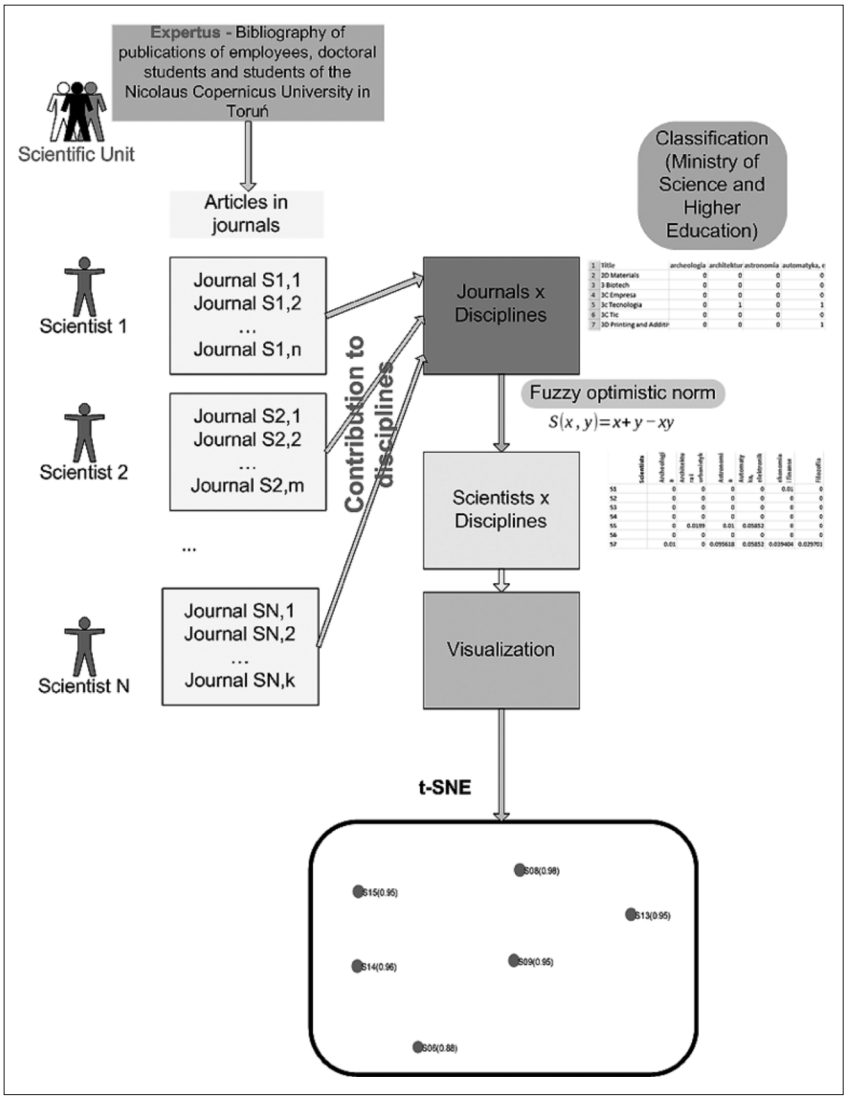

Fig. 3. The stages of the research

4.3. Optimistic fuzzy aggregation norm

Being a fuzzy variable, A must be between 0 and 1. The function S must increase with the scientist's contribution to the disciplines by adding the following article, so the authors used an optimistic fuzzy aggregation norm (Sokolov et al., 2018). One of examples of the function S is the following function (Fig. 4), for x,y ∈ [0,1].

$$S(x,y) = x + y - xy \qquad (2)$$

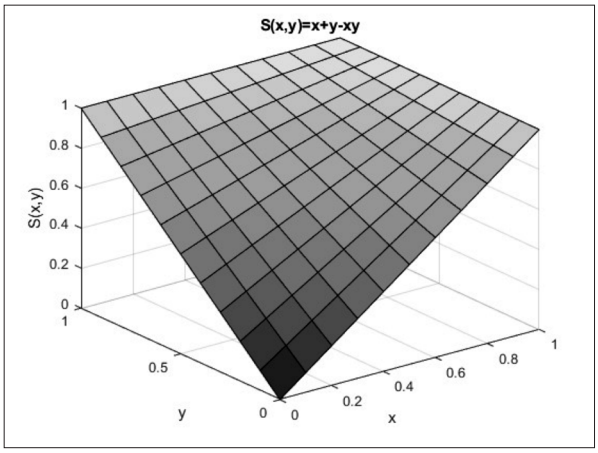

Fig. 4. 3D graph of function

To demonstrate the method, let the abovementioned value A be equal to A = 0.01. If we use the optimistic fuzzy aggregation norm as a function of scientific contribution, then

$R(D_3) = R(D_3,J_1) + R(D_3,J_2) - R(D_3,J_1) \cdot R(D_3,J_2) = A + A - A \cdot A = S(A,A) = 0.0199.$

In such a way, the relation between scientists and disciplines is established.

The value of the variable A is chosen so that the estimated scientist's contribution to the disciplines representative of the research team would be as diverse as possible. It is set experimentally. If it is observed that the total scientist's contribution to the disciplines is approximately the same and close to 1 (maximum value), the value of variable A should be reduced.

4.4. T-SNE algorithm

T-distributed Stochastic Neighbor Embedding (t-SNE) is a technique based on machine learning developed by Laurens van der Maaten and Geoffrey Hinton (2008). This technique is non-linear and well suited for embedding multidimensional data for visualization in a 2 or 3-dimensional space. This method uses a t-distribution rather than a Gaussian distribution to compute the similarity between two points in the low-dimensional space (Maaten & Hinton, 2008). Explicitly, it models each high-dimensional object with a two- or three-dimensional point in such a way that there is a high probability of similar objects being modeled by nearby points, and non-similar objects are modeled by distant points.

The advantage of t-SNE visualization relies on reducing the tendency to crowd points together in the center of a map. It better reveals structure on many different scales. This algorithm has a wide range of applications, including computer security research, music analysis, cancer research, bioinformatics, and biomedical signal processing.

4.5. Expert assessment

The two authors are the members of considered scientific units, and they are knowledgeable about these teams, which impacts the internal relations between researchers: for example, "who collaborates with whom?". In this case, they are the experts because they have more

knowledge than others about the research interests of a particular scientist. It is worth mentioning that the considered teams have got only a small number of members (18 and 16). Hence, the authors-experts can estimate the results of visualization taking into account the context of relationships between the scientists.

The number of co-authorship links is called *a degree* in network analysis (Barabasi, 2010). The larger the degree, the closer the connection (and/or the disciplinary similarity) between the authors is.

However if we focus on the disciplinary similarity, we will face a serious problem. Portals such as Research Gate or Academia.edu use specialized algorithms very effectively to search for researchers with similar interests. In the case explored in this article, the authors played the role of experts and evaluated the results of visual representation of contribution to science and co-authorship of these scientists.

4.6. Tools

For bibliographic data cleansing, processing and extraction R scripts were used. In the Matlab environment, the weights for the scientists-disciplines table were calculated. The t-SNE technique helps to visualize the relations between scientists in 2D space. Co-authorship graph layout was provided by Gephi platform (https://gephi.org) with embedded appearance tools. Data and results were communicated and discussed between the authors in the Microsoft Excel environment.

5. Visualization results

Below are the visualization layouts made by both algorithms: t-SNE and force-directed graph for two analyzed groups. The t-SNE maps illustrate disciplinary relationships, the graphs – collaboration in publishing. According to the mapping principle, the distances between data points reflect the similarity between the researchers in the terms of disciplinary activity. The links on the co-authorship graph show the ties between the co-authors. Accordingly, the node size depends on the number of the articles, so the larger the data point, the more significant scientific the output of the considered scientist is.

5.1. The co-authorship graph

Figure 5 presents co-authorship networks within both groups. The researchers' names are coded by letters: "S" – for the employees within the department, and "O" – for the external scholars. The links between nodes indicate the existence of joint publications. As the amount becomes higher, the edges thicken. The size of the node is proportional to the number of the papers written by the given researcher. Thanks to linked components, we can track the subgroups of collaborators and the people individually conducting research (separate, regularly arranged nodes in Fig. 5). As can be expected, there are more collaboration units in the multidisciplinary group GROUP_M.

By the application of fuzzy optimistic aggregation norm S, the total contribution to science of these scientists was calculated and is presented in the graphs (Fig. 6) by the size

of circles representing scientists. A large size of a circle (the high contribution to science) can be achieved in two cases:

(1) The scientist is devoted to one discipline and achieved excellent results publishing in many journals assigned to this discipline.
(2) The scientist's output is interdisciplinary and he/she has published papers in many journals assigned to many disciplines.

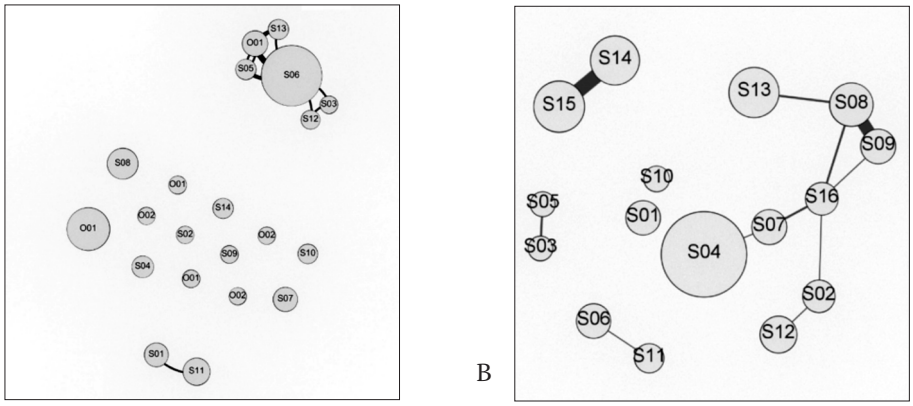

Fig. 5. Co-authorships graph layout for GROUP_H (A) and GROUP_M (B)

The more papers published, the higher the value of contribution to science. To reduce the 47-dimensional space to 2-dimensional one, a modern t-SNE technique was applied. Thanks to such visualization the relationships between points (scientists) can be easily considered and discussed.

5.2. T-SNE space

The disciplinary similarity of the researchers within the groups are presented by t-SNE space in Figure 6. The smaller the distance between the nodes, the more significant the disciplinary similarity between the respective scientists. For each scientist belonging to one of these two teams, with the application of the S norm, contributions to each discipline and finally to science were calculated, and the result is additionally presented in brackets beside the node.

6. Discussion

The classical approach used in bibliometrics is based on the co-authorships analysis. If we refer to the t-SNE maps, the name-coding reveals regularities in both arrangements. Close relationships on the co-authorship map (Fig. 5) mean that there are publications available by the pair of researchers, identified by the linkage. What we can observe in Figure 6 is that the scientists represented by the large circles play a central role in co-authorship and are attractors (see Buskell, 2017) to other scientists. So, they attract other researchers to collaborate with them. On the other hand, the close research interests do not always cause collaboration in

publishing. The co-authorship graph shows no permanent collaboration between researchers, but only intermittent connections, which assist with selected study themes and directions. Nor does it manifest the scientific interests of researchers and their relationships.

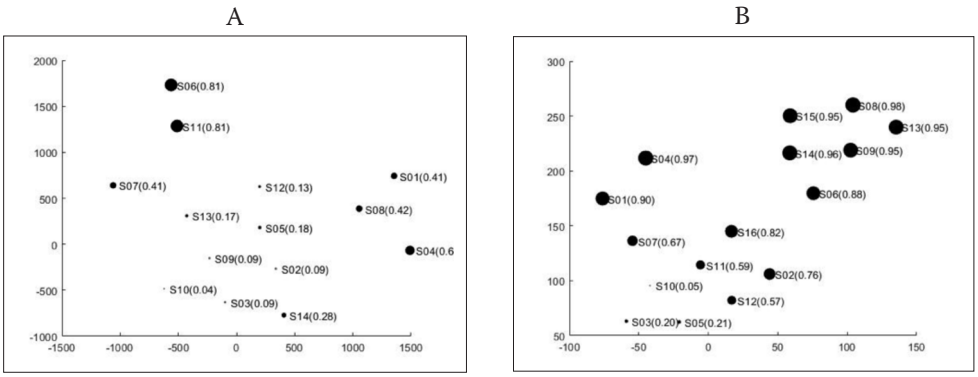

Fig. 6. The disciplinary similarity of scientists
in t-SNE space for GROUP_H (A) and GROUP_M (B)

Quantitative comparison of two spatial arrangements of the sets of scientists is presented in Table 3 and Table 4. For n=18 nodes, it is theoretically possible to extract $n(n-1)/2 = 153$ links. A close relationship here means direct neighboring between the nodes, a distant relationship – all other pairs. Table 3 shows an extreme case: researchers who publish joint articles but are affiliated with different disciplinary and vice versa – represented by the most significant numbers (6, 6), which are in diagonal cells. In the case of GROUP_H, which is a homogenous computer science team, all scientists work in the same discipline but in different specializations, and they do not cooperate.

Tab. 3. Proximity matrix of two visual representations for GROUP_H

	Close relationships on t-SNE map	**Far relationships on t-SNE map**
Linkage on co-authorship map	3	6
No linkage on co-authorship map	6	4

Tab. 4. Proximity matrix of two visual representations for GROUP_M

	Close relationships on t-SNE map	**Far relationships on t-SNE map**
Linkage on co-authorship map	7	3
No linkage on co-authorship map	2	11

It is noticeable that they do not create teams which involve members of different specializations. In the case of GROUP_M, two approaches became complementary because they show the similar relationships: a linkage reflects the similarity in disciplines and vice versa, what can be observed in Table 4 when right diagonal numbers dominate (7, 11).

The co-authorship data reflects the real state in particular conditions, depending on the collaboration period. For example, the data can indicate a common study within the grants, at a particular stage of the researcher's career, affected by individual predispositions and other factors. The disciplinary profile of a scientist can give valuable information about the researcher's interests, their history, and development, that qualifies as a stable and independent parameter describing their scientific output.

The authors learned about the t-SNE algorithms and experienced within interpreting the visualizations. Relationships between data can be found on the periphery of the visual layout; non-clustered data remains in the center in the form of a regular grid. They can be interpreted as data with no or loose relationships.

The applied optimistic fuzzy aggregation norm, defined by the formula (2), is not linear, and the S norm does not imply linearity. Moreover, to make direct conclusions about complementing both approaches in a study of collective profiles, we need to operate on a big data set. In the future, research will be extended to include the data from several departments.

7. Conclusion

Authors proposed a combined approach to evaluate the researchers' disciplinary profiles and represent their similarity according to their interests. This approach is a combination of two unique relationships disciplinary profiles and co-authorships of the researchers within one team.

The disciplinary similarity of researchers allows predicting their future cooperation, even in the case of scientists who have not written a joint article yet. Thus, the chance for their collaboration can be estimated. If the disciplinary similarity is high, it can be expected that the scientists will write a joint article in the future. Moreover, it encourages them to do research together.

The cooperation in an interdisciplinary group is distributed according to the disciplinary interests of members; however, in a homogeneous group, there are no observed direct dependencies in collaborators selection. This conclusion from the pilot study is essential for future planning research.

To sum up the authors conclude that co-authorship and disciplinary similarity characterize researchers in two different ways. The disciplinary similarity profile can show the collaboration potential because scientists with a high level of disciplinary similarity attract young scientist or are a center of a group of researchers. Fuzzy logic is developed to deal with different kinds of uncertainty and estimation in social sciences, so fuzzy relations, such as relations between disciplines and journals, or between disciplines and authors, are the best way to represent such subjective coupling in the multidisciplinary world of science. The application of fuzzy logic to calculate the contribution to discipline/science or discipline similarity is possible because all discussed values belong to the interval [0,1]. Hence, they are easy to compare and there is no need for continuous scaling. In this context, the assumed hypothesis, i.e., that two combinations of relationships give the best results of a real disciplinary profile of the selected group, was confirmed.

Fuzzy logic used for the design of intelligent systems, which is an area entering to information sciences (Babik, 2016), can be a natural way to find levels of connection between two

spaces in scientometrics such as disciplines, journals and authors. In this paper, not all problems and research questions are fully developed, mainly because of the limited data sample, but this way of solving these problems seem to be appropriate, and the authors are going to perform a series of further research activities on the disciplinary similarity of researchers.

The proposed approach favors scientists having significant output. For young researchers, the results can be misleading because they have published few articles so far, so perhaps they will derive a high disciplinary profile of co-authorship. However, thanks to the presented data, it is possible to predict their interests in the future.

(1) In the case of poor data about co-authorships, the mapping does not reflect the real state of scientific interest. In this case, the authors recommend using the disciplinary profile.
(2) Robust disciplinary dissemination (Ministerial WoS categorization) makes it difficult to describe the primary domain, and therefore, the authors' analysis can be based only on disciplines.
(3) One can be tempted to say that Ministerial WoS/Scopus categorization of journals is not appropriate for study researchers disciplinary profiles.

References

Babik, W. ed. (2016). *Nauka o informacji*. Warszawa: Wydaw. SBP.

Barabasi, A. (2010). Network Science [online] [10.08.2019], http://networksciencebook.com/

Boyack, K.W., Klavans, R., Börner, K. (2005). Mapping the Backbone of Science. *Scientometrics*, 64(3), 351–374, http://dx.doi.org/10.1007/s11192-005-0255-6

Buskell, A. (2017). Cultural Attractor Theory and Explanation. *Philosophy, Theory and Practice in Biology*, 9 (13), http://dx.doi.org/10.3998/ptb.6959004.0009.013

Expertus (n.d.). Expertus – Bibliografia publikacji pracowników, doktorantów i studentów Uniwersytetu Mikołaja Kopernika w Toruniu [online]. Biblioteka Uniwersytecka w Toruniu [12.08.2019], http://bg.cm.umk.pl/splendor/umk/

Fields, C. (2015). Co-authorship Proximity of A. M. Turing Awarded John von Neumann Medal Winners to the Disciplinary Boundaries of Computer Science. *Scientometrics*, 104(3), 809–825, https://doi.org/10.1007/s11192-015-1575-9

Huang, M-H.; Chang, C-P. (2015). A Comparative Study on Detecting Research Fronts in the Organic Light-Emitting Diode (OLED) Field Using Bibliographic Coupling and Co-Citation. *Scientometrics*, 102(3), 2041–2057, https://doi.org/10.1007/s11192–014-1494–1

Jarneving, B. (2005), A Comparison of Two Bibliometric Methods for Mapping of the Research Front. *Scientometrics*, vol. 65(2), 245–263, https://doi.org/10.1007/s11192-005-0270-7

Kokowski, M. (2015). Science of Science (naukoznawstwo) in Poland. The Changing Theoretical Perspectives and Political Contexts – A Historical Sketch from the 1910s to 1993. *Organon* [online], (47), 147–237 [23.10.2019], http://www.cyfronet.krakow.pl/~n1kokows/Kokowski-Organon-2015.pdf

Komunikat MNSW (2019). *Komunikat Ministra Nauki i Szkolnictwa Wyższego z dnia 31 lipca 2019 r. w sprawie wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych wraz z przypisaną liczbą punktów* [online]. Ministerstwo Nauki i Szkolnictwa Wyższego [23.10.2019], http://www.bip.nauka.gov.pl/akty-prawne-mnisw/komunikat-ministra-nauki-i-szkolnictwa-wyzszego-z-dnia-31-lipca-2019-r-w-sprawie-wykazu-czasopism-naukowych-i-recenzowanych-materialow-z-konferencji-miedzynarodowych-wraz-z-przypisana-liczba-punktow.html

Kulczycki, E., Rozkosz, E. (2017). Publication Patterns in the Social Sciences and Humanities in Flanders and Poland. In: *Proceedings of ISSI 2017 Wuhan: 16th International Society of Scientometrics and Informetrics Conference*. Wuhan, China, 16–20 October 2017, 95–104.

Kulczycki, E. Stan praktyk publikacyjnych polskich czasopism naukowych w 2017 roku. [online] Report 2018 [10.08.2019], http://excellence-project.zut.edu.pl/

Leydesdorff, L., Wagner, C., Bornmann, L. (2019). Diversity Measurement: Steps towards the Measurement of Interdisciplinarity? *Journal of Informetrics*, 13(3), 904–905, https://doi.org/10.1016/j.joi.2019.03.016

Maaten L. van der, Hinton G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9, pp. 2579–2605.

Merigó, J.M., Pedrycz, W., Weber, R., la Sotta, C. (2018). Fifty Years of Information Sciences: A Bibliometric Overview. *Information Sciences*, 432, 245–268, https://doi.org/10.1016/j.ins.2017.11.054

MNSW (2018). Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 20 września 2018 r. w sprawie dziedzin nauki i dyscyplin naukowych oraz dyscyplin artystycznych [online]. DzU, poz. 1818, http://prawo.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20180001818

Osińska, V., Kowalska, M., Osiński, Z. (2018). The Role of Visualization in the Shaping and Exploration of the Individual Information Space: Part 1. *Knowledge Organization*, 45 (7), 547–558.

Shibata, N., Kajikawa, Y., Takeda, Y., Matsushimaet, K.(2009), Comparative Study on Methods of Detecting Research Fronts. *Journal of the American Society for Information Science and Technology*, 60(3), 571–580, https://doi.org/10.1002/asi.20994

Small, H., Garfield, E. (1986) The geography of science: disciplinary and national mappings. *Journal of Information Science*, 11(4), 147–159, https://doi.org/10.1177/016555158501100402

Sokolov, O., Osińska, W., Mreła, A., Duch, W. (2018). Modeling of Scientific Publications Disciplinary Collocation Based on Optimistic Fuzzy Aggregation Norms. In: *Advances in Intelligent Systems and Computing. Vol. 853. Information Systems Architecture and Technology: Proceedings of 39th International Conference on Information Systems Architecture and Technology* – ISAT 2018, 145–156 [online]. Springer Link [10.08.2019], https://link.springer.com/chapter/10.1007/978-3-319-99996-8_14

Yu, D., Shi, Sh. (2015). Researching the Development of Atanassov Intuitionistic Fuzzy Set: Using a Citation Network Analysis. *Applied Soft Computing* 32, 189–198, https://doi.org/10.1016/j.asoc.2015.03.027

Nieliniowa ocena podobieństwa profili dyscyplinarnych naukowców. Studium przypadku

Abstrakt

Cel/Teza: Autorzy oceniają podobieństwo dyscyplinarne badaczy wybranych jednostek akademickich o różnej naukowej specjalizacji i przedstawiają model badania różnorodności dyscyplinarnej jednostek naukowych. Założeniem artykułu jest to, że wiedza o profilach dyscyplinarnych badaczy może być wykorzystana do stworzenia zespołów interdyscyplinarnych lub jednego zespołu dyscyplinarnego o ukierunkowanych specjalizacjach.

Koncepcja/Metody badań: Metoda oparta jest na wizualizacji oraz porównaniu przestrzeni dyscyplinarnej i współautorstwa. Do obliczenia wag dyscyplinarnych każdego rozważanego czasopisma w bazie danych zastosowano logikę rozmytą i optymistyczną rozmytą normę agregacji. Do wizualizacji zastosowano nowy algorytm redukcji wymiarów t-SNE. Osiągnięte wyniki zostały zweryfikowane przy użyciu wiedzy ekspertów.

Wyniki i wnioski: W ocenie współpracy naukowej relację współautorstwa można uzupełnić profilami dyscyplinarnymi badaczy wyznaczonymi za pomocą rozmytej normy agregacji. Dzięki ciągłości działalności wydawniczej badaczy proponowana miara oparta na profilu dyscyplinarnym jest stabilna.

Ograniczenia badań: Próba badawcza jest ograniczona ponieważ autorzy rozważają dwa zespoły naukowców i ich publikacje z dwóch baz danych. Wybrano czasopisma z listy WoS/Scopus, ponieważ analizowani badacze tam publikują artykuły. Ponadto, podczas łączenia tych dwóch baz danych pojawiły się problemy z dopasowaniem tytułów czasopism.
Zastosowanie praktyczne: Autorzy zaproponowali model oceny podobieństwa dyscyplinarnego naukowców, a następnie oszacowania potencjału ich współpracy.
Oryginalność/Wartość poznawcza: Podejście to stosuje algorytmy rozmytej logiki do kwantyfikacji zainteresowań naukowych i jest kolejnym rzadkim przypadkiem praktycznego zastosowania algorytmów sztucznej inteligencji (logika rozmyta) w badaniach informatologicznych w szerokim kontekście.

Słowa kluczowe

Logika rozmyta. Mapowanie t-SNE. Naukometria. Profil interdyscyplinarny. Współautorstwo.

VESLAVA OSIŃSKA is an Associate Professor at the Nicolaus Copernicus University in Torun. She has a physics background and holds a PhD in information sciences. She is specializing in the mapping of information and data from both professional databases and the Internet. She is lecturing information visualization and graphic design, information architecture, and database management. She is certificated by ECDL training. Veslava Osińska is collaborating with Places @ Spaces International Project and is its ambassador in Poland. She is a local coordinator of the Polish Chapter of ISKO and the member of Polish Information Processing Society. More at her blog: www.wizualizacjainformacji.pl.

Contact to the Author:
wieo@umk.pl
Institute of Information and Communication Research
Nicolaus Copernicus University in Torun
Wladyslawa Bojarskiego 1
87-100 Torun, Poland

OLEKSANDR SOKOLOV, Professor, Dr of Science, PhD. He took his degree in Applied Mathematics at Kharkiv Aviation Institute, and earned his PhD in Control Theory at Institute of Machinery Building of Academy of Science of Ukraine, and earned his Dr of Science in Control Theory at National Aerospace University, Kharkiv. Currently he is a professor in the Nicolaus Copernicus University in Toruń in the Department of Informatics, Faculty of Physics, Astronomy and Informatics. His research interests are: Fuzzy Logic, Decision Support Systems, Intelligent Control Systems, Test Theory, Multiagent systems, Simulation of biological systems.

Contact to the Author:
osokolov@fizyka.umk.pl
Faculty of Physics, Astronomy and Informatics
Nicolaus Copernicus University in Torun
Grudziadzka 5
87-100 Torun, Poland

ALEKSANDRA MREŁA, PhD, took MSc degree in Mathematics at the Higher Pedagogical School in Bydgoszcz (Poland), and earned PhD in Mathematics at the University of Łódź, Poland. Currently she is an Assistant Professor in the Institute of Informatics, Kazimierz Wielki University in Bydgoszcz. Her research interests are: Fuzzy Logic, Decision Support Systems, Test Theory, Mathematics.

Contact to the Author:
Institute of Informatics
Kazimierz Wielki University
Kopernika 1
85-074 Bydgoszcz, Poland

Analysis of the Thematic Overlap Between Library and Information Science and Other Subdisciplines of the Social Communication and Media Sciences in Poland

Zbigniew Osiński
ORCID 0000-0003-4484-7265
Department of Digital Humanities, Faculty of Humanities,
Maria Curie-Skłodowska University in Lublin

Abstract

Purpose/Thesis: The recent decision to join three previously separate disciplines – library and information science, media studies, and cognition and social communication science, into a single discipline of social communication and media sciences prompted the author to investigate if joining of these disciplines according to the compulsory categorization published by the OECD, is supported by an overlap in their fields of research, or by a similarity in their methods of conducting it.
Approach/Methods: An analysis of the review articles devoted to the research fields of all three disciplines, and of the information regarding the research interests of the journals affiliated with them, as published on the journals' websites, allowed the author to establish their thematic scope. The results of this analysis were compared with bibliographic data and sets of keywords found in the affiliated journals. The comparison relied on an analysis of citations, and of coexistence of specialized terms.
Results and conclusions: The analysis of the review articles suggested that the basic research fields of library and information science and of the media studies and cognition and social communication science are aligned and complement each other. This conclusion was further supported by the analysis of the guidelines for the potential contributors provided on the websites of the investigated journals. However, the analysis of the bibliographic data and of the keyword sets gave an entirely different idea of the relation between the studied disciplines, indicating that there is no significant thematic overlap between them. Nevertheless, this might be due to the quality of this particular data sample, and to the methods' susceptibility to data disruption.
Originality/Value: The article proves that there is an overlap between library and information science, and the social communication and media sciences. Furthermore, it shows the limits of the citation method and of the specialized terms coexistence method, resulting from the practices of the authors and the editorial teams of some of the journals discussed. The article shows that all quantitative studies of the state of scholarship in a given discipline in Poland must be conducted with great care, and their results should not be the only basis for conclusions.

Keywords

Bibliographic coupling method. Cognition and social communication sciences. Library and Information Science. Media studies. Social communication and media sciences. Specialized terms coexistence method.

Received: 23 September 2019. Reviewed: 21 October 2019. Revised: 22 October 2019. Accepted: 18 November 2019.

1. Introduction

The regulation issued by the Ministry of Science and Higher Education in Poland on September 20th, 2018 regarding the new classification of the fields of science and scientific disciplines as well as artistic disciplines (*Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 20 września 2018 r. w sprawie dziedzin nauki i dyscyplin naukowych oraz dyscyplin artystycznych*) joined three previously separate disciplines, library and information science (humanities), media studies, and cognitive and social communication science (social sciences) into a single new discipline, social communication and media sciences (social science). The decision was justified as an improvement to the previous artificial narrowing of the research fields, which would adjust the Polish classification of the fields of science and scientific disciplines to the international standards, particularly those established by the OECD, and allow it to meet the demands of the new methodology of the evaluation of research conducted at particular institutions (KDN, 2018). It bears reminding that the OECD classification differs from the classification applied by the most important international bibliographic databases, Web of Science and Scopus, on which the Ministry's list of scoring journals and rules of evaluation of scientific disciplines are based. Both of these bibliographic databases clearly distinguish between Communication Science, and Library and Information Sciences (Information Science & Library Science in WoS). It therefore becomes pertinent to ask if the joining of library and information science with media studies and cognitive and social communication sciences might be justified by an overlap in their research fields; if the academics from these disciplines take similar approaches to their research; and if these disciplines can be sensibly identified as a new joint discipline.

2. Literature review

We find in Polish scholarship studies devoted to the thematic scope of all three disciplines under discussion. Because of the focus and goals of this study, which is concerned with the Polish academic environment, the literature review prioritized the most recent publications of Polish academics.

Tomasz Globan-Klas (2008) described the object of study and the research field of a discipline which he called media and social communication studies. Among its primary research interests he included mass media and mass communication, i.e. communication mediated by the use of public and collective media. In his estimation, the rise of the Internet brought the object of study and the research field of media and social communication studies increasingly closer to these previously specific to the study of individual and group communication (the psychology of interpersonal communication), which impacts the study of communication codes (linguistics; cognitive and social communication science; semiotics; nonverbal communication), for the study of immediate group communication (rhetoric, theater studies), and for the specialized communication sciences (computer science, telecommunication).

Arguing for a formal recognition of the discipline of communication and social media studies, Janusz W. Adamowski (2009), listed history of media and journalism, sociology

of journalism, legal basis of media activity, linguistics of media, the effect of media on the socio-economy and politics, and communication in politics, business, and culture among the discipline's research interests.

For Małgorzata Lisowska-Magdziarz (2013), the media and social communication science distinguishes itself by posing research questions which would not be prioritized by other disciplines, although they might feature as complementary questions, subordinated to their respective research interests. In the media and social communication science, these questions are not a means to another end, but the focus of research. The research field of this discipline encompasses the author of mediated content and its the recipient; the mediation and content of mass media; the relations between authors and recipients; the comprehension of mediated content; the effect of media on individuals and collectives; the influence of media on human behavior; the source/sender institutions and organizations; the technology and means of communication; the economy of media; the position and role of mass media in the culture and society. With the development of the digital media and of the Internet, the field extended to include the issues of the virtual, the visual, transmediality, transculturality, multimediality, the convergence of media, interactivity and mediatization. These issues are studied within an empirical framework, presuming the existence of an empirically knowable and intersubjectively describable reality, which it is possible to analyze following the precisely described methods of constructivism, which focuses on the processes and ways in which the authors and the recipients of mediated content construe their world, creating their meanings.

Marek Jabłonowski and Tomasz Gackowski (2012) distinguished the following fields of research in the media studies: theory of mass media; the study of the influence of media; the study of the content of media; the language of media; the history of media and journalism; the economy of media; media public relations and marketing; media management and logistics; the aesthetics of media; political communication; media audiences.

According to Marek Jabłonowski and Wojciech Jakubowski (2014), media studies are simultaneously within the research field of the arts disciplines (the linguistics, literature studies, history, culture and anthropology, and theology of media), and of the social sciences (the political, psycho-social, and economic conditions of the operation of the mass media, and their effect on the audience). Furthermore, they stated that media studies are concerned not only with the mass media (press, radio, television, Internet), but also the individual who communicates and mediates, and who is susceptible to the media's influence. The research on these subjects has three distinct aspects: cognitive, communicological, and medialogical. The authors posit that in Poland, cognitive and social communication science has an existence separate from media studies. Three strands of research are to be distinguished in the mass media: history of media, theory of media, and several sub-fields: anthropology of media, pedagogy of media, philosophy of media, media law. Furthermore, the authors presented a detailed list of research areas of media studies, based on *Nauka o Mediach* (*Media Studies*), an official publication prepared by J.W. Adamowski, M. Jabłonoswski and K.A. Wojtaszczyk for the Central Commission for Academic Degrees and Titles. These included: the circulation of information facilitated by media; the relation between the consciousness of the recipients of the mediated content, the mediated content, and the purposes of the authors of the mediated content; the use of media in the changing society and media environment; legal, economical, and political conditions of the functioning of

media; the specifics of the structures within media institutions; developing tools allowing an estimation and prediction of the influence of media on the consciousness of the recipients and on the public opinion; constructing models representing the functioning of media systems; developing tools for the analysis of the mediated content as a historical source; describing media history; the study of the offer of particular sources of mediated content; analysis of the media convergence.

The communication science was defined by Bogusława Dobek-Ostrowska (2006), among others. According to her, its research field encompasses all elements, stages, and aspects of the social communication process, immediate (interpersonal communication) as well as mediated (with the use of media). Thus defined, the communication science is characterized by its interdisciplinary methodological approach, as it borrows from the methodology of sociology, psychology, psychiatry, history and mathematics.

However, Emanuel Kulczycki (2008) claims that all issues identified as objects of study of the communication science are in fact objects of study for other academic disciplines, such as sociology, linguistics, psychology, ethics, and so on, while the communication science itself is only one of the research fields shared by these disciplines. He stated that there is a vast difference between an interdisciplinary research field and an independent academic discipline, with distinct aims and methodology. He argued that (2011):

> Communication science is not to be identified with communicology, nor with the discipline of communication (to be understood as a consideration of the communication process, which is characterized by its status as an autonomous discipline). What is called 'communication science' (and identified with communicology) in Poland, is a type of an interdisciplinary research method or rather, a type of a research field shared by multiple disciplines, studied within the framework of specific subdisciplines (e.g. psychology or communication sociology).

It is worth mentioning that all scholars discussed thus far join media studies with communication science, without consideration of cognitive science, which is more commonly considered a part of philosophy and psychology. If epistemology, cognition, and psychology are not involved in cognitive and social communication science, only social communication science remains. Furthermore, social communication is identified with an immediate interpersonal communication (face-to-face), which is an object of study in linguistics and psychology. Thus, media studies should be joined only with the study of mediated social communication (Goban-Klas, 2009). Meanwhile, J.W. Adamowski suggested that cognitive and communication sciences have more in common with culture studies than they do with media studies (2009).

Regarding the information science, Barbara-Sosińska-Kalata (2013) defined following issues as the research interests of the discipline: the analysis of the state and the development of information and knowledge resources within different branches and specializations; information architecture and usability of internet websites; digital archives, libraries, and repositories; user studies – information needs, information literacy, and information behaviors; information barriers; history and contemporary nature of the book and of the library; quantitative research of information use – bibliometrics, scientometrics, webometrics; library science; information economy; information ethics; scholarly communication; information in digital and social media; knowledge organization – folksonomy, indexing, metadata, ontology, theory of classification and knowledge organization; theory of information; methodology and terminology of information science; data mining and text

mining; big data; systems for the automatic content identification and extraction; expert systems and artificial intelligence; technology of information processing and publishing; information services; digitization and visualization of information; information retrieval and evaluation; information management; information sources. This set of research interests was met with acceptance of the Polish academic community, which became evident when it gave the basis for a textbook *Nauka o informacji* (*Information Science*) published in 2016 with Wiesław Babik as the head editor.

Writing in *Nauka o informacji*, Mirosław Górny (2016) stated that every research field of information science might be put into one of four broad categories:

– the discovery of qualities and characteristics of information objects and information processes;
– the study of the functions of information objects and information processes;
– the clarification of the mechanisms of information processes, and all other processes occurring in the information infrastructure;
– the design and modernization of information systems.

Hanna Batorowska (2015) highlighted the new challenges to the information science posed by the development of information and communication technology, which included the study of information culture, development of information literacy, information education, and information ecology.

The introduction of the Ministry's new classification of academic disciplines opened a discussion of a potential relation between the previously separate disciplines, now joined into social communication and media sciences. Marek Jabłonowski and Tomasz Mielczarek (2018) opined that the joining of media studies, cognitive and social communication science, and library and information science did not open new research fields, nor did it suggest new research questions. For example, the book, which is the main object of study for library science, is of interest to media studies as well, because it might be considered as a medium for communication of content, as it is to the cognitive and social communication science, which views it as a source of knowledge and as a means to communicate content. Information, the central focus of the information science, has also been studied by the scholars affiliated with the other two disciplines, as an effect of the operation of media, and as a basis for knowledge and communication. Furthermore, the approaches to, and methods of conducting research within all three disciplines show similarities. Jabłonowski and Mielczarek concluded that the joining of these disciplines will not disrupt their paradigms and methodologies, but instead it will allow for a creation of interdisciplinary research teams who will view given objects of study from multiple perspectives. In the future, the objects of study, research methods and approaches, and paradigms, may converge.

