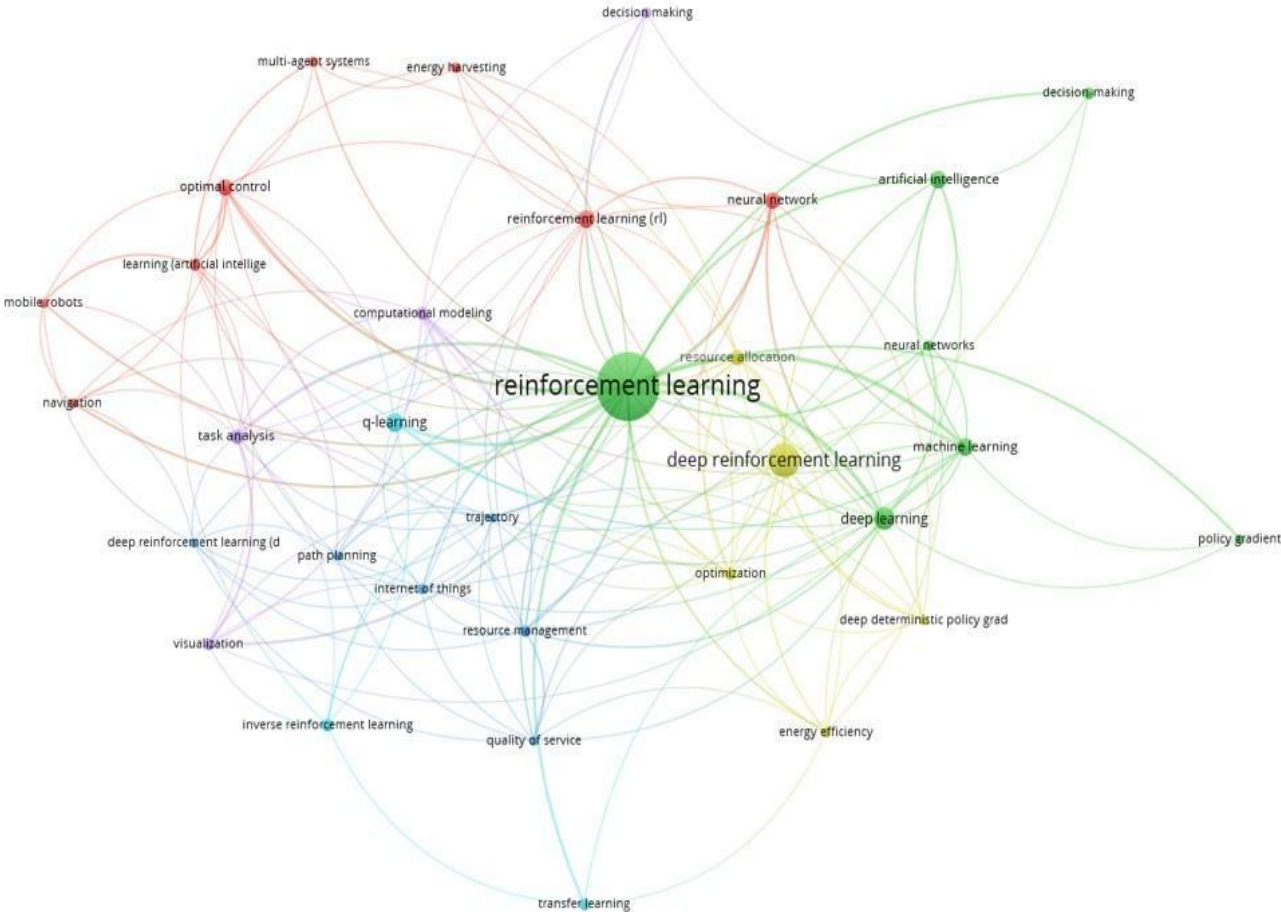


**SCIENTIFIC INDEX SEARCH**



**MINISTRY OF SCIENCE AND HIGHER EDUCATION  
OF THE RUSSIAN FEDERATION**

**ITMO UNIVERSITY**

**M.Ya. Marusina**

# **SCIENTIFIC INDEX SEARCH**

**TRAINING MANUAL**

**RECOMMENDED FOR MASTER'S PROGRAMS**

**ITMO**

**Saint Petersburg**

**2023**

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**УНИВЕРСИТЕТ ИТМО**

**М.Я. Марусина**

# **НАУЧНЫЙ ПОИСК**

**УЧЕБНОЕ ПОСОБИЕ**

**РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО  
по направлению подготовки 27.04.03, 27.04.04  
в качестве учебного пособия для реализации основных профессиональных  
образовательных программ высшего образования магистратуры**

**Санкт-Петербург**

**2023**

Рецензент(ы):

Максимова Татьяна Геннадьевна, доктор экономических наук, профессор, профессор факультета инфокоммуникационных технологий Университета ИТМО.