3. Aim, methods

As discussed in the introduction, the aim of this study is to establish whether the decision to join the disciplines of information and library science, media studies, and cognitive science, into one discipline in the new Polish system of classification, based on the OECD system, might be justified by their sharing of study objects, and by overlaps between the

areas of, and approaches to, previously conducted research. To answer these questions we may follow the example of B. Sosińska-Kalata (2013) and conduct a thematic analysis of the articles published in the journals considered to be the most important sources of new scholarship within the studied disciplines. We may also use a shortcut and analyze the content of review articles concerned with the research interests of these disciplines, and the thematic scope of the abovementioned select journals, available on their websites. The author chose the second approach, and studied the previously discussed review articles. Regarding the journals, they had to meet two criteria to be chosen for further study: they had to be included on the Polish Ministry's list of scoring journals in 2018 and they had to made their content fully available online, at least the content published between 2010–2018. Thus, the following journals were selected:

(1) concerned with library and information science: Annales Universitatis Paedagogicae Caracoviensis. Studia ad Bibliothecarum Scientiam Pertinentia (The Pedagogic Yearbook[s] of the Cracow University. The Current Studies in the Library Science); Praktyka i Teoria Informacji Naukowej i Technicznej (Practice and Theory of Scientific and Technical Information), Przegląd Biblioteczny (Library Review), Studia o Książce i Informacji (Book and Information Studies), Toruńskie Studia Bibliologiczne (Toruń Bibliological Studies), Zagadnienia Informacji Naukowej – Studia Informacyjne (Issues in Information Science – Information Studies).
(2) concerned with media studies and cognitive and social communication sciences: Central European Journal of Communication, Media – Kultura – Komunikacja Społeczna (Media – Culture – Social Communication), Naukowy Przegląd Dziennikarski (Journalism Research Review Quarterly), Rocznik Bibliologiczno-Prasoznawczy (Yearbook of Bibliology and Press Studies), Rocznik Prasoznawczy (Yearbook of Press Studies), Studia Medioznawcze (Media Studies), Zeszyty Prasoznawcze (Media Research Issues).

The author chose not to analyze the articles published by Polish scholars in foreign journals, as they are so rare they would not contribute to the results of the study.

The collected data was compared with the results of a quantitative research to see if the results of the qualitative study (the analysis of the review articles and the websites) would be supported by the quantitative methods, which are increasingly popular in humanities and social sciences. The comparison was further motivated by E. Kulczycki's (2008) observation that overlaps between research fields of various disciplines did not indicate that they are related. For example, 'the child,' or 'the family,' are studied within both sociology and medicine, which are disciplines separated by a great distance. Many disciplines shares research objects, but in most cases, it does not inspire collaborative interdisciplinary research, because the differences in methodology and research aims are too vast. Therefore, the author relied on the quantitative data, the analysis of citations (although initially, the bibliographic coupling method was considered), and the specialized terms coexistence method, to verify the initial conclusions based on the qualitative data.

The method of generating bibliographic couplings is based on the premise that the works which occupy identical bibliographic positions will have similar content. The more bibliographic positions they share, the more similar their content should be. The method involves searching for publications which share at least one such bibliographic position and creating thematic clusters, which would contain the works presumed to

contain similar content (Kessler, 1963). However, because the instances in which the groups of articles containing at least one article from the discipline of library and information science and one from media studies and cognitive and social communication sciences, sharing at least one bibliographic position, were exceedingly rare, the author did not use this method.

Instead, the author prioritized searching for the articles whose bibliographies included at least one work published in one of the journals under the consideration (see the list above), i.e. the method of citation analysis. It was assumed that if citing the works from one journal in the articles published in another journal would be regular, rather than incidental, it would suggest a thematic overlap between these journals, and, implicitly, between the academic disciplines with which they are affiliated.

The specialized terms coexistence method involves browsing for terms featuring in the titles, keywords, abstracts and full texts of the compared publications, or groups thereof (a journal would be such a group). The use of the same specialized terms without variations in their referents, indicates a thematic overlap. The higher the number of shared specialized terms, the stronger the overlap. The author analyzed the keywords of all the articles whose bibliography contained at least one work published in one of the studied journals. He thus collected two sets of data, one for the journals associated with the library and information science, and the other with the media studies and cognition and social communication science. The frequency with which the specific specialized terms appeared, with a division between single words and longer phrases, was calculated using the internet application Calculla (https://calculla.pl/licznik_slow).

4. Results

The analysis of the articles discussed in the literature review section allowed the author to put forward a thesis that the basic research fields of the library and information science, and of media studies and the cognitive and social communication sciences, are aligned and complementary (Tab. 1).

The discussed disciplines are concerned with sources and resources of information, which include media, and the users thereof, as well as with their influence on the individual, the society, and the communication processes. Although historically, they have focused on different objects of study – books, recorded information, and libraries in the case of the library and information science, and press, radio, and television in the case of the other two disciplines, all three analyze the Internet. The information user and the participant in the communication processes, i.e., the individual and the society, constitute another object of study all three disciplines share.

This observation is confirmed by the analysis of the guidelines for the potential authors, as published on the websites of the studied journals (Tab. 2).

The data collected in Table 2 shows that, in a very general sense, the journals representative of the disciplines under discussion, are concerned with similar research fields and objects of study. They differ where they prioritize different specific aspects of these fields and objects. Journals and online social media might be studied either as sources of information, or as species of mass media and mass communication tools. Linguistic

aspects might be discussed either in relation to information architecture, or to the construction of messages in traditional or digital media. All three disciplines consider how the individual and the society might be affected – either by information processes, or by media and mass communication processes. Among the interests of the researchers is the issue of digitalization, either in the libraries, or in the digital media. The architecture of information is studied as a part of a wider discussion regarding data repositories and websites, as well as an aspect of a media. Information and digital literacies are aspects of both the information environment and the media environment. The academics research mass communication in the domains of politics and economy, as well as in the domain of science. The development of culture is analyzed with the use of case studies from media, but also from libraries. Significant differences between the three disciplines emerge only in the details of their research questions.

The citation analysis was conducted on a data sample of 1432 articles published in the journals discussed above, between years 2010–2018, available online (Tab. 3).

Tab. 1. The basic research fields of the analyzed disciplines

Information & library science	**Media studies and cognition & social communication science**
The history of the library and of the book The functioning of libraries, publishing houses, repositories, and other information resources Information theory Information ethics Information technology; media Legal aspects of information Linguistic aspects of information Information literacy, information competences Information users Information processes and their quality The organization and architecture of information The relations between authors and recipients Digitalization, multimedia, and the Internet Scientific communication Information society Quantitative research – bibliometrics, webometrics Information economy Information culture	The history of media and journalism The functioning of media Theory of media and communication Media ethics Media technology Media law Linguistic aspects of media's functioning (Linguistics of media) Media pedagogy Media audiences Communication and cognitive processes; their quality The organization and architecture of media content The relations between media creators and audiences Digitalization, multimedia, and the Internet Political and marketing communication The social influence of media Information society Quantitative research – audience measurement, reaching target audiences, changing preferences Media economy Culture in media; media in culture

Tab. 2. The thematic scope of the articles published in the studied journals

Discipline	Journal	Thematic scope
Library & information science	Annales Universitatis Paedagogicae Cracoviensis. Studia ad Bibliothecarum Scientiam Pertinentia	scientific information; book and library history; journals, literary culture
	Praktyka i Teoria Informacji Naukowej i Technicznej	the field of scientific information (information science) and related ideas; current theoretical and practical issues which are of interest to the creators and users of information (considered in interdisciplinary perspective)
	Przegląd Biblioteczny	contemporary research trends in library science, book studies, bibliography and scientific information
	Studia o Książce i Informacji	theory and practice of bibliology; the aesthetics of the book; modern media forms, contemporary publishing market in Poland and in the World
	Toruńskie Studia Bibliologiczne	issues in the field of social communication and media, including bibliology and information science and related sub-disciplines, such as media studies, archival studies and document management, journalism, media art
	Zagadnienia Informacji Naukowej	information science in relation to library science, archival science, museology and other disciplines researching preservation and access to scientific and cultural heritage; information and knowledge management; traditional and digital scholarly communication; information and knowledge organization; metadata theory and practice; Web 2.0; Semantic Web; information architecture; information websites usability; human-computer interaction; natural language processing; information retrieval; information use and information users behavior; social response to modern information technologies; digital humanities; information and digital literacy; information policy; information ethics

Media studies, Cognition and social communication sciences	Central European Journal of Communication	an international forum for empirical, critical and interpretative, quantitative and qualitative research examining the role of communication in Central Europe and today's world; media and communication studies
	Media – Kultura – Komunikacja Społeczna	press, radio, television, new media; social communication in its theoretical and practical aspects; media politics, media law, media economics; practical and ethical aspects of a journalist's work; the media as an integral element of the culture system and culture in the media
	Naukowy Przegląd Dziennikarski	journalism (press, television, radio, Internet), its history, theory and the newest trends
	Rocznik Bibliologiczno-Prasoznawczy	no information provided
	Rocznik Prasoznawczy	discussions of the newest theories, concepts and problems relating to the media studies; articles concerned with the widely understood questions of media studies, and particularly with the history of the press, theory of mass communication, media systems and the culture of language
	Studia Medioznawcze	the identity of the media studies; new media; ethics and public relations; law in media; history of media; media in Poland – press, radio, television; media environments abroad
	Zeszyty Prasoznawcze	social, psychological, political, linguistic, legal, economic, technological, organizational and professional aspects of mass communication, and in particular of the printed press, radio, television and other media, journalism, advertising, propaganda and public opinion, in the past and present, in Poland and abroad

Tab. 3. Quantitative results of the citation analysis of the articles published in the journals studied, as cited in other articles published in these journals

No.	The journal name (abbreviated)	Number of the articles studied	Number of the articles citing others	Number of citations	Included: self-citations	Articles with a citation number of 1–2	Articles with a citation number >3
1	ZIN	123	48	91	26	36	12
2	PTINT	152	44	75	18	38	6
3	PB	139	51	99	19	41	10
4	TSB	150	24	41	14	19	5
5	SKI	25	4	4	0	4	0
6	AUPC	70	20	30	6	17	3
7	MKKS	128	19	42	4	11	8
8	ZP	227	55	115	18	41	14
9	SM	55	19	39	4	13	6
10	RBP	58	21	46	8	16	5
11	RP	43	5	7	1	5	0
12	NPD	139	23	44	12	19	4
13	CEJC	123	5	12	1	4	1
	Total	1432	338	645	131	262	76

Full names of the journals featured in the table: ZIN – Zagadnienia Informacji Naukowej, PTINT – Praktyka i Teoria Informacji Naukowej i Technicznej, PB – Przegląd Biblioteczny, TSB – Toruńskie Studia Bibliologiczne, SKI – Studia o Książce i Informacji, AUPC – Annales Universitatis Paedagogicae Cracoviensis. Studia ad Bibliothecarum Scientiam Pertinentia, MKKS – Media – Kultura – Komunikacja Społeczna, ZP – Zeszyty Prasoznawcze, SM – Studia Medioznawcze, RBP – Rocznik Bibliologiczno-Prasoznawczy, RP – Rocznik Prasoznawczy, NPD – Naukowy Przegląd Dziennikarski, CEJC – Central European Journal of Communication.

The citation analysis firstly made manifest the limitations of this method, whose cause are the authors themselves and the editing teams of some of the journals studied. It is clear that the authors are reluctant to cite articles published in Polish journals (as well as articles by Polish authors published in foreign journals). Out of the 1432 articles studied, only 338 cites texts published in the Polish journals featured in the present research. These articles usually contain only one or two references to other papers published in the studied journals. The simplest explanation of this phenomenon, i.e., that the majority of the authors publishing in Polish journals concerned with the library and information science, media studies, and cognition and social communication sciences, is concerned with questions so original that no other publications on the subject might be found, is highly unlikely. A study of these articles' bibliographies clarifies the situation: it seems that the authors tend to refer to monographs and chapters of collaborative works (these citations account for 70-90% of the bibliographies). Unfortunately, these types of publications are much more rarely available online, and, because of the faults in the functioning of the legal deposit system, they are available only in a few libraries. Thus, it is much more difficult to access

them, which makes a citation analysis of bibliographies of texts contained in such publications extremely time consuming and inefficient. This constitutes a serious limitation to the method of citation analysis. Further problems are generated by absence of the list of references in the end of articles in several journals, or particular issues of these journals, and by the confusion of primary and secondary sources in bibliographies that are attached to the articles. Without a careful reading of the article in its entirety, it is difficult to tell whether a particular position in the list of references (even if it is an article published in a peer-reviewed journal) was used as a secondary or a primary source. In a number of articles secondary sources constituted only a small part of their references. It is possible to describe certain articles studied as completely lacking in reference to secondary sources, as their lists of references contained only a list of the primary sources used.

Reading the texts concerned with the research questions familiar to the author of the present study, suggested another tendency which might disrupt the results of a citation analysis. It is only signaled here, as the author did not conduct satisfactory quantitative research. However, it was apparent that many of the articles studied did not refer to the most important works of a similar, or complementary, thematic scope, published in Polish journals. It is therefore difficult to rely on the results achieved by the method of citation analysis.

Accordingly, the results presented in Table 4 should be approached with an appropriate degree of care. As a result of the limitations discussed above, they might not give an accurate view of real bibliographic couplings between the journals representative of the disciplines studied, and thus, between the disciplines themselves.

Tab. 4. A number of citations of articles published in the studied journals (columns) in the articles in these journals (rows). The journals' full names are abbreviated as they were in Table 3

		ZIN	**PTINT**	**PB**	**TSB**	**AUPC**	**SKI**	**ZP**	**SM**	**RBP**	**NPD**	**MKKS**	**CEJC**	**RP**	**razem**
1	ZIN	57	10	17	4			2	1						91
2	PTINT	30	16	25	2	1		1							75
3	PB	19	4	72	2			2							99
4	TSB	3	1	23	9			4	1						41
5	AUPC	5	1	8	3	7		4		2					30
6	SKI	1		2				1							4
7	ZP			2				81	25	3	1			3	115
8	SM							8	25	2	2	1	1		39
9	RBP			1	1			29	8	3			3	1	45
10	NPD			1				17	14	6	2	1	1	2	44
11	MKKS							10	22	2	1	4		3	42
12	CEJC							4	4				4		12
13	RP							4	2					1	7
	Total	115	32	151	21	8	0	167	102	18	6	6	9	10	

The data presented in Table 4 might suggest that the library and information science has no distinct thematic relation to the media studies, or the cognition and social communication sciences. If they cite other articles at all, the authors of the texts published in the journals affiliated with the discipline of the library and information science cite only the articles published in the journals affiliated with their own discipline. The same is true for the works published in the journals affiliated with media studies, and for those affiliated with the cognition and social communication sciences. Furthermore, the authors of the texts published in *ZIN*, *PB*, *ZP*, and *SM* tend to cite the articles published in the journals in which they publish. The references to articles published in journals associated with other disciplines are only incidental. However, it would be unadvisable to take these results on their face value, as they have been collected with the use of the method which has serious limitations, as discussed above.

The analysis of the frequency with which keywords feature in the articles relied on a set of words and phrases derived from the keywords of these articles which cite other articles published in one of the studied journals at least once. A set of more than eight hundred keywords and key phrases was established for both library and information science, and the media studies with the cognition and social communication sciences. A majority appeared only rarely, in less than ten works. The twenty keywords featured most frequent (between 10 and 100 times) in both groups of journals are featured in Table 5; the eleven most frequent phrases (more than 10 times) – in Table 6.

Tab. 5. The most frequent keywords, from the most to the least frequent (nouns in the singular)

The library and information science	**Media studies, and the cognition and social communication sciences**
Informacja (Information)	Prasa (Press)
Biblioteka (Library)	Media (Media)
Zarządzanie (Management)	Komunikacja (Communication)
Bibliografia (Bibliography)	Radio (Radio)
Wiedza (Knowledge)	Społecznościowe (Social)
Nauka (Science)	Publiczne (Public)
System (System)	Dziennikarstwo (Journalism)
Badania (Research)	Kultura (Culture)
Książka (Book)	Rynek (Market)
Prasa (Press)	Analiza (Analysis)
Bibliometria (Bibliometrics)	Badania (Research)
Cyfrowy (Digital)	Słowa (Words)
Czasopisma (Journals)	Wolność (Freedom)
Komunikacja (Communication)	Czasopisma (Journals)
Analiza (Analysis)	Dane (Data)
Dane (Data)	Dyskurs (Discourse)
Dzieci (Children)	Internet (Internet)
Hasła (Headings)	Lokalnie (Locally)
Informatologia (Information science)	Okładka (Cover)
Język (Language)	Treść (Content)

Table 6. The most frequent key phrases, from the most to the least frequent

The library and information science	**Media studies & cogntivie and social communications sciences**
Nauka o informacji (Information science) Zarządzanie informacją (Information management) Biblioteki akademickie (Academic libraries) Kompetencje informacyjne (Information literacy) Komunikacja naukowa (Scientific communication) Język haseł przedmiotowych (Subject headings) Rozwój dyscyplin naukowych (The development of scientific disciplines) Wyszukiwanie informacji (Information retrieval) Języki informacyjno-wyszukiwawcze (Information retrieval languages) Metody ilościowe w informatologii (Quantitative methods in the information science) Czasopisma naukowe (Scientific journals)	Media społecznościowe (Social media) Nowe media (New media) Wolność słowa (Freedom of speech) Analiza zawartości (Content analysis) Fotografia prasowa (Press photography) Prasa polska (Polish press) Prasa lokalna (Local press) Rynek prasy (Press market) Komunikacja marketingowa (Marketing communication) Komunikacja elektroniczna (Electronic communication) Przetwarzanie danych (Data processing)

The data from both tables indicates that there is only a small overlap between the research fields of these disciplines, noted on the level of single words with a vast scope of reference: press, journal, communication, analysis, data. When we include more precisely identified research fields, described by two or three words long phrases, then we may conclude that each of the two groups of the disciplines has different research interests. However, these sets of key words and phrases might have been affected similarly to the date collected with the method of data citation. The analysis might have been disrupted by the absence of standards for keyword creation, resulting in rather loose descriptions of the articles. Occasionally, the authors will refer to one phenomenon, i.e. excess of information, information overload (information and library sciences), as "the abundance of news" (media studies, cognitive and social communication sciences). However, it should be noted that the indicated limited thematic overlap does correspond to reality. The proximity of data collected with the citation analysis method to those collected with the analysis of specialized terms coexistence confirms as much. Nevertheless, the author remains aware that this convergence may have been achieved with the use of imperfect data samples.

5. Conclusion

To answer the question posed in the Introduction, the author has to admit that the joining of the library and information science together with media studies and the cognitive and social communication sciences, as ruled by the Polish Ministry's regulation, is justified by the similarity between their thematic scopes declared in descriptions of these disciplines.

The analysis of review articles, which consider the thematic scope of the disciplines suggests that, despite frequent application of different concepts and terms, and prioritizing different detail issues in their research, more generally understood, their research fields and objects of studies overlap. Within all three disciplines, Polish researchers were concerned with information objects (sources, resources, tools), among which we should include media and communication tools, as well as communication processes, which encompass mass communication and the influence of media on the society. The researchers studied infrastructure – of information, media, and communication. Every discipline accounted for the individual – a user, a recipient and the author of information, news and media messages. Furthermore, shared research fields and approaches have already emerged in their practice. The study of phenomena and processes which occur online is gaining popularity in all three disciplines. The researchers are increasingly drawn to interdisciplinary topics, inductive approaches, and methods based on quantitative data analysis, frequently collected by at least partially-automatized methods. Therefore, it can be stated that all three disciplines can co-create a new scientific discipline called the social communication and media sciences. However, the practice observed in all three disciplines that researchers mainly publish in journals related to their own discipline and rely solely on the papers published therein, is without a doubt detrimental to the integration of these disciplines.

The research presented in the current study achieved a yet another aim, already suggested in the description of the research process: it has indicated that all investigations of the state of scholarship in any discipline in Poland which rely on quantitative methods, must be conducted with great care, and their results should not be the main basis for drawing conclusions. It has been shown that there are numerous factors disrupting a collection of representative sample of bibliographic data and keywords. The studies based on such data sets may turn out to be misleading and falsifying the picture of the examined reality. The methods applies in the current study are highly dependent on the quality of the used data.

References

Adamowski, J.W. (2009). O pilnej potrzebie formalnego powołania w Polsce nauki o komunikowaniu i mediach społecznych jako dyscypliny naukowej. *Studia Medioznawcze*, 4(39), 20–24.

Batorowska, H. (2015). Nauka o informacji (informatologia) z perspektywy nowych wyzwań edukacyjnych. *Edukacja-Technika-Informatyka*, 3, 73–80.

Dobek-Ostrowska, B. (2006). *Komunikowanie polityczne i publiczne*. Warszawa: Wydaw. Naukowe PWN.

Goban-Klas, T. (2008). Nauki o mediach i komunikacji społecznej jako nowa dyscyplina nauk humanistycznych. *Studia Medioznawcze*, 2(33), 11–19.

Goban-Klas, T. (2009). Emergencja nowej dyscypliny: nauki o mediach i komunikacji społecznej. *Studia Medioznawcze*, 4(39), 14–19.

Górny, M. (2016). Nauka o informacji jako dyscyplina naukowa. W: W. Babik (red.) *Nauka o informacji*. Warszawa: Wydaw. SBP.

Jabłonowski, M., Gackowski, T. (2012). Tożsamość nauk o mediach. Obszary, perspektywy, postulaty. *Studia Medioznawcze*, 2(49), 15–24.

Jabłonowski, M., Jakubowski, W. (2014). Status teoretyczny nauk o mediach – kilka uwag do dyskusji. *Studia Medioznawcze*, 4(59), 11–24.

Jabłonowski, M., Mielczarek, T. (2018). Komunikowanie społeczne i media – federacja, a nie inkorporacja. *Studia Medioznawcze*, 4(75), 13–27.

KDN (2018). *Nowe rozporządzenie ws. dyscyplin – to rzetelna ocena badań naukowych* [online]. Konstytucja dla nauki [18.09.2018], https://konstytucjadlanauki.gov.pl/nowe-rozporzadzenie-ws-dyscyplin-to-rzetelna-ocena-badan-naukowych

Kessler, M.M. (1963). Bibliographic Coupling Between Scientific Papers. *American Documentation*, 14(1), 10–25.

Kulczycki, E. (2011). Kulturowo-obiektywne istnienie procesu komunikacji jako warunek projektowania autonomicznej dyscypliny komunikacji. *Kultura i Historia* [online], 21, [18.09.2018], http://www.kulturaihistoria.umcs.lublin.pl/archives/2997.

Kulczycki, E. (2008). Status komunikologii – przyczynek do dyskusji. *Homo Communicativus*, 3, 27–37.

Lisowska-Magdziarz, M. (2013). Metodologia badań nad mediami – nurty, kierunki, koncepcje, nowe wyzwania. *Studia Medioznawcze*, 2(53), 27–42.

Sosińska-Kalata, B. (2013). Obszary badań współczesnej informatologii (nauki o informacji). *Zagadnienia Informacji Naukowej – Studia Informacyjne*, 2(102), 9–41.

Analiza związków tematycznych bibliologii i informatologii z pozostałymi subdyscyplinami nauk o komunikacji społecznej i mediach w Polsce

Abstrakt

Cel/Teza: Połączenie w ramach jednej dyscypliny naukowej – nauk o komunikacji społecznej i mediach – trzech dotychczas odrębnych dyscyplin: bibliologia i informatologia (nauki humanistyczne), nauka o mediach oraz nauki o poznaniu i komunikacji społecznej (nauki społeczne) skłoniło autora do szukania odpowiedzi na pytania: czy połączenie tych dyscyplin, bazujące na systematyce OECD, jest uzasadnione podobieństwem problematyki badawczej? Czy w dotychczasowej praktyce naukowej pojawiły się wspólne obszary i podejścia badawcze?

Koncepcja/Metody badań: Zakres tematyczny poszczególnych nauk ustalono poprzez analizę treści artykułów przeglądowych poświęconych problematyce badawczej poszczególnych dyscyplin naukowych oraz informacji o zakresie tematycznym czasopism związanych ze wspomnianymi dyscyplinami, podawanych na ich stronach WWW. Ustalone w ten sposób informacje skonfrontowano z danymi bibliograficznymi i zestawami słów kluczowych, pochodzącymi z czasopism związanych z badanymi naukami, za pomocą metody analizy cytowań oraz metody współwystępowania specjalistycznej terminologii naukowej.

Wyniki i wnioski: Analiza artykułów przeglądowych pozwoliła na postawienie tezy mówiącej, że podstawowe obszary badawcze bibliologii i informatologii z jednej strony, a nauki o mediach oraz nauk o poznaniu i komunikacji społecznej z drugiej, są ze sobą zbieżne i nawzajem się uzupełniają. Powyższą konstatację potwierdza też analiza informacji dla potencjalnych autorów artykułów, zamieszczona na stronach WWW badanych czasopism. Jednak analiza danych bibliograficznych i zestawów słów kluczowych nie potwierdziła tego wniosku, dała zupełnie odmienny obraz – braku istotnych związków tematycznych pomiędzy badanymi dyscyplinami. Jednakże taki wniosek może być skutkiem ograniczeń zbiorów danych zastosowanych w obu metodach i podatności tychże metod na zaburzenia danych.

Oryginalność/Wartość poznawcza: Główna wartość poznawcza pracy sprowadza się do identyfikacji zakresu związków bibliologii i informatologii z naukami o komunikacji społecznej i mediach. Ponadto wykazano istnienie ograniczeń do stosowania metody analizy cytowań oraz metody współwystępowania specjalistycznej terminologii naukowej, stwarzanych przez samych autorów i redakcje części

czasopism. Udowodniono, że wszelkie analizy stanu dyscyplin naukowych, w których stosuje się metody ilościowe, w polskich realiach muszą być prowadzone z dużą ostrożnością, a ich wyniki nie powinny stanowić głównej podstawy do wyciągania wniosków.

Słowa kluczowe

Analiza cytowań. Analiza współwystępowania terminów specjalistycznych. Bibliologia i informatologia. Bibliotekoznawstwo i nauka o informacji. Medioznawstwo. Metoda powiązań bibliograficznych. Nauki o komunikacji społecznej i mediach. Nauki o poznaniu i komunikacji społecznej.

Dr habil. ZBIGNIEW OSIŃSKI is Professor at the Department of Digital Humanities at the Maria Curie-Skłodowska University Lublin. He specializes in information science and digital humanities. His most important publications include: M. Górny, M. Kisilowska, E. Głowacka, Z. Osiński: Mechanisms of the formation and evolution of personal information spaces in the humanities *(Poznań, 2017); M. Kowalska, V. Osińska, Z. Osiński: The Role of Visualization in the Shaping and Exploration of the Individual Information Space (*Knowledge Organization*, 2018); Z. Osiński: Information infrastructure of contemporary humanities and the digital humanities development as a cause of creating new information barriers. A Polish case (*Digital Scholarship in the Humanities*, 2019).*

Contact to the Author:
zbigniew.osinski@gmail.com
Department of Digital Humanities
Faculty of Humanities
Uniwersytet Marii Curie-Skłodowskiej
pl. Marii Curie-Skłodowskiej 4
20-031 Lublin, Poland

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

Zbigniew Osiński
ORCID 0000-0003-4484-7265
Department of Digital Humanities, Faculty of Humanities,
Maria Curie-Skłodowska University in Lublin

Abstract

Purpose/Thesis: Many countries increasingly use bibliographic databases while devising new scientific policies to analyze and diagnose the state of a scientific discipline. Previous studies on the suitability of data from Web of Science and Scopus databases for this purpose gave ambiguous results. Their authors did not always account for an important issue – the quality of data from these databases. The aim of the article is to analyze the quality of data downloaded in an automated manner from the resources of the mentioned databases.

Approach/Methods: The author used a qualitative method of data verification which consisted of automatic acquisition of data about journals from the Web of Science and Scopus databases, and then in their qualitative analysis. The analysis consisted of a comparison of data on journals representing of library and information science (LIS) retrieved from both databases and of the comparison between the qualitative data taken from the studied databases and the data from other, domain focused bibliographic databases; of comparing the acquired data with the information available on the websites of indexed journals and of the comparison of the method used by the producers of the abovementioned databases used to classify the journals as related of LIS, with the thematic scope of the discipline, as agreed upon by scholars.

Results and conclusions: It was found that in the case of the examined discipline, automated data acquiring poses a risk of obtaining a low credibility set of data. Most problems are caused by the incompleteness of data and errors in disciplinary classifying journals, articles and authors.

Originality/Value: It was shown that, contrary to the claims of the decision-makers of Polish science, in its present form, the studied bibliographic databases have only negligible usefulness for monitoring the state and development tendencies of LIS. Methodological problems created by both databases, presented in this article, may also have an impact on generating a reliable and objective picture of other scientific disciplines. The changes in the sphere of the functioning of WoS and Scopus, apparent for several years, have not dealt with the already existing problems and inconveniences.

Keywords

Bibliographic data. Qualitative analysis. Quantitative analysis. Scientific discipline. Scopus. Web of Science.

Received: 28 June 2019. Reviewed: 10 October 2019. Revised: 23 October 2019. Accepted: 19 December 2019.

1. Introduction

Systematic analysis and diagnosis of the state of science and of the directions of its development is one of the tasks of library and information science (LIS), a discipline has been joined with social communication and media sciences since 2019 by the Polish ministry's regulation. It is also a concern for the individual researchers. The analysis of the state of their discipline allows them to choose to publish in those journal, which will ensure that their work will be received by the widest possible audience with a potential interest in the themes of their work. It is also an important part of science management. An awareness of the topics in international research, of published studies and journals, as well as of active researchers and their research is a necessary research competence. A relatively high number of scientific institutions and researchers, as well as a massive amount of publications, together with the scattering of information, make it difficult for an individual to grasp the development of a given discipline in its entirety. Therefore, it is necessary that the specialists in LIS, in possession of suitable methods and research tools, systematically monitor the development of given scientific disciplines. They should also perfect methods for analysis and diagnosis of the scientific disciplines, so that they may be of use to the researchers and those in charge of science management. The gravity of the situation increases with the development of the science policy introducing the element of evaluation of the quality of research, which relies on the resources and tools historically designed to serve the researchers for the monitoring of a given scientific discipline, especially on bibliometric data and indicators. May such an evaluation be objective and thorough, considering the abundance of available data bases, as well as the functionality of tools, and their dynamic development?

The scientific literature concerned with this issue does not offer definitive answers to these questions. The dilemma has been explored in the work of Bjorn Hammarfelt and Alexander D. Rushforth (2017), where they focused on the use of bibliometrics to evaluate candidates for academic positions. They argued that the bibliometric indicators based on scientific publications and citations to these publications registered in Web of Science (WoS) and Scopus should be used only very carefully, as a supplement to an expert's assessment. They showed that easily available indicators, such as IF and h-index, do not consitute a sufficient basis for an assessment, because, depending on the discipline, these indicators may not correspond to the quality of the scientific output. Hammarfelt and Rushforth suggested that better results may be achieved with an aid of more sophisticated bibliometric indicators. A.A.M Prins, Rodrigo Costas, Thed N. van Leeuwen and Paul F. Wouters (2016) demonstrated that in case of the arts and humanities, the WoS database does not contain enough journals to make its bibliometric data useful for any assessment. Jorge Manana-Rodrigues (2015) questioned the choice to use the SCImago Journal & Country Rank indicator based on the data from the Scopus database, as there are serious gaps in the collections of the journals registered there. Elizabeth S. Vieira and Jose A.N.F. Gomes (2016) concluded that the assessment of candidates for academic positions based on bibliometric indicators brings the same results as the judgement based on a peer review of the candidates' scientific output in 75% of the cases considered. However, the authors did not verify the quality of the data from the abovementioned databases by comparing it with other sources of scientific information.

A study of the works on the topic of bibliometrics inspires several questions: What is the quality and reliability of the data from the largest bibliographic databases? Which of

these databases provides more reliable information where Library and Information Science is concerned? What methodological issues a user of these databases will face if they want to conduct an analysis and diagnosis of the development of a given scientific discipline? The author of this article decided to answer these questions, and to test and compare the usefulness of data from the two most popular bibliographic databases, Web of Science and Scopus, to analyze and diagnose the state of the Library and Information Science, within which he conducts his own research. Furthermore, the author set out to verify the quality of the data from the two databases by a comparison with other sources of scientific information.

Although scholars have already discussed faults of both databases such as favoring of the English language, domination of big publishers – especially those based in the Anglo-Saxon countries, errors in bibliographic descriptions, underrepresentation of journals based outside Northern America and Western Europe, insufficient usefulness for research of humanities and most social sciences, and so on, the author wished to see if this criticism provoked the producers of these databases to introduce any corrections to their operations. He was inspired by the recent news regarding the important changes to the process of creation and supplementation of the resources in these two databases, the number of the journals considered, and the functionalities made available to the users (e.g., in 2015 the list of sources indexed in Web of Science was extended as to include a group of regional journals; furthermore, Web of Science Core Collection was created, together with an index of Emerging Sources Citation; in 2016, the Clarivate Analytics company became independent from the Thomson Reuters conglomerate as an owner and the operator of WoS; in the same year, Scopus created a new evaluation metric, CiteScore). It seems that we should ask if these recent changes made the databases more efficient as tools for the analysis and diagnosis of science, and if the criticisms of these databases in scientific publications have been taken into account.

These databases have been chosen as the focus of this article because the research literature shows an increasing interest in the possibilities they offer for the analysis and diagnosis of the state of the science. Undoubtedly, it relates to the growth of their resources and the increasing popularity of these databases among the researchers, as well as their role in formulating of regulations in many countries and generating evaluations of research institutions and individual researchers. The act reforming Polish system of higher education introduced a rule that the articles published in the journals indexed in the major international bibliographic databases (Web of Science and Scopus) will be taken into consideration in the evaluation of scientific output. Therefore, the quality of data, as well as the resources and functionality of these databases became crucial for the research evaluation, and thus for the research itself.

2. The literature review

2.1. The Web of Science and Scopus databases

The review of literature shows that the analysis and comparison of the contents made available by Web of Science and Scopus enjoy a significant success. Ten years ago, Eric Archambault and others (2009) established that there is a high correlation between the

results of a comparison of a scientific output of an institution or a country based on the data from both databases. They concluded that the two databases' indicators of scientific productions and publication on the level of a country and of an institution show similar values. Around the same time, Elizabeth Vieira and Jose Gomes published a study (2009) comparing the data related to the scientific output of two typical Portuguese universities retrieved from Web of Science and Scopus. They found that 2/3 of the scientific publications searched is available in both databases. However, 1/3 of the scientific publications is available only in one of the two databases, even the texts that had a significant impact on the development of science. Mehmet A. Abdulhayoglu and Bart Thijs (2018) observed a similar trend, as they found that 74% of the articles indexed by WoS is indexed by Scopus; 92%, when only cited publications are considered.

However, recent studies increasingly focus on the faults of the two databases. Philippe Mongeon and Adele Paul-Hus (2016) established that both WoS and Scopus take into account only a small part of the existing scientific journals. In 2015, they compared these databases' resources with those of Urlich's (the most complete index of journals and serial publications in the world), and found that WoS included only 20% of all journals, and Scopus c. 30%. It was not a surprise that the bibliographic citation databases register only a part of all journals in the world, focusing on those whose quality, as indicated by bibliometric indicators, is highest. The problem lies in the suitability of the selection of the registered journals for representing a given discipline, a region or a country. Mongeon and Paul-Hus demonstrated that journals affiliated with humanities and social sciences were by far the worst represented, as these databases included only less than 20% of the journals indexed by Urlich's. The authors estimated that the situation had remained unchanged for a decade, and contributed to the databases' negligible usefulness for a bibliometric analysis of humanities and social sciences. Fiorenzo Franceschini, Domenico Maisano and Luca Mastrogiacomo (2016) established that in both databases, the bibliographic descriptions of the articles in the category of engineering-manufacturing contain as many as 10 thousand errors, which had a negative impact on the attempts to browse the articles and the data on their citation number. The authors observed that such errors and their effects might seriously harm the image of an output of an individual researcher, an institution, a discipline, or a country.

Diego Chayarro, Ismael Rafols and Puay Tang (2018) showed that the selection of the journals registered in the WoS database is dictated not only by universal criteria, applicable to every journal, such as the editing standards and the rules of scientific assessment. Considering the case of journals published in Spanish and Portuguese, they established that the selection is also influenced by particular criteria, such as the country and language of publication, as well as the discipline with which the journal is affiliated. The journals published in languages other than English, and affiliated with humanities and social sciences, were much less represented. The authors estimated that this phenomenon had a negative impact on the capacity of the data from WoS to represent many countries, languages, and disciplines.

Anne-Wil Harzing and Satu Alakangas (2016) compared the data from 2013–2015 regarding 146 researchers affiliated with five different scientific disciplines, as available in WoS, Scopus, and Google Scholar. They established that in case of the researchers affiliated with humanities and social sciences, Google Scholar provides information about a number of articles four times as high as the number found in WoS and Scopus, and an average number of citations more than ten times as high. They found that the works of humanities scholars

were cited much more rarely (between fifty and eight times less, depending on the source of data – the highest disparity occurring in the WoS database) than the works of those affiliated with life science and science. This difference was indicated by all bibliometric tools used in the study. Harzing and Alakangas highlighted that different sources of information provided different views of the relations between the disciplines.

Oi Wang and Ludo Waltman (2016) conducted an analysis of systems of disciplinary classification of journals employed in both databases, and estimated their accuracy. With a method of direct citation relations between journals, they established that in both databases, a big part of the journals belongs to either of the two groups: (1) classified as affiliated with a given discipline, but without a significant relation to it; (2) classified as not affiliated with a given discipline, but with a significant relation to it. They judged that both databases are insufficiently precise in its systems of disciplinary classification of journals, and that a big part of the journals in both databases, but especially in the Scopus database, seems to be associated with too many different disciplinary categories. Earlier studies by Abdullah Abrizah and others (2013) showed, that the abovementioned problem affects every fourth journal in the category of Information Science and Library Sicence (IS&LS) in WoS and Library and Information Science (L&IS) in Scopus.

Alberto Martin-Martin, Enrique Orduna-Malrea and Emilio D. Lopez-Cozar (2018) showed that a relatively high number of much-cited (as per Google Scholar) articles related to social sciences and humanities is not accounted for in WoS and Scopus databases. The gaps are serious enough to undermine the usefulness of these databases for formulating bibliometric indicators-based assessments regarding these scientific disciplines.

2.2. The discipline of library and information science in the Web of Science and Scopus databases

The bibliometric analysis of the scientific discipline of library and information science (LIS) enjoys a long tradition and persisting popularity. More than ten years ago, Lonkman Meho and Kiduk Yang's study (2007) showed that there are serious problems with generating rankings for this discipline. They established that the Scopus database, as compared to WoS, significantly alters the ranking of researchers, especially those in the middle of the list, and that to achieve a more precise and complete view of the impact various researchers have on the development of LIS a complementary use of both databases, and additionally, of Google Scholar, would be required.

Isola Ajiferuke and Dietmar Wolfram (2010) described the impact of a given researcher on the development of science by measuring the ch-index, i.e. a method of estimating author research impact using the number of citers per publication an author's research has been able to attract. They estimated that for LIS scholars, it is a more accurate indicator than the general citation number, or the h-index. William H. Walters and Esther I. Wilder (2016) demonstrated that the development of LIS has been significantly impacted by research from the disciplines of computer science and management, by scholars from the USA, United Kingdom, Spain, China, Canada, and Taiwan. According to the research of Yu-Wei Chang (2018), based on the data from the WoS database, LIS is becoming increasingly interdisciplinary. Articles written by at least one author affiliated with another discipline constitute almost a half of the discipline's scientific output.

Apparently, scholars affiliated with medical sciences are to have a significant impact on the development of LIS.

Carlos G. Figuerola, Francisco J. Garcia Marco and Maria Pinto (2017) reviewed the projects from the previous decade that relied on the quantitative data from the Library and Information Science Abstracts (LISA). The review showed that the dominant methods employed in quantitative research were the bibliometric analysis of citation and co-authorship, statistical data, and terms co-occurrence method. These methods were characterized by the automated manner of data collection and analysis. The article inspires questions regarding the quality and representative capacity of the data subject to analysis: Are automatically collected data complete and reliable? To what extent does the content of the data in the analyzed bibliographic database correspond with the actual output within the studied discipline? These questions become increasingly urgent as the number of scientific publications related to LIS grows, and the role of multi- and interdisciplinary research for the development of this discipline increases. Does the content of the bibliographic databases keep up with the swiftly accumulating publications? Do automated methods of data collection and analysis, and statistic techniques, account for the multi- and interdisciplinarity of research?

The review of literature shows that the verification of the quality of data from the WoS and Scopus databases has not received the attention it merits, at least not in the studies concerned with library and information science.

3. Methods and results of the study of the usefulness of the data from the WoS and Scopus databases for the assessing the state of library and information science

A definite majority of the studies using the bibliometric indicators and the largest bibliographic databases, relied on the automated methods of data collection and analysis. These methods are not immune to errors in bibliographic descriptions, or errors in the assignment of journals to disciplinary or subject categories, and in the result of errors in recognition of the proper scope of data acquiring. Therefore, the author employed the method of qualitative verification of the automatically collected data about journals from the Web of Science and Scopus databases (the full description of the considered journals is provided in Appendix), and following, of qualitative analysis. The qualitative analysis involved a comparison of datasets regarding the journals affiliated with the library and information science from both of the databases studied; a comparison of the quantitative data collected from the databases studied with the data collected from other, domain-focused bibliographic databases; a comparison of the data with the information available on the websites of the journals indexed, and a comparison of the categorization of the journals, articles and authors as affiliated with given disciplines, employed by the producers of the databases, with the research scope of LIS accepted by the researchers. The aim of the analysis was to establish the quality and completeness of the data, and to identify the problems with a systematic collection and employment of such data.

3.1. The journals: their number in the databases, publishers, languages, disciplines and recognition

The 2017 Journal Citation Report available in the Web of Science database includes 87 journals assigned to the category of Information Science & Library Science (IS&LS), and 11 journals, which WoS puts in different categories, but which are assigned to the category of Library and Information Science (L&IS) in the Scopus database. These journals have their ascribed impact factors (IF). Furthermore, the Core Collection database contains data regarding further 47 journals, without providing their IF, the articles from which are assigned to the IS&LS category, out of which WoS puts 24 in other categories, but which belong in the Scopus category of L&IS. To the journals in the last group the author added three open access journals published in Spanish and Portuguese, indexed in the SciELO Citation Index database, and four from the Medline database (life sciences). In total, 176 journals registered in the WoS database were subject to analysis.

It is impossible to collect data about the journals without an IF indicator in any automated manner. The author had to type the name of the journal into the search tool and choose the field "Publication Name". The result of such a search is a list of bibliographic data of the articles published in a given journal and a set of data: a citation number of each article published in a given year, the number of articles published according to the rules of the open access, the number of publications of a particular type (article, book review, editorial material, note), subject category of the articles, last names and institutional affiliations of the authors with a number of the articles. Unfortunately, in the case of interdisciplinary journals, the function showing the subject categories of the articles does not work properly, as almost every article is put in every category simultaneously, and there is no possibility of distinguishing those which are concerned with IS&LS. Of course, an analogous set of data is a result of a search for a journal with an IF measurement, but in the case of such a journal, its name is an active hyperlink to the following information: the IF for the last two and the last five years, scientific categories, ranking, publisher, ISSN.

Among the 176 journals studied, as many as 96 (54.5%) were published by ten great publishing conglomerates (Taylor & Francis – 28, Emerald – 21, Elsevier – 12, Springer – 12, SAGE – 7, Wiley-Blackwell – 5, IGI Global – 3, Palgrave Macmillan – 3, Walter De Gruyter – 3, Brill – 2). The definite majority of the journals included – 159 (90.5%) publishes articles in English (a few journals also publishes texts in other languages, mostly in French, Spanish, and German). The group of journals in languages other than English is dominated by Spanish (10) and Portuguese (6). Two journals publish texts in both of these languages. If we take into consideration where the publisher of the journal is based, two countries dominate: USA – 68 (38.5%) and Great Britain – 51 (29%), followed by the Netherlands – 11 (7%), Germany – 8 (5%), Spain – 8 (5%), and Canada – 5 (3%). In a dozen or so other countries, there operate only singular publishing houses registered at the WoS database.

Among 134 journals (76%) categorized as IS&LS (out of which 87 have the IF calculated, and 47 do not), 18 have been additionally included in the Management category (out of which one has been also assigned to Computer Science), three to Communication, three to Interdisciplinary / Multidisciplinary, three to Education, two to Computer Sciences, and on to each of the following categories: Biomedical, Ethics, Geography, History, History of Social Sciences and Law (in total, 35 articles has been assigned to one of these categories).

Among 42 journals (24%) which WoS does not classify in IS&LS, but which Scopus does classify as L&IS), 10 has been additionally assigned to the category of Computer Science (out of which two have been also categorized as related to Engineering, and two – to Chemistry), seven – Science Technology (out of which two have been also assigned to Social Sciences), three – History, three – Humanities Multidisciplinary, two – Communication, two – Literature, two – Education, two – Language, Linguistics, two – Music, two – Social Sciences Interdisciplinary, and one to each of the following: Asian Studies, Biology, Law, Mathematics + Psychology, Medical Ethics, Multidisciplinary; one more journal has not been assigned to any category.

The Scopus database makes it easier to find a full list of the journals affiliated with a given scientific discipline. It is sufficient to use the function Sources – Enter subject area and choose a specific discipline, which will provide the user with a list of journals from that discipline indexed at Scopus. 208 journals have been indexed in the category of Library and Information Science (as per data from March 2019). Additionally, 28 journals whose are indexed as relating to this discipline by the WoS database have been assigned to other categories. Therefore, 236 journals from Scopus have been taken into consideration by the present study. Only two journals indexed in WoS are not indexed at all by Scopus. Every journal included has calculated an indicator based on citations. The following data on every journals is available: title and publisher, ISSN, subject area, the number of citations, CiteScore, SJR and SNIP, a link to the website, an information regarding the employment of open access rules, the number of texts published within a given year with different types of texts distinguished (article, editorial, review, note, conference paper), titles and authors of the given texts, the authors and their affiliations. The tool supposed to distinguish the scientific discipline which a given article is related to does not function properly in Scopus when multidisciplinary journals are concerned, as it did not work in WoS; all articles are assigned to all categories. However, Scopus offers a possibility of retrieving articles by specific keywords, which WoS did not enable. It makes it easier to select articles according to their subject scope, and to select authors concerned with specific research questions.

Among the 236 journals, 118 (59%) is published by the big publishing group (Taylor & Francis – 45, Emerald – 23, Elsevier – 12, Springer – 12, SAGE – seven, Wiley-Blackwell – six, Palgrave Macmillan – four, Walter De Gruyter – four, IGI Global – three, Brill – two), with the rest published by universities and scientific societies. Here, too, English definitely dominates, featuring in 215 journals (91%), out of which only 19 also publishes texts in other languages (mainly French, Spanish, and German). The remaining 9% is published in Spanish, French, German, and Portugese (several journals publishes texts in several languages). When the question of where the publisher is based is concerned, as in WoS, USA – where 93 (39.5%) publishers are based and Great Britain – 59 (25%), dominate. The list of the countries that follow is similar to that at WoS as well: Netherlands – 15 (6.5%), Spain – nine (4%), Germany – nine (4%), France – six (2.5%) and Canada – six (2.5%). The position of France on the list is the first significant difference between the sets of journals from these databases, as WoS does not index any French journals associated with the discipline. The second difference is that Scopus features journals from more countries, where only several journals (between one and three) are published. It features 27 such journals, as compared to WoS's 15.

Among the 208 journals assigned to L&IS, only 67 (32%) publishes exclusively articles associated with this discipline. Other journals publish articles associated with several disciplines, out of which most often next to L&IS occurs additionally: Computer Science – 32, and 16 further assigned also to either Decision Sciences, Education, Law, Chemistry, Business, Management and Accounting, or Engineering; Business, Management and Accounting – nine, and further two also assigned to Computer Science; Education – nine, and further three assigned also to Computer Science; Medicine and Health Profession – six; Communication – five, and one further assigned also to History; History – four, and one further assigned also to Communication. Among the 28 journals which WoS assigned to LS&IS, and which Scopus does not assigned to L&IS, most is assigned to Computer Science, Business, Management and Accounting, Decision Sciences, Communication, or Engineering.

To estimate the overlap between the set of the journals in both databases assigned to the subject area of LIS, and the set of journals considered as related to library and information science published in the world, the data from other international bibliographic databases was used. The specialist database Library, Information Science & Technology Abstracts (LISTA) provided by Ebsco accounts for 470 peer-reviewed scientific journals (https://www.ebsco.com/products/research-databases/library-information-science-and-technology-abstracts). The analysis of publishers of these journals shows that the journals published by the large publishing groups, discussed above, constitute a much smaller part of all journals than it did at WoS and Scopus: they are only 39% of all journals (185). Therefore, LISTA classifies many more journals (285) published by university presses, small scientific publishing houses and scientific societies, than WoS (80) and Scopus (118), as scientific and meeting the standards of quality. It means that both databases neglect to index the majority of scientific journals related to the LIS discipline, which are published outside the large publishing groups. The second database specializing in LIS, Library and Information Science Abstracts (LISA), provided by the ProQuest company, features 440 journals published in 20 languages and 45 countries (http://proquest.libguides.com/lisa). The number of countries and languages suggests that WoS (which features journals from only 24 countries) and Scopus (34 countries) do not take into consideration a large amount of scientific activity and publications of LIS scholars. Hence, the information from LISTA and LISA confirms Philippe Mongeon and Adele Paul Hus's (2016) thesis that WoS and Scopus take into account, respectively, only every fourth and every third, scientific journal related to LIS.

In Poland, it has been a long held view that journals featured in the Journal Citation Reports of the WoS database, and ascribed an IF value, are more prestigious. Currently, a position in the Scopus database is becoming a similar mark of prestige, which further depends on the indicators based on the citation numbers. A question arises if the two databases agree where a given journal ranks, and therefore, how prestigious it is. To answer it, two rankings were compared: first, based on the IF indicator, and second, on CitesSore. To compare them, 60 journals with the highest CiteScore were considered – this limit was imposed because a large part of the journals below the 60th position is not ascribed the IF value in the WoS database. The comparison showed that among the 60 journals with the highest CitesScore at Scopus, there are eight who do not have an IF value ascribed by WoS, i.e., they do not belong to the 98 most prestigious journals indexed there. Therefore, there is a major disagreement when it comes to the prestige of the 13% journals from the studied group. The following nine journals (15%) are ranked very differently by the two databases,

positioned more than 10 places apart. 18 (30%) journals have a more or less similar position in both rankings (a difference between 6 to 10 positions), and 25 (c. 42%) are ranked very similarly (the difference is five or less). Therefore, it seems that both databases tend to agree when determining the prestige, as measured by their citability, of a given journal from the L&IS (IS&LS) subject area.

3.2. The disciplinary classification in the databases versus the real subject scope of the journals

As mentioned above, a big part of the Polish academic community, as well as those in charge of science management in Poland, is convinced that the most valuable texts are published in journals who are attributed an Impact Factor by the Web of Science database. Until 2018, there functioned a special list of the journals indexed by WoS, created by the Ministry of Science and Higher Education (the so-called "A-list"), publishing in which was considered more prestigious, and which allowed the author to score more points in the evaluation of individual researchers and research institutions. After the introducing of a new legislation, *Prawo o szkolnictwie wyższym i nauce* (Law on Higher Education and Research) on 20th of July in 2018, the select journals from the A-list were incorporated into a new ministry list of scoring journals, which also includes journals from the Scopus database, and a set of Polish journals, which score much fewer points. The value of publications is supposed to be determined by the IF and a set of indicators from the Scopus database – CiteScore, SNIP and SJR (Komunikat MNSW, 2019). In such a model of evaluation, scientific prestige and the estimation of the quality of scientific output are determined by bibliometric indicators which depend on the citation number the journal is attributed on the basis of citations by other journals from the given database. Therefore, to establish which researchers and institutions have the most valuable output and which articles contribute to it, one would have to analyze the data regarding the journals from the group with the highest IF, or the highest CiteScore. Here, however, we have to ask if these criteria allow an unproblematic evaluation of publications from the discipline of LIS. The differences and problems discussed above should raise our doubts. To answer this question, the journals with the high IF assigned to the IS&LS category were examined to see if they actually published texts related to the discipline. The classification of disciplines employed at WoS was verified by an analysis of guidelines for the potential authors published on the journals' websites. The information on the subject scope of a given journal was selected, and then compared with the subject scope accepted by the specialists within the studied discipline. The 20 journals with the highest score for 2017 (i.e., the first quartile, Q1, in the IS&LS category) were selected for the comparison. The system of disciplinary classification employed by WoS was compared with the analogous classification employed by Scopus.

The thematic scope for research within the discipline of LIS which served as a model for comparison was established basing on two articles. The first study (Milojević et al., 2011) presented the results of thematic analysis of more than 10 thousand articles published between 1988 and 2007 in 16 journals associated with the LIS discipline. It was established that this discipline studies five main areas: the functioning of libraries, the world of information, the use of bibliometrics for the evaluation of science, information behavior, and bibliography. The specific issues included: public and academic libraries, digital libraries,

information systems, information competences, online services, knowledge management, scientific publications, the productivity of the researchers, citing, bibliometric indicators, information retrieval, catalogues and databases, classifications and internet search engines. The second article is a result of the research of Barbara Sosińska-Kalata (2013). She conducted a thematic analysis of the articles published in the journals considered to be the most important for the information science (a part of LIS)[1]. She established that information science is concerned with the following issues: the analysis of the state and the development of information and knowledge resources within different branches and specializations; information architecture and usability of internet websites; digital archives, libraries, and repositories; user studies – information needs, information literacy, and information behaviors; information barriers; history and contemporary nature of the book and of the library; quantitative research of information use – bibliometrics, scientometrics, webometrics; library science; information economy; information ethics; scholarly communication; information in digital and social media; knowledge organization – folksonomy, indexing, metadata, ontology, theory of classification and knowledge organization; theory of information; methodology and terminology of information science; data mining and text mining; big data; systems for the automatic content identification and extraction; expert systems and artificial intelligence; technology of information processing and publishing; information services; digitization and visualization of information; information retrieval and evaluation; information management; information sources. This set of research interests was accepted by the Polish academic community, which showed when it became the basis for a textbook *Nauka o informacji* (*Information Science*) published in 2016 with Wiesław Babik as the head editor.

If we compare the information from the websites of the journals studied (Table 1) with this set of research interests, we find that only journals no. 2, 3, 11, 12, 14, 15, 17 and 21, i.e., 36% of the journals featured in the table, might be considered as distinctly concerned with the discipline. In the following three quartiles of the journals assigned to IS&LS, the proporions are reverse: 66% of the journals categorized by WoS shows a distinct relation to the discipline of library and information science, while the remainder mostly publishes texts on the technical, medical, educational and business applications of the digital solutions (devices, apps, and systems), and on business management, which are mostly assigned to Computer Science, Communication, and Management.

Although LIS is interdisciplinary to a large extent, it does have specific research problems. It conducts its research and describes results thereof with methods borrowed from other disciplines, however its specific methodologies and techniques often differ from that used by the other disciplines. A study of technological aspects of the functioning of a given database is a different process than a study of linguistic aspects of its indexing and searching tools. A study of the management of a given information resource as a basis for generating decisions in business practice is different from the study of the process of creating, processing and publishing such a resource as a specific information structure. The

[1] The research of B. Sosińska-Kalata was concerned with the journals considered to be the most representative for the information science, which is a part of the designated research area of the library and information science, i.e., *Journal of the American Society for Information Science*, *Journal of Information Science* and *Journal of Documentation* and *Annual Review of Information Science and Technology*, published until 2011.

same research object might feature in different studies and journals, which does not mean that these studies and the journals publishing their results belong to the same discipline. A superficial analysis of the names of the journals, articles and books might suggest a substantial thematic overlap between library and information science, and disciplines such as computer science, engineering, or management. However, it is often a mistaken impression. The disciplinary classification systems employed by the largest databases might be similarly misleading. Indexing a given article as related to LIS, as well as to computer science, might be justified in certain instances, but not in others. Therefore, relying on the disciplinary classification employed by WoS for the diagnosis and evaluation of research conducted in the discipline of LIS risks basing these diagnoses and evaluations on bibliographic data of publications which are not representative of the discipline.

Tab. 1. Subject scope of the journals assigned to the IS&LS category from the first quartile, as indicated on their websites in the guidelines for authors

No.	Journal title	Subject scope as indicated on the journal's website in the guidelines for authors	WoS classification	Scopus classification
1	2	3	4	5
1	MIS Quarterly	development of IT-based services, the management of IT resources, and the use, impact, and economics of IT with managerial, organizational, and societal implications	IS&LS, management	computer science, decision sciences, business, management and accounting
2	Journal of Information Technology	technology and the management of IT – including strategy, change, infrastructure, human resources, sourcing, system development and implementation, communications, technology developments, technology futures, national policies and standards	IS&LS, management	L&IS, business, management and accounting; computer science
3	International Journal of Information Management	information management in learning organizations, business intelligence, security in organizations, social interactions and community development, knowledge management, information design and delivery, information for health care, Information for knowledge creation, legal and regulatory issues, IS-enabled innovations in information, content and knowledge management, philosophical and methodological approaches to information management research	IS&LS	L&IS, computer science

1	2	3	4	5
4	Journal of Strategic Information Systems	strategic management, business and organizational issues associated with the introduction and utilization of information systems, and considers these issues in a global context; organizational transformation on the back of IT; information systems/business strategy alignment; inter-organizational systems; global issues and cross-cultural issues; the impact and significance of emerging IT	IS&LS, management	business, management and accounting, computer sciences, decision sciences
5	Journal of the American Medical Informatics Association	clinical care, clinical research, translational science, implementation science, imaging, education, consumer health, public health, and policy	IS&LS	medicine
6	Information Systems Journal	information systems – research, practice, experience; articles that integrate technological disciplines with social, contextual and management issues	IS&LS	computer science
7	Government Information Quarterly	intersection of policy, information technology, government, and the public; how policies affect government information flows and the availability of government information; the use of technology to create and provide innovative government services; the impact of information technology on the relationship between the governed and those governing; and the increasing significance of information policies and information technology in relation to democratic practices	IS&LS	L&IS, sociology and political science, law
8	Journal of Computer--Mediated Communication	social science research on communicating with computer-based media technologies; work by scholars in communication, business, education, political science, sociology, psychology, media studies, information science	IS&LS, communication	computer science
9	Information and Management	research in the information systems field and managers, professionals, administrators of organizations which design, implement and manage Information Systems Applications; to collect and disseminate information on new and advanced developments in the field of information systems	IS&LS, management	computer science, decision sciences, business, management and accounting

1	2	3	4	5
10	Telematics and Informatics	the social, economic, political and cultural impacts and challenges of information and communication technologies	IS&LS	law, communication, engineering, computer science
11	Journal of Informetrics	bibliometrics, scientometrics, webometrics, and altmetrics, studying informetric problems using methods from other quantitative fields	IS&LS	L&IS, computer science
12	Information Processing and Management	research in information science, information searching, human information behavior, the areas of web searching, online advertising, public relations, communication, management information systems, computational economics, computational advertising, web analytics, online news, bibliometrics, scientometrics, health informatics, experimental processes related to digital libraries, knowledge management systems, multimedia processing, human--computer interfaces	IS&LS	L&IS, engineering, decision sciences
13	International Journal of Computer--Supported Collaborative Learning	education, computer science, information technology, psychology, communications, linguistics, anthropology, sociology, and business, investigate how to design the technological settings for collaboration and how people learn in the context of collaborative activity	IS&LS, education	education, computer science
14	Social Science Computer Review	artificial intelligence, business, computational social science theory, computer-assisted survey research, computer-based qualitative analysis, computer simulation, economic modeling, electronic modeling, electronic publishing, geographic information systems, instrumentation and research tools, public administration, social impacts of computing and telecommunications, software evaluation, world-wide web resources for social scientists	IS&LS, interdisciplinary	L&IS, law, computer science
15	European Journal of Information Systems	European perspective on the theory and practice of information systems; a critical view on technology, development, implementation, strategy, management and policy	IS&LS	L&IS, computer science

1	2	3	4	5
16	Journal of the Association for Information Systems	the field of information systems – it inclusive in topics, level and unit of analysis, theory, method and philosophical and research approach, reflecting all aspects of information systems globally	IS&LS	computer science
17	Journal of the Association for Information Science and Technology	research that focuses on the production, discovery, recording, storage, representation, retrieval, presentation, manipulation, dissemination, use, and evaluation of information and on the tools and techniques associated with these processes	IS&LS	L&IS, computer sciences
18	Journal of Management Information Systems	forum for the presentation of research that advances the practice and understanding of organizational information systems; the gap between theory and practice of management information systems	IS&LS, management	business, management and accounting, decision sciences, computer sciences
19	Journal of Knowledge Management	HR, learning & organization studies, information & knowledge management	IS&LS, management	business, management and accounting
20	Journal of Enterprise Information Management	information & knowledge management in enterprise	IS&LS, computer science, management	computer science. decision sciences, business management and accounting
21	Research Evaluation	evaluation of activities concerned with scientific research, technological development and innovation	IS&LS	L&IS, education
22	Qualitative Health Research	health care and further the development and understanding of qualitative research in health-care settings	IS&LS Interdisciplinary biomedical	medicine

This risk might be lowered with the use of the Scopus database, which offers a more accurate disciplinary classification of scientific journals than WoS, most often according with the real subject profile of the particular journals. As many as 13 of the 22 journals included by WoS in the category of IS&LS are not featured in the analogous category, L&IS, by Scopus. The information from the websites of the journals considered shows that the Scopus classification corresponds much more closely to the actual content of these journals. Additionally, the database allows to filter by keyword, which allows a more precise browsing of articles.

3.3. The key researchers concerned with a given research area

The gravity of the methodological problems generated by the imprecise thematic classification of the journals becomes apparent with the attempts to identify the key researchers

in a given discipline. In the Web of Science database, such an attempt to identify the key researchers (i.e., those publishing the highest number of articles and cited most often, thus having the largest impact on the discipline) concerned with the issues of library and information science, began with the use of the "Advanced Search" function. The author searched for SU – Research Area: SU=Information Science & Library Science, with the limiting of the results to Document types = Article. The search yielded 142 272 records. The use of the function "Refine Results = Web of Science Categories" showed that a large part of the group of the articles was assigned also to at least one other category (Fig. 1). The elimination of the articles classified in other categories reduced the number of records to 65 199. However, there was no option to automatically analyze to what extent did the eliminated articles relate to Information Science & Library Science, and to what extend did they relate to other scientific disciplines. The comparative analysis of the categories in WoS and the real thematic scope of journals conducted earlier suggests that the articles in other categories, such as Computer Science, Communication, and Management, should be considered as related to IS&LS only after sufficient deliberation; however, the interdisciplinary nature of LIS justifies categorizing many articles related to it as related to other scientific disciplines as well, among them these just mentioned. To satisfactorily resolve this matter, the only solution is to examine the title and abstract of every article included in these categories, which in total number more than 77 thousand. This is impossible, not in the least because of the time constraints.

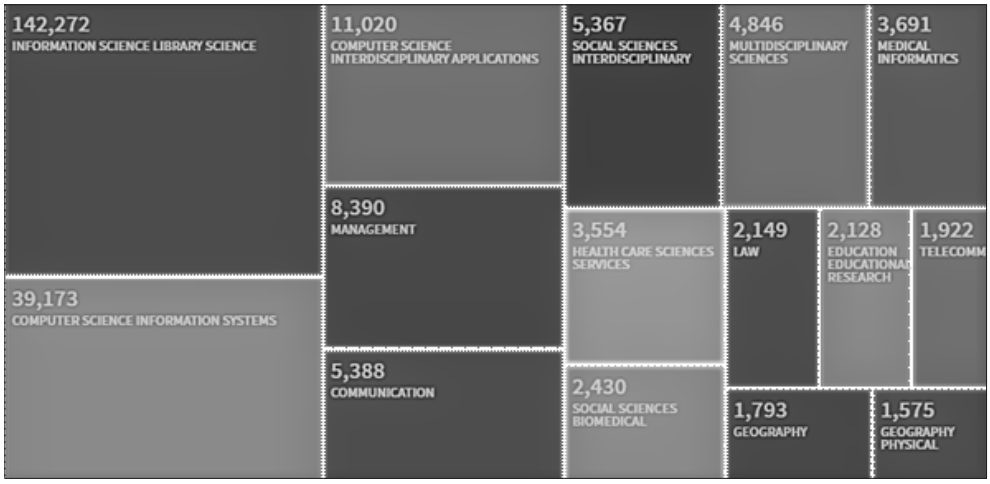

Fig. 1. The main categories to which the articles found as a result of the search SU=Information Science & Library Science are classified (Source: WoS)

The option to "Refine Result = Authors" allows to automatically show 100 authors who published the highest number of texts. The number of their publications, before elimination of the articles assigned also to other categories, is between 55 and 264 per author (25 authors who published the highest number of texts – Fig. 2).

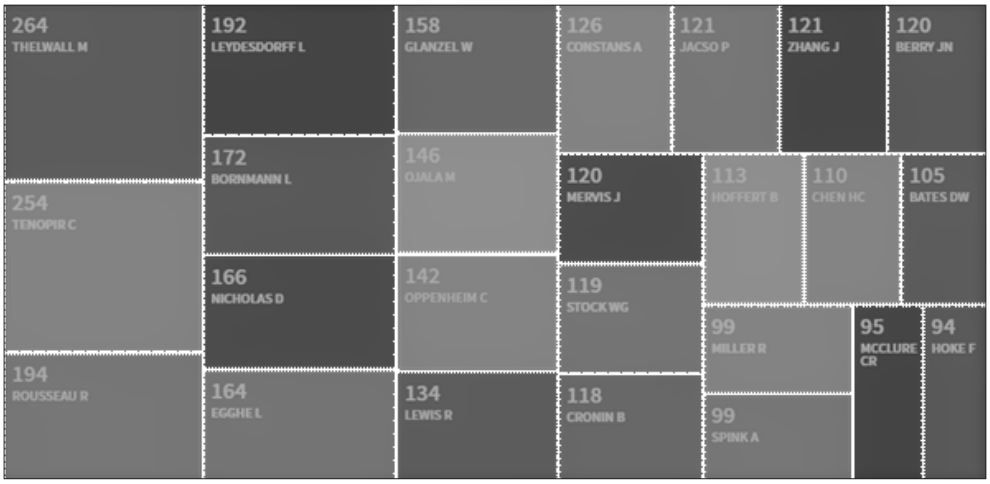

Fig. 2. The first 25 authors who published the highest number of texts in the journals in the category of Information Science & Library Science. The data before the elimination of the articles classified in other categories (Source: WoS)

After eliminating the articles assigned also to other categories, the number of the publications per author fell to a 28 – 216 range (Fig. 3). The members of the group change as well. Among the first 25 names, only eight recurs (C. Tenopir, J.N. Berry, W.G. Stock, B. Hoffert, M. Ojala, R. Miller, C.R. McClure, D. Nicholas), and the following two (M. Thelwall, C. Oppenheim) fall below 25th position, but remain above the 100th.

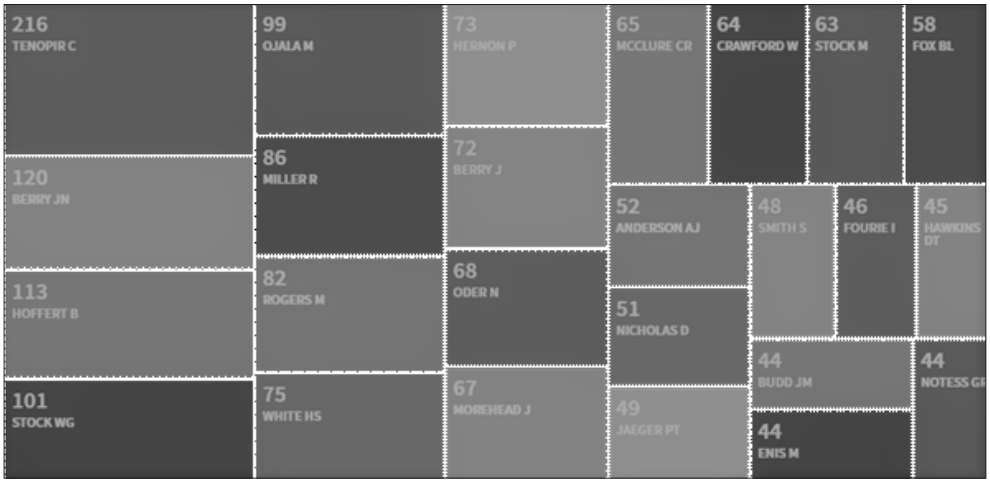

Fig. 3. The first 25 authors who published the highest number of texts in the journals in the category of Information Science & Library Science. The data after the elimination of the articles classified in other categories (Source: WoS)

WoS allows for a creation of a list of 500 authors who published the highest number of articles assigned to a given category. However, it is not obvious that whoever published the most articles is the leading researcher in a given discipline. The scientific position of an author might be verified by a study of the citability of their work.

To analyze the citations, a list of 100 authors with the highest number of publications was used, initially assuming that it would feature the authors with the largest impact, because of the citability of their work (additionally, WoS generates citability reports for sets of up to 10 thousand records). This list included 3993 articles, which were cited 15 400 times (14 246 without auto-citations) in 11 917 articles (11 247 without auto-citations) indexed in WoS Core Collection. It gives an average of 3.86 citation per article. WoS allows to rank given articles according to the number of their citations. The first 10 articles were cited between 111 and 463 times. The first 50 were cited 40 or more times; 100 – eight times; 1000 – three times. 911 articles had a number of citations above the average (4 and more). The most cited authors would be found in this group. Unfortunately, WoS does not enable such an automated search. The author had to create a set of publications by given authors (e.g. from the list of 100 with the highest number of articles published, or 100 whose articles had the highest citation numbers), and then to add up the citation numbers and to compare the average citation numbers of their work. A set of each author's publications was found with the use of the command AU=last name, first letters of the name. It was possible to generate a citation report for a thus generated set. The data on 25 authors who published the highest number of articles is presented in Table 2.

Tab. 2. Authors with the highest number of articles in the IS&LS category

No.	Author	Numbers of articles published	Citation number / without self-citations	h-index	Average citation number per article	Time of publishing
1	**2**	**3**	**4**	**5**	**6**	**7**
1	Tenopir C.	216	999/928	15	4.63	1999–2018
2	Berry J.N.	120	53/53	5	0.44	1998–2017
3	Hoffert B.	113	44/43	3	0.39	1999–2018
4	Stock W.G.	101	68/49	3	0.67	1999–2018
5	Ojala M.	99	72/64	3	0.72	1993–2012
6	Miller R.	86	81/81	4	0.94	1994–2013
7	Rogers M.	82	31/31	4	0.38	1998–2017
8	White H.S.	75	290/271	9	3.82	1980–1999
9	Hernon P.	73	793/752	16	10.86	1992–2011
10	Berry J.	72	26/26	3	0.36	1988–2007
11	Oder N	68	73/73	5	1.07	1991–2010
12	Morehead J.	67	35/27	2	0.52	1976–1995
13	McClure C.R.	65	693/668	13	10.66	1993–2012

1	2	3	4	5	6	7
14	Crawford W.	64	60/59	3	0.94	1993–2012
15	Stock M.	63	1/0	1	0.02	1994–2013
16	Fox B.L.	58	28/28	2	0.48	1998–2017
17	Anderson A.J.	52	12/12	1	0.23	1979–1998
18	Nicholas D.	51	579/531	15	11.35	1998–2017
19	Jaeger P.T.	49	1918/1800	19	39.14	1999–2018
20	Smith S.	48	41/41	4	0.85	1999–2018
21	Fourie I.	46	134/108	7	2.91	1998–2017
22	Hawkins D.T.	45	323/303	10	7.18	1983–2002
23	Budd J.M.	44	456/449	12	10.13	1999–2018
24	Enis M.	44	3/3	1	0.07	2012–2018
25	Notess G.R.	44	49/47	3	1.11	1992–2003

The data makes clear that the number of publications of a given author does not correspond to their impact, as indicated by the citation number. Only seven authors from the list of 25 with the highest number of publications published articles with an above-average citability.

The authors with the highest impact might be identified by an analysis of the most commonly cited articles (Tab. 3). However, this method does not guarantee reliable results, either.

Tab. 3. The authors of the most commonly cited articles published in the journals classified as related to IS&LS (ranked according to the citation number of the most cited work)

No.	Author	Number of articles published	Number of citations / without self-citations	h-index	Average citation number per article	Time of publishing
1	2	3	4	5	6	7
1	Pawlak Z.	1	6622/6622	1	6622	1982
2	Layne K.	1	841/841	1	841	2001
3	Lee J.W.	8	852/848	3	106.5	2001–2016
4	Lee D.T.	3	618/618	2	206	1980–1984
5	Schachter B.J.	1	573/573	1	573	1980
6	Taylor R.S.	7	500/500	4	71.43	1996–2015
7	Bertot J.C.	43	1286/1238	14	29.91	1997–2016
8	Jaeger P.T.	49	1918/1800	19	39.14	1999–2018
9	Grimes J.M.	1	463/463	1	463	2010
10	Eppler M.J.	1	417/417	1	417	2004

1	2	3	4	5	6	7
11	Mengis J.	1	417/417	1	417	2004
12	van Dijk J.	4	419/419	3	104.75	2003–2017
13	Hacker K.	1	368/368	1	368	2003
14	Glanville J.M.	2	386/386	2	193	2006–2014
15	Lefebvre C.	4	409/405	3	102.25	2006–2014
16	Miles J.N.V.	1	367/367	1	367	2006
17	Belkin N.J.	2	375/374	2	187.5	1980–1987
18	Heeks R.	7	729/728	5	104.14	2002–2018
19	Savolainen R.	29	837/820	11	28.86	2001–2017
20	Boulos M.N.K.	3	346/346	2	115.33	2006–2009
21	Wheeler S.	2	339/339	1	169.5	2007
22	Gandomi A.	1	330/330	1	330	2015
23	Haider M.	1	330/330	1	330	2015
24	Lin J.C.C.	1	327/327	1	327	2000
25	Lu H.P.	2	366/366	2	183	1994–2000

This method of data collection from the WoS database is only minimally useful for the identification of the researchers crucial for the development of a discipline. Among the 25 researchers who published the most commonly cited articles, there are only three (P.T. Jaeger, J.C. Bertot, R. Savolainen), whose h-index indicated an above-average impact on other researchers. The remainder is found on the list because singular, but commonly cited works; these citations are not necessarily in texts in the category of IS&LS. It is unclear if these works had a real impact on library and information science. To verify this, the author analyzed their thematic scope, and publication. Below is the list of the first 10, excluding the three already mentioned.

(1) Pawlak Z.: Rouhh Sets. *International Journal of Computer & Information Sciences*, 6622 citations. An article on mathematical sciences, published in a journal related to widely understood information science, which is no longer published, and is not indexed neither in the Journal Citation Reports, nor in the Scopus database. In the 1980s, the journal was replaced by the *International Journal of Parallel Programming*, which is related to computer science.