Учебное пособие является междисциплинарным и может быть использовано студентами различных направлений подготовки, связанных с информационными технологиями. Формирует общекультурные и профессиональные компетенции. В пособии рассмотрены практические вопросы применения методов наукометрии, представлены возможности и ограничения библиометрических показателей при принятии решений об эффективности научной деятельности ученых, исследовательских коллективов, организаций и стран. Приведена основная информация о базах данных, размещенных на платформе Web of Science, и информация о Российском индексе научного цитирования (РИНЦ). Учебное пособие будет полезно студентам, аспирантам и научным работникам, а также всем, кто интересуется оценкой показателей уровня развития науки и техники.

**ИТМО**

**Университет ИТМО** – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2023

© Марусина М.Я., 2023

M.Ya. Marusina. Scientific Index Search. – St.Petersburg: ITMO University, 2023.  
– 52 p.

Reviewed by:

Tatyana Gennadievna Maksimova, Doctor of Economic Sciences, Professor,  
ITMO University.

This training manual is interdisciplinary and can be used by postgraduates of various fields related to information technology. It is aimed at forming general and specific professional competencies. The manual discusses some practical issues of applying methods of Scientometrics. It describes the opportunities and limitations of bibliometric indicators in evaluating the effectiveness of scientific activities of individual scholars, research teams, institutions and countries. It provides some basic information about databases hosted on the Web of Science platform, as well as about the Russian Science Citation Index (RSCI). The training manual will be useful for postgraduate students and researchers, as well as anyone interested in the assessment indicators in science and technology.



**ITMO University** is Russia's leading university in the field of information and photonic technologies, one of the few Russian universities that received the status of a national research university in 2009. Since 2013, ITMO University has been a participant in the program to increase the competitiveness of Russian universities among the world's leading research and educational centers, known as the 5 in 100 project. The goal of ITMO University is to become a world-class research university, entrepreneurial in type, focused on the internationalization of all areas of activity.

© ITMO University, 2023

©Marusina M.Ya., 2023

# CONTENTS

<b>INTRODUCTION</b>	6
<b>CHAPTER 1. SCIENTIFIC PUBLICATIONS AS A MEANS OF COMMUNICATION</b>	7
1.1. Scientific Journals. History and Characteristics	7
1.2. Publications Databases as Search and Analysis Tools	12
1.3. Databases Hosted on the Web of Science Platform	15
1.4. Russian Science Citation Index	17
1.5. Test Questions	20
<b>CHAPTER 2. INDICATORS IN BIBLIOMETRY</b>	21
2.1. Bibliometric Tools. Analytical add-ins of the Web of Science Core Collection Database	21
2.1.1. Principles of Bibliometric Databases Arrangement	21
2.1.2. Analytical Tools: Journal Citation Reports, Essential Science Indicators and InCites	23
2.2. Impact Indicators: Indicators of Article Impact	24
2.2.1. Journal Impact Factor	24
2.2.2. Five-year Impact Factor of the Journal. Efficiency Index	27
2.3. Indicators of the Relative Influence of Articles: Relative, Normalized Impact Indicators	29
2.3.1. Relative Impact Factor	29
2.3.2. Rank Indicators	31
2.3.3. High Citation	32
2.3.4. Citation Profiles	33
2.3.5. “Weighted” Indicators	34
2.4. Scientific Journal Indicators: SJR and SNIP	35
2.5. Test Questions	37
<b>CHAPTER 3. ASSESSMENT OF THE EFFECTIVENESS OF SCIENTIFIC ACTIVITY</b>	38
3.1. Tendencies of Developing the Methods of Assessment of Scientific Activity	38
3.2. Methods for Assessing the Effectiveness of Scientists	40
3.2.1. Hirsch Index	40
3.2.2. Beamplots: a New Standard for Citation Statistics	44
3.3. Altmetrics: Evaluation Methods Other than Citation Analysis	46
3.4. Test Questions	47
<b>CONCLUSION</b>	48
<b>ABBREVIATIONS</b>	49
<b>LIST OF SOURCES USED</b>	50

## INTRODUCTION

The information provided in the training manual is the basis for studying some practical issues of applying **Scientometrics** methods for assessing the effectiveness of scientific activities.