(2) Layne K.: Developing Fully Functional E-government: A Four Stage Model. *Government Information Quarterly*, 841 citations – a description of the development of e-administration and a proposal for a model of its development stages, published in the journal assigned to IS&LS; however, it mostly publishes works concerned with political sciences and administration.

(3) Lee D.T., Schachter B.J.: 2 Algorithms for Constructing a Delaunay Triangulation. *International Journal of Computer & Information Sciences*, 574 citations – an article on mathematical sciences, published in the journal related to widely understood information science, but focused mostly on the issues of computer science; as discussed above, the journal is no longer published.

(4) Taylor R.S.: Question-Negotiation and Information Seeking in Libraries. *College & Research Libraries,* 476 citation – the article, as well as the journal where it was published, are representative of LIS.
(5) Eppler M.J., Mengis J.: The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *Information Society*, 417 citations – a review article concerned with the information overload, which is a research problem in LIS as well as in some other disciplines; however, it was published in a journal specializing mostly in political sciences, cultural studies, and computer science.
(6) van Dijk J., Hacker K.: The Digital Divide as a Complex and Dynamic Phenomenon. *Information Society*, 368 citations – a discussion of factors influencing the phenomenon of digital exclusion, which is a subject of interest in LIS as well as in some other disciplines; however, it was published in a journal specializing mostly in political sciences, cultural studies, and computer science.
(7) Glanville J.M., Lefebvre C., Miles J.N.V: How to Identify Randomized Controlled Trials in MEDLINE: Ten Years On. *Journal of the Medical Library Association*, 367 citations – the article, and the journal where it was published, are representative of LIS; the journal specializes in problems of medical information and medical librarianship.
(8) Belkin N.J.: Anomalous States of Knowledge as a Basis for Information-Retrieval. *Canadian Journal of Information Science – Revue Canadienne Des Sciences De L'Information*, 363 citations – the article, as well as the journal where it was published, are representative of LIS.
(9) Heeks R.: Information Systems and Developing Countries: Failure, Success, and Local Improvisations. *Information Society*, 342 citations – the article presents models and theories of functioning of information system in the developing countries, and the related problems; it belongs in the research area of LIS, however the article was published in a journal specializing in political sciences, cultural studies, and computer science.
(10) Boulos M.N.K., Wheeler S.: The Emerging Web 2.0 Social Software: An Enabling Suite of Sociable Technologies in Health and Health Care Education. *Health Information and Libraries Journal*, 339 citations – employment of social media in health care and health education is a subject within the research area of LIS, and the article was published in a journal assigned to IS&LS, focusing on the problems of health and medical information (accordingly, it was assigned to the category Medicine and health profession as well).

The search for the most cited articles in the IS&LS category indexed in WoS brought up only three articles with an undeniable connection to LIS (a study of information users and their information behaviors) and a significant impact on the discipline (articles no. 4, 7, and 8). Five articles had some connection to LIS, but were published in journals which in fact do not focus on the discipline's research problems (no. 2, 5, 6, 9, and 10). Two articles were published in a journal who has not been a platform for scientific communication of the LIS researchers for a long time.

Therefore, the application of the WoS function which allows for an automated search of the authors who published the highest number of articles, and for the articles which have the highest number of citations, is not always reliable, and does not give an accurate

view of the discipline studied. It is only when the detailed information regarding every author and article is studied that the researchers and articles most representative of LIS might be selected.

The function "Search (Documents, Authors, Affiliations, Advanced)" of the Scopus database does not allow for an automated search of the articles or authors related to a given scientific discipline. The author could only use the "Sources – Enter subject area" option and choose "Library and Information Sciences". However, as a result he received only a list of journals assigned to the category. To find the information on the authors in every journal requires a separate process. After selecting a journal known to be publishing articles related to the studied discipline, the "View all documents" function had to be selected. Among various data provided, Scopus showed a list of authors and a number of the works they published. The search by "Author name" limits the result to the works of the given author. To find the authors of the articles which had the largest impact on the given discipline, the author assumed that they would publish their works in the journals with the highest impact measurements. For the purpose of the current study, the author examined the authors from the journals with the highest CiteScore indicator, assigned to the L&IS category in Scopus and verified as representative of the discipline by an analysis of the content on their websites. The examination of every journal was performed with following commands: "View all documents", "Limit to – Subject area: Social Science", and then "Sort on: Cited by (highest)". Then, the "Author details" command was used to find the number of their texts in the Scopus database, among them those included in the Social Sciences category; their h-index; the part of their articles assigned to given categories. Table 4 presents the results of the search of 10 most commonly cited authors in the eight LIS journals with the highest CiteScore.

The data presented in Table 4 shows that it is impossible to distinguish the most influential authors from the set of the journals studied, as only two of them appear more than once – M. Thelwall and J.D. Roessner appear twice. It seems that every journal relies on their particular set of authors to supply texts for publishing. Searching for researchers with the greatest impact on the discipline among authors of articles most often cited in individual periodicals also proved to be problematic. This group of authors also includes researchers who published singular but highly cited texts, however in articles assigned to disciplines other than the one which the author or the journal publishing these texts represent. In the studied set, it was 11 of the 78 researchers in the analyzed group. Each of these researchers is attributed a low (single-digit) h-index by Scopus, which does not allow to identify any of them as an influential author. Furthermore, the authors who publish the most cited articles in the studied LIS journals rarely focus on this discipline in their research. According to the Scopus algorithms, only 16 out of 78 researchers studied focuses on social sciences (unfortunately, Scopus does not indicate whether it belongs to the more specific L&IS sub-category), and more than half, i.e. 44 authors focus mostly on computer science. As Chang observed (2018), the researchers from other disciplines who publish in the LIS journals most often use scientometrics to analyze trends occurring in the disciplines they are interested in; as well as explore problems related to information technology, information issues in economics, and information issues in medicine, which is generally confirmed by the data presented in Table 4.

Tab. 4. The authors of the most cited articles from the L&IS category
(ranked according to the citation number of the most commonly cited work)

No.	Journal title	Author	The number of the articles/ including the articles in the Social Sciences category	The citation number	h-index	The dominant category of the author's output as per Scopus
1	2	3	4	5	6	7
1	Journal of Information Technology	Baskerville R.L.	177/42	5953	36	computer science
		Wood-Harper A.T.	11/3	1040	7	computer science
		Chan Y.E.	53/20	2962	18	computer science
		Reich B.H.	45/10	3117	18	computer science
		Markus M.L.	74	7589	34	computer science
		Axline S.	3/1	546	3	computer science
		Petrie D.	3/1	546	3	computer science
		Tanis C.	2/1	826	2	computer science
		Jeyaraj A.	29/19	971	8	computer science
		Rottman J.W.	23/6	989	11	business, management and accounting
2	International Journal of Information Management	Lin J.C-C.	26/11	1816	14	computer science
		Lu H.	90/28	4366	28	computer science
		Gandomi A.	7/1	475	3	engineering
		Haider M.	19/10	671	8	engineering
		Sultan N.	23/13	806	7	social sciences
		Yates D.	52/12	1018	11	computer science
		Paquette S.	16/11	654	7	social sciences
		Edmunds A.	1/1	340	1	computer science 50%, social science 50%
		Morris A.	114/66	1823	20	social sciences
		Trkman P.	53/14	1592	19	computer science

1	2	3	4	5	6	7
3	Journal of Informetrics	Alonso S.	69/6	3314	22	computer science
		Cabrerizo F.J.	85/7	2623	23	computer science
		Herrera-Viedma E.	373/29	17088	67	computer science
		Herrera F.	589/32	38007	95	computer science
		Moed H.F.	116/74	4854	36	computer science
		Prabowo R.	9/5	371	6	computer science
		Thelwall M.	381/245	11187	55	computer science
		Wagner C.S.	37/21	506	15	social sciences
		Roessner J.D.	37/15	882	13	business, management and accounting
		Bobb K.	7/7	270	3	social sciences
4	Information Processing and Management	Salton G.A.	103/22	13317	33	computer science
		Buckley C.	42/10	9400	28	computer science
		Jansen B.J.	200/70	7953	40	computer science
		Spink A.H.	206/111	7496	42	computer science
		Saracevic T.	83/45	4880	25	computer science
		Sokolova M.V.	97/10	1424	12	computer science
		Lapalme G.	89/20	1682	15	computer science
		Radev D.R.	97/46	4633	32	computer science
		Jing H.	10/8	698	7	social sciences
		Styś M.	1/1	02	1	computer science 50%, social sciences 50%
5	Social Science Computer Review	Walther J.B.	90/56	8672	36	social sciences
		D'Addario K.P.	1/1	307	1	computer science 50%, social sciences 50%
		Crawford S.D.	15/6	761	10	medicine
		Couper M.P.	142/86	7116	45	social sciences
		Lamias M.J.	3/3	650	3	social sciences
		Zhang W.	18/17	617	10	social sciences
		Johnson T.J.	70/58	238	22	social sciences
		Seltzer T.	15/14	869	8	social sciences
		Bichard S.L.	14/11	579	8	social sciences
		Barrett L.F.	197/16	15291	60	psychology

1	2	3	4	5	6	7
6	European Journal of Information Systems	Walsham G.	84/34	6257	31	computer science
		Petter S.C.	41/10	2703	12	computer science
		Delone W.H.	28/7	11595	15	computer science
		McLean E.R.	70/11	11769	19	computer science
		Heijden H.V.D.	1/1	468	1	computer science 50%, social sciences 50%
		Verhagen T.	24/5	1197	14	computer science
		Creemers M.	2/1	509	2	computer science
		Zhu K.X.	33/9	4064	16	computer science
		Kraemer K.L.	150/68	10496	42	computer science
		Xu S.X.	26/6	2242	12	business, management and accounting
7	Journal of the Association for Information Science and Technology	Hamari J.	46/14	2556	20	computer science
		Sjoklint M.	4/1	246	4	computer science
		Ukkonen A.	40/1	670	11	computer science
		Costas R.	71/36	1381	20	computer science
		Zahedi Z.	12/6	347	5	computer science
		Wouters P.F.	54/31	1358	18	computer science
		Bornmann L.	292/188	6450	40	computer science
		Mutz R.	59/30	1699	20	computer science
		Thelwall M.	382/246	11260	55	computer science
		Peters I.	39/27	532	13	computer science
8	Research Evaluation	Aksnes D.W.	31/15	1164	15	computer science
		Laudel G.	31/26	833	14	social sciences
		Lissoni F.	40/12	2837	16	business, management and accounting
		Llerena P.	45/6	1182	13	business, management and accounting
		McKelvy M.	87/15	1185	15	business, management and accounting
		Sanditov B.	10/2	180	6	business, management and accounting
		Martin B.R.	84/31	4593	30	business, management and accounting
		Taxt R.E.	2/1	181	2	veterinary, agricultural and biological sciences
		Porter A.L.	248/85	4503	33	business, management and accounting
		Roessner J.D.	37/15	888	13	business, management and accounting

The study shows that the automatic data filtering functions in the Scopus and WoS databases are hardly useful for identifying key researchers in LIS. Without a tedious, immediate analysis of the texts of specific articles and the scientific output of the researchers, it is not possible to accurately classify their work as related to a given discipline, and to properly evaluate it.

4. Conclusion

The method of qualitative analysis employed in the study has limitations which might change the results of the study. The first limitation comes from the necessity of accepting a closed set of research interests as definitive of the LIS discipline. Modifications of this set will without a doubt change a part of the data. However, in the case of this interdisciplinary research area strongly affected by other disciplines and employing their research methods, it is difficult to avoid this problem. The author followed the classification of research fields in information science established in the study by Barbara Sosińska-Kalata (2013), not only because it is commonly accepted in Poland and in accord with the classification proposed by the team under the direction of Stas Milojević (2011) commonly employed in research. It was also important that it is sufficiently detailed, facilitating a classification of a thematic scope of particular journals and articles. This limitation did not have an impact on the realizing the primary goal of the current study, i.e., indicating methodological problems in the use of the WoS and Scopus databases, with their filtering and analytical tools, by researchers and those in charge of science management to evaluate the state of library and information science's development, as well as of the quality of the scientific output of the LIS researchers.

The second limitation comes from the quality of the data studied. As it has already been mentioned, Fiorenzo Franceschini, Domenico Maisano and Luca Mastrogiacomo (2016) established that bibliographic records contain multiple errors which disrupt the results of searches. We should also remember the studies of Qi Wang and Ludo Waltman (2016), and of Abdullad Abrizah and others (2013), which show that the producers of the most important multidisciplinary bibliographic databases make mistakes in assigning journals to scientific disciplines they are supposed to represent. The results presented in this article confirmed that mistakes had been made in assigning not only journals to disciplines, but also authors. Therefore, it is necessary to verify the data retrieved from the WoS and Scopus databases with qualitative methods, e.g. analysis of the thematic profile, or content, of the given authors' publications.

The third limitation comes from the quantity of data and the sample selection. In practice, the time constraints on the qualitative analysis, and the limits to an article's length imposed by the editing team, make it impossible to examine and discuss all data regarding the articles, authors, and citations. Therefore, the author had to narrow down the quantity of the analyzed data with a use of a formal criterium, i.e. selecting the most commonly cited publications. However, the attached risk is minimal in the case of the research focused on the identification of methodological issues, which is possible even with a small data sample.

Both databases studied offer an automatic search function, and the refinement and analysis of the search results, which the researchers conducting bibliometric and scientometric

analyses are happy to use, but which, as the study has shown on the case of LIS, do not ensure a collection of sufficiently representative and reliable datasets for such analyses. Only when the quantitative data is accompanied by a qualitative analysis based on the information from various sources, it is possible to conduct a reliable and holistic evaluation of the usefulness of the bibliographic databases for an analysis and diagnosis of the state of a scientific discipline. In the case of LIS, the author faced many obstacles to a collection of full and reliable information. Both databases offer only a very limited set of data to conduct a reliable analysis and diagnosis of the LIS discipline. The journals published by universities and scientific societies, or published outside the USA and the Great Britain, are definitely underrepresented. The WoS database does not enable an automated data acquiring of journals that do not have a measured IF. It is impossible to filter the articles published in multidisciplinary journals by their research area categories, because all articles are assigned to all disciplines the journals is associated with. A similar problem occurs in the Scopus database, but Scopus does enable a filtering of the articles by the keywords. WoS mistakenly classified a big group of journals as belonging to LIS. Neither database allows an automated, but still reliable identification of the researchers with the highest impact on the development of the discipline, which is made even more difficult by its interdisciplinary character. A similar problem occurs when the user searchers for the most commonly cited work closely related to the given discipline. Scopus does not allow for a precise assignment of a researcher to a research area category. Neither database enables an automated and reliable assignment of a researcher to particular research interests.

The results of the quantitative study of the state and development of a discipline reliant on the data acquired from the WoS and Scopus databases and processed automatically by filtering and statistical tools, should be approached only with great care. Without verifying if the journals, articles, authors and citations studied have a real relation to the given scientific discipline, or if the databases' system of disciplinary classification corresponds to the real subject scope of the journals included, the results might be false. In the case of LIS, it is very easy to receive data presenting an inaccurate view of the discipline, especially when using the WoS database, as its category of IS&LS is too broad, and includes many journals which have only a tangential relation to the discipline. There is no point of constructing a view of LIS on the basis of data acquired in the most part from the journals which publish texts concerned mostly, if not exclusively, with computer science, communication sciences, business, management and accounting. All lists of influential journals and authors, and measurements of publication numbers and their citability become unreliable as a result of confusing the data related to LIS with the data related to other disciplines. Because of the different citation models, number of researchers and frequency of publishing in the disciplines assigned to one category with LIS, the journals, articles and researchers which are actually crucial for LIS become invisible in the category containing less related data. It is worth mentioning that the studies discussed in this article additionally showed that the data from WoS and Scopus highlights the technological research paradigm in LIS, despite the development of other methodologies.

In their current state, the studied multidisciplinary databases, especially WoS, have only a negligible usefulness for the diagnosis of the state and development of LIS. The methodological issues discussed in this article may prevent generating an accurate and objective view of the discipline. The changes to the organization of resources and functioning of

WoS and Scopus, which have been taking place for a few years now, did not resolve the previously existing issues.

Appendix: The list of journals included in the study

Note: In the column "IF 2017" (impact factor) in the absence of calculated IF there was entered the information about indexing in the Core Collection (cc) or in another database included in the Web of Science. The "-" means that the database does not include a specific journal.

L&IS – Library and Information Science; IS&LS – Information Science and Library Science.

No.	Journal title	Cite-Score 2017	IF 2017	Publisher	Headquarters	Language of publication	Scopus system of disciplinary classification	WoS system of disciplinary classification
1	2	3	4	5	6	7	8	9
1	*Accountability in Research*	1.05	1.400	Taylor & Francis	United Kingdom	English	L&IS	medical ethics
2	*African Journal of Library Archives and Information Science*	0.30	0.286	Archlib & Information Services	Nigeria	English	education, medicine	IS&LS
3	*AIB Studi*	0.39	cc	Associazione Italiana Biblioteche	Italy	English Italian	L&IS	IS&LS
4	*American Archivist*	0.6	cc	Society of American Archivists	USA	English	L&IS	IS&LS
5	*Analecta Hibernica*	0.0	–	Irish Manuscripts Commission	Ireland	English	L&IS	–
6	*Anales de Documentacion*	0.32	cc	University of Murcia	Spain	Spanish Portuguese	arts and humanities	IS&LS
7	*Annals of Library and Information Studies*	0.39	cc	National Institute of Science Communication and Information Resources	India	English	L&IS	IS&LS
8	*Archival Science*	1.52	–	Springer	USA	English	history	–
9	*Archivaria*	0.51	Medline	Association of Canadian Archivists	Canada	English	L&IS	science technology, social sciences

1	2	3	4	5	6	7	8	9
10	*Archives*	0.00	cc	British Records Association	United Kingdom	English	L&IS	IS&LS
11	*Archives and Manuscripts*	0.52	cc	Taylor & Francis	United Kingdom	English	computer science	history
12	*Aslib Journal of Information Management*	2.01	1.461	Emerald	United Kingdom	English	L&IS	IS&LS
13	*Australian Academic and Research Libraries*	0.92	0.818	Taylor & Francis	United Kingdom	English	history	IS&LS
14	*Behavioral and Social Sciences Librarian*	0.8	–	Taylor & Francis	USA	English	L&IS	IS&LS
15	*Biblios*	0.08	cc	University of Pittsburgh	USA	English	L&IS	–
16	*BiD*	0.04	cc	Universitat de Barcelona	Spain	Portuguese Spanish	history	IS&LS
17	*Bilgi Dunyasi*	0.06	–	University and Research Librarians Association, Ankara	Turkey	English French Portuguese Spanish	L&IS	IS&LS
18	*Boletim do Arquivo da Universidade de Coimbra*	0.00	cc	Imprensa da Universidade de Coimbra	Portugal	English Turkish	communication	–
19	*Bottom Line*	0.24	cc	Emerald	United Kingdom	Portuguese	L&IS	history
20	*Bulletin des Bibliotheques de France*	0.01	–	Ecole nationale superieure des sciences de l'information et des bibliotheques	France	English	computer science	IS&LS
21	*Bulletin. John Rylands University Library of Manchester*	0.11	Medline	John Rylands University Library	United Kingdom	French	L&IS	–
22	*Canadian Journal of Information and Library Science*	0.42	0.243	University of Toronto Press	Canada	English	L&IS	science technology

1	2	3	4	5	6	7	8	9
23	*Canadian Journal of Program Evaluation*	0.3	cc	Canadian Journal of Program Evaluation	Canada	English French	general social science	IS&LS
24	*Cataloging and Classification Quarterly*	0.67		Taylor & Francis	USA	English French	L&IS	social sciences
25	*Ciencia da Informacao*	0.01	cc	Brazilian Institute for Information in Science and Technology	Brazil	English	L&IS	IS&LS
26	*Collection Building*	0.6	–	Emerald	United Kingdom	Portuguese Spanish	L&IS	–
27	*Collection Management*	0.27	cc	Taylor & Francis	USA	English	L&IS	IS&LS
28	*College and Research Libraries*	1.7	cc	Association of College and Research Libraries	USA	English	conservation	IS&LS
29	*College and Research Libraries News*	0.46	1.626	Association of College and Research Libraries	USA	English	L&IS	IS&LS
30	*College and Undergraduate Libraries*	0.57	–	Taylor & Francis	USA	English	business, management and accounting	–
31	*Communications in Information Literacy*	1.44	cc	Communications in Information Literacy	USA	English	L&IS	IS&LS
32	*Community and Junior College Libraries*	0.3	cc	Taylor & Francis	USA	English	L&IS	IS&LS
33	*Computers in Libraries*	0.25	–	Information Today	USA	English	arts and humanities	–
34	*Computers in the Schools*	1.06	–	Taylor & Francis	USA	English	L&IS	–
35	*Cuadernos.info*	0.43	cc	Pontificia Universidad Catolica de Chile	Chile	English	computer science	education
36	*Cybermetrics*	5.50 (2016)	SciELO	Centro de Informacion y Documentacion Cientifica	Spain	Spanish	L&IS	social sciences

1	2	3	4	5	6	7	8	9
37	*Data Base for Advances in Information Systems*	0.56	–	Association for Computing Machinery	USA	English	L&IS	–
38	*DESIDOC Journal of Library and Information Technology*	0.42	0.400	Defence Scientific Information & Documentation Centre	India	English	L&IS	IS&LS
39	*Development and Learning in Organizations*	0.21	cc	Emerald	United Kingdom	English	L&IS	IS&LS
40	*Digital Library Perspectives*	0.59	–	Emerald	United Kingdom	English	L&IS	–
41	*D–Lib Magazine*	0.86	cc	Corporation for National Research Initiatives	USA	English	business, management and accounting	IS&LS
42	*Document Numerique*	0.13	–	Lavoisier	France	English	L&IS	–
43	*Documentaliste: Sciences de l'Information*	0.02	–	Association des Professionnels de l'Information et de la Documentation	France	English French	L&IS	–
44	*East Asian Publishing and Society*	0.27	–	Brill	USA	English French	education	–
45	*EContent*	0.02	cc	Online Inc.	USA	English	L&IS	Asian studies
46	*Education and Information Technologies*	1.3	0.039	Springer	USA	English	education	IS&LS
47	*Education for Information*	0.55	cc	IOS Press	Netherlands	English	L&IS	education
48	*Electronic Library*	0.99	cc	Emerald	United Kingdom	English	education	IS&LS
49	*Ethics and Information Technology (55)*	1.62	0.800	Springer	Netherlands	English	L&IS	education
50	*European Journal of Information Systems (17)*	4.23	1.080	Palgrave Macmillan	United Kingdom	English	education	IS&LS

1	2	3	4	5	6	7	8	9
51	*Evidence Based Library and Information Practice*	0.29	3.197	University of Alberta	Canada	English	L&IS	IS&LS
52	*Fontes Artis Musicae*	0.03	cc	Music Library Association	Switzerland	English	computer science	ethics
53	*Gazette des Archives*	0.01	cc	Association des Bibliothecaires Francais	France	English French German	L&IS	IS&LS
54	*Government Information Quarterly*	5.82	–	Elsevier	United Kingdom	French	education, computer science	IS&LS
55	*Grey Journal*	0.11	4.009	GreyNet	Netherlands	English	L&IS	music
56	*Health Information and Libraries Journal*	1.02	–	Wiley–Blackwell	USA	English	communication	–
57	*Ibersid*	0.06	1.190	Universidad de Zaragoza	Spain	English	L&IS	IS&LS
58	*IC Revista Cientifica de Informacion y Comunicacion*	0.00	cc	Editorial Universidad de Sevilla	Spain	Spanish	Computer science,	–
59	*IEEE Transactions on Information Theory*	3.33	cc	Institute of Electrical and Electronics Engineers	USA	English Spanish Portuguese	Business, management and accounting	IS&LS
60	*IFLA Journal*	0.71	cc	Sage	USA	English	L&IS	IS&LS
61	*Informacao and Sociedade – Estudos*	–	cc	Univ. Federal Campina Grande	Brazil	English	L&IS	communication
62	*Informacion, Cultura y Sociedad*	0.13	0.159	Instituto de Investigaciones Bibliotecologicas	Argentina	Portuguese	business, management and accounting	computer science, engineering
63	*Informacios Tarsadalom*	0.08	SciELO	Infonia	Hungary	Spanish	L&IS	IS&LS
64	*Information – Wissenschaft und Praxis*	0.1	0.023	Walter de Gruyter	Germany	Hungarian	education, computer science	IS&LS
65	*Information and Culture*	–	cc	Univ. Texas Press	USA	German	L&IS	science technology, social sciences

1	2	3	4	5	6	7	8	9
66	*Information and Learning Science*	1.01	0.229	Emerald	United Kingdom	English	L&IS	IS&LS
67	*Information and Management*	5.24	cc	Elsevier	Netherlands	English	L&IS	IS&LS
68	*Information and Organization*	3.15	3.890	Elsevier	United Kingdom	English	communication	computer science
69	*Information Communication and Society*	4.09	1.857	Taylor & Francis	United Kingdom	English	L&IS	IS&LS
70	*Information Design Journal*	0.07	3.084	John Benjamins Publishing Company	Netherlands	English	communication, engineering	History of social sciences
71	*Information Development*	0.71	–	Sage	USA	English	L&IS	IS&LS
72	*Information Discovery and Delivery*	0.21	0.905	Emerald	United Kingdom	English	computer science	IS&LS
73	*Information Processing and Management*	4.23	cc	Elsevier	United Kingdom	English	L&IS	management
74	*Information Research – an International Electronic Journal*	0.84	3.444	University of Borås	Sweden	English	education	IS&LS
75	*Information Resources Management Journal*	0.39	0.762	IGI Global	USA	English	L&IS	management
76	*Information Retrieval*	2.18	cc	Springer	Netherlands	English	education, computer science	communication sociology
77	*Information Services and Use*	0.39	1.488	IOS Press	Netherlands	English	L&IS	–
78	*Information Society*	1.86	–	Taylor & Francis	USA	English	computer science	IS&LS
79	*Information Systems Journal*	4.22	1.889	Wiley–Blackwell	USA	English	L&IS	IS&LS

1	2	3	4	5	6	7	8	9
80	*Information Systems Management*	1.5	4.267	Taylor & Francis	United Kingdom	English	computer science	IS&LS
81	*Information Systems Research*	3.7	1.255	Institute for Operations Research and the Management Sciences	USA	English	L&IS	IS&LS
82	*Information Technology and Libraries*	0.88	2.301	Amer. Library Assoc.	USA	English	computer science	IS&LS
83	*Information Technology and Management*	1.79	0.968	Springer	USA	English	L&IS	computer science
84	*Information Technology and People*	2.35	1.635	Emerald	United Kingdom	English	L&IS	–
85	*Information Technology for Development*	1.66	1.639	Taylor & Francis	United Kingdom	English	music	IS&LS
86	*Informing Science*	1.04	1.387	Informing Science Institute	USA	English	L&IS	IS&LS
87	*Insights*	0.45	–	United Kingdom Serials Group	United Kingdom	English	history	computer science
88	*Interlending & Document Supply*	0.73 (2015)	cc	Emerald	United Kingdom	English	L&IS	IS&LS
89	*International Information and Library Review*	0.24	0.242	Taylor & Francis	USA	English	sociology and political science, law	management
90	*International Journal of Computer–Supported Collaborative Learning*	3.09	cc	Springer	USA	English	L&IS	IS&LS
91	*International Journal of Data Mining and Bioinformatics*	0.74	3.273	Inderscience Enterprises	Switzerland	English	L&IS	IS&LS

1	2	3	4	5	6	7	8	9
92	*International Journal of Geographical Information Science*	3.0	0.652	Taylor & Francis	United Kingdom	English	medicine,	management
93	*International Journal of Information Management*	5.78	2.370	Elsevier	United Kingdom	English	health professions	IS&LS
94	*International Journal of Information Science and Management*	0.26	4.516	Regional Information Center for Science and Technology	Iran	English	L&IS	IS&LS
95	*International Journal of Law and Information Technology*	0.53	–	Oxford University Press	United Kingdom	English	communication, computer science	–
96	*International Journal of Metadata, Semantics and Ontologies*	0.51	cc	Inderscience Enterprises	United Kingdom	English	L&IS	IS&LS
97	*International Journal of Multimedia Information Retrieval*	1.03	–	Springer	Germany	English	communication, cultural studies, linguistics and language	IS&LS
98	*International Journal of the Book*	0.07	cc	Common Ground Research Networks	USA	English	L&IS	science technology, social sciences
99	*International Journal on Digital Libraries*	1.67	–	Springer	Germany	English	computer science	IS&LS
100	*Internet Reference Services Quarterly*	0.89	cc	Taylor & Francis	USA	English	L&IS	education
101	*Investigacion Bibliotecologica*	0.23	–	Univ. Nacional Autonoma Mexico	Mexico	English	–	biology

1	2	3	4	5	6	7	8	9
102	*Issues in Science and Technology Librarianship*	0.36	0.212	Association of College and Research Libraries	USA	Spanish	L&IS	IS&LS
103	*JLIS.it*	0.00	–	Universita di Firenze	Italy	English	communication	geography
104	*Journal of Academic Librarianship*	2.32	cc	Elsevier	USA	English	L&IS	IS&LS
105	*Journal of Access Services*	0.34	1.459	Taylor & Francis	USA	English	computer science	–
106	*Journal of Archival Organization*	0.08	–	Taylor & Francis	USA	English	–	law
107	*Journal of Business and Finance Librarianship*	0.4	–	Taylor & Francis	USA	English	L&IS	–
108	*Journal of Chemical Information and Modeling*	3.9	–	American Chemical Society	USA	English	education, computer science	computer science
109	*Journal of Cheminformatics*	3.98	cc	Chemistry Central	United Kingdom	English	computer science, decision sciences, business, management and accounting	–
110	*Journal of Classification*	2.83	3.893	Springer	Germany	English	L&IS	IS&LS
111	*Journal of Computer–Mediated Communication*	5.97	1.214	Wiley–Blackwell	USA	English	business, management and accounting; computer science	–
112	*Journal of Digital Information Management*	0.24 (2016)	4.000	Digital Information Research Foundation	India	English	L&IS communication	IS&LS

1	2	3	4	5	6	7	8	9
113	*Journal of Documentation*	1.44	–	Emerald	United Kingdom	English	L&IS	–
114	*Journal of Education for Library and Information Science*	0.0	1.157	Association for Library and Information Science Education	USA	English	L&IS	IS&LS
115	*Journal of Educational Media and Library Science*	0.22	cc	Tamkang University	Taiwan	English	L&IS	IS&LS
116	*Journal of Electronic Resources in Medical Libraries*	0.51	–	Taylor & Francis	USA	English	L&IS engineering, decision sciences	–
117	*Journal of Electronic Resources Librarianship*	0.32	–	Taylor & Francis	USA	English	L&IS	–
118	*Journal of Enterprise Information Management*	3.59	–	Emerald	United Kingdom	English	L&IS	–
119	*Journal of Global Information Management*	1.44	2.482	IGI Global	USA	English	business, management and accounting	chemistry, computer science
120	*Journal of Global Information Technology Management*	0.72	0.613	Taylor & Francis	USA	English	L&IS	chemistry, computer science
121	*Journal of Health Communication*	1.97	1.000	Taylor & Francis	USA	English	computer science	mathematics
122	*Journal of Hospital Librarianship*	0.25	1.648	Taylor & Francis	USA	English	L&IS	psychology

1	2	3	4	5	6	7	8	9
123	*Journal of Information and Computational Science*	0.17 (2016)	Medline	Binary Information Press	China	English	computer science	IS&LS
124	*Journal of Information and Knowledge Management*	0.6	–	World Scientific Publishing	USA	English	computer science, political sciences, cultural studies	communication
125	*Journal of Information and Organizational Sciences*	0.55	cc	University of Zagreb	Croatia	English	computer science	–
126	*Journal of Information Ethics*	0.1	cc	McFarland and Company	USA	English	L&IS	IS&LS
127	*Journal of Information Literacy*	0.68	cc	CILIP Information Literacy Group	United Kingdom	English	computer science	IS&LS
128	*Journal of Information Science*	2.09	–	Sage	United Kingdom	English	L&IS	education
129	*Journal of Information Science and Engineering*	0.53	1.939	Academia Sinica	Taiwan	English	decision sciences, computer science	–
130	*Journal of Information Science Theory and Practice*	0.0	0.237	Korea Institute of Science and Technology Information	Korea	English	L&IS	–
131	*Journal of Information Technology*	3.83	–	Palgrave Macmillan	United Kingdom	English	computer science	–
132	*Journal of Information Technology Teaching Cases*	0.21	4.535	Palgrave MacMillan	Switzerland	English	business, management and accounting	IS&LS
133	*Journal of Informetrics*	3.52	–	Elsevier	Netherlands	English	communication	computer science, management

1	2	3	4	5	6	7	8	9
134	*Journal of Interlibrary Loan, Document Delivery and Electronic Reserve*	0.0	3.484	Taylor & Francis	USA	English	computer science	IS&LS
135	*Journal of Knowledge Management*	3.12	–	Emerald	United Kingdom	English	L&IS	IS&LS
136	*Journal of Librarianship and Information Science*	1.2	2.551	Sage	United Kingdom	English	computer science	IS&LS communication
137	*Journal of Library Administration*	0.77	1.098	Taylor & Francis	USA	English	computer science, public administration	none of the results
138	*Journal of Library and Information Services in Distance Learning*	0.42	cc	Taylor & Francis	USA	English	L&IS	–
139	*Journal of Library Metadata*	0.43	–	Taylor & Francis	United Kingdom	English	L&IS	science technology, social sciences
140	*Journal of Management Information Systems*	3.22	–	Taylor & Francis	United Kingdom	English	L&IS	computer science
141	*Journal of Map and Geography Libraries*	0.72	2.744	Taylor & Francis	USA	English	L&IS	IS&LS
142	*Journal of Organizational and End User Computing*	1.47	cc	IGI Global	USA	English	education, computer science	–
143	*Journal of Scholarly Publishing*	0.46	0.744	Univ. Toronto Press	Canada	English	L&IS	IS&LS

1	2	3	4	5	6	7	8	9
144	*Journal of Strategic Information Systems*	3.82	0.447	Elsevier	Nether-lands	English	computer science, biochemi-stry	computer science
145	*Journal of the Ameri-can Medical Informatics Association*	4.11	4.313	Oxford Univ. Press	United Kingdom	English	L&IS	–
146	*Journal of the Asso-ciation for Information Science and Technology*	3.36	4.270	Wiley–Blackwell	USA	English	geography, computer science	IS&LS manage-ment
147	*Journal of the Asso-ciation for Information Systems*	4.14	2.835	Assoc. Informa-tion Systems	USA	English	L&IS	–
148	*Journal of the Australian Library and Information Association (Australian Library Journal)*	0.52	2.839	Taylor & Francis	USA	English	computer science	IS&LS
149	*Journal of the Medical Library Association*	1.14	0.500	Medical Library Assoc.	USA	English	L&IS	–
150	*Journal of Web Libra-rianship*	0.69	1.541	Taylor & Francis	USA	English	decision sciences, business, manage-ment and accounting	IS&LS manage-ment
151	*Knowledge Cultures*	0.00	cc	Addleton Acade-mic Publishers	USA	English	L&IS	IS&LS
152	*Knowledge Manage-ment Re-search and Practice*	1.51	–	Palgrave Mac-millan	United Kingdom	English	law	IS&LS
153	*Knowledge Organiza-tion*	0.57	0.864	Ergon–Verlag	Germany	English	L&IS	–

1	2	3	4	5	6	7	8	9
154	*Language Documentation and Conservation*	0.50	0.59	University of Hawaii Press	USA	English German	computer science	–
155	*Language Resources and Evaluation*	1.15	cc	Springer	Netherlands	English	L&IS	IS&LS management
156	*Law Library Journal*	0.45	0.656	American Association of Law Libraries	USA	English	engineering, computer science	IS&LS
157	*Learned Publishing*	1.12	0.583	Wiley–Blackwell	USA	English	L&IS	IS&LS management
158	*Lecture Notes in Control and Information Sciences*	0.36	1.632	Springer	USA	English	history, literature and literary theory, communication	IS&LS
159	*Legal Reference Services Quarterly*	0.21	cc	Taylor & Francis	USA	English	L&IS	IS&LS management
160	*LIBER Quarterly*	0.73	–	Association of European Research Libraries	Netherlands	English	L&IS	IS&LS
161	*Library*	0.41	–	Oxford University Press	United Kingdom	English	L&IS	IS&LS
162	*Library and Information Science*	0.07	cc	Mita Soc. Library Information Science	Japan	English	L&IS engineering	IS&LS
163	*Library and Information Science Research*	1.7	0.300	Elsevier	USA	English Japanese	L&IS conservation, computer science	IS&LS
164	*Library Collections, Acquisition and Technical Services*	0.25	1.372	Taylor & Francis	United Kingdom	English	L&IS	IS&LS
165	*Library Hi Tech*	0.9	0.333	Emerald	United Kingdom	English	education	IS&LS
166	*Library Hi Tech News*	0.33	0.759	Emerald	United Kingdom	English	L&IS	–
167	*Library Journal*	0.02	–	Reed Business Information	USA	English	L&IS	IS&LS management

1	2	3	4	5	6	7	8	9
168	*Library Leadership and Management*	0.23	0.458	American Library Association	USA	English	L&IS	IS&LS
169	*Library Management*	0.76	–	Emerald	United Kingdom	English	business, management and accounting	language
170	*Library Philosophy and Practice*	0.33	cc	University of Idaho Library	USA	English	L&IS	linguistics
171	*Library Quarterly*	1.02	–	Univ. Chicago Press	USA	English	chemistry, computer science	computer science
172	*Library Resources and Technical Services*	0.43	0.913	Amer. Library Assoc.	USA	English	L&IS	IS&LS
173	*Library Review*	0.94	0.657	Emerald	United Kingdom	English	chemistry, computer science	law
174	*Library Trends*	0.4	cc	Johns Hopkins Univ. Press	USA	English	L&IS mathematics decision sciences, psychology	IS&LS
175	*Libres*	0.46	0.474	Curtin University of Technology	Australia	English	computer science	computer science, engineering
176	*Libri*	0.52	cc	Walter De Gruyter	Germany	English	L&IS	–
177	*Logos*	0.08	0.500	Brill	Netherlands	English German	computer science, business, management and accounting	–
178	*Malaysian Journal of Library and Information Science*	0.6	cc	Univ. Malaya	Malaysia	English	L&IS	humanities multidisciplinary
179	*Manuscripta Orientalia*	0.33	0.425	Thesa Publishers	Russia	English	computer science	IS&LS

1	2	3	4	5	6	7	8	9
180	*Masaryk University Journal of Law and Technology*	0.17	–	Masaryk University	Czech Republic	English	L&IS	IS&LS
181	*Medical Reference Services Quarterly*	0.79	–	Taylor & Francis	USA	English	education	IS&LS
182	*Methis*	0.00	Medline	University of Tartu Press	Estonia	English	L&IS conservation, archeology, computer science	IS&LS
183	*MIS Quarterly*	8.33	–	Univ. Minnesota	USA	English	L&IS	–
184	*MIS Quarterly Executive*	1.6	5.430	Indiana Univ.	USA	German	health	IS&LS
185	*Music Reference Services Quarterly*	0.18	1.862	Taylor & Francis	USA	Russian	L&IS	–
186	*New Review of Academic Librarianship*	1.18	–	Taylor & Francis	United Kingdom	Estonian	computer science	IS&LS
187	*Notes*	0.14	–	Music Library Association	USA	English	computer science. decision sciences, business management and accounting	–
188	*Notes and Queries*	0.04	cc	Oxford University Press	United Kingdom	English	business management and accounting,	IS&LS
189	*Online Information Review*	2.01	cc	Emerald	United Kingdom	English	decision sciences,	IS&LS
190	*Pakistan Journal of Information Management and Libraries*	0.19	1.675	University of the Punjab	Pakistan	English	computer sciences	IS&LS

1	2	3	4	5	6	7	8	9
191	*Papers of the Bibliographical Society of America*	0.11	–	Bibliographical Society of America	USA	English	computer science, decision sciences	IS&LS
192	*Performance Measurement and Metrics*	0.51	cc	Emerald	United Kingdom	English	L&IS communication, public health	IS&LS
193	*Perspectivas em Ciencia da Informacao*	0.24	cc	Escola de Ciencia da Informacao da UFMG	Brazil	English	L&IS	IS&LS
194	*Portal: Libraries and the Academy*	1.31	cc	Johns Hopkins Univ. Press	USA	English	medicine	humanities multidisciplinary
195	*Preservation, Digital Technology and Culture*	0.02	1.473	Walter de Gruyter	Germany	English	L&IS	IS&LS
196	*Proceedings of the Association for Information Science and Technology*	0.46	–	John Wiley and Sons	USA	English	computer science	–
197	*Profesional de la Informacion*	1.17	–	El Profesional de la Informacio	Spain	Portuguese	L&IS	–
198	*Program – Electronic Library and Information Systems*	1.30	1.318	Emerald	United Kingdom	English	computer science	science technology, social sciences
199	*Prologue*	0.02	1.170	National Archives and Records Administration	USA	English	L&IS	–
200	*Public Library Quarterly*	0.43	cc	Taylor & Francis	USA	English	computer science	IS&LS management
201	*Public Services Quarterly*	0.34	cc	Taylor & Francis	USA	Spanish	L&IS	IS&LS management
202	*Qualitative Health Research*	2.22	–	Sage	USA	English	philosophy	–

1	2	3	4	5	6	7	8	9
203	*Records Management Journal*	1.18	2.413	Emerald	United Kingdom	English	L&IS	–
204	*Reference and User Services Quarterly*	0.42	cc	Amer. Library Assoc.	USA	English	L&IS	music
205	*Reference Librarian*	0.58	0.377	Taylor & Francis	USA	English	computer science	literature
206	*Reference Services Review*	1.2	–	Emerald	USA	English	L&IS	IS&LS
207	*Research Evaluation*	2.79	cc	Oxford Univ. Press	United Kingdom	English	computer science	–
208	*Restaurator – International Journal for The Preservation of Library and Archival Material*	0.29	2.449	Walter De Gruyter	Germany	English	L&IS	humanities multidisciplinary
209	*Revista Cubana de Informacion en Ciencias de la Salud*	0.29	0.344	Centro Nacional De Informacion De Ciencias Medicas	Cuba	English	computer science,	IS&LS
210	*Revista Espanola de Documentacion Cientifica*	0.83	SciELO	Consejo Superior Investigaciones Cientificas	Spain	English	decision sciences	IS&LS
211	*Revista General de Informacion y Documentacion*	0.23	0.632	Universidad Complutense de Madrid	Spain	English	L&IS	IS&LS
212	*Revue Francaise d'Histoire du Livre*	0.00	cc	Librairie Droz SA	France	English	business, management and accounting; computer science	–
213	*School Library Media Research*	0.27 (2016)	–	American Library Association	USA	English	L&IS	–
214	*Science and Technology Libraries*	0.58	–	Taylor & Francis	USA	Spanish	education	IS&LS

1	2	3	4	5	6	7	8	9
215	*Scientific Data*	6.08	cc	Nature Publishing Group	United Kingdom	Spanish	L&IS	IS&LS
216	*Scientist*	0.03	5.305	Labx Media Group	Canada	Spanish	computer science	computer science
217	*Scientometrics*	2.72	0.537	Springer	Netherlands	French	L&IS	history
218	*Scire*	0.09	2.173	Universidad de Zaragoza	Spain	English	business, management and accounting	IS&LS
219	*Script and Print*	0.17	cc	Australian and New Zealand Student Services Association	Australia	English	L&IS	–
220	*Scriptorium*	0.16	–	Centre d'Etude des Manuscrits	Belgium	English	L&IS	IS&LS Interdisciplinary biomedical
221	*Serials Librarian*	0.42	cc	Taylor & Francis	USA	English	public administration	IS&LS
222	*Serials Review*	0.35	cc	Taylor & Francis	United Kingdom	English	L&IS	IS&LS
223	*Slavic and East European Information Resources*	0.07	0.310	Taylor & Francis	USA	Spanish Portuguese	L&IS	–
224	*Social Science Computer Review*	2.96	–	Sage	USA	English	business, management and accounting, decision sciences, computer sciences	IS&LS
225	*Social Science Information*	0.52	3.253	Sage	United Kingdom	French German Spanish Italian	L&IS	IS&LS
226	*Technical Services Quarterly*	0.12	0.571	Taylor & Francis	USA	English	business, management and accounting,	IS&LS
227	*Telecommunications Policy*	2.14	cc	Elsevier	United Kingdom	English	computer sciences	science technology

1	2	3	4	5	6	7	8	9
228	*Telematics and Informatics*	4.33	2.087	Elsevier	Netherlands	English	education,	IS&LS
229	*Terminology*	0.42	3.789	John Benjamins Publishing Company	Netherlands	English	engineering	IS&LS
230	*Transactions of the Cambridge Bibliographical Society*	0.0	0.389	Cambridge Bibliographical Society	United Kingdom	English French	business, management and accounting,	–
231	*Transinformacao*	0.33	–	Pontificia Universidade Catolica Campinas	Brazil	English	computer sciences, decision sciences	–
232	*Tuna*	0.00	0.255	Eesti Arhivaaride Uhing	Estonia	English	medicine	IS&LS
233	*VINE Journal of Information and Knowledge Management Systems*	1.27	cc	Emerald	United Kingdom	English	L&IS	multidisciplinary
234	*Vjesnik Bibliotekara Hrvatske*	0.15	cc	Hrvatsko Knjiznicarsko Drustvo	Croatia	English French Spanish	computer sciences	IS&LS multidisciplinary
235	*VOEB–Mitteilungen*	0.05	–	Universitätsbibliothek Graz	Austria	French	computer science	IS&LS
236	*Weblogy*	0.77	–	University of Aix–Marseille	France	Portuguese	L&IS	
237	*World Patent Information*	0.88	–	Elsevier	United Kingdom	English	L&IS	IS&LS
238	*Zeitschrift für Bibliothekswesen und Bibliographie*	0.08	cc	Vittorio Klostermann	Germany	Estonian	medicine	

References

Abdulhayoglu, M.A., Thijs, B (2018). Use of Locality Sensitive Hashing (LSH) Algorithm to Match Web of Science and Scopus, *Scientometrics*, 116(2), 1229–1245.

Abrizah, A., Zainab, A.N., Kiran, K., Raj, R.G. (2013). LIS Journals Scientific Impact and Subject Categorization: A Comparison Between Web of Science and Scopus, *Scientometrics*. 94(2), 721–740.

Ajiferuke, I., Wolfram, D. (2010). Citer Analysis as a Measure of Research Impact: Library and Information Science as a Case Study. *Scientometrics*. 83(3), 623–638.

Archambault, É., Campbell, D., Gingras, Y., Larivière, V. (2009). Comparing Bibliometric Statistics Obtained from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology*, 60(7), 1320–1326.

Chang, Y-W. (2018). Examining Interdisciplinarity of Library and Information Science (LIS) Based on LIS Articles Contributed by Non-LIS Authors. *Scientometrics*, 116(3), 1589–1613.

Chavarro, D., Rafols, I., Tang, P. (2018). To What Extent Is Inclusion in the Web of Science an Indicator of Journal 'Guality'? *Research Evaluation*, 27(2), 106–118.

Figuerola, C.G., García, Marco, F.J., Pinto, M. (2017). Mapping the Evolution of Library and Information Science (1978–2014) Using Topic Modeling on LISA. *Scientometrics*, 112(3), 1507–1535.

Franceschini, F., Maisano, D., Mastrogiacomo, L. (2016). Empirical Analysis and Classification of Database Errors in Scopus and Web of Science. *Journal of Informetrics*, 10(4), 933–953.

Hammarfelt, B., Rushforth, A.D. (2017). Indicators as Judgment Devices: An Empirical Study of Citizen Bibliometrics in Research Evaluation. *Research Evaluation*, 26(3), 169–180.

Harzing, A-W, Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A Longitudinal and Cross-Disciplinary Comparison. *Scientometrics*, 106(2), 787–804.

Komunikat MNSW (2019). *Komunikat Ministra Nauki i Szkolnictwa Wyższego z dnia 31 lipca 2019 r. w sprawie wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych wraz z przypisaną liczbą punktów* [online]. Ministerstwo Nauki i Szkolnictwa Wyższego [23.10.2019], http://www.bip.nauka.gov.pl/akty-prawne-mnisw/komunikat-ministra-nauki-i-szkolnictwa-wyzszego-z-dnia-31-lipca-2019-r-w-sprawie-wykazu-czasopism-naukowych-i-recenzowanych-materialow-z-konferencji-miedzynarodowych-wraz-z-przypisana-liczba-punktow.html

Manana-Rodriguez, J. (2015). A Critical Review of SCImago Journal & Country Rank. *Research Evaluation*, 24(4), 343–354.

Martin-Martin, A., Orduna-Malea, E., Lopez-Cozar, E.D. (2018). Coverage of Highly-Cited Documents in Google Scholar, Web of Science, and Scopus: A Multidisciplinary Comparison. *Scientometrics*, 116(3), 2174–2188.

Meho L. I., Yang K. (2007). Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125.

Milojević, S., Sugimoto, C., Yan, E., Ding, Y. (2011). The Cognitive Structure of Library and Information Science: Analysis of Article Title Words. *Journal of the American Society for Information Science and Technology*, 62(10), 1933–1953.

Mongeon, P., Paul-Hus, A. (2016). The journal Coverage of Web of Science and Scopus: A Comparative Analysis. *Scientometrics*, 106(1), 213–228.

Prins, A.A.M, Costas, R., van Leeuwen, T.N., Wouters, P.F. (2016). Using Google Scholar in Research Evaluation of Humanities and Social Science Programs: A Comparison with Web of Science Data. *Research Evaluation*, 25(3), 264–270.

Sosińska-Kalata, B. (2013). Obszary badań współczesnej informatologii (nauki o informacji). *Zagadnienia Informacji Naukowej – Studia Informacyjne*, 51(2), 9–41.

Wang, Q., Waltman, L. (2016). Large-Scale Analysis of the Accuracy of the Journal Classification Systems of Web of Science and Scopus. *Journal of Informetrics*, 10(2), 347–364.

Walters, W.H., Wilder, E.I. (2016). Disciplinary, National, and Departmental Contributions to the Literature of Library and Information Science, 2007–2012. *Journal of the Association for Information Science and Technology*, 67(6), 1487–1506.

Vieira, E.S., Gomes, J.A.N.F. (2009). A Comparison of Scopus and Web of Science for a Typical University. *Scientometrics*, 81(2). 587–600.

Vieira, E.S., Gomes, J.A.N.F. (2016). The Bibliometric Indicators as Predictors of the Final Decision of the Peer Review. *Research Evaluation*, 25(2), 170–183.

Przydatność danych pochodzących z baz Web of Science i Scopus do analizowania stanu dyscypliny naukowej. Przypadek dyscypliny library and information science

Abstrakt

Cel/Teza: Wykorzystywanie baz bibliograficznych do analizowania i diagnozowania stanu nauki coraz częściej stanowi stały element polityki naukowej wielu państw. Dotychczasowe badania przydatności danych z baz Web of Science i Scopus do tego celu nie dały jednoznacznych wyników. Ich autorzy nie zawsze odnosili się do ważnej kwestii – jakości danych pochodzących ze wspomnianych baz. Celem artykułu jest analiza jakości danych pobieranych w sposób zautomatyzowany z zasobów wymienionych baz.

Koncepcja/Metody badań: Autor posłużył się metodą jakościowej weryfikacji danych polegającą na początkowo zautomatyzowanym pobraniu danych o czasopismach z baz Web of Science i Scopus, a następnie na poddaniu ich analizie jakościowej. Analiza ta polegała na: wzajemnej konfrontacji danych o czasopismach reprezentujących Library and Information Science pobranych z obu baz; skonfrontowaniu danych ilościowych pobranych z badanych baz z danymi pochodzącymi z innych, tematycznych baz danych bibliograficznych; porównaniu pobranych danych z informacjami dostępnymi na stronach WWW indeksowanych czasopism oraz na skonfrontowaniu przyporządkowywania czasopism, artykułów i autorów do dyscyplin naukowych, stosowanego przez redakcje wspomnianych baz, z przyjętym przez badaczy zakresem tematycznym dyscypliny Library and Information Science.

Wyniki i wnioski: Ustalono, że w przypadku badanej dyscypliny automatyczne pobieranie danych stwarza ryzyko uzyskania zbioru o niskiej wiarygodności. Najwięcej problemów stwarza niski poziom kompletności danych oraz błędy w kategoryzowaniu czasopism, artykułów i autorów.

Oryginalność/Wartość poznawcza: Wykazano, że wbrew twierdzeniom decydentów polskiej nauki, w obecnym kształcie badane bazy bibliograficzne jedynie w niewielkim stopniu przydatne są do monitorowania stanu i tendencji rozwojowych badanej dyscypliny naukowej. Wykazane w niniejszym artykule problemy metodyczne stwarzane przez obie bazy mogą rzutować także na generowanie rzetelnego i obiektywnego obrazu innych dyscyplin naukowych. Zmiany w obszarze funkcjonowania WoS i Scopus, obserwowane od kilku lat, nie rozwiązały istniejących już wcześniej problemów i niedogodności.

Słowa kluczowe

Analiza ilościowa. Analiza jakościowa. Dane bibliograficzne. Dyscyplina naukowa. Scopus. Web of Science.

Dr habil. ZBIGNIEW OSIŃSKI is Professor at the Department of Digital Humanities at the Maria Curie-Skłodowska University Lublin. He specializes in information science and digital humanities. His most important publications include: M. Górny, M. Kisilowska, E. Głowacka, Z. Osiński: Mechanisms of the formation and evolution of personal information spaces in the humanities (Poznań, 2017); M. Kowalska, V. Osińska, Z. Osiński: The Role of Visualization in the Shaping and Exploration of the Individual Information Space (Knowledge Organization*, 2018); Z. Osiński: Information infrastructure of contemporary humanities and the digital humanities development as a cause of creating new information barriers. A Polish case (*Digital Scholarship in the Humanities*, 2019).*

Contact to the Author:
zbigniew.osinski@gmail.com
Department of Digital Humanities
Faculty of Humanities
Uniwersytet Marii Curie-Skłodowskiej
pl. Marii Curie-Skłodowskiej 4

The Evolution of Researchers' Bibliography: From Systematic Organization to Citation

Viviane Couzinet
ORCID 0000-0003-4807-2014
LERASS, University of Toulouse III Paul Sabatier – IUT, France

Regina Marteleto
ORCID 0000-0002-3439-0217
Brazilian Institute of Information in Science and Technology, Rio de Janerio, Brazil

Icléia Thiesen
ORCID 0000-0002-3137-8933
Federal University of the State of Rio de Janeiro, Brazil

Abstract

Purpose/Thesis: This paper analyses the publications of two authors from different countries through a number of citations identified by Google Scholar. Jean Meyriat and Edson Nery da Fonseca are outstanding researchers in France and Brazil, respectively, both contributing to the formation and advancement of information science in their countries. The "traditional" bibliography and the development trajectory of their scientific achievements were considered in this analysis.

Approach/Methods: The analysis of the researchers' publications identified by Google Scholar based on their names helps to understand the role that search engines play in the evaluation of science and its effect on information seeking and in the evaluation of the scientific production. The qualitative research is based on a bibliological analyses focusing on the way in which the written production of the selected authors is reported and highlighted. The carried out study is exploratory in nature, so it proposes to raise questions and to emit hypotheses that could suggest directions for further research.

Results and conclusions: The uniformity of controlled metadata used in traditional bibliographies make it possible to follow the progress of an author's thought. The same is true of other activities he has performed in and out of the field and, it is also true for what has been produced about this author. Searching for information in Google Scholar can show a researcher's production and publications about her/him in the order of citations but not in the order of the progress of the researcher's scientific achievements or development of research focused on these achievements. Placed in the context of the analysis of the scientific field as it was defined by Bourdieu (1976), the observation of the visibility of the production Jean Meyriat and Edson Nery da Fonseca, two major actors in information science in France and Brazil , by Google Scholar, shows a tendency to underemphasize the role they played.

Practical implications: The functioning of the scientific field has specific effects on the production of research. The search engines such as Google Scholar, made the research output more accessible on Internet. However, the criteria they prioritize, i.e., the number of citations, ignores important aspects of a scientific career. Qualitative analysis that considers the networks of sociability and the development trajectory of researchers' scientific achievements may reveal relevant contributions to their scientific field.

Originality/Value: The results reveal the importance of research methods that combine qualitative and quantitative procedures, thus avoiding the risks of rendering scientific careers invisible.

Keywords

Bibliography. Edson Nery da Fonseca. Google Scholar. Information science. Jean Meyriat. Scientific field.

Received: 5 May 2019. Reviewed: 13 July 2019. Revised: 13 September 2019. Accepted: 6 December 2019.

1. Introduction

It is commonplace knowledge that bibliography has been used to assess an author's scientific production since the 2nd century AD when Galen published *De libris propiis*. The rules of bibliographic writing have evolved over with an aim to identify sources of information as to make them more available for sharing. For this reason, the work of French scholars and librarians is different from the work associated with the book science tradition, as well as more generally, with the written text science and bibliology. Both specialized and general bibliographies meet the requirements of an internal organization which provides access to the needed documents and defines which subject areas are covered by the given bibliography.

A bibliography which records a single researcher's production might be organized with a specific goal in mind, e.g. it may highlight his or her expertise in the subject, or association with publications recognizable in the researcher's community. However, search engines which have taken over printed bibliographies, organize their records according to different measures of visibility, such as the number of citations. The object of a search is also the basis for the evaluation of the researcher's production. At this point, it is worthwhile to ask what view these engines afford us, and how do they change scientific practices.

To answers these questions, we examine the modalities of the functioning of the scientific field (Bourdieu, 1976), and define the bibliography and the systematic knowledge organization it proposes (Meyriat, 1957; 1993). The analysis of the record of researches' publications provided by Google Scholar will show us the role that search engines play in the evaluation of scientific production and how they change both the process of evaluation and of information seeking.

2. Literature review

2.1. Scientific field

The notion of «scientific field» was developed by Pierre Bourdieu. Bourdieu argued that to understand an intellectual production, analysis must not be limited to the content and social context of production (Marteleto & Carvalho, 2015). It must take into account the intermediate universe into which the agents (as Bourdieu named actors in the field) are incorporated. This universe is a social and institutional world that obeys its own laws (Bourdieu, 1997). It is a microcosm with a certain autonomy but it does not escape the macrocosm which surrounds it. To understand the processes of information production and dissemination in a field of knowledge, or a scientific field, it is then necessary to take into consideration the common language, classifications and agents that are constitute an

epistemic community, as well as their links with the outside world and with other social fields (Nascimento & Marteleto, 2007).

The scientific field is a place of struggle for scientific domination. It shares fundamental concepts, methods, verification procedures and tacit norms which comprise a «scientific habitus». The scientific authority defined as a technical capacity and a social power is inseparable from the recognition given by peers, who are also competitors. This assumes that the recognition of value, i.e., prestige, scientific authority, depends on other agents who, in order to grant it, will examine and discuss the production. Thus, groups in contact draw their legitimacy from their strength, the perpetuation of which is an important aspect. Thus achieved, the scientific authority is a symbolic capital (Bourdieu, 1976).

There are two kinds of capital in the scientific field: institutional and institutionalized. The former refers to the eminent positions and the power over the means of production; the latter refers to the personal prestige obtained by the contributions to the progress of science (Bourdieu, 1997). Productivity and prestige are part of a trajectory constituted by an academic origin (high school, prestigious university), which fosters academic aspirations and provides opportunities for development in that direction. The social capital is also acquired through scientific output, particularly in scientific journals evaluated by peer-competitors (Bourdieu, 1976). Therefore, an analysis of the evolution of a researcher's production should begin with a record which identifies it: a bibliography.

2.2. Bibliography

The word bibliography has two meanings. It may refer to a list recording the written production of a particular person, country, time period or theme, or to the activity of developing the notes which constitute such a list. Various types of bibliography in the first sense have been distinguished according to their geographic coverage – international or national, to their theme, the field covered by their content or by their selective or exhaustive elaboration method (Malclès, 1984). Although for a long time it was reserved only for books, since then it has been extended to all forms of writing (Meyriat, 1993) and its role in knowledge circulation has been emphasized (Couzinet & Fraysse, 2018; Malclès 1984; Meyriat, 1993; Provansal, 1997). In France, the production of bibliographies constitutes a separate profession, and it is considered as an art. Meyriat (1993) and Estivals (1993) introduced an understanding of the "science of writing", i.e. bibliology, with bibliography as one of its methods. However the separation between bibliology and bibliography is not always maintained (McKenzie, 1991), with both bibliography and bibliology occasionally considered as science.

Bibliographical notes record data about the author, the title, the source, the date and the pagination of the text. They may be supplemented by an abstract summary or other comments. Bibliography is presented in the alphabetical order by the authors' last names, or by themes. Indexes multiply the possible entries. Bibliographies are considered to be the ancestors of data bases. The computer has facilitated the multiplication of the entries while preserving the quality of the data. Bibliographical records facilitate organization of knowledge and thus support the researcher's work.

2.3. Research evaluation

A researcher's scientific production as recorded by a bibliography goes through a critical evaluation of the peer-competitors before being accepted by a mediator (Bourdieu 1976). "Pure" scientific capital is acquired mainly through recognized contributions to the progress of science, inventions, or discoveries" (Bourdieu, 1997, 29), circulating in the scientific field through media. It is then possible to establish the list of publications for each researcher to disseminate his personal contribution. The tools for the circulation of science, the researchers' bibliographies, are therefore also a means to position a researcher on a scale of production and thus to evaluate it quantitatively. In this way, a «capital of authority» (Bourdieu, 1976) is built.

Nevertheless, the functioning of the scientific field affects the way the research is conducted. This is what Merton called "the Matthew effect in science", which refers to a phenomenon where the greater the scientific capital is, the more the scientific production is accepted by the peers (Merton, 1968). This is particularly relevant to ordering articles in journals to produce syntheses. The authority of the requested author contributes to the journal's reputation.

Currently, researchers' production circulates on the Internet through search engines. How do they make the researchers' production visible? To answer this question, we examined the results of searches conducted using Google Scholar.

3. Research methods

3.1. Google Scholar

Google Scholar was launched in November 2004 by Google as a search service for research journal articles (mainly but not exclusively peer-reviewed), scientific books and academic theses. It also indexes conference proceedings, guides, reports and other texts researchers may produce. It is possible to choose the date, sort by relevance, to include patents and quotes, and to create an alert. The results provide the title, the author's first and last name (sometimes confusing the two), the number of citations received, with an option to access the document which cites the original text. Furthermore, it is possible to record the data in a personal library, to write the following reference to specific standards and to access publications that Google assumes are of interest. A link to the original text, if it has been deposited in an archive, is present. Often a short extract from the summary accompanies the record. A notice may indicate a change of language or an editorial unavailability. In this it is difficult to establish from where the enumerated data is coming (Rovira et al., 2018), even if clear agreements have been reached with libraries.

The records are ranked according to the number of citations received. The tests we have conducted show many errors, which suggest that the ranking is untrustworthy. The top results do not give a comprehensive view of the author's scientific output.

A progress report has been prepared by Martin-Martin et al. (2017), who in a recent research based on a bibliometric analysis conducted on 64 000 records, questioned the possibility of using Google Scholar for identifying the most cited documents. They showed

that the publication date of the documents cited has a low impact on the classification and that the language of the documents does not affect the classification either (Martin-Martin et al., 2017).

Google Scholar is well-adapted to our research as the variety of the documents it indexes corresponds to the various forms of publications by the authors' whose research output is analyzed. It should be stressed that in the time when they published their most influential works, there were only few journals in information science in France and Brazil.

For the purposes of this study, the observation consisted of searching by author's name in French and Portuguese records from any time period, sorted according to their relevance. In order to make the analyzed problems more significant, we have chosen two scholars as well as practitioners who contributed to development of information science in France and in Brazil. Each has a well-established international reputation. This research work has been undertaken within a Franco-Brazilian cooperation network Mediations and Social Use of Knowledge and Information (MUSSI).

3.2. A bibliological analysis

The research is based on a bibliological analysis. It analyzes the way in which the written production of the selected authors is reported and highlighted. The approach is comparative, as it takes into account the representation modes of the bibliographic lists described as "traditional" – the printed bibliographies and the bibliographical databases and the data available through Google Scholar. The study we have carried out is exploratory in nature: it raises questions and suggests hypotheses that could inspire further research.

This research is qualitative. The context of the authors' scientific production is a point emphasized in the analysis, our initial hypothesis being that although science is increasingly globalized, a researcher who has only a limited geographical recognition can nevertheless play a major role in the development of a science because of other factors, which we take into consideration. Our central question is whether such research needs to be continued.

3.3. Selected authors

Jean Meyriat (1921–2010), a former student of the Ecole Normale Supérieure, had been a professor at the Insitute d'Études Politiques (IEP) and Director of Studies at the Ecole Pratique des Hautes Études (EPHE) in Paris. Since the 1950s he had been working with the UNESCO on documentation issues in his capacity as Director of Documentation Services at the Fondation Nationale des Sciences Politiques (FNSP). Within the Center, which under his leadership became a center for library and documentation, he organized production of many bibliographic tools drawn from journals articles, which at the time had been an unprecedented practice. These bibliographic tools covered the major disciplines at the IEP, international law and comparative law. In 1952, he co-founded the first FNSP research laboratory, the Center for the Study of the International Relations Studies, which operates to this day attached to the National Scientific Research Center (Fr. *Centre national de la recherche scientifique* – CNRS). From 1961 to 1965, Meyriat chaired the Mediterranean Social Science Research Council.

In addition to his documentation-oriented activities, Meyriat was a lecturer at the Institute of Advanced Studies in Latin America, where he directed many doctoral theses on

the Latin American politics. He had been a visiting professor at the Colegio de México. He was convinced that library and documentation activities must be founded on research, as the techniques used by the information professionals "depend, for their dynamism and thriving and acquirement of a fundamental knowledge" (Meyriat, 1994, 42). Accordingly, he added theoretical reflection to this other achievements, and taught a course "Theory of Documentation" at the request of the historian Fernand Braudel, from 1962.

His work focused on the development of a social theory of the document, consolidation of the vocabulary of this new discipline in France, establishment of its research methods, and supervising of multiple doctoral theses. His contributions to the field of law, Latin America studies and information science were tremendous. He gave priority to the latter from 1975–1980 onwards by leading a research group on the written word and document (1979–1982) with Robert Estivals, then a young doctor working on book history and a researcher associate at the CNRS. The work they had undertaken gave rise to a major text, *Document, documentation, documentology* (Meyriat, 1981), which marked France's entry into information science. This seminal text, which serves as a reference for all the French social approaches to the document, has recently been translated into Portuguese by Brazilian colleagues (see Meyriat, 2016). A part of Jean Meyriat's writing has been re-issued with his biography and testimonials of people who worked with him, and a research study, which highlighted his action or developed his reflection, (Couzinet & Rauzier, 2001). According to the bibliographical record from the book, Meyriat had published 127 texts (39 journal articles, 22 contributions in conference proceedings, eight book chapters, six professional guides, 48 articles in the *International Encyclopedia of Bibliology*, four articles in the *Encyclopedia of Information and Documentation*). In addition, he had produced standards, courses for the Commission of the European Communities, reports for various academic bodies or for UNESCO, training programs and extensive bibliographic work in political science. He had been the founder of several repertories and bibliographic journals in this discipline. He was also responsible for the creation and animation of several scientific societies, including the International Association of Schools of Information Science (AIESI). Its international action has been recognized by UNESCO which awarded him with the Aristotle Gold Medal (1992) for its contribution to international scientific cooperation and by Professor PN's Foundation Kaula (1992) who awarded him with the gold medal as a recognized world-wide scholar for his entire work in information science.

Other publications were found after the publication of *Jean Meyriat, théoricien et praticien de l'information-documentation* (Couzinet & Rauzier, 2011) and Meyriat continued to publish until his death in 2010. In particular, an article published in the journal *Sciences de la société* extended his reflection on the document by proposing a new analytical method (Meyriat, 2006).

Various meetings had marked his career, especially the one with Robert Escarpit and Robert Estivals as they were preparing a collective work *Le Livre français* (Cain & Martin, 1972)[1]. Together, they began a scientific and institutional construction of an academic

[1] This book was edited by Julien Cain, a historian, director of the Bibliothèque nationale de France and by Henri-Jean Martin, also a historian. It was published in 1972 by the National Printing Services. Jean Meyriat whose activity and theoretical reflection had then centered on journals is the author of a chapter devoted to the "non books" (319–331).

discipline named information and communication sciences, combining the information science and the science of communication.

Meyriat had been involved in the institutionalization of information and communication sciences (in 1975) and had created training courses from undergraduate to doctoral level (Couzinet, 2017). In particular, he chaired the National Council of Universities (Fr. *Conseil national des universités* – CNU), a body that developed an understanding of the "science of writing", i.e. bibliology, with bibliography as one of its methods, defined the contours and the criteria of membership of the discipline, managed the careers of academics in the discipline and awards qualifications, i.e. the possibility of applying for a post in higher education after the doctorate. Moreover, he established links between university education and the professional bodies (Couzinet, 2000; 2017). He is considered to have been a major theoretician in IS and documentation in France.

Information science in Brazil derived from a broad sociotechnical network of multiple institutional actors in the cross-conformation of the discipline. This network consisted of researchers and information professionals, but also of publications, meetings, agency programs, institutional policies of state agencies. As a teacher, writer and librarian, Edson Nery da Fonseca (1921–2014) had been an important node of this network.

He graduated as a librarian at Biblioteca National do Brasil in Rio de Janeiro in 1946. Devoted to the teaching of "disciplina do livro", he was interested in the training of an interdisciplinary librarian with technical knowledge and a cultural base which he considered essential. During his long career, he worked in various institutions, such as the Brazilian Institute of Bibliography and Documentation (now IBICT – Por. *Instituto Brasileiro de Informação en Ciência e Tecnologia*), where he headed the bibliography department. At the new University of Brasilia, he was a professor and the founder of library science course and the founder of the Central Library. He witnessed the emergence of the information science in the early 1970s, when a master-research level training was created at IBICT, as well as the establishing of the journal *Ciência da Informação*, which consolidated the emerging field (Thiesen, 2010). He was a great critic of library practices, which he said ignored the advances and transformations of scientific practices; he argued that one

> must consider library science as one of the sciences of information and consider all knowledge from an interdisciplinary a point of view (Fonseca, 1988, 102)[2].

In addition to being a librarian and a bibliographer, a teacher and a columnist, Fonseca was an intellectual. He belonged to a network of intellectuals from different institutions and of different profiles. His production, however, was not limited to the field of library science, documentation, and information science. His intellectual curiosity had enabled him to be involved in a network of sociability made up of writers, filmmakers, journalists, academics and had inspired him to activities beyond libraries and universities. Several authors contributed to the book published on his eightieth birthday, edited by Mota and Verri, including the eminent sociologist Gilberto Freyre who entitled his tribute *A Master of Masters* (Mota & Verri, 2001).

Fonseca's professional experience was further enriched by his work at the National Book Institute (Por. *Instituto Nacional do Livro* – INL), the Brazilian Association of Technical

[2] Own translation [IT].

Standards (Por. *Associação Brasileira de Normas Técnicas* – ABNT) and the Joaquim Nabuco Foundation in Recife. However, he never forgot his relationship with the northeast region, where he established libraries and found library courses.

The primary themes of Fonseca's works are bibliography and its precursors, documentation, teaching history of librarianship and scientific information. His long public life had allowed the author to witness many changes in these domains of knowledge which he had influenced over fifty years he had been active.

Fonseca published a number of works between 1942 and 2001, according to a repertoire compiled by the librarians from Pernambuco State, Cordelia Cavalcanti and Lucia Gaspar and presented by Mota and Verri (2001). Their book showcases the extent and variety of Fonesca's output, as it refers to 15 books, 20 booklets, 73 collaborations in individual and collective works, eight entries in consultation papers, 23 prefaces, afterwords, presentations and introductions, 198 articles, bibliographies and magazines, 420 newspaper collaborations, four publishers, 20 individual and collective works organizations, seven interviews, four discographies, three participations in filmography, and 103 items published about Fonesca.

We are therefore dealing with two quite different figures, both of whom had made major contributions to the evolution of training and research in their respective countries.

The analysis of the production of Meyriat and Fonseca in Google Scholar was conducted according to the criteria stated above. The order of the retrieved bibliographic descriptions is random. For example, the most cited text of Meyriat, which is a major contribution to the information science, appears only in the 8th position on the days of the search. We checked the first pages of the results provided by Google Scholar between April 16th 2019 and April 30th, and on May 1st 2019. The observations made raised some questions.

4. Results and discussion

4.1. Scientific controversies and fashion effects

Unlike the bibliography in its old form, which exhaustively or selectively lists the production of a researcher in a chronological order which enables the reader to follow the progression of the work on a given theme, Google Scholar gives priority to the works which are cited by others. This preference is problematic. The reference is isolated from its citation context: we may ask then if it is a "true" reference, i.e. bearing a call to an author of which one recognizes the expertise or has produced a remarkable breakthrough in a field, is it a negative criticism, that is to say a rejection of the results based on a method equally questionable? Is it a "recognition quotation" of an author citing or the desire to fit into a movement? Scientific controversies inspire many works which inflate the number of citations.

Furthermore, identifying the status of an issue, or publishing a book of synthesis lead to a large number of citations. They are a part of readings prior to any research, and therefore they are cited more often than other works in the field. Some research topics are considered as more "fashionable", e.g. the research of environmental issues has been very popular recently, and as such, it has been attracting more funding. Although we do not intend to dismiss the urgent need to study the environment, we question if it is comparable with the fundamental research aiming to establish a discipline.

4.2. Globalization and marginalization

If the professional field and the scientific field of information are related, to be in a recognized academic discipline, to train high-level practitioners and researchers at the university, it is necessary to think about their theoretical foundations. For the last three decades of the 20th century, industries and the society in general have demanded a political and organizational response to their increasing need for scientific and technical information. The development of computer science led to proliferation of studies which applied it to practical problems using bibliometric analysis. At the time they were well received by journals and on conferences and seminars, while qualitative or theoretical analyzes of research output were marginalized. Meyriat's project in France was to build information science as academic discipline, to ensure its position, build its base, produce theories, develop its concepts and methods, which was the most important issue between 1975 and 1990. However, as more scholars move into bibliology from the hard sciences, especially physics, the increasing trend has been to overlook the theoretical part of scientific work. This could explain why the works of the scholars being the subject of this study are cited only rarely in Google Scholar (1710 citations in political science, law and information science for Meyriat, a researcher with a 46-year career) and reflects the Matthew effect in science (Merton, 1968). The number of citations of Fonseca's work retrieved by Google Scholar is much bigger (3080). However, he was active on many fields, as librarian-bibliographer, teacher, columnist, and intellectual who was a part of a wide social network. His production had not been limited to the field of librarianship and information science.

4.3. Splitting up or career path

The uniformity of the controlled metadata of traditional bibliographies make it possible to follow the progress of an author's thought. The same is true of all other activities he has conducted inside and outside the field and, of what has been produced about this author. A search conducted through Google Scholar can give a view of the author's work, ordered by a number of citations, but not a view of how the author's though has progressed. Thus it is very difficult to draw a complete research activity by searching in Google Scholar.