The manual describes the history of the concept of using references to published papers for scientific index search and the implementation of this concept in the **Science Citation Index** search tool. It discusses the opportunities and limitations of bibliometric indicators when evaluating the effectiveness of research activity of scientists, scientific teams, institutions and countries. On the other hand, it is important to judiciously combine formal Scientometrics indicators and expert assessments of subject matter professionals in the assessment of scientific activity.

The first chapter considers scientific publication as a means of scientific communication. It describes the evolution of scientific journals and of the principles of forming databases of scientific publications and citation indices as tools for search and analysis. It gives some basic information about the databases hosted on the **Web of Science platform** as well as about the **Russian Science Citation Index (RSCI)**.

The second chapter of the manual examines bibliometric indicators that assess the citation. The following analytical tools are analyzed: **Journal Citation Reports**, **Essential Science Indicators** and **InCites**. The use of impact factor and its variations: *relative citation*, *rank metrics*, *"weighted" indicators* is described.

The third chapter provides some information on the techniques for assessing the effectiveness of scientific activities, the main indicators of citation and the quality of journals. It describes the distinctive features of publications in various scientific fields and some approaches to comparing citations in various fields of knowledge. Information about **Hirsch index**, **Beamplots**, and the **Almetrics** is provided.

The materials of this training manual give students advanced knowledge in Scientometrics. The manual was developed in accordance with current information on the indicators of the development of science.

The manual will be useful for undergraduate, postgraduate students, researchers, as well as for anyone interested in Scientometrics.

# CHAPTER 1. SCIENTIFIC PUBLICATIONS AS A MEANS OF COMMUNICATION

The aim of any research activity is to discover some new knowledge and create some new technologies to improve the quality of human life. To use the research findings it is necessary to communicate them to consumers and other producers of new scientific results [1].

In ancient times and in the Middle Ages, the exchange of knowledge took place by means of personal communication and correspondence, and the most important results were presented in the form of scientific treatises, which resulted from the many years of scientists' activity [1]. Later, scientific journals began to be formed to convey scientific information to the audience concerned. The system of conferences, symposiums, seminars and other forms of scientific meetings started to take shape. At present, open sources play a special role in the Internet.

Patents fix the priority of the researcher in the creation of a new technology, a material, and some useful models and thus have their special niche in the system of scientific communication being full-fledged objects of commodity and monetary relations.

## 1.1. Scientific Journals. History and Characteristics

The exponential growth in the number of people involved in scientific research and thus, in the number of scientific journals and scientific communities in the 1960s resulted in both finding reliable relevant information and selecting a platform to publish the results becoming challenging.

The first scientific journals presented in Figure 1.1 appeared in the 17th century and provided a list of news about books published, soon reports on scientific findings began to appear. In March 1665, the first issue of "*The Philosophical Transactions of the Royal Society*" [1] was published. This journal has been produced without interruption from its inception to the present.

Scientific journals acquired their modern and familiar form by the mid-19th century. In 1987 the first electronic journals appeared. The main function of a scientific journal since the release of the first ones has always been scientific communication. According to *V. G. Belinsky*, "...a journal is neither science nor scholarship in itself, but, so to speak, a factor of science and scholarship, a mediator between science and the scientists. No matter how long an article may be, it will never translate the whole system of any kind of knowledge. It can only present the results of this system to draw the attention of scientists and to notify the audience as soon as possible of something meaningful having happened" [2].