The increasing order of the citations, even if not always respected, can confuse the reader. Results of the search for Meyriat's works present a statistics guide as one of his cited publications. It appears in the first position (with 76 citations), while his article about document, with 124 citations appears in the eighth position. In results for Fonseca's work, his publication with G. Freyre, including some recipes for cooking, are presented between the most cited. In both cases, retrieved data is the mark of various interests of these two authors related to their belonging to their time, to the professional, the academic and intellectual world. This may lead to confusion: for example, the first item in Google Scholar page for Fonseca, a book *Açúcar: uma sociologia do doce, com receitas de bolos e doces do Nordeste do Brasil* (*Sugar: a sociology of sweets with cake and candy recipes from Northeast Brazil*)[3], authored by sociologist Gilberto Freyre, had its first edition in 1939 and has been reissued

[3] Freyre: *Açúcar: uma sociologia do doce, com receitas de bolos e doces do Nordeste do Brasil*. São Paulo: Cia das Letras, 1997 (1939).

over the years. In this book, the author continues to analyze the importance of Portuguese, indigenous and African influences on the emergence of Northeastern Brazilian cuisine, and emphasizes the effect of the sugarcane monoculture on the socio-cultural creation of the state of Pernambuco, where it originated. Freyre's close friend of many decades, Edson Nery da Fonseca is the author of a biobibliography of Freyre attached to the work. He has been listed in Google Scholar as a co-author and not as a contributor, what a traditional bibliography would have corrected.

Nevertheless, Fonseca's most cited publications refer to the history of librarianship (197 citations), bibliography and bibliometric (190 citations), which have recently become timely as la result of the current interest in research related to the information science. He had commitments to politics and institutions. As far as Meyriat is concerned, the negligible number of references to his scientific work in information science results from the lack of the visibility of the journal where he had published most of his writing. Coming from a *Grande Ecole*, he was a professor in an institution close to the center of political power: this connection made the communication and information science, which development he led, more recognized and prestigious. As Bourdieu says, everything happens only if his "institutional scientific capital" (Bourdieu, 1976) becomes recognized. Fonseca was closer to the profile of the academic intellectual, who united scientific aspirations with information-based practices, as he was actively involved in library policies and programs in his state of Nordeste, and participated in the creation of new lines of research for the new discipline in Brazil.

The scientific contribution of these two precursors of the information field in France and Brazil, seems to be forgotten in the time of citation-counting. Publication in journals that have become minor or in less known or vanished publishing houses, books that are out of print, is superseded by the current circulation on the Internet. Only fragments of the journey are taken into consideration. Therefore, we are right to ask if their invisibility in Google Scholar means that these authors have been completely forgotten.

5. Conclusion

Taking into account the differences between the French and Brazilian experiences, we may say that the two authors are considered to be major founders of the field of the documentation and in information science. If, as Martin-Martin et al. argue,

> Google Scholar can be used to reliably identify the most highly-cited academic documents. Given its wide and varied coverage, Google Scholar has become a useful complementary tool for bibliometrics research dealing with the identification of the most influential scientific works in the state of our research (Martin-Martin et al., 2017, 162),

then the disorder of the classification and the writing of the bibliographic records do not accurately represent the fundamental role played by figures nowadays recognized in their discipline.

The evaluation of their scientific output cannot be limited to quantitative criteria. It is necessary that the works published by barely visible publishers are better known; they are analyzed and cited by researchers interested in the themes to which they have devoted themselves, in particular to develop the state of the art; also to work and research questions

that show developments and place current research within a long disciplinary time frame. The role of Meyriat and Fonseca in the establishment of the field of information science is still to be examined and analyzed.

The questions raised by this qualitative research show that a further study will increase the visibility of the role played by Fonseca and Meyriat in the institutionalization of the information science in France and Brazil, its training processes, the sociability network in the field and scientific debate inspired by their ideas. Other scholars who have now passed away might be the subject of similar studies, which will be a significant contribution to the historical study of the information science.

References

Bourdieu, P. (1976). Le champ scientifique. *Actes de la recherche en sciences sociales*, 2(2/3), 88–104.

Bourdieu, P. (1997). *Les usages sociaux de la science: pour une sociologie clinique du champ scientifique*. Paris: INRA Éditions.

Cain, J., Martin, H.-J. (eds.). (1972). *Le Livre français*. Paris: Imprimerie nationale.

Couzinet, V. (2000). *Médiations hybrides : le documentaliste et le chercheur*. Paris: ADBS Éditions.

Couzinet, V., Rauzier, J.M. (collab.). (2001). *Jean Meyriat, théoricien et praticien de l'information-documentation*. Paris: ADBS Éditions.

Couzinet, V. (2017). Construction des sciences de l'information et de la communication françaises: rôle de Jean Meyriat (1921–2010). In: W. Mustafa El Hadi (ed.). *Fondements épistémologiques et théoriques de la science de l'information-documentation. Actes du 11e colloque ISKO France 2017* (40–58). London: ISTE.

Couzinet, V., Fraysse P. (2018). Bibliographie et bibliographes en France: des revues pour la circulation de la science. *Bibliothecae.it*, 7(2), 172–197.

Estivals, R. (ed.) (1993). *Les sciences de l'écrit: Encyclopédie internationale de bibliologie*. Paris: Retz.

Fonseca, E.N. da. (1988). *Ser ou não ser bibliotecário e outros manifestos contra a rotina*. Brasília: ABDF.

McKenzie, D. F. (1991). *La bibliographie et la sociologie des textes*. Paris: Cercle de la librairie.

Malclès, L-N. (1984). *Manuel de bibliographie*. Paris: Presses Universitaires de France.

Marteleto, R., Carvalho, L. (2015). Health as a Knowledge Domain and Social Field: Dialogues with Birger Hjørland and Pierre Bourdieu. *Knowledge Organization*, 42(8), 581–590, https://doi.org/10.5771/0943-7444-2015-8-581

Martin-Martin, A., Orduna-Malea, E., Harzing A-W, Delgado Lopez-Cozar E. (2017). Can We Use Google Scholar to Identify Hightly-Cited Documents? *Journal of Informetrics*, 11(1), 152–163, http://dx.doi.org/10.1016/j.joi.2016.11.008

Merton, R.K. (1968). The Matthew Effect in Science. *Science*, 159(3810), 56–63.

Meyriat, J. (1957). *Rapport sur les principes généraux du travail bibliographique international* [online]. UNESCO Digital Library [06.12.2019], https://unesdoc.unesco.org/ark:/48223/pf0000127383_fre.

Meyriat, J. (1981). Document, documentation, documentologie. *Schéma et schématisation*, 14, 51–63.

Meyriat, J. (1993). La bibliographie. In: R. Estivals, J. Meyriat, F. Richaudeau (eds.), *Les Sciences de l'écrit, encyclopédie internationale de bibliologie* (548–553). Paris: Retz.

Meyriat, J. (1994). Y a t'il une place pour une théorie de la documentation? *Revue de bibliologie*, 40, 39–45.

Meyriat, J. (2006). Pour une compréhension plurisystémique du document (par intention). *Sciences de la société*, 68, 11–28.

Meyriat, J. (2016). Documento, documentação, documentologia. *Perspectivas em Ciencia da Informação*, 21(3), 240–253, http://dx.doi.org/10.1590/1981-5344/2891

Mota, A., Verri, G.M.W. (2001). *Interpretação de Edson Nery da Fonseca*. Recife: Bagaço.
Nascimento, D. M., Marteleto, R. (2007). Social Field, Domains of Knowledge and Informational Practice. *Journal of Documentation*, 64(3), 397–412, https://doi.org/10.1108/00220410810867605
Provansal A. (1997). Bibliographie. In : S. Cacaly (ed.), *Dictionnaire encyclopédique de l'information et de la documentation* (67–71). Paris: Nathan.
Rovira, C., Guerrero-Salé, F., Codina, L. (2018). Las citas recibidas como principal factor de posicionamento SEO en la ordenación de resultados de Google Scholar. *El Profesional de la información*, 27(3), 559–569, http://dx.doi.org/10.3145/epi.2018.may.09
Thiesen, I. (2010). L'information entre réalité et imaginaire: approches historiques au Brésil. In: V. Couzinet, R. Marteleto (eds.), *Médiations documentaires: entre réalités et imaginaires: actes de la première journée scientifique internationale du réseau MUSSI* (53–70). Toulouse: Université de Toulouse 3.

Ewolucja bibliografii badaczy: od systematycznej organizacji do cytowań

Abstrakt

Cel/Teza: Artykuł analizuje publikacje naukowe dwóch badaczy z różnych krajów, wykorzystując cytowania zidentyfikowane przez Google Scholar. Jean Meyriat i Edson Nery da Fonseca byli wybitnymi naukowcami we Francji i w Brazylii. Obaj mieli duży wkład w kształtowanie i rozwój informacji naukowej i bibliotekoznawstwa w swoich krajach. Tradycyjna bibliografia oraz kierunki naukowego rozwoju obu badaczy zostały opisane w niniejszym artykule.
Koncepja/Metody badań: Przeprowadzono analizę publikacji obu badaczy na podstawie danych z Google Scholar, co pomogło prześledzić rolę, jaką odgrywają wyszukiwarki w ewaluacji nauki i ich wpływ na wyszukiwanie informacji oraz ocenę dorobku naukowego. Analiza jakościowa opierała się na analizie bibliologicznej, skupionej na sposobie, w jaki piśmiennictwo obu autorów jest przedstawiane i które elementy są uwypuklane. Niniejsza analiza ma charakter poszukiwawczy, sugeruje kolejne pytania badawcze i hipotezy, które mogą wskazać kierunki dalszych badań.
Wyniki i wnioski: Jednolite, kontrolowane metadane stosowane w tradycyjnych bibliografiach pozwalają na prześledzenie rozwoju myśli badawczej danego autora. To samo dotyczy innych podejmowanych przez niego aktywności, zarówno w ramach własnego pola badawczego, jak i poza nim, a także innych publikacji, które danego autora dotyczą. Wyszukiwanie w Google Scholar daje dostęp do informacji na temat publikacji autora i publikacji jego dotyczących wyłącznie w kolejności liczby cytowań, a nie w porządku pozwalającym na prześledzenie naukowych osiągnięć autora czy rozwoju jego badań. Obserwacja widoczności dorobku dwóch wybitnych informatologów, Jean Meyriata (Francja) i Edsona Nery de Fonesci (Brazylia), na podstawie analizy wyszukiwań w Google Scholar, usytuowana w kontekście analizy pola badawczego, zdefiniowanego przez Bourdieu (1976), ukazuje tendencję do umniejszania roli, jaką odegrali.
Zastosowanie praktyczne: Sposób funkcjonowania danej dyscypliny badawczej wpływa na prowadzenie w niej badań. Wyszukiwarki, takie jak Google Scholar ułatwiają dostęp do opisów rezultatów badań naukowych. Jednakże kryterium, jakie stosują do szeregowania wyników, czyli liczba cytowań, skutkuje pominięciem innych ważnych aspektów dorobku naukowego danego autora. Analiza jakościowa, obejmująca powiązania między autorami oraz kierunki naukowego rozwoju, może ukazać właściwy wkład danego autora w rozwój dyscypliny.
Oryginalność/Wartość poznawcza: Przedstawione wyniki badań uwypukliły znaczenie stosowania w procesach wyszukiwawczych zarówno metod ilościowych, jak i jakościowych, aby zapobiec ryzyku pominięcia różnych aspektów dorobku naukowego.

Słowa kluczowe

Bibliografia. Edson Nery da Fonseca. Google Scholar. Jean Meyriat. Nauka o informacji. Pole badawcze.

VIVIANE COUZINET is Professor emeritus in Information and Communication Sciences at the University of Toulouse III-Paul Sabatier. She holds a PhD in Information and Communication Sciences from the University of Bordeaux III Michel de Montaigne, France. Her main areas of interest are the mediation process in hybrid contexts (scientific mediation and documental mediation); information representation in "mediators" objects and informational practices; knowledge organization as "mediators" objects of knowledge; social theory of document; documentologic approach (methods, concepts and theorization)

Contact to the Author:
viviane.couzinet@iut-tlse3.fr
Laboratoire d'études et de recherches appliquées en sciences sociales (LERASS)
Equipe : Médiations en information-communication spécialisée (MICS)
Université Toulouse III-Paul Sabatier-IUT
115 B Route de Narbonne
BP 67 701
F- 31077 Toulouse Cedex 4, France

REGINA MARTELETO, PhD, is a tenured researcher at the Brazilian Institute of Information in Science and Technology (IBICT) and a core member of the postgraduate program in Information Science at the Federal University of Rio de Janeiro (UFRJ). She is the coordinator of Culticom, a research group on Culture and Info-communicational processes: www.culticom.org. Her main areas of interest are: culture and information; knowledge, information and society; information and communication on health; info--communicational mediation in social networks; the subject, interpretations, and languages of information in contemporaneity; social theory, epistemology and interdisciplinarity in the field of information.

Contact to the Author:
regina.mar@ibict.br
PPGCI/IBICT-UFRJ
Rua Lauro Muller, 455 – 4o. andar
CEP 22290-160 – Rio de Janeiro, RJ, Brazil

ICLÉIA THIESEN, PhD, is a tenured professor and researcher of Information Science at the Federal University of the State of Rio de Janeiro (UNIRIO) at the Department of History. Her main areas of interest are: relations between information, memory and history; sensitive documents in exception regimes; institutional memory; political uses of the past; info-communicational devices aimed at retrieving sensitive information in archives.

Contact to the Author:
icleia.thiesen@unirio.br
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Av. Pasteur, 458 – Prédio do CCH – subsolo – sala 20
CEP 22280-240 – Urca – Rio de Janeiro, Brazil

Patent Applications for Electronic Publishing Market (2014–2018). Selected Issues

Adam Jachimczyk
ORCID 0000-0003-2917-6926
Department of Bibliography and Documentation,
Faculty of Journalism, Information and Book Studies,
University of Warsaw

Abstract

Purpose/Thesis: The article analyses patent applications for inventions in the area of electronic publishing. Its investigates how the number of patent applications changed in the years 2014–2018, which companies were the most active in the field, and what was their providence. The issues of crucial importance to the inventors were identified by the study of the symbols used by International Patent Classification.
Approach/Methods: The analysis was based on the data from the base Lens.org from years 2014–2018. In total, 1733 patent applications related to electronic publishing were identified.
Results and conclusions: The study recorded an increasing activity of Chinese and Korean inventors, and a low number of patent applications filed by European firms. The applicants focus on specific functions facilitating interaction between the user and digital devices; a relatively high number of applications is concerned with electronic commerce, and with the use of electronic publishing in education.
Originality/Value: An analysis of patent applications may help to identify specific phenomena occurring in the field of information technologies. It records the recent trends in research, and identifies the countries and agents which lead the development of widely understood electronic publishing.

Keywords

Ebooks. Electronic books. e-Publications. Patents.

Received: 18 October 2019. Reviewed: 24 November 2019. Accepted: 10 December 2019.

1. Introduction

Electronic publishing (e-publishing) refers to publications of different types (books, press, journals, websites) available in many formats (HTML, EPUP, MOBI, PDF) on various devices (designated reading devices, or less specialized devices, such as personal computers, tablets, or smartphones) (Burke, 2013).

The emergence of portable devices (e.g. Sony Data Discman, or NuvoMedia's Rocket) in the 1990s constituted a significant development for the devices enabling access to electronic publishing. Because of their limitations (small screens, low resolution, limited memory), they were not popular with the users (Herther, 2008).

In the first decade of the 21st century, the spread of electronic paper produced by E Ink company inspired further development of the electronic publishing market and the devices facilitating access to it. By the end of the first decade, at least 90% of the producers of

e-readers used the electronic paper (Harris, 2010). Furthermore, many models of e-readers entered the market. In 2010 alone, the produces introduced 15 new devices (Manley & Holley, 2012).

Another factor which contributed to the growth of e-publishing was the appearance of devices (mobile phones and slightly larger devices, so-called tablets) with touch-screens, which became an alternative to the reading devices with e-ink screens, as they supported multimedia content and made it possible to handle richly illustrated publications. With the development of these devices, bookstores started to sell more digital publications, while the publishers gradually resigned from burdensome security technologies, which restricted the use of e-publications, e.g. by restricting the number of devices on which a purchased copy could be read.

E-publications have been on the offer of big- and small-scale publishers for several years now. Although the sales remain small in comparison to those of the print books, they have been increasing in the recent years. Only in the last years the growth became slightly slower (Milliot, 2018; Springer, 2017).

2. The methods of the analysis of patent claims

According to the definition published by the Patent Office of the Republic of Poland (UPRP), a patent is "...the exclusive right to use the invention for a specified period, for profit (industrial, commercial) in the territory of a State or States, granted by the competent authority of the state, regional or international" (UPRP, n.d.-a). The exclusive right to a given invention ensures that the author is protected from unfair competition from other agents wishing to use the technological solution in their products. It also creates opportunities for extra profits from selling, or licensing the patent to be used by other agents (Redl et al., 2016).

The process of acquiring a protected status for the invention is relatively long. It starts with filing an application in the appropriate intellectual property office. The period between filing the application and the moment it is published might be longer than ten months (in Poland, this process takes between 12 and 18 months) (Redl et al., 2016; Śnieżko, n.d.; UPRP, n.d.-c). The decision regarding the patent itself might be made within years of filing the application (Nauka w Polsce, n.d.). We should bear in mind that some of the declared and patented solutions will never be employed, e.g. because of the limited possibility of industrial use (Burak, 2015).

The general availability of sources of information on patents, such as the databases of national intellectual property offices, the Espacenet (European Patent Office), or Patentscope (World Intellectual Property Organization) allows for a comprehensive analysis of patent applications. The research is concerned with, e.g., the extent of technological innovation in specific countries (Hicks et al., 2001; Nikzad, 2014; Meyer et al., 2003), in specific organizations (Kang, 2015; Martínez & Rama, 2012), or generally, in a given field (Jana et al., 2012; Vandeberg & Boon, 2009). The researchers also analyze patent applications to identify technologies to be developed in the future (Daim et al., 2006; Trappey et al., 2011). Patent analysis is also commonly used in the so-called competitive intelligence. The study of patent applications allows to determine the quality and to identify the assets of other companies, in order to carry out a fusion or an acquisition, or to identify the persons which

play the key role in the development sector of these companies (Breitzman & Mogee, 2002; Breitzman & Thomas, 2002).

The researchers employ various methods to analyze the patent applications. The most basic method comprises quantitative analysis of the meta data contained in the file (the country where the application was filed, the date the application was filed, last name of the applicant, the date the application was published, the date the patent was granted, the symbol of type classification). The researchers combine it with an analysis of abstracts or full texts of the patent files to identify the nature of a given technical solution (Abraham & Moitra, 2001; Singh et al., 2018).

The researchers also seek to identify the most relevant data set for the analysis. Basing on the patent applications in the automotive sector, Xie and Miyazaki (2013) established that browsing the databases with the use of keywords included in the title of the patent application, the abstract, and the patent claim, yields the best results. International Patent Classification (IPC) may also be used to identify the patent applications in a given sector (Abraham & Moitra, 2001). The symbols of IPC may also be correlated with other data. In their study of Apple's technological innovation, Jun and Sung Park (2013) correlated the ICP symbol with the year of the patent application to identify the issues which in the recent years have been most important for the company.

We should also mention the studies which analyze the references to other patents in applications, e.g. in order to trace the technological development in a given sector (Gui et al., 2019), and those which record the references to websites of specific organizations as an indicator of their technological level (Orduna-Malea et al., 2017). The researchers use the techniques of text mining, information visualization, or natural language processing to analyze the unstructured data contained in the abstract, the patent claim, or the description of the invention (Abbas et al., 2014).

3. Objectives and methods

The e-publishing market should be considered from the point of view of technological infrastructure (hardware and software) which enables the user to access the digital content. Patent applications filed at the appropriate intellectual property office may serve as an indicator of technological development, and as a measure of innovation.

For this purpose, the patent applications from years 2014–2018 were reviewed. This period was chosen to allow tracking specific phenomena related to the development of technology connected with e-publishing, which occurred in an already formed market where a relatively high number of publishers as well as producers of devices and software enabling the access to the digital content were already operating. The goals of the analysis were:

(1) to identify the countries where the highest numbers of patent applications come from;
(2) to establish the agents who filed the highest numbers of applications and were granted the highest number of grants, and the time at which they made the applications;
(3) to establish the thematic scope of the patent applications.

A quantitative analysis was based on the select metadata describing the patent applications: the date the application was filed, the title of the claim, the name of the applicant,

the type of the application (an application, a granted patent), the number of priority, the ICP symbol.

The subject of the applications was established with a reference to the ICP symbols. All symbols used to characterize the matter of the applications were identified. The qualities of the inventions were identified basing on the frequency with which certain ICP symbols occurred in the description. The subject analysis of applications was supplemented by the examples of claims concerned with the issues specific to different groups distinguished by ICP.

3.1. The selection of research data

The data was collected from the base Lens.org, which was developed as a result of a collaboration between an Australian non-profit organization Cambia and Queensland University of Technology. The database collects the data on patent applications from European Patent Office, United States Patent and Trademark Office, World Intellectual Property Organization and Intellectual Property Australia. The dataset comprises *c.* 120 millions of records from 65 jurisdictions registering the filed patent applications (per data from October 31st, 2019) (Lens, n.d.). The website also publishes a set of tools for the analysis of the found patent claims.

The data for the analysis were collected on November 19th, 2019. To find patent applications, the following command was used. Following the results of the study conducted by Xie and Miyazaki (2013), keywords were used as search terms to browse patent names, abstracts, and claims, as applicants should state technical qualities of the invention, or describe its application in the patent claim (UPRP, n.d.-b).

title:(Ebook OR "e-book" OR "digital book" OR ereader* OR epub OR "electronic books" OR "electronic publication" OR "electronic publications" OR "e-reader" OR "electronic book") OR abstract:(Ebook* OR "e-book" OR "digital book" OR ereader* OR epub OR "electronic books" OR "electronic publication" OR "electronic publications" OR "e-reader" OR "electronic book") OR full_text:(Ebook* OR "e-book" OR "digital book" OR ereader* OR epub OR "electronic books" OR "electronic publication" OR "electronic publications" OR "e-reader" OR "electronic book") OR claims:(Ebook* OR "e-book" OR "digital book" OR ereader* OR epub OR "electronic books" OR "electronic publication" OR "electronic publications" OR "e-reader" OR "electronic book")*

The use of keywords as search terms, rather than ICP classification, to identify patent applications seems more justified, especially in a study on constantly developing technologies which are not assigned a strictly defined category in the abovementioned system (Xie & Miyazaki, 2013). The extensive search query was used to identify the largest possible set of files related to the widely understood area of e-publishing. It was not an infallible method. Lack of control over the terminology of the applications was an obstacle to precise browsing of very large sets of patent files. Montecchi et al. (2013) noted following problems in the descriptions of patents:

- the varying degree of detail, depending on the specific style of the author of the description and the application of different terms to refer to one issue. In the case of e-publishing, the variations included ereaders and e-readers, ebook and e-book;
- imprecise, incorrect terminology; occasionally new coinages used to describe the claimed inventions; lack of a standardized terminology for the new technologies;

– the use of automated translation, which might disrupt the meaning of the original text.

The search results were narrowed down to patent applications filed between 2014 and 2018. In total, the set comprised 3721 records. The duplicated descriptions of the files, filed at different intellectual property offices, were eliminated. When a duplicate was identified, the earliest record was preserved. The descriptions of patents which did not relate to the search command were eliminated as well. In the end, 1733 records remained as the subject of the following analysis.

The names of applicants, recorded differently in different files (e.g. Google was referred to as Google INC and Google LLC), were standardized for the purposes of the study. Furthermore, if it was established that two agents were related, the patent was assigned to the mother company. This was the case with Audible, which is a subsidiary of the Amazon corporation. The tables below employ the phrasing used in the application. In the case of the Kobo company, all applications where it was identified as the applicant, were assigned to its current iteration, Rakuten Kobo Inc., as it has been referred to since the Canadian producer was acquired by the Japanese corporation Rakuten.

The analysis had certain limitations. Because the search command was formulated in English, the applications in other languages were not taken into account. Other limitations derived from the problems discussed above, i.e. the imprecise terminology, or an incorrect choice of keywords used to browse the data bases. Furthermore, the analysis did not account for the applications which related to e-publishing, but did not include any of the abovementioned key words in the title, the abstract, or the patent claim.

4. Results

Within the studied period, the average of 350 applications per year was filed. Only in 2014 and in 2015 did this number reach above the average. The deviation in 2018 should be seen as a result of the time spent between the moment an application is filed and the moment it is published, as the regular period of upwards of ten months has not yet passed (Fig. 1).

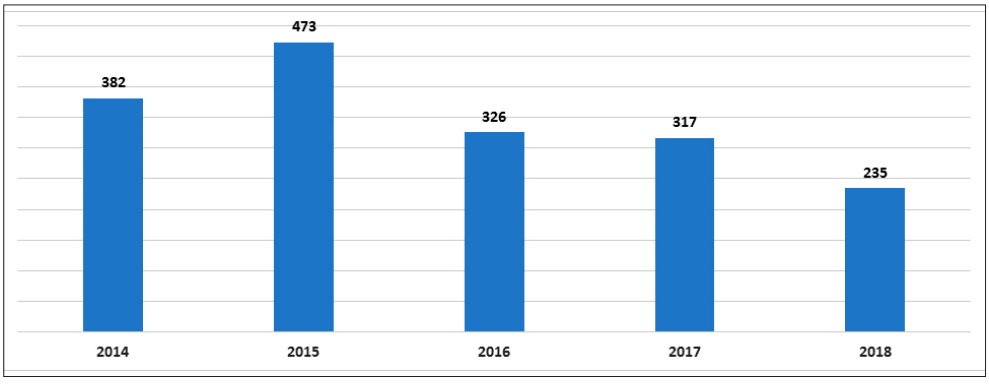

Fig. 1. The number of patent applications per year in the years 2014–2018

4.1. The providence of the application

The providence of the application might be identified in the base Lens.org, which designates it as the priority number, assigned in the moment of filing the application (Espacenet, n.d.). However, it is not a perfectly precise information, as agents might file applications in various countries (Tab 1).

Tab. 1. Countries, where (more than 10) patent applications were file

Providence	Number of applications	% (n=1733)
China	773	44.60
United States	502	28.97
South Korea	193	11.14
Japan	118	6.81
Taiwan	53	3.06
Great Britain	23	1.33
European Patent Office*	16	0.92
India	14	0.81

*The number of applications filed at the European Patent Office.

For the sake of comparison, a list of 43 agents which filed at least five applications in the period studied has been prepared, with a consideration for their country of origin (Tab. 2).

Tab. 2. The country of origin of the agents with the highest number of patent applications

Country	Number of agents	% (n=43)
China	19	44.19
United States	10	23.26
Japan	7	16.28
South Korea	4	9.30
France	1	2.33
Netherlands	1	2.33
Taiwan	1	2.33

In the case of China, South Korea and Taiwan, the results were similar. The variations observed in the case of Japan and USA might be explained with a reference to the number of the agents filing the applications. In USA, there probably has been a higher number of agents, who filed less than five applications; the opposite was the case in Japan, where a higher number of agents filed more than five applications.

Inevitably, Tables 1 and 2 do not record the activity of corporations operating on the international scale, which often acquire companies operating in other countries and become

the owners of their patents. Rakuten Kobo Inc. might serve as an example, as it was established as a result of an acquisition of a Canadian producer of e-book readers by Japanese corporation Rakuten (BBC, 2011).

The analysis of data shows the countries, which begin to dominate the discussed field. These are China, which is the country of origin for almost 45% patent applications, United States, South Korea, Japan, and Taiwan. In this technological race, Europe is visibly losing distance, as only a relatively small number of applications comes from its countries. The countries featured in the Table 1 are the source of almost 98% of the applications filed in the period discussed.

4.2. Applicants

In total, 1013 entities (companies and private persons) that filed a patent application were identified. This set includes 797 entities which filed only one application. 199 entities filed between two and nine applications. Numerous applications were jointly filed by more than one entity, e.g. Univ Peking Founder Group Co. and Beijing Founder Electronics Co., Ltd filed 17 joint applications.

Tab. 3. The most active entities (who filed more than 10 applications)

Entity	Number of applications	Number of patents
Rakuten Kobo Inc.	103	16
IReader Tech Co., Ltd	99	51
Amazon Tech Inc.	65	64
Samsung Electronics Co., Ltd	29	12
IBM	28	18
Alibaba Group Holding Ltd	22	2
China Mobile Comm Corp	18	1
Univ Peking Founder Group Co.	17	5
Beijing Founder Electronics Co., Ltd	17	5
Microsoft Corp	16	4
Dainippon Printing Co., Ltd	16	6
Beijing Xiaomi Technology Co.	16	5
Google LLC	15	8
Beijing Qihoo Tech Co., Ltd	14	4
Tencent Tech (Beijing) Company Ltd	12	4
Beijing Jingdong Century Trading Co., Ltd	12	4
Woongjin Thinkbig Co., Ltd	10	1

Rakuten Kobo dominates the number of applications (Tab. 3). In the last five years, Chinese company IReader has also begun to show an increased activity in filing the patent applications. Amazon, which recently reached dominance over the e-publishing market,

especially in the United States, filed slightly less applications. Amazon's advantages are the introduction of innovative hardware solutions, e.g. the Kindle reader, whose first iteration entered the market in 2007, and a wide offer of digital publications which might be accessed with this device. The group featured in Table 3 also includes the leaders of the IT market, Alibaba, Google, IBM, Microsoft, and Samsung, among others.

It should be mentioned that Sony and the bookstore chain Barnes & Noble filed a relatively low number of applications (11 in total). Even in the beginning of the second decade of the 21st century, these companies were trying to match Amazon, and to introduce their own e-readers onto the market, e.g. the Nook reader of Barnes & Noble (Griffey, 2012). Other IT giants (Apple, Lenovo), filed 13 applications in the period under discussion.

The situation looks slightly different if we consider the number of patents granted. Almost 27% of the applicants considered received the patent they applied for. Here, Amazon dominates with exclusive rights to 64 inventions. IReader boasts a slightly lower number (51) of rights. Together, Amazon, IReader, and Rakuten Kobo have rights to almost 29% of the inventions.

Tab. 4. The number of patent applications filed by the most active entities in years 2014–2018

Entity	2014	2015	2016	2017	2018	Total
Rakuten Kobo Inc.	67	35	–	–	1	103
IReader Tech Co., Ltd	–	1	8	35	55	99
Amazon Tech Inc.	30	28	6	1	–	65
Samsung Electronics Co., Ltd	4	20	4	1	–	29
IBM	5	10	7	6	–	28
Alibaba Group Holding Ltd	2	1	5	10	4	22
China Mobile Comm Corp	–	–	1	2	15	18
Univ Peking Founder Group Co. Beijing Founder Electronics Co., Ltd	5	4	3	5	–	17
Microsoft Corp	6	2	1	7	–	16
Beijing Xiaomi Technology Co.	4	7	3	2	–	16
Dainippon Printing Co., Ltd	9	2	4	1	–	16
Google Inc	10	3	1	1	–	15
Beijing Qihoo Tech Co., Ltd	–	14	–	–	–	14
Tencent Tech (Beijing) Co., Ltd	2	4	3	2	1	12
Beijing Jingdong Century Trading Co., Ltd	–	2	9	1	–	12
Woongjin Thinkbig Co., Ltd	5	4	1	–	–	10

Amazon, IReader, Alibaba and Tencent, among others, filed the applications most regularly. Table 4 shows that in 2018, Amazon filed no applications. Between 2016 and 2018, Rakuten Kobo filed only one application, as did Google. However, we should not assume that these companies became significantly less active. The unusual activity of IReader is striking: 90% of their applications were filed between 2017 and 2018.

4.3. Subject of the applications

The analysis of the thematic scope of the patent applications was based on the IPC symbols attached to the description. A complete IPC symbol is made up of a symbols denoting section, class, subclass, group, and subgroup (UPRP, 2012). For example, GO6Q 30/02 represents a file assigned to section G (Physics), class 06 (Computing; calculating; counting), subclass Q (Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for); subgroup 30/02 (Marketing, e.g. market research and analysis, surveying, promotions, advertising, buyer profiling, customer management or rewards; Price estimation or determination).

The table includes every symbol which featured at least 50 times. It is not a perfect indicator of the thematic scope of the application because of the superficiality of some descriptions. Less than 6% of the descriptions used only one symbol, G06F17/20 (now reclassified as G06F16/00-G06F16/958), which gives only a general indication of the thematic scope of the application (Information retrieval; Database structures therefor; File system structures therefor – Organization or management of web site content, e.g. publishing, maintaining pages or automatic linking). Only one ICP symbol was used to describe almost 33% files; only 1% were not assigned any symbols.

More elaborate symbols are slightly more informative. Numerous applications were put in the subgroup G06F3/048 (Interaction techniques based on graphic user interfaces [GUI], which "covers subject matter where the focus is placed on the way the user can interact with the displayed data") (IPC Publication, n.d.). This group includes descriptions of 490 files with symbols G06F3/0481, G06F3/0483, G06F3/0484, G06F3/0488.

These are the applications concerned with, e.g. browsing websites in digital publications proposed by Rakuten Kobo (Parker & Landau, 2016), IReader (Chen et al., 2018) or Google Beavers et al., 2016). The group also includes a project designed by Rakuten Kobo, enabling a display of content from various sources on a divided screen (Landau, 2015), a description of a patent application solving the problem of viewing electronic illustrated books (Murase, 2017), and a description of a prototype of an e-book for visually impaired readers (Jin et al., 2015).

The technological solutions designed with the thought of visually impaired readers deserve particular consideration. The development of e-publishing created new opportunities to access books and press. Text is being converted to speech or enlarged; the electronic devices are employing the Braille alphabet (Junus & Booth, 2012). Browsing by terms 'blind' and 'Braille' yields 10 results describing devices designated for the visually impaired users.

Browsing by terms navigation, scrolling, turn, turning yielded around 50 patent applications, which are concerned with various ways of viewing digital content.

The group G06/F17/00 (Digital computing or data processing equipment or methods, specially adapted for specific functions) contains descriptions of 260 applications assigned subgroup G06/F17/21 (Text processing), G06/F17/24 (Editing, e.g. insert / delete), G06/F17/27 (Automatic analysis, e.g. parsing, orthographic correction). IReader's proposition to use hand-writing in the e-readers is particularly striking (Cheng et al., 2018). Applications featured in this group were also concerned with the use of EPUB format in e-readers, e.g.

Androids (Li, 2017), and with a solution allowing annotating digital content (Heo, 2016). In total, around 20 applications were concerned with the use and modification of the EPUB format in digital devices.

The subgroup G06F3/01 (Input arrangements or combined input and output arrangements for interaction between user and computer), whose symbol was used to describe 86 applications, includes, among others, a patent application describing an electronic book capable of displaying three-dimensional models (Gonzalez, 2018), or IBM's proposal regarding a display of additional information on the object mentioned in an electronic publication (Ekambaram & Rakshit, 2018).

Tab. 5. Thematic scope of the patent applications
(taking into account the symbols occurring at least 50 times)

ICP symbol	Number	Description
G06F17/30	256	information retrieval; database structures therefor; file system structures therefor – organization or management of web site content, e.g. publishing, maintaining pages or automatic linking
G06F3/0483	196	interaction with page-structured environments, e.g. book metaphor
G06F3/0488	120	using a touch-screen or digitizer, e.g. input of commands through traced gestures
G06Q50/10	120	services
G06F3/0484	115	for the control of specific functions or operations, e.g. selecting or manipulating an object or an image, setting a parameter value or selecting a range
G06F17/21	109	text processing
G06F17/24	94	editing, e.g. insert/delete
G06Q30/06	91	buying, selling or leasing transactions
H04L29/08	86	transmission control procedure, e.g. data link level control procedure
G06F3/01	86	input arrangements or combined input and output arrangements for interaction between user and computer
G06F15/02	75	manually operated with input through keyboard and computation using a built-in program, e.g. pocket calculators
G06Q30/02	66	marketing, e.g. market research and analysis, surveying, promotions, advertising, buyer profiling, customer management or rewards; price estimation or determination
G09B5/06	63	electrically-operated educational appliances with both visual and audible presentation of the material to be studied
G06F3/0481	59	based on specific properties of the displayed interaction object or a metaphor-based environment, e.g. interaction with desktop elements like windows or icons, or assisted by a cursor's changing behavior or appearance
G06F17/27	56	automatic analysis, e.g. parsing, orthographic correction
H04L29/06	50	characterized by a protocol

Descriptions of 120 applications assigned the symbol of subgroup G06Q50/10 (Services) are concerned with various aspects of e-publishing, such as the systems of immediate acquisition (I Sa et al., 2017), methods of creating multi-lingual electronic publications (Yun, 2016), or the means of recommending e-books (Seo, 2016). Generally, more than 30 applications which are concerned specifically with the implementation of a recommended function in electronic devices was found browsing by the term recommend.

Applications from the group G06Q30/00 (Commerce, e.g. shopping or e-commerce) include more than 90 applications from the subgroup G06Q30/06 (Buying, selling or leasing transactions). IReader proposed a system enabling the user to resale e-books (Zou & He, 2019); Keydo Communication – a system of reviewing digital books using big data (Zou & He, 2019), and Optim – a method of displaying advertisement that does not disrupt the viewing of the e-book (Sugaya, 2016).

Engineers also work on the application of e-publishing and e-books in advertisement. Browsing the analyzed set by the term 'advertisement' yielded 12 results. Before 2014, such solutions were proposed by Amazon, Microsoft, and Yahoo! (Johnson, 2014). Interestingly, Amazon filed no applications related to advertisement in the period discussed.

The group G09B5/00 (Electrically-operated educational appliances) contained more than 60 applications assigned to the sub-group G09B5/06 (Electrically-operated educational appliances with both visual and audible presentation of the material to be studied). Arbordale Publishing suggests an educational platform transforming e-books into interactive, multilingual publications, using technology supporting reading (German, 2019). Another file is concerned with e-books supporting foreign language learning (Gao, 2015).