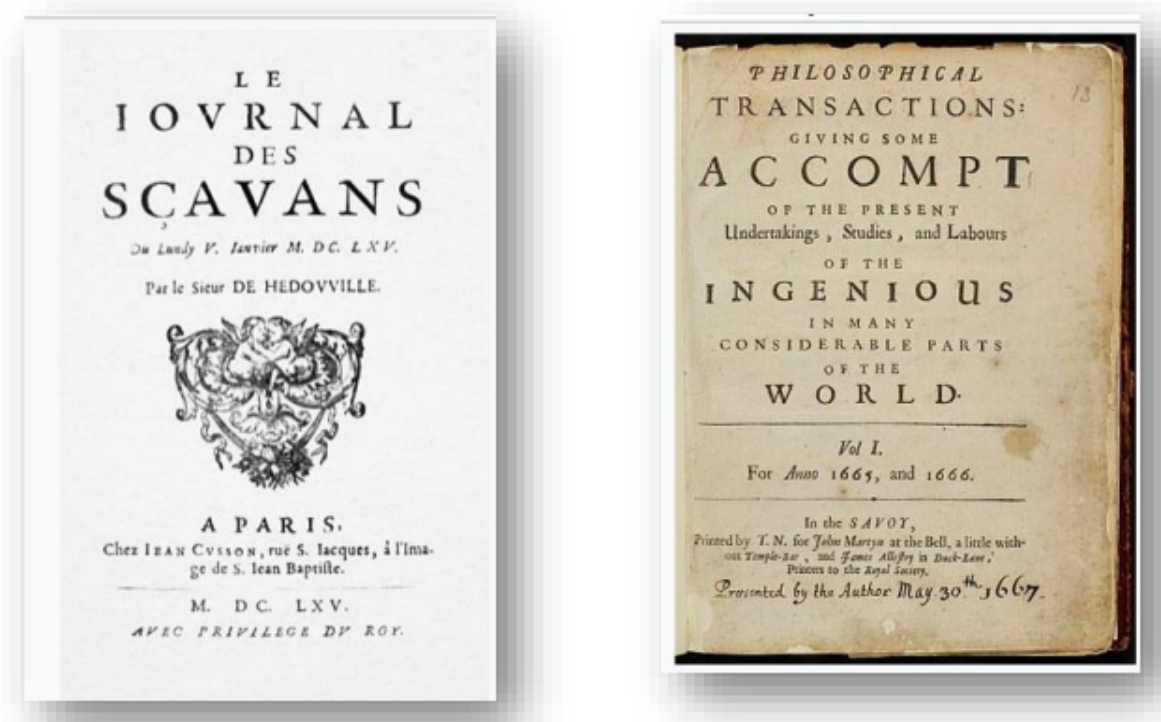


Figure 1.1. First scientific journals published in 1665

There are currently more than 200,000 journals worldwide [1], but not all are scientific. A scientific journal is a journal in which all articles are pre-reviewed. Of the 200,000 journals, about 70 to 100,000 can be referred to as scientific.

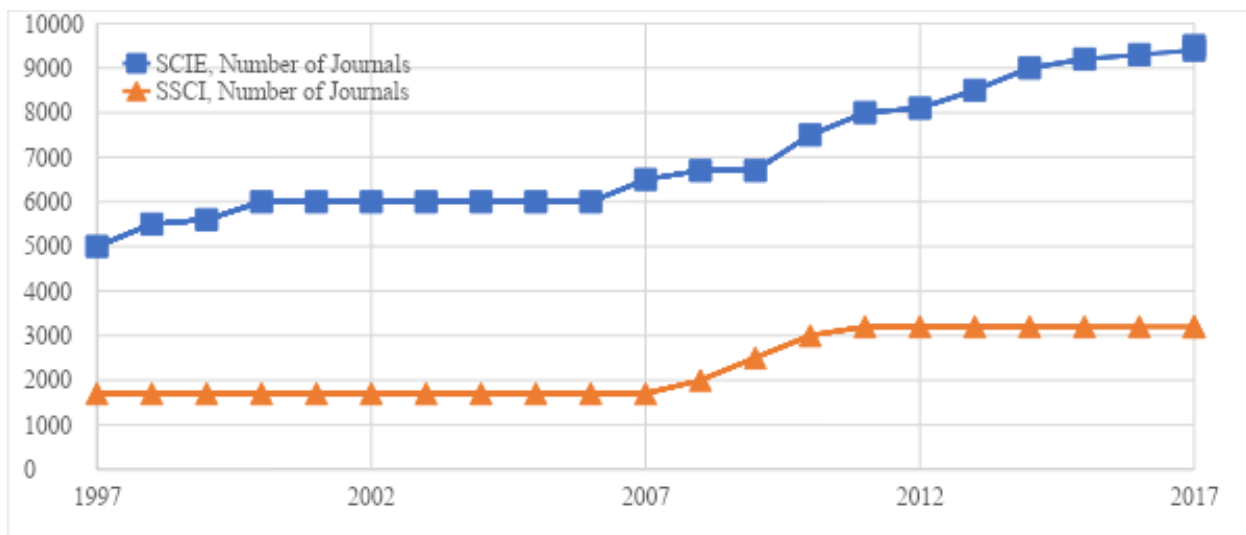


Figure 1.2. Change in the number of journals in the Science Citation Index Expanded and Social Science Citation Index Web of Science Core Collection [1]

The evolution of the number of journals over the 20 years from 1997 to 2017 is presented in Figure 1.2 **SCIE - Science Citation Index Expanded** - is the index of scientific citation in natural and technical sciences, and **SSCI - Social Sciences Citation Index** - is the citation index in social sciences.