More than 130 applications are concerned with transmitting digital information (H04L29/08, H04L29/06). Patent applications propose a system of e-book recommendation based on the technology of data transmission via cloud computing (Yang et al., 2018), a means of sending an e-book as a gift (Landau, 2015), and a system of sharing digital books (Chen, 2018).

5. Conclusion

The number of technological solutions concerned with e-publishing has been stable over the last few years. Inventors from East Asia, especially from South Korea and China, are increasingly active, while the European agents are filing less applications. Chinese corporation IReader is particularly active. A slightly lower number of application has been filed by other leading producers on the e-publishing market: Amazon, Rakuten Kobo, and especially Sony and Barnes & Noble. The number of all agents filing applications should be considered: more than 1000 corporations and individual persons proposed new technological solutions related to e-publishing.

Smaller activity of several companies does not mean that they do no work in the area. Implementation of a given solution might take several years; only now the new inventions of such producers as Barnes & Noble and Sony do begin to enter the market (Kozlowski, 2018).

The crucial question is which of the proposed solutions will start a new trend in the development of the e-publishing infrastructure. It is likely that majority of the proposed innovations will belong the group of the so-called incremental innovations, i.e., solutions

which do not cause a break-through in the production technology, but only introduce newer, improved, in the creators' opinion, functions to the already established technologies (Majewska & Szulczyńska, 2012). In the case of the analyzed group, basing on the IPC symbols, it might be observed that the applicants focus mostly on certain functions facilitating the interaction between the user and the electronic devices. However, applications concerned with e-commerce, and the use of e-publishing in education are as numerous.

References

Abbas, A., Zhang, L., & Khan, S. U. (2014). A Literature Review on the State-of-The-Art in Patent Analysis [online]. *World Patent Information*, 37, 3–13, https://doi.org/10.1016/j.wpi.2013.12.006

Abraham, B. P., Moitra, S. D. (2001). Innovation Assessment Through Patent Analysis. *Technovation*, 21(4), 245–252, https://doi.org/10.1016/S0166-4972(00)00040-7

BBC (2011). Retail Giant Buys Kobo Ebook Firm [online]. BBC News [16.11.2019], https://www.bbc.com/news/business-15664947

Beavers, A., Johnston, C., Paglia, M., Armstrong, B., Streu, N. S. (2016). *Patent Nr US2016041700 (A1)* [online]. Espacenet [12.11.20109], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20160211&DB=EPODOC&locale=en_EP&CC=US&NR=2016041700A1&KC=A1&ND=4

Breitzman, A. F., Mogee, M. E. (2002). The Many Applications of Patent Analysis. *Journal of Information Science*, 28(3), 187–205, https://doi.org/10.1177/016555150202800302

Breitzman, A., Thomas, P. (2002). Using Patent Citation Analysis to Target/Value M&A Candidates. *Research-Technology Management,* 45(5), 28–36, https://doi.org/10.1080/08956308.2002.11671519

Burak, A. (2015). *Statystykami patentów można wykazać dowolne tezy* [online]. Obserwator Finansowy: Ekonomia, debata, Polska, Świat. [12.11.2019], https://www.obserwatorfinansowy.pl/tematyka/makroekonomia/statystykami-patentow-mozna-wykazac-dowolne-tezy/

Burke, P. S. (2013). *e-Publikacje w InDesign ® CS6: Projektowanie i tworzenie publikacji cyfrowych dla tabletów, czytników, smartfonów i innych urządzeń*. Warszawa: APN Promise.

Chen, J., Han, F., Yuan, Y., Lei, X., Chen, Y., Lu, J. (2018). *Patent Nr CN107608596 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20180119&DB=EPODOC&locale=en_EP&CC=CN&NR=107608596A&KC=A&ND=4

Chen, X. (2018). *Patent Nr CN108462757 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20180828&DB=EPODOC&locale=en_EP&CC=CN&NR=108462757A&KC=A&ND=4

Cheng, C., Tao, H. (2018). *Patent Nr CN108959201 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20181207&DB=EPODOC&locale=en_EP&CC=CN&NR=108959201A&KC=A&ND=4

Daim, T. U., Rueda, G., Martin, H., Gerdsri, P. (2006). Forecasting Emerging Technologies: Use of Bibliometrics and Patent Analysis. *Technological Forecasting and Social Change,* 73(8), 981–1012, https://doi.org/10.1016/j.techfore.2006.04.004

Ekambaram, V., Rakshit, S. K. (2018). *Patent Nr US2018107638 (A1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20180419&DB=EPODOC&locale=en_EP&CC=US&NR=2018107638A1&KC=A1&ND=4

Espacenet (n.d.). Espacenet – Priority Number [online]. Espacenet – European Patent Office [12.11.2019], https://worldwide.espacenet.com/help?locale=en_EP&method=handleHelpTopic&topic=prioritynumber

Gao, L. (2015). *Patent Nr CN204257022 (U)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20150408&DB=EPODOC&locale=en_EP&CC=CN&NR=204257022U&KC=U&ND=4

German, L. B. (2019). *Patent Nr US2019196675 (A1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20190627&DB=EPODOC&locale=en_EP&CC=US&NR=2019196675A1&KC=A1&ND=4

Gonzalez, D. R. S. (2018). *Patent Nr US2018059880 (A1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20180301&DB=EPODOC&locale=en_EP&CC=US&NR=2018059880A1&KC=A1&ND=4

Griffey, J. (2012). E-Readers Now, E-Readers Forever! *Library Technology Reports*, 48(3), 14–20.

Gui, B., Ju, Y., Liu, Y. (2019). Mapping Technological Development Using Patent Citation Trees: An Analysis of Bogie Technology. *Technology Analysis & Strategic Management*, 31(2), 213–226.

Harris, S. (2010). Emergence of the E-book. *Nature Photonics,* 4(11), 748–749, https://doi.org/10.1038/nphoton.2010.247

Heo, I. (2016). *Patent Nr US2016170591 (A1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20160616&DB=EPODOC&locale=en_EP&CC=US&NR=2016170591A1&KC=A1&ND=4

Herther, N. K. (2008). The Ebook Reader Is Not the Future of Ebooks. *Searcher*, 16(8), 26–40.

Hicks, D., Breitzman, T., Olivastro, D., Hamilton, K. (2001). The Changing Composition of Innovative Activity in the US — a Portrait Based on Patent Analysis. *Research Policy,* 30(4), 681–703, https://doi.org/10.1016/S0048-7333(00)00147-5

I Sa. Y., Kim, D. H., Kim, H. S., Park, B. G. (2017). *Patent Nr KR101715814 (B1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20170320&DB=EPODOC&locale=en_EP&CC=KR&NR=101715814B1&KC=B1&ND=4

IPC Publication (n.d.). [online]. Międzynarodowa Klasyfikacja Patentowa [14.11.2019], http://mkp.uprp.pl/classifications/ipc/ipcpub/

Jin, H., Wei, C., Peng, P., Zhou, Z. (2015). *Patent Nr CN104318821 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20150128&DB=EPODOC&locale=en_EP&CC=CN&NR=104318821A&KC=A&ND=4

Johnson, B. (2014). Would I Read an Ad-Supported Ebook? Probably. *Computers in Libraries*, 34(4), 10–14.

Jun, S., Sung Park, S. (2013). Examining Technological Innovation of Apple Using Patent Analysis. *Industrial Management & Data Systems*, 113(6), 890–907, https://doi.org/10.1108/IMDS-01-2013-0032

Junus, S. G. R., Booth, C. (2012). E-books and E-readers for Users with Print Disabilities. *Library Technology Reports*, 48(7), 22–28.

Kang, B. (2015). The Innovation Process of Huawei and ZTE: Patent Data Analysis. *China Economic Review*, 36, 378–393, https://doi.org/10.1016/j.chieco.2014.12.003

Kozlowski, M. (2018, styczeń 2). *E-Reader Trends and Statistics for 2018* [online]. Good e-Reader. [17.11.2019], https://goodereader.com/blog/electronic-readers/e-reader-trends-and-statistics-for-2018

Landau, B. (2015). *Patent Nr US2015234786 (A1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20150820&DB=EPODOC&locale=en_EP&CC=US&NR=2015234786A1&KC=A1&ND=4

Landau, B. (2015). *Patent Nr US2015317050 (A1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20151105&DB=EPODOC&locale=en_EP&CC=US&NR=2015317050A1&KC=A1&ND=4

Lens (n.d.). About The Lens [online]. The Lens – Free Open Patent and Scholarly Search [12.11.2019], https://about.lens.org/

Li, C. (2017). *Patent Nr CN107220221 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20170929&DB=EPODOC&locale=en_EP&CC=CN&NR=107220221A&KC=A&ND=4

Majewska, M., Szulczyńska, U. (2012). Innowacje przyrostowe jako źródło postępu technologicznego gospodarce opartej na wiedzy. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Ekonomiczne Problemy Usług*, 87, 1, 101–109.

Manley, L., Holley, R. P. (2012). History of the Ebook: The Changing Face of Books. *Technical Services Quarterly*, 29(4), 292–311, https://doi.org/10.1080/07317131.2012.705731

Martínez, C., Rama, R. (2012). Home or Next Door? Patenting by European Food and Beverage Multinationals. *Technology Analysis & Strategic Management*, 24(7), 647–661.

Meyer, M., Utecht, J. T., Goloubeva, T. (2003). Free Patent Information as a Resource for Policy Analysis. *World Patent Information*, 25(3), 223.

Milliot, J. (2018). E-book Sales Fell 10% in 2017 [online]. *Publishers Weekly*, [12.11.2019], https://www.publishersweekly.com/pw/by-topic/digital/content-and-e-books/article/76706-e-book-sales-fell-10-in-2017.html

Montecchi, T., Russo, D., Liu, Y. (2013). Searching in Cooperative Patent Classification: Comparison Between Keyword and Concept-Based Search. *Advanced Engineering Informatics*, 27(3), 335–345, https://doi.org/10.1016/j.aei.2013.02.002

Murase, S. (2017). *Patent Nr US2017038931 (A1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20170209&DB=EPODOC&locale=en_EP&CC=US&NR=2017038931A1&KC=A1&ND=4

Nam, K. D. (2019). *Patent Nr KR101951757 (B1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20190225&DB=EPODOC&locale=en_EP&CC=KR&NR=101951757B1&KC=B1&ND=4

Nauka w Polsce (n.d.). Wiceprezes UP RP wyjaśnia, dlaczego na patent czeka się średnio 3,6 roku [online]. Nauka w Polsce [17.11.2019], http://naukawpolsce.pap.pl/aktualnosci/news%2C4127-91%2Cwiceprezes-up-rp-wyjasnia-dlaczego-na-patent-czeka-sie-srednio-36-roku.html

Nikzad, R. (2014). Canadian Worldwide Patent Activity: An Industrial Level Analysis. *World Patent Information*, 38, 12–18, https://doi.org/10.1016/j.wpi.2014.03.001

Orduna-Malea, E., Thelwall, M., Kousha, K. (2017). Web Citations in Patents: Evidence of Technological Impact? *Journal of the Association for Information Science & Technology*, 68(8), 1967–1974.

Parker, N., Landau, B. (2016). *Patent Nr US2016202868 (A1)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20160714&DB=EPODOC&locale=en_EP&CC=US&NR=2016202868A1&KC=A1&ND=4

Redl, G., Bogin, L., Parczewski, R. (2016). *Jak skutecznie patentować*. Warszawa: Narodowe Centrum Badań i Rozwoju.

Seo, Y. T. (2016). *Patent Nr KR20160129684 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20161109&DB=EPODOC&locale=en_EP&CC=KR&NR=20160129684A&KC=A&ND=4

Springer, R. (2017). The State of Ebooks 2017 [online]. *EContent Magazine*, [16.11.2019], http://www.econtentmag.com/Articles/Editorial/Feature/The-State-of-Ebooks-2017–115709.htm

Sugaya, S. (2016). *Patent Nr JP2016184246 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20161020&DB=EPODOC&locale=en_EP&CC=JP&NR=2016184246A&KC=A&ND=4

Śnieżko, A. (n.d.). *Jak chronić wynalazki w Polsce, w Europie i na świecie Podstawowe pojęcia i procedury* [online]. WTS Patent Attorneys, [16.11.2019], http://wtspatent.pl/wp-content/uploads/2014/05/pl1.pdf

Trappey, C. V., Wu, H.-Y., Taghaboni-Dutta, F., Trappey, A. J. C. (2011). Using Patent Data for Technology Forecasting: China RFID Patent Analysis. *Advanced Engineering Informatics*, 25(1), 53–64, https://doi.org/10.1016/j.aei.2010.05.007

UPRP (n.d.-a) Czym jest patent na wynalazek i prawo ochronne na wzór użytkowy? [online]. Urząd Patentowy Rzeczypospolitej Polskiej [16.11.2019], https://uprp.pl/czym-jest-patent-na-wynalazek-i-prawo-ochronne-na-wzor-uzytkowy/Lead05,150,1696,4,index,pl,text/

UPRP (n.d.-b). Informacje podstawowe [online]. Urząd Patentowy Rzeczypospolitej Polskiej, [23.11.2019], https://uprp.pl/podstawowe-informacje-o-ochronie-prawnej-wynalazkow-i-wzorow-uzytkowych/Lead05,23,1692,4,index,pl,text/

UPRP (n.d.-c). Procedura krajowa [online]. Urząd Patentowy Rzeczypospolitej Polskiej [16.11.2019], https://uprp.pl/procedura-krajowa/Lead05,24,1708,4,index,pl,text/

UPRP (2012). Wskazówki dla użytkowników MKP [online]. Urząd Patentowy Rzeczpospolitej Polskiej [16.11.2019], https://www.uprp.pl/uprp/redir.jsp?place=GalleryStats&id=64545

Vandeberg, R. L. J., Boon, W. P. C. (2009). Anticipating Emerging Genomics Technologies: The Role of Patents and Publication for Research and Policy Strategies. *Genomics, Society & Policy*, 5(2), 1–21.

Xie, Z., & Miyazaki, K. (2013). Evaluating the Effectiveness of Keyword Search Strategy for Patent Identification. *World Patent Information*, 35(1), 20–30.

Yang, S., Wang, S., Zhang, H. (2018). *Patent Nr CN108228668 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20180629&DB=EPODOC&locale=en_EP&CC=CN&NR=108228668A&KC=A&ND=4

Yun, K. O. (2016). *Patent Nr KR20160140527 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20161207&DB=EPODOC&locale=en_EP&CC=KR&NR=20160140527A&KC=A&ND=4

Zou, Y., He, L. (2019). *Patent Nr CN109389470 (A)* [online]. Espacenet [12.11.2019], https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20190226&DB=EPODOC&locale=en_EP&CC=CN&NR=109389470A&KC=A&ND=4

Wybrane aspekty rynku publikacji elektronicznych w świetle zgłoszeń patentowych w latach 2014–2018

Abstrakt

Cel/Teza: W artykule dokonano analizy zgłoszeń patentowych dotyczących publikacji elektronicznych. Celem badań była odpowiedź na pytanie, jak w latach 2014–2018 kształtowała się liczba wniosków patentowych, z jakich krajów pochodziły firmy zgłaszające wynalazki oraz jakie firmy były najbardziej aktywne na tym polu. Na podstawie analizy symboli Międzynarodowej Klasyfikacji Patentowej starano się także wyodrębnić pewne kluczowe zagadnienia, nad którymi pracują wynalazcy.
Koncepcja/Metody badań: Źródłem analizy były dane z bazy Lens.org obejmujące lata 2014–2018. Łącznie zidentyfikowano 1733 opublikowanych wniosków, dotyczących publikacji elektronicznych oraz urządzeń zapewniających do nich dostęp.
Wyniki i wnioski: Zwraca uwagę duża aktywność chińskich i koreańskich wynalazców oraz niska liczba zgłoszeń patentowych europejskich firm. Wnioskodawcy koncentrują się głównie na pewnych funkcjach ułatwiających interakcję użytkownika z urządzeniami cyfrowymi, ale stosunkowo liczne są także zgłoszenia wynalazków związanych z handlem elektronicznym czy wykorzystaniem publikacji elektronicznych do celów edukacyjnych.
Oryginalność/Wartość poznawcza: Analiza zgłoszeń patentowych może być pomocna w identyfikowaniu określonych zjawisk zachodzących obszarze technologii informacyjnych. Wskazuje zarówno kierunki prac badawczych, jak i kraje i podmioty, które odgrywają wiodącą rolę w rozwoju szeroko rozumianego obszaru publikacji elektronicznych.

Słowa kluczowe

Ebooki. Książki elektroniczne. Patenty. Wydawnictwa elektroniczne.

ADAM JACHIMCZYK PhD is Assistant Professor at the Department of Bibliography and Documentation at the Faculty of Journalism, Information and Book Studies at University of Warsaw. His research interests encompass the application of information technology in information activities. His publications include: Internetowe zasoby bibliograficzne instytutów badawczych oraz jednostek Polskiej Akademii Nauk (2016). In: J. Franke i J. Woźniak-Kasperek (red.) Bibliografia: historia, teoria, praktyka: praca zbiorowa. Warszawa: SBP; with M. and Z. Chrapek: Web directories: selected features and their impact on directory quality (2016). Program-Electronic Library and Information Systems, vol. 50 no 3; and Otwarte dane badawcze. Casus polskich instytutów badawczych (2015). Zagadnienia Naukoznawstwa, nr 4.

Contact to the Author
a.jachimczyk@uw.edu.pl
Department of Bibliography and Documentation
Faculty of Journalism, Information and Book Studies
University of Warsaw
Nowy Świat 69
00-046 Warsaw

Citation Type Analysis for *Zagadnienia Informacji Naukowej – Studia Informacyjne* (2016–2017)[1]

Marcin Roszkowski
ORCID 0000-0001-7396-4685
Department of Information Studies, Faculty of Journalism, Information and Book Studies, University of Warsaw

Abstract

Purpose/Thesis: The article studies the types of citations in research articles published in the journal *Zagadnienia Informacji Naukowej – Studia informacyjne (ZIN)*.
Approach/Methods: The study applied a method of analysis of citations, supplemented by a typology of citations established by B. Peritz. The research set consisted of 34 research articles published in ZIN in 2016–2017. The study required a manual identification and categorization of the citations which featured in the set, and correlating their types with the location in the article and the type of the article.
Results and conclusions: The results of the study showed significant variations in the distribution of the citations in the article, depending on the type of the article. A major part of the identified citations featured in the state of the art, and in general discussions. The authors studied often referred to other works to establish a general critical consensus regarding the research problems they analysed, to support their own hypotheses, or to suggest directions for further research. More than a half of the articles studied made no citations of methodological character.
Originality/Value: The article analyses the functions of the citations in the rhetoric structure of a research article concerned with information science, and offers a methodological critique of the research processes and tools.

Keywords

Citation analysis. Bibliometrics. Citation types. IMRaD. Zagadnienia Informacji Naukowej.

Received: 16 November 2019. Reviewed: 26 November 2019. Accepted: 16 December 2019.

1. Introduction

Citations in scholarly publications are an exponent of intertextual relations between the citing and cited texts. The basic research method employed in the analysis of such relations is citation analysis, which is defined as a bibliometric research method, "thanks to which, one studies various aspects of the information flow between citing and cited texts" (Żbikowska & Skalska-Zlat, 2017, 152).

[1] The article uses the research data collected by Aleksandra Wolińska for her Master's dissertation, *Citation Analysis in the Journal Zagadnienia Informacji Naukowej. Studia informacyjne*, 2016–2017 (*Analiza cytowań w czasopiśmie Zagadnienia Informacji Naukowej z lat 2016–2017*), completeded in the academic year 2018/2019 under the supervision of prof. dr. hab. Jadwiga Woźniak-Kasperek and dr. Marcin Roszkowski.

The premise of citation analysis is a particular epistemological attitude, which assumes that the number of citations of a given text reflects its influence on a given discipline, or a domain, and thus corresponds to its quality; as such, it allows quantification of many aspects of cognitive and social structure of science (Wade, 1975, 429). This method assumes a possibility of an objective evaluation of a publication's impact on the scientific discourse in a given discipline, based on the assumption that quantity corresponds to quality. Aside from recording the very fact a citation occurred, citation analysis accounts for further variables, such as age, carrier, and type of the cited work, the number and character of which determine the value of this kind of bibliometric research.

If the functionality of citations is considered, they might be seen as tools used in an exploration of information and evaluation of research activity (Di Iorio et al., 2013a). The first view assumes that an analysis of the network emergent in the bibliographic connections between a given article and the works cited is a source of information complementary to databases. This understanding of bibliographic citations imagines the reader following up on the citations with an aim of finding publications with an similar thematic scope. This technique fits within the exploratory model of information seeking termed "berrypicking", proposed by M. Bates (1989). The second view is based on a quantitative attitude, characteristic for bibliometrics. In this case, the citation number of a given publication, the citation number of all works of a given author, or the sum of citation numbers of all articles published in a given a journal, is an exponent of their quality, which is expressed with specific bibliometric indicators (e.g. Hirsch index for the authors, and impact factor for the journals). This view also allows a qualitative approach, i.e., basing on the works cited in a given text, one may attempt to establish their significance for a given research issue in relation to the previous codified knowledge in the given research area.

However, the method of citation analysis has its limitations, which come from its premises as an empirical approach based on positivist epistemology. Because the data analysed in such a study may not be complete, or representative, generalizing conclusions should be accepted only with care (Osareh, 1996, 220; Smith, 1981, 93). The crucial issue here is a functional and rhetorical understanding the nature of a bibliographic citation. Linda C. Smith (1981, 87–89) identified five assumptions frequently underlying citation analysis, and which may define its limitations.

(1) Citation of a document implies use of that document by the citing author.

According to Smith, this premise comprises two parts: (a) the author cited all, or the most important texts, which they used in the preparation for the writing of their article, and (b), all positions in the attached bibliography were used in the text. Even if the author made no errors in constructing of the bibliographic apparatus in the form of a list of cited texts, it is impossible to determine the influence of a cited work on the author's research without analysing the content of both the citing text, and the cited works.

(2) Citation of a document (author, journal, etc.) reflects the merit (quality, significance, impact) of that document (author, journal, etc.).

The premise is based on the previously mentioned assumption that a number citation of a given text corresponds to its status and significance in the scientific discourse. Without an additional insight into the reasons behind the choice to cite a given text, citation number should be understood rather as an orientation point.

(3) Citations are made to the best possible works.

According to Smith, if we assume that authors cite the most significant works from a given discipline in their own work, then we should also consider the availability of these works to the author, and its impact on the author's choice of source materials. Furthermore, Smith argues that the availability of source material is as important a factor in the citing author's decision process as their quality. In the time of open sciences and the transfer of information flow online, it seems that physical access to secondary sources should not determine the choice of source materials to such an extent.

(4) A cited document is related in content to the citing document; if two documents are bibliographically coupled, they are related in content; and if two documents are co-cited, they are related in content.

Again, there arises a question of reasons for citing a given text and the function of bibliographic citation. Smith proposes various interpretations of thematic overlap based on the relation of citation and co-citation. She refers to the results of research conducted by J. Barlup (1969), who questions scholars about the thematic overlap between their works and the texts alongside which they were cited. Barlup's results indicated that in 72% cases, the authors said that there was a strong thematic overlap, and in 5%, that there was none. Without a qualitative citation analysis, the citing relation might be interpreted only in terms of probability of a thematic overlap between the citing and cited texts, and between the texts cited simultaneously.

(5) All citations are equal.

According to Smith, citation analysis accepts the assumption that all citations, aside from self-citations, have the same status within the citing text. However, such an approach to analysis of intellectual influence of the cited works on the citing text is insufficient and too general. We should consider the function the cited texts fulfil in the rhetoric structure of the citing text. Accepting a structure based on the IMRaD model (Introduction, Methods, Results and Discussion), we might pose questions regarding the rhetorical function of citations in relation to these sections. Smith also refers to the construction of the bibliographic citation (a reference to a whole publication, a part of it, or a verbal citation) and the frequency with which a given work is cited within one text, thus suggesting the possibilities for a more concrete understanding of the citations' significance.

The five issues discussed by Smith constitute five research problems related to citation analysis as a research method, which derive from the necessity of qualitative variables to citation analysis. This means that we should ask questions regarding types of citations within a research publication, and the function they fulfil in its rhetoric and discursive construction.

2. Typologies of bibliographic citations

The researchers suggested the limitations of citation analysis deriving from its lack of insight into the nature of citing as early as in the 1960s, i.e., since the beginnings of the development of bibliometrics. H. White (2004) argues that the paper *Can Citation Indexing*

Be Automated by E. Garfield (1965) constituted first attempt to establish a typology of citations for the purpose of a qualitative analysis. Garfield's typology corresponds to the motivations which might dictate the authors' construction of citations. It was based on the analysis of linguistic means of expressing the context in which the citation occurred. Garfield did not characterize a research sample; he only suggested that he based his research on the publications in the Science Citation Index, i.e., on the works from the exact sciences. Garfield's scheme accounts for 15 types of citations, which feature valorizing attitudes, and pragmatical approaches, instrumental to the use of existing scholarship. Although Garfield's typology was the first attempt to qualitatively systematize bibliographic citations, White (2004, 107) argues that it has never been applied to empirical research.

In 1975, M. J. Moravcsik and P. Murugesan (1975) proposed a multi-dimensional categorization of bibliographic citations for the purpose of citation analysis of the articles from the journal *Physical Review*. Their citation typology was constructed before the empirical research itself. The conceptual core of their typology was formed by four questions posed in citation analysis, which indicated four interpretative dimensions (Moravcsik & Murugesan, 1975, 88; Swales, 1986, 41):

(1) Is the citation conceptual or operational? Does the citing work refer to another in connection with a concept or theory, or is the citation made in connection with a technique or a method?
(2) Is the citation organic or perfunctory? In other words, is the referring work built on the foundations provided by the citation, or is it an alternative to it?
(3) Is the citation evolutionary or juxtapositional? Is the cited work needed for the understanding of the citing work or is it mainly an acknowledgement that some other work in the same general area has been performed?
(4) Is the citation confirmative or negational? Is there any dispute about the correctness of the findings proposed in the cited work?

In 1977, Ina Spiegel-Rösing (1977) published the results of research which employed citation analysis to study 66 texts published in the journal *Science Studies* in the years 1971–1974. The thematic scope of the journal fit within social sciences, and encompassed the issues of science studies, the organisation and infrastructure of science, as well as science communication. Basing on her empirical research, the author identified 13 types of citations which occurred in the data sample. Spiegel-Rösing's scheme included types of citations which indicated the function which the cited work fulfils in the discursive structure of the article, as well as the author's evaluation of the cited text.

B.C. Pertiz (1983) established a typology of bibliographic citations with an intention to make it adequate to the specifics of the scientific discourse in the disciplines of social sciences, and to allow an easy application to empirical research. The starting point of her work was a categorization of bibliographic citations proposed by T. Hodges (1972). His scheme consisted of eight citation types, which mostly described the functions which the cited work fulfilled in the citing text; their order corresponded to the logic of an argument presented in a scientific publication. The main limitation of Peritz's typology was the emphasis on the specifics of bibliographic citations in the publications recording the results of empirical research.

Nanba, Kando and Okamura (2000) established a citation typology to meet the needs of an information system for the automatized classification of scientific publications. They

established their own typology, which simplified M. Weinstock's scheme (1971), based on the linguistic analysis of the fragments of publications where the citations featured:

(1) Type B: citations indicating theories, methods, and concepts of other authors.
(2) Type C: citations indicating the problems or gaps in related scholarship.
(3) Type O: citations other than B and C.

The research of Nanba, Kando and Okumura identified the phrases employed by the authors in the citation process, which were then classified according to these types.

This short review of citation typologies should also mention the attempt to formalize the citation types into an ontology. CiTO, Citation Typing Ontology (Shotton, 2010), was established within the framework of semantic publishing, which is an interdisciplinary research area focused on the application of semantic technologies to the information flow in science. CiTo distinguishes 96 types of bibliographic citations, which are classified either as rhetorical or as factual. The first category is further divided into positive, negative, and neutral. CiTO offers a detailed set of citation types, which makes it a very expressive tool, but might be prove challenging when applied in empirical research.

This section discusses only a few significant attempts to determine the nature of bibliographic citation in scientific publications. At least two crucial problems might be identified here: the means of constructing such typologies, and their application in empirical research. The scholars cited here took various approaches to the first problem (empirical, rationalist), and to the specificity of the scientific discourse particular to a given discipline, which impacts the types and functions of bibliographic citations. The second problem is related to conducting the citation analysis using information technologies which would account for qualitative factors. Because manual citation analysis is time-intensive, the researchers direct a large part of their attention and energy towards the automatization of this process.

3. The aim and methodology of research

The main aim of the research presented in this study was the characterization of various types of bibliographic citations in the research articles published in the journal *Zagadnienia Informacji Naukowej – Studia Informacyjne* (*ZIN*) in the years 2016–2017. We studied those texts which presented the results of the authors' own research, published in the section 'Theses'. Our method relied on the application of citation analysis (Smith, 1981), supplemented by a qualitative analysis, i.e., an application of an existing citation typology. For the purposes of this study, we used the typology established by B. Peritz (1983), which emphasizes the functions the citations fulfil in the rhetorical structure of a scientific publication. The Peritz's scheme was established to be easily applicable to a manual citation analysis, mostly in the study of texts from the discipline of social sciences. This scheme consists of eight types of citations:

C1. Setting the stage for the present study. Citations to texts in order to justify the proposed research aims.

C2. Background information. Citations to texts presenting general knowledge on the subject of the issues discussed in the citing text, as well as factual information.

C3. Methodological. Citations of works describing methodological issues featuring in the citing text.

C4. Comparative. Citations to other studies in order to compare the results presented in the citing text.

C5. Argumental, speculative, hypothetical. Citations to other texts in order to support the proposed hypotheses and to suggest directions for further research.

C6. Documentary. Citations to sets of research data, i.e., raw data.

C7. Historical. Citations to the work of other scholars in order to reconstruct the history of research of a given issue or as a sign of respect for their pioneering work in a given area.

C8. Casual. Citations indicating thematically connected works, without the comparative aspect.

Every bibliographic citation in the data sample was manually classified as one of the eight types from Peritz's model. Additionally, its position in the citing text was noted, with the IMRaD model as a starting point. Although the guidelines for the authors publishing in ZIN do not oblige the authors to format their texts according to the IMRaD scheme, they require that the authors prepare a structured abstract referencing the key elements of this model. Therefore, it is likely that the body of an article published in ZIN will contain sections typical for the IMRaD model, even if they will not be explicitly identified as such by their headings. Therefore, we assumed that it is possible to analyze the structure of a research article published in ZIN using the IMRaD model and thus to determine the position of the citation in relation to the model's elements. We also assumed that the type of the research presented in a given article may have an impact on the type of bibliographic citations. To account for this, we employed a simplified typology of research publications, which consists of four types of publications (1) theoretical (presenting theoretical and methodological considerations); (2) empirical (presenting results of the research based on the collected research material); (3) review (presenting results of a study of critical literature); (4) other. The fourth category had a complementary character, and was introduced to close the set.

The accepted two-year long range of the research material was determined by the main research aim, i.e., characterization of functions which the citations fulfilled in the citing texts. Therefore, we resigned from an in-depth quantitative citation analysis, because the narrow time span would not justify a generalization of the results achieved.

The basic metadata regarding the articles and their appendiced bibliographies was collected from the Central European Journal of Social Sciences and Humanities (CEJSH). Then, a simplified bibliographic record of every article was entered into a spreadsheet, supplemented by the information regarding the bibliographic citations identified in its content. The primary source of the information employed in the citation analysis were the texts of the articles published electronically as pdf files. Each citation was interpreted in the context of its appearance, which necessitated a familiarity with the fragment of the article where it appeared.

An additional research aim was to test the usefulness of Pertiz's typology for the empirical research. It also involved the issue of its adequacy to the publications from the discipline of information science, as well as possible interpretative problems in the process of categorizing the citations.

The methodological premises accepted in this study have their limitations. Firstly, subjective factors may influence the process of categorizing bibliographic citations, which was conducted by one person, a student in the second year of Library and Information Science MA at University of Warsaw. There is a possibility that a classification conducted

by a larger number of people would bring different results. Secondly, Pertiz's typology has limitations as well. The author showed that although her typology has a wide application to citation analysis of articles from the generally understood discipline of social sciences, its recommended application is to the texts presenting the results of empirical research, as it is less suitable for the citation analysis of articles concerned with history and methodology alone (Peritz, 1983, 304). Because the types established by Peritz are highly general, and only C6 clearly refers to citing quantitative data, it has a high research value. Therefore, we chose to apply it in our citation analysis.

4. Results

In the years 2016–2017, ZIN published 34 research papers in total, out of which 16 (47%) were classified as presentation of the results of empirical research, 11 (32%) as theoretical reflection, and seven (21%) as a review of scholarship. The authors cited 1022 works in total, with the total number of citations being 1825. It means that certain works were cited more than once in a single text. Therefore, in the further discussion of results we will define the citation number in a given article as a total number of citations in the text, rather than a number of positions in the bibliography. The results of quantitative analysis show that in 71% cases a given work was cited only once, in 13% it was cited twice, in 5% – three times, in 4% – four times, and in 2% – five times. In 54 cases (5%), a text was cited more than five times.

Table 1 presents the interquartile range of the dispersion of the citations per article. This method allows for an analysis of the dispersion of the citations above and below the median value, which equalled 41 citations per article. The lowest number of citations (2) was noted for the article by Mariusz Luterek, *Polish Public Libraries as Intermediaries in Accessing Information and Public Services (e-Government) in the Opinion of Librarians*, which recorded the results of a survey conducted among librarians. The highest number of citations (236) occurred in a text by Łukasz Opaliński, *Bibliometric Methods to Foresee and Assess the Development of Scientific Disciplines. Literature Analysis. Part 2. Comparisons, Hybrid and Statistical Methods, Analysis of Patents and Main Paths of Literature Development and Other Original Approaches in Terms of Predictive Methodology*[2]. The text was an in-depth analysis of the studies on the application of quantitative methods to predicting the development of scientific disciplines.

The data presented in Table 1 indicates that the value of the first quartile was 25, which means that 25% of the texts in the research set, cited 25 or less works. The value of the upper quartile was 62, which means that 75% of the texts in the research set cited 62 articles, or less. The average citation number in the data sample was 53, but the high variance in the number of the works cited (between 2 and 236) and this, a high value of standard deviation (50), does not allow us to use this number to characterize the research set.

[2] Oryg. *Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych. Analiza piśmiennictwa. Część 2. Badania porównawcze, hybrydowe, statystyczne, analizy dokumentów patentowych, ścieżek rozwoju dyscyplin oraz pozostałe oryginalne podejścia metodologiczne.*

Tab. 1. Statistical dispersion of citations per article

Minimum value	2
Lower quartile (Q1)	25
Median (Q2)	41
Upper quartile (Q3)	62
Maximum value	236

We correlated the number of the citations with the type of the article citing to see the distribution of citations in articles presenting different types of research processes. The results show that the average citation number was highest for the review articles (116), which is an understandable consequence of the method employed in such publications. The average citation number for theoretical texts was 32, and for presentations of results of empirical research was 41. The number for theoretical texts was lower than the median number for the whole data set, while the number for the presentations of results of empirical research was equal to the median.

Another variable we accounted for in our research was the position of the citation relative to the structure of the article. According to the methodological premises presented in the previous section, we accepted the IMRaD model as a starting point. Figure 1 shows the percentage distribution of citations taking into account the types of the articles and location of the citation.

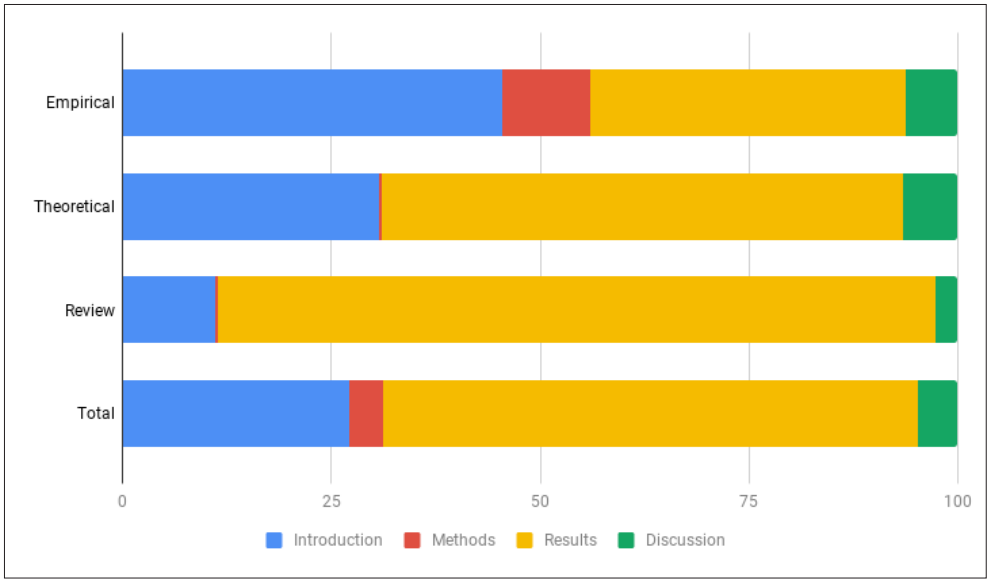

Fig. 1. Percentage distribution of citations taking into account the types of the articles and location of the citation

The analysis of the percentage distribution of citations taking into account the types of the articles and location of the citation shows that 64% of citations occur in the section devoted to the presentation of the results of the conducted research. Almost one third (27%) of the citations to secondary sources occurs in the introduction. A relatively small part of the citations occurs in sections on the methodology (4%) and conclusions drawn from the conducted research (5%). However, this image changes when we correlate the position of the citation with the type of the article. We see significant variations in the distribution of the citations relative to the structure of the article. In review articles, a relatively high proportion of citations (86%) occurs in the section presenting the results of research, which is understandable, as it is the secondary sources which constitute the material for the research presented in these publications. 62% of the citations occur in this section of theoretical text, and 38% in the articles presenting results of empirical research. The section devoted to the presentation of conclusions drawn from the research features the least citations to secondary sources, irregardless of the type of the article.

After analyzing the citations to secondary sources occurring in the body of the citing article, we classified every citation as one of the types identified in the Peritz's typology. Table 2 presents the quantitative and percent distribution of the citation types in the data sample.

Tab. 2. Citation distribution according to the types

Citation type	Number of occurrences	Percentage of occurrences
C1	192	11
C2	129	7
C3	62	3
C4	89	5
C5	148	8
C6	75	4
C7	832	46
C8	269	15

The highest number of citations belongs to the type C7, the category of historical citations, which contribute to a review and discussion of the issues in a given field. They constituted almost a half (46%) of all the citations recorded in the data sample. This citation type indicates that the author considers the cited work as influential for the growth of knowledge in a given area. Peritz (1983, 305) notes that it might be difficult to distinguish between a historical (C7) and a preparatory (C1) citation. She suggests that type C7 is not directly related to the research questions posed in the citing article, which would be characteristic for type C1. The following most dominant citation types were casual citations (C8 – 15%), and preparatory citations (C1 – 11%), which mostly serve to identify and justify the research problem. 8% of the citations belongs to type C5, indicating a citation to secondary sources to support the author's hypotheses and to suggest directions for further research. A relatively small proportion of the citations belongs to type C4 – comparative (5%), C6 – documentary (4%) and C3 – methodological (3%).

It should be mentioned that not all citation types occurred in every article analysed, which might influence the interpretation of the data from Table 2. All types of citations (C1–C8) featured only in two articles. Four was the average number of citation types occurring in a single article. Table 3 shows the distribution of citation types per article.

Tab. 3. Distribution of citation types per article

Citation type	The number of articles where the citation type occurred	The percentage of articles where the citation type occurred
C1	21	62
C2	26	76
C3	14	41
C4	11	32
C5	26	76
C6	18	53
C7	34	100
C8	24	71

The data presented in the Table 3 indicates that only historical citations (C7), contributing to a review and discussion of issues within a given field, occurred in all articles in the research set. The authors often referred to secondary sources to give a view of a general knowledge on the subject of the research problems they studied (C2) and to support their hypotheses and suggest the directions for further research (C5). Surprisingly, more than a half of articles (59%) featured no methodological citations (C3).

To achieve a more in-depth understanding of the distribution of the citation types, we decided to correlate this variable with the type of the publication. Table 4 shows the percent distribution of the citations in the articles of various types.

Tab. 4. Percent distribution of the citation types in the articles of various types

	Empirical	Theoretical	Review
C1	16.25	15.38	4.22
C2	5.94	16.81	3.98
C3	8.28	1.42	0.50
C4	12.50	1.14	0.62
C5	8.91	10.83	6.58
C6	6.88	3.13	2.48
C7	31.41	41.03	60.50
C8	9.84	10.26	21.12

It is evident that the distribution of preparatory citations (C1) varies between review articles, and the empirical and theoretical articles: they constitute a much smaller portion of all citations in review articles. Theoretical texts contain a higher number of citations

presenting general knowledge of the issues they discuss (C2) than other types of the articles. Methodological citations (C3) occur in the empirical articles more often than in others, as do comparative citations (C4). We noted a relatively low variance in the distribution of the argumental citations (C5). Documentary citations (C6) feature in empirical articles more often, which is understandable if we consider the specific methodology of these texts. The specific methodology also explains the prevalence of historical citations (C7) in review articles. The casual citations (C8), indicating citations to thematically similar works, occurs mostly in the review articles.

5. Conclusion

The presented results allow us to formulate conclusions of two kinds, firstly those immediately related to the results, and secondly, those related to the methodology.

Despite a small size of the data sample, the achieved results allow us for a certain degree of generalization, which makes it possible to sketch a profile of the articles from the discipline of information science published in *ZIN*, accounting for the functions the citations they feature fulfil. The results of our research show significant variations in the distribution of the citations relative to the structure of the article depending on the type of the article. The citations in the works presenting results of an empirical research are usually concentrated in the introduction and in the discussion of the results, whereas in the review and theoretical articles, they occur in the section presenting the results of the research. We have also noted that only a small portion of all citations occurred in the section concerned with the methodology. The results we achieved are mostly aligned with the citation distribution relative to the sections of IMRaD in scientific publications following the scheme of John Swales (Campbell, 2013; Swales, 2004). On the basis of a study of the genres of scientific publications, Swales concluded that a high number of the citations is characteristic for the introductions, and for the sections presenting the achieved results; a low number – for the section devoted to the methodology, and irregular – for the discussion and results sections. The results of our study showed that Swales' model holds up only for the introductory and methodological sections. Our research showed that the highest portion of citations in the analysed set of the articles occurred in the section presenting the achieved results, and the lowest – in the section devoted to the discussion and conclusions. However, we should be careful with the interpretation of these results. We should bear in mind that the model of an article based on the IMRaD structure was first established in relation to the publications from the disciplines of exact sciences, and is not necessarily the best framework for an approach of rhetorical structure of the publications from the discipline of information science.

With regards to the function the citations fulfil in an article, we may suggest several conclusions. The authors do not always justify their selection of a research question by situating it in the larger context, and do not always indicate the source of their methodological consideration. The methodological citations are particularly marginalized in theoretical and review articles. However, authors often refer to works presenting general knowledge, whether to define the terms they are borrowing, or to refer to scientific laws and theories; they always refer to the historical aspect of the problems they discuss. It is apparent that

they wish to participate in the scientific discourse through a confrontation of their theses with the work of other scholars. However, they rarely compare the results of their research.

As far as the methodology is concerned, we found Pertiz's typology an efficient research tool allowing for a categorization of all the citations found in the articles studied. For the purpose of the current study, we accepted a working uncertainty coefficient. We marked the citations which posed difficulties to an interpretation following Pertiz's model; they constituted 3.5% of all citations identified in the data sample. It shows that Peritz's typology is a relatively appropriate research tool for the analysis of the types of bibliographic citations. Considering that we noted a small uncertainty while categorizing the citations, we argue that a manual study of citation types requires a several iterations for the identification and elimination of interpretative problems, as suggested earlier by other authors (Di Iorio et. al., 2013a; 2013b).

References

Barlup, J. (1969). Relevancy of Cited Articles in Citation Indexing. *Bulletin of the Medical Library Association*, 57(June), 260–263.

Bates, M. J. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Information Review*, 13(5), 407–424, https://doi.org/10.1108/eb024320

Campbell, K.S. (2013). The Genre of Research Articles [online]. Pros Write, [10.11.2019], https://proswrite.com/2013/04/01/the-genre-of-research-articles

Di Iorio, A., Nuzzolese, A. G., Peroni, S. (2013a). Characterising Citations in Scholarly Documents: The CiTalO Framework. In: P. Cimiano, M. Fernández, V. Lopez, S. Schlobach, J. Völker (eds.). *The Semantic Web: ESWC 2013 Satellite Events. ESWC 2013. Lecture Notes in Computer Science, vol 7955* (66–77). Cham: Springer, https://doi.org/10.1007/978–3-642–41242-4_6

Di Iorio, A., Nuzzolese, A. G., Peroni, S. (2013b). Towards the Automatic Identification of the Nature of Citations. In: A.G. Castro, C. Lange, P. Lord, R. Stevens (eds.). *Proceedings of the 3rd Workshop on Semantic Publishing (SePublica 2013)* (63–74). Birmingham.

Garfield, E. (1965). Can Citation Indexing Be Automated? *Statistical Association Methods for Mechanized Documentation, Symposium Proceedings*, 269, 189–192.

Hodges, T. L. (1972). *Citation Indexing: Its Potential for Bibliographical Control.* Berkeley: University of California.

Moravcsik, M. J., Murugesan, P. (1975). Some Results on the Function and Quality of Citations. *Social Studies of Science*, 5(1), 86–92.

Nanba, H., Kando, N., Okumura, M. (2000). Classification of Research Papers Using Citation Links and Citation Types: Towards Automatic Review Article Generation. *Advances in Classification Research Online*, 11(1), 117–134, https://doi.org/10.7152/acro.v11i1.12774

Osareh, F. (1996). Bibliometrics, Citation Anatysis and Co-Citation Analysis: A Review of Literature II. *Libri*, 46(4), 217–225, https://doi.org/10.1515/libr.1996.46.4.217

Peritz, B. C. (1983). A Classification of Citation Roles for the Social Sciences and Related Fields. *Scientometrics*, 5(5), 303–312, https://doi.org/10.1007/BF02147226

Shotton, D. (2010). CiTO, the Citation Typing Ontology. *Journal of Biomedical Semantics* [online], 1, S6, https://doi.org/10.1186/2041–1480-1-S1-S6

Smith, L. C. (1981). Citation Analysis. *Library Trends*, 30(1), 83–106.

Spiegel-Rösing, I. (1977). Bibliometric and Content Analysis. *Social Studies of Science*, 7(1), 97–113.

Swales, J. (1986). Citation Analysis and Discourse Analysis. *Applied Linguistics*, 7(1), 39–56, https://doi.org/10.1093/applin/7.1.39

Swales, J. M. (2004). *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press, https://doi.org/10.1017/CBO9781139524827

Wade, N. (1975). Citation Analysis: A New Tool for Science Administrators. *Science*, 188(4187), 429–432, https://doi.org/10.1126/science.188.4187.429

Weinstock, M. (1971). Citation Indexes (reprint). In: *Encyclopedia of Library and Information Science*, vol. 5 (16–41). New York: Marcel Dekker.

White, H. D. (2004). Citation Analysis and Discourse Analysis Revisited. *Applied Linguistics*, 25(1), 89–116, https://doi.org/10.1093/applin/25.1.89

Żbikowska, A., Skalska-Zlat, M., eds. (2017). Analiza cytowań. In: *Encyklopedia książki. T. 1: Eseje, A–J*. (152–153). Wrocław: Wydaw. Uniwersytetu Wrocławskiego.

Analiza typów cytowań bibliograficznych w *Zagadnieniach Informacji Naukowej* w latach 2016–2017

Abstrakt

Cel/Teza: Celem artykułu jest zbadanie typów cytowań bibliograficznych w artykułach badawczych opublikowanych w czasopiśmie *Zagadnienia Informacji Naukowej – Studia Informacyjne* (ZIN).
Koncepcja/Metody: W badaniach wykorzystano metodę analizy cytowań bibliograficznych rozszerzoną o typologię cytowań opracowaną przez B. Peritz. Przedmiotem badań były 34 artykuły badawcze opublikowane w ZIN w latach 2016–2017. Koncepcja badań zakładała manualną identyfikację i kategoryzację cytowań bibliograficznych, które wystąpiły w zbiorze badawczym oraz korelację typów cytowań z miejscem wystąpienia w strukturze artykułu oraz typem artykułu.
Wyniki i wnioski: Wyniki badania pokazały istotne różnice w dystrybucji cytowań w strukturze artykułu naukowego w zależności od jego typu. Dominującym typem cytowań w zbiorze badawczym były odwołania w częściach artykułu stanowiących przegląd piśmiennictwa i omówienie zagadnień. Autorzy często odwoływali się do innych prac w celu wskazania na istniejącą wiedzę ogólną na temat opisywanych problemów badawczych oraz w celu wsparcia stawianych przez siebie hipotez i wyznaczania dalszych kierunków badań. W ponad połowie artykułów nie wystąpiły cytowania o charakterze metodologicznym.
Oryginalność/Wartość poznawcza: Wartość poznawczą artykułu stanowi z jednej strony analiza rzeczywistych funkcji, jakie pełnią cytowania w strukturze retorycznej publikacji naukowych z obszaru informatologii, a z drugiej – krytyka metodologiczna zarówno procesu, jak i narzędzia badawczego.

Słowa kluczowe

Analiza cytowań. Bibliometria. IMRaD. Typy cytowań. Zagadnienia Informacji Naukowej.

MARCIN ROSZKOWSKI, PhD is Assistant Professor at the Department of Information Studies at the Faculty of Journalism, Information and Book Studies of University of Warsaw. He is a member of International Society for Knowledge Organization (ISKO) and DBpedia ontology committee. His research interests encompass issues in knowledge organisation and representation of information in the online environment with a particular focus on the conceptual modelling of information systems, as well as on metadata and online ontologies. His most important publications include: Organizacja informacji i wiedzy (2016, co-author B. Sosińska-Kalata; published in: W. Babik (ed.) Nauka o informacji); The Role of Digital Libraries as Virtual Research Environments for the Digital Humanities *(2016, co-author W. Mustafa El Hadi; published in: J. A. C. Guimarães, S. Oliveira Milani, V. Dodebei (eds.),* Advances in Knowledge Organization*, vol. 15. Ergon Verlag);* Kartoteka haseł wzorcowych jako usługa sieciowa – automatyczna identyfikacja nazw

osobowych z wykorzystaniem kartoteki VIAF *(2016; published in: J. Woźniak-Kasperek, J. Franke (eds.),* Bibliografia – teoria, praktyka, dydaktyka*).*

Contact to the Author:
m.roszkowski@uw.edu.pl
Department of Information Studies
Faculty of Journalism, Information and Book Studies
University of Warsaw
Nowy Świat 69
00-046 Warsaw

Wskazówki dla autorów

Redakcja *Zagadnień Informacji Naukowej – Studiów Informacyjnych* przyjmuje wyłącznie teksty wcześniej nieopublikowane i niezłożone do druku w innych czasopismach lub pracach zbiorowych. Przyjmowane są: oryginalne rozprawy i prace badawcze, recenzje oraz sprawozdania z konferencji i innych wydarzeń naukowych.

Teksty artykułów są recenzowane zgodnie z zasadami *double-blind peer review*. Zapewnienie anonimowości tekstów przekazywanych do recenzji wymaga, aby w tekście artykułu w żadnym miejscu nie była umieszczona informacja umożliwiająca identyfikację autora.

Każdy artykuł recenzowany jest na podstawie jednolitego formularza przez dwóch recenzentów dobieranych spośród specjalistów problematyki w nim poruszanej. Każda recenzja zawiera jednoznaczne wskazanie czy tekst rekomendowany jest do publikacji w *Zagadnieniach Informacji Naukowej*. Podstawowymi kryteriami oceny artykułu są: zgodność tematu z profilem czasopisma, wartość merytoryczna, organizacja logiczna i forma językowa tekstu.

O przyjęciu tekstu do publikacji autorzy informowani są w ciągu 10 tygodni od otrzymania go przez Redakcję. Redakcja przyjmuje wyłącznie teksty przygotowane zgodnie z zasadami przedstawionymi poniżej. Teksty należy nadsyłać na adres e-mail: **zin@uw.edu.pl**

1. Zasady ogólne

1.1. Format

Wszystkie pliki (tekst artykułu, materiały ilustracyjne) należy przesyłać jako dokumenty edytora MS WORD w formacie RTF. Zaleca się stosować w tekście czcionkę Times New Roman 12 pkt, interlinię 1.5. Tytuł artykułu należy wyróżnić czcionką Times New Roman 16 pkt. **Nie należy używać automatycznych stylów.**

Materiały ilustracyjne, wstawione w treść artykułu, dodatkowo należy przesyłać również w formacie JPG. Załączniki powinny być ponumerowane według kolejności występowania w tekście oraz zawierać nazwę, np.: *1. Tab. 1. Poziomy metadanych* albo *3. Rys. 1. Mapa myśli*.

1.2. Długość tekstu

Artykuł nie powinien przekraczać 40 000, a recenzja lub sprawozdanie 14 000 znaków (ze spacjami).

1.3. Strona tytułowa

Autorzy artykułów proszeni są o przygotowanie odrębnej strony tytułowej, zawierającej:

- **tytuł artykułu (w językach polskim i angielskim)**
- **dane autora (imię i nazwisko, afiliacja – w językach polskim i angielskim, identyfikator ORCID)**
- **adres *e-mail***
- **adres do korespondencji**
- **notę biograficzną autora (patrz niżej)**
- **abstrakt ustrukturyzowany (patrz niżej)**
- **słowa kluczowe (patrz niżej)**
- **oświadczenie o oryginalności tekstu (patrz niżej).**

Zgodnie z zasadami przeciwdziałania zjawiskom *ghostwritingu* i *guest authorship* Redakcja prosi również, aby na tej stronie ujawnione zostały nazwiska i afiliacje wszystkich osób, które przyczyniły się do powstania artykułu, ich rola i udział w przygotowaniu publikacji (kto jest autorem koncepcji, założeń, metod itp. wykorzystywanych w pracy zgłoszonej do druku; procentowy udział w przeprowadzonych badaniach i opracowaniu artykułu). Redakcja prosi także o podanie informacji o źródłach finansowania publikacji, wkładzie instytucji naukowo-badawczych, stowarzyszeń i innych podmiotów (*financial disclosure*).

1.4. Nota biograficzna autora / autorów

Na stronie tytułowej należy umieścić zwięzłą notę biograficzną (ok. 70 słów) każdego autora artykułu. Nota powinna zawierać następujące informacje: tytuł / stopień naukowy lub zawodowy autora, aktualne miejsce pracy i zajmowane stanowisko; specjalności naukowe lub zawodowe, najważniejsze publikacje (max. 3). Opisy publikacji powinny być sporządzone zgodnie z zasadami APA Style 6th.

1.5. Abstrakt ustrukturyzowany

Na stronie tytułowej należy umieścić abstrakt w języku polskim o objętości ok. 100 słów (ok. 1 tys. znaków) oraz jego przekład na język angielski. W abstrakcie należy wyróżnić co najmniej cztery spośród następujących kategorii informacji:

- **Cel/Teza | Purpose/Thesis** *(obowiązkowo)*
- **Koncepcja/Metody badań | Approach/Methods** *(obowiązkowo)*
- **Wyniki i wnioski | Results and conclusions** *(obowiązkowo)*
- **Ograniczenia badań | Research limatations** *(opcjonalnie)*
- **Zastosowanie praktyczne | Practical implications** *(opcjonalnie)*
- **Oryginalność/Wartość poznawcza | Originality/Value** *(obowiązkowo)*

1.6. Słowa kluczowe

Na stronie tytułowej artykułu należy umieścić od 4 do 10 słów kluczowych, w formie fraz nominalnych w mianowniku liczby pojedynczej, których pierwszy wyraz zapisany jest wielką literą, uporządkowanych alfabetycznie, rozdzielonych kropkami. Słowa kluczowe należy podać w językach polskim i angielskim.

1.7. Oświadczenie o oryginalności tekstu

Na stronie tytułowej artykułu należy umieścić oświadczenia autora /autorów, że tekst przedstawiany Redakcji *Zagadnień Informacji Naukowej – Studiów Informacyjnych* nie był dotychczas opublikowany ani zgłoszony do publikacji w żadnym innym czasopiśmie lub pracy zbiorowej. Jeśli tekst był prezentowany na konferencji, należy podać jej szczegółowe dane wraz z ewentualnymi informacjami o publikacji materiałów konferencyjnych. Jeśli artykuł jest częścią przygotowywanej do druku książki, należy podać jej dane oraz planowany termin publikacji.

2. Zasady opracowania artykułu

2.1. Organizacja i podział tekstu

Tekst artykułu powinien być podzielony na podrozdziały zaopatrzone w tytuły. W pierwszej części pod nagłówkiem **Wprowadzenie** zaleca się umieścić informacje wprowadzające w problematykę prezentowaną w artykule. W części ostatniej – pod nagłówkiem **Wnioski** lub **Zakończenie** – wnioski końcowe i podsumowanie przedstawionych rozważań.

Dopuszcza się stosowanie do trzech poziomów podziału tekstu, każdy wyodrębniony własnym śródtytułem i opatrzony oznaczeniem numerycznym zgodnie z następującymi regułami:

1. Pierwszy poziom podziału
1.1. Drugi poziom podziału
1.1.1 Trzeci poziom podziału

2.2. Przypisy

Nie stosuje się przypisów bibliograficznych. Odesłania do wykorzystanej literatury należy przygotować zgodnie z edytorskimi standardami tekstu naukowego APA 6th (patrz niżej).

Przypisy zawierające komentarze, dygresje, objaśnienia i inne dodatkowe informacje należy umieszczać na dole strony i numerować liczbami arabskimi; zaleca się ograniczenie liczby przypisów do niezbędnego minimum.

2.3. Pisownia tytułów w tekście artykułu

Tytuły wystaw, konferencji, programów itp. powinny być zapisane w cudzysłowie. Tytuły publikacji (książek, czasopism, artykułów itp.) należy wyróżnić kursywą.

2.4. Wyróżnienia w tekście

W tekście można stosować wyróżnienia za pomocą czcionki półgrubej (bold).

2.5. Materiały ilustracyjne i ich oznaczanie w tekście

Materiały ilustracyjne (tabele, wykresy itp.) powinny być przygotowane w odcieniach szarości lub kolorystyce czarno-białej. Wszystkie tego typu materiały należy oznaczyć wskazaniem rodzaju materiału (np. Tabela, Rysunek, Fotografia, Wykres), jego numeru w tekście oraz jego tytułu (np. Tab. 1. Poziomy metadanych). W odpowiednich miejscach tekstu artykułu należy umieścić odesłania do informacji prezentowanych w formie ilustracji, używając w tym celu skrótu określenia rodzaju ilustracji oraz jej numeru (np. zob. Tab. 1, zob. Wykr. 5).

2.6. Cytowanie wykorzystanej literatury w tekście i bibliografia załącznikowa

Cytowania w tekście i bibliografię załącznikową należy przygotować zgodnie ze standardami edytorskim publikacji naukowych APA 6th. W bibliografii załącznikowej mogą być umieszczone wyłącznie opisy publikacji cytowanych w tekście artykułu.

Publikacje należy cytować w tekście używając odsyłaczy w formie: (nazwisko, rok wydania), np. (Dembowska, 1991); gdy publikacja ma dwóch autorów należy podać obydwa nazwiska połączone znakiem ampersand (nazwisko1 & nazwisko2, rok), np. (Cisek & Sapa, 2007); gdy publikacja ma trzech i więcej autorów należy podać nazwisko pierwszego autora, skrót *et. al.* i rok wydania (nazwisko1 et al., rok), np. (Berners-Lee et al., 2001); gdy publikacja jest pracą zbiorową, należy podać nazwisko redaktora, skrót red. i rok wydania (nazwisko, red., rok), np. (Kocójowa, red., 2005). Jeśli w publikacji nie wskazano nazwiska autora lub redaktora, należy podać pierwszy wyraz tytułu i rok wydania (Wyraz, rok), np. (Biblioteki, 1976). Odwołania do określonych stron cytowanych tekstów należy podawać w formie: (Dembowska, 1991, 15), albo (Cisek & Sapa, 2007, 40–42), (Dervin & Nilan, 1986, 3) albo (Kocójowa, red., 2005, 18).

Opisy bibliograficzne wykorzystanych publikacji należy umieścić na końcu tekstu w układzie alfabetycznym, bez numeracji pozycji, pod nagłówkiem **Bibliografia**.

Opisy autorskich książek i artykułów umieszcza się pod nazwiskiem pierwszego autora. Opisy prac zbiorowych należy umieszczać pod nazwiskiem redaktora, po którym podaje się skrót *red.* lub *ed.* Jeśli w publikacji nie wskazano autora lub redaktora pracy zbiorowej, jej opis należy umieścić pod pierwszym wyrazem tytułu.

Tytuły książek i czasopism należy zapisać kursywą, tytuły artykułów w czasopismach i artykułów lub rozdziałów w książkach – czcionką prostą.

W opisach artykułów w pracach zbiorowych stosuje się oznaczenie skrótu „W" dla publikacji w języku polskim i „In" dla publikacji w językach obcych.

Opisy prac tego samego autora powinny być uporządkowane według chronologii wstępującej, a w każdym z nich należy powtórzyć nazwisko i inicjał (inicjały) autora. Prace tego samego autora opublikowane w tym samym roku należy uporządkować w kolejności alfabetycznej tytułów i oznaczać wg zasady:

Dembowska, M. (1976a) ...,
Dembowska, M. (1976b) ..., itd.

2.6.1 Przykłady redagowania opisów bibliograficznych

KSIĄŻKA

Breslin, J.G., Passant, A., Decker, S. (2009). *The Social Semantic Web*. Berlin: Heidelberg: Springer Verlag.

Dembowska, M. (1991). *Nauka o informacji naukowej: organizacja i problematyka badań w Polsce*. Warszawa: IINTE.

PRACA ZBIOROWA

Bellardo Hahn, T., Buckland, M., eds. (1998). *Historical Studies in Information Science*. Medford, NJ: Information Today.

Biblioteki (1976). *Biblioteki publiczne województwa toruńskiego: informator*. Toruń: Wojewódzka Biblioteka Publiczna i Książnica Miejska im. M. Kopernika.

Kocójowa, M., red. (2005). *Profesjonalna informacja w Internecie*. Kraków: Wydaw. UJ.

ARTYKUŁ W CZASOPIŚMIE

Dervin, B., Nilan, M. (1986). Information Needs. *Annual Review of Information Science and Technology*, 21, 3–31.

Osińska, V. (2010). Rozwój metod mapowania domen naukowych i potencjał analityczny w nim zawarty. *Zagadnienia Informacji Naukowej*, 96(2), 41–51.

ARTYKUŁ W PRACY ZBIOROWEJ

Rayward, W.B. (1998). Visions of Xanadu: Paul Otlet (1868–1944) and Hypertext. In: T. Bellardo Hahn & M. Buckland (eds.). *Historical Studies in Information Science* (65–80). Medford, NJ: Information Today.

Gawrysiak, P. (2000). W stronę inteligentnych systemów wyszukiwawczych. W: Cz. Daniłowicz (red.) *Multimedialne i sieciowe systemy informacyjne* (59–69). Wrocław: Oficyna PWr.

ARTYKUŁ W CZASOPIŚMIE ELEKTRONICZNYM

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. *Scientific American* [online], May, [30.06.2013], http://www.scientificamerican.com/article.cfm?id=the-semantic-web

Bartalesi, V., Meghini, C. (2016). Using an Ontology for Representing the Knowledge on Literary Texts: The Dante Alighieri Case Study. *Semantic Web* [online], 8(3), 385–394, http://doi.org/10.3233/SW-150198

Miller, H. (2013). Big-Data in Cloud Computing: A Taxonomy of Risks. *Information Research* [online], 18(1), [15.07.2013], http://informationr.net/ir/18-1/paper571.html

HASŁA ENCYKLOPEDYCZNE

Psychology of Culture Contact (1926). *Encyclopaedia Britannica*, Vol. 1, 13th ed. (765–771). London and New York, NY: Encyclopaedia Britannica.

Iluminatorstwo (1971). *Encyklopedia Wiedzy o Książce* (911–952). Wrocław – Warszawa – Kraków: Zakł. Narod. im. Ossolińskich.

Big Data (2013, November 12). *Wikipedia, The Free Encyclopedia* [online] [12.11.2013], http://en.wikipedia.org/w/index.php?title=Big_data&oldid=581347727

Autorskie artykuły encyklopedyczne należy opisywać tak jak artykuły w pracach zbiorowych.

DOKUMENT Z WITRYNY INSTYTUCJI, ORGANIZACJI LUB OSOBY PRYWATNEJ

Aristotle (2009). *Organon*. From 1a to 164 a according to Bekker numbers [online]. Translated under the editorship of W.D. Ross. Internet archive [29.10.2013], http://archive.org/stream/AristotleOrganon/AristotleOrganon-collectedWorks_djvu.txt

MNiSW (2011). *Narodowe Centrum Nauki w Krakowie. Nadchodzi czas nauki* [online]. Ministerstwo Nauki i Szkolnictwa Wyższego, [15.07.2013], http://www.nauka.gov.pl/?id=2268

Smith, B. (2004). *Ontology and Information Systems* [online]. The Buffalo University, Department of Philosophy, [15.07.2013], http://ontology.buffalo.edu/ontology.doc

US NLM (2004). *Medical Subject Headings* [online]. US National Library of Medicine. National Institutes of Health, [15.07.2013], http://www.nlm.nih.gov/mesh/meshhome.html

Guidelines for Authors

ZIN – *Studia Informacyjne* (ZIN – *Information Studies*) accepts only manuscripts that have not been published before and are not under consideration for publication anywhere else. Following types of paper may be submitted for publication: original papers, book reviews, conference (and other events) reports.

Each manuscript is reviewed under a double-blind peer review process. In order to ensure the anonymity of the review process, please do not place any information in the text that could be used to identify the author.

Each manuscript is reviewed by two referees, selected on the basis of necessary expertise in the subject area under review. The review report is based on standard form containing a statement whether the manuscript is recommended for publication. Criteria for acceptance include appropriateness to the field of the Journal, scientific merit, proper text organization and correct language use.

The final decision about publication of manuscript will be sent to Author within 10 weeks after text submission. Manuscript should be formatted according to guidelines listed below and submitted via e-mail: **zin@uw.edu.pl**

1. General guidelines

1.1. Format

All files should be submitted in RTF (Rich Text Format) files, including text and illustrative content. All pages must be typed and 1.5 spaced using 12-point Times New Roman font. The title of the manuscript should be typed 14-point font. Please do not use any preformatted styles.

Illustrative content inserted in the article, should be send also in JPG format. Attachments should be numbered in order of occurrence and include the title, for example: *1. Tab. 1. List...* or *3. Fig. 1. System....*

1.2. Extent

Manuscript should be no longer than 40,000 characters (including spaces), review and report no longer than 14,000 characters.

1.3. Title page

Authors should prepare **separate title page**, which include:

- **title of the paper,**
- **the name(s) of the author(s) with appropriate affiliations and the ORCID numbers,**
- **the e-mail address of the corresponding author,**
- **address for correspondence,**
- **biographic note (see below),**
- **structured abstract (see below),**
- **keywords (see below),**
- **statement of originality (see below).**

According to the Journal policy against *ghostwriting* and *guest authorship*, authors are requested to list on title page names and affiliations of each person that contributed to the text (author of the idea, methods, etc. used in the submitted manuscript; percentage of contribution to the research process and text compilation). Authors are also requested to describe sources of founding that have supported the work and the financial involvement of research institutes, associations and other entities (*financial disclosure*).

1.4. Author(s) biographic note

Title page should include concise biographic notes (about 70 words) of each author : academic degree or professional position, current place of work and position, area of interest, the most important publications (max. 3).

1.5. Structured abstract

An abstract (about 100 words or 1000 characters) should be included with each submission and placed on the title page. Abstract should be formatted according to categories listed below. Author should identify at least four mandatory sections:

- **Purpose/Thesis** *(mandatory)*
- **Approach/Methods** *(mandatory)*
- **Results and conclusions** *(mandatory)*
- **Research limitations** *(optional)*
- **Practical implications** *(optional)*
- **Originality/Value** *(mandatory)*

1.6. Keywords

Title page should include keywords (4 to 10) as a noun phrases in singular form, where first element is capitalized. Keywords in alphabetical order should be delimited by full stop.

1.7. Statement of originality

Author(s) should include on title page statement that submitted text has not been published before and is not under consideration for publication anywhere else. If the paper was presented at a scientific meeting, provide detailed information about the event and the conference proceedings. If the paper will be the part of the author's book, provide its details and planned publishing date.

2. Manuscript format and preparation

2.1. Body of the paper

The text should be organized into entitled sections and subsections. Text should start with **Introduction**, giving an overview and stating the purpose and end with **Conclusion**, giving the summary of the author contributions to the study.

Author may use three levels of headings. Each heading should have its own title and number according to the following pattern:

1. First-level heading
1.1. Second-level heading
1.1.1 Third-level heading

2.2. References

Bibliographic citations are not allowed in footnotes. The reference list should be prepared according to APA 6-th Edition citation style (see below). Footnotes can be used only to give additional information or commentary. Footnotes to the text are numbered consecutively with Arabic numerals. It is recommended to limit the amount of footnotes per page.

2.3. Titles in the body of the text

Titles of exhibitions, conferences, programmes, etc should be written within double quotation marks. Use italics for publication titles (books, journals, papers, etc.).

2.4. Emphasis

Bold face should be used to emphasize certain words or passages.

2.5. Illustrative content

All illustrations (tables, charts, figures etc.) should be converted to greyscale. All illustrations should be cited in the text properly to their form (Table, Figure, Photograph, etc.) and have title and consecutive number (e.g. Tab. 1. Metadata levels). Use abbreviation in the text when refereeing to the illustrative content (e.g. see Tab. 1, see Fig. 5).

2.6. Citations and reference list

Use APA 6-th Edition as a citation and reference list format. The references list should only include works that are cited in the text.

Cite references in the text by name of the author(s) and year of publication in parentheses: (Name, Year of publication), eg. (Dembowska, 1991). If there are two authors, put their names with ampersand (&) mark

between: (Name & Name, Year of publication), eg. (Cisek & Sapa, 2007). If there are more than two authors, put the name of the first one followed by abbreviation *et al.*: (Name et al., Year of publication), eg. (Berners-Lee et al., 2001). Edited books are cited by the name(s) of the editor(s) followed by abbreviation *ed(s).*: (Name, ed., Year of publication), eg. (Bellardo Hahn & Buckland, eds., 1998). If there is no author or editor information, put the first word from the title and the year of publication : (Word, Year of publication), eg. (Biblioteki, 1976). Use the following pattern when referring to specific pages in the cited publications: (Dembowska, 1991, 15) or (Cisek & Sapa, 2007, 40–42) or (Bellardo Hahn & Buckland, eds., 1998, 18).

Place the reference list at the end of the text under the heading **References**. Reference list should be in alphabetical order without numbering.

List the references (books and journal articles) in alphabetical order by authors' last names. Citations of edited books list under the name of editor followed by abbreviation Ed.. If there is no author or editor information, list the publication under the first word from the title.

Use italics for book titles and regular font for titles of papers and book chapters. Use abbreviation In: when referring to book chapters in citations.

If there are two or more items by the same author(s), list them in order of year of publication (reverse date order). If two or more works are by the same author(s) within the same year, list them in alphabetical order by title and distinguish them by adding the letters a, b, c, … to the year of publication:

Dembowska, M. (1976a) ....

Dembowska, M. (1976b) ...., etc.

2.6.1 References List Examples

BOOK

Breslin, J.G., Passant, A., Decker, S. (2009). *The Social Semantic Web*. Berlin: Heidelberg: Springer Verlag.

Dembowska, M. (1991). *Nauka o informacji naukowej: organizacja i problematyka badań w Polsce.* Warszawa: IINTE.

BOOK (EDITED)

Bellardo Hahn, T., Buckland, M., eds. (1998). *Historical Studies in Information Science*. Medford, NJ: Information Today.

Biblioteki (1976). *Biblioteki publiczne województwa toruńskiego: informator*. Toruń: Wojewódzka Biblioteka Publiczna i Książnica Miejska im. M. Kopernika.

JOURNAL ARTICLE

Osińska, V. (2010). Rozwój metod mapowania domen naukowych i potencjał analityczny w nim zawarty. *Zagadnienia Informacji Naukowej*, 96(2), 41–51.

Dervin, B., Nilan, M. (1986). Information Needs. *Annual Review of Information Science and Technology*, 21, 3–31.

BOOK CHAPTER

Rayward, W.B. (1998). Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext. In: T. Bellardo Hahn & M. Buckland (eds.). *Historical Studies in Information Science* (65–80). Medford, NJ: Information Today.

ELECTRONIC JOURNAL ARTICLE

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. *Scientific American* [online], May, [30.06.2013], http://www.scientificamerican.com/article.cfm?id=the-semantic-web

Bartalesi, V., Meghini, C. (2016). Using an Ontology for Representing the Knowledge on Literary Texts: The Dante Alighieri Case Study. *Semantic Web* [online], 8(3), 385–394, http://doi.org/10.3233/SW-150198

Miller, H. (2013). Big-Data in Cloud Computing: A Taxonomy of Risks. *Information Research* [online], 18(1), [15.07.2013], http://informationr.net/ir/18-1/paper571.html

ARTICLE IN ENCYCLOPEDIA

Psychology of Culture Contact (1926). *Encyclopaedia Britannica*, Vol. 1, 13th ed. (765–771). London and New York, NY: Encyclopaedia Britannica.

Iluminatorstwo (1971). *Encyklopedia Wiedzy o Książce* (911–952). Wrocław – Warszawa – Kraków: Zakł. Narod. im. Ossolińskich.

Big Data (2013, November 12). *Wikipedia, The Free Encyclopedia* [online] [12.11.2013], http://en.wikipedia.org/w/index.php?title=Big_data&oldid=581347727

Article in encyclopedia with author information describe as book chapter.

ELECTRONIC DOCUMENT FROM WEBSITE

MNiSW (2011). *Narodowe Centrum Nauki w Krakowie. Nadchodzi czas nauki* [online]. Ministerstwo Nauki i Szkolnictwa Wyższego, [15.07.2013], http://www.nauka.gov.pl/?id=2268

Smith, B. (2004). *Ontology and Information Systems* [online]. The Buffalo University, Department of Philosophy, [15.07.2013], http://ontology.buffalo.edu/ontology.doc

US NLM (2004). *Medical Subject Headings* [online]. US National Library of Medicine. National Institutes of Health, [15.07.2013], http://www.nlm.nih.gov/mesh/meshhome.html

Adres Wydawnictwa
ul. Konopczyńskiego 5/7
00-335 Warszawa, tel. 22 827 52 96

Prenumerata i sprzedaż
Dział Promocji i Kolportażu SBP
Al. Niepodległości 213, 02-086 Warszawa, tel. 22 608 28 24

Wydawnictwo Naukowe i Edukacyjne SBP – Warszawa 2019
Ark. wyd. 12,5.
Skład i łamanie: Justyna Grzymała-Łuszcz


WYDAWNICTWO
SBP