The graphs have roughly the same trend, during the 20 years the number of journals for both indices doubled.

It is interesting to note that the number of journals presented in the Journal Citation Index doubling during 20 years corresponds to the hypothesis of *Derek John de Solla Price* [3]. According to D. Price, here are the periods of doubling the following characteristic for different spheres of activity:

100 years for the number of personalities mentioned in national biographical handbooks;

50 years for employment (workforce), population and the number of universities;

20 years for aggregate national product, important scientific discoveries, the number of outstanding physicists, number of known chemical elements, precision of instruments, the number of college enrollments per 1,000 population [1];

15 years for the number of Bachelors of Arts and Bachelors of Science, the number of scientific journals, the number of members of scientific societies, the number of known chemical compounds, the total number of Abstract Journals in all branches [16];

10 years for the number of known asteroids, the number of publications on non-Euclid geometry, the number of publications on X-rays, the number of publications on experimental psychology [3] and of publications in some other areas.

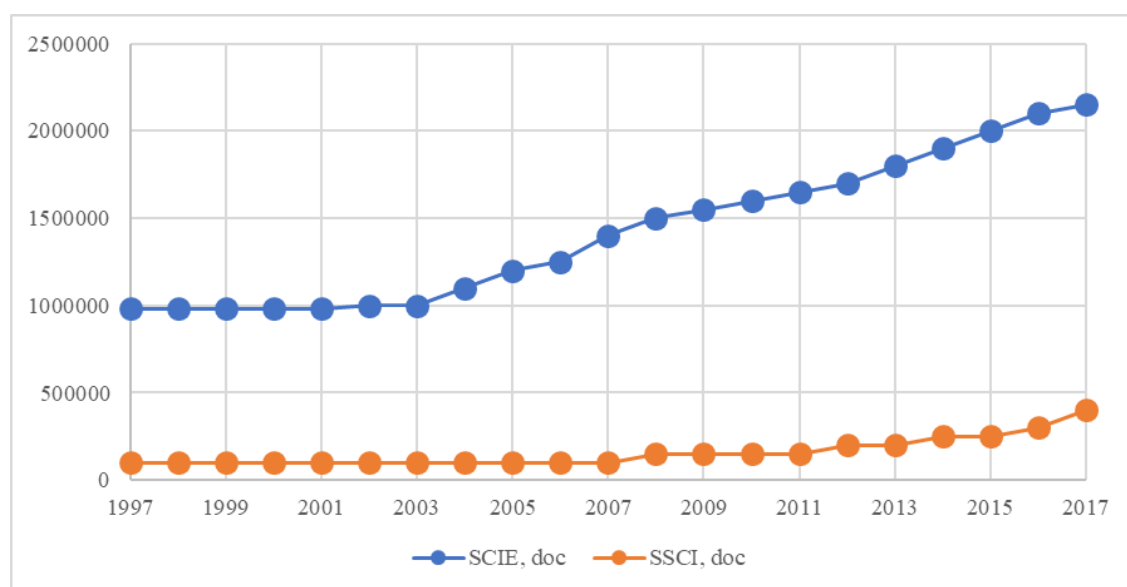


Figure 1.3. Number of articles in the Science Citation Index Expanded and Social Science Citation Index

Figure 1.3 shows the evolution of the number of articles for two indices: scientific and social citations.

The gradual increase in the number of articles published on average in one journal corresponds to a gradual increase of that of scientific journals or to their number remaining steady. On the Other hand, a sharp decline in the number of articles per journal corresponds to a sharp increase in the number of journals indexed (2008).

Thus, new journals appear when the subject matter of the existing journals ceases to meet the requirements of an evolving science and there is a need for a new discussion platform in the emerging scientific fields.

We live in an era of information explosion. According to **Clarivate**, there are more than 100,000 scientific journals in the world, more than 110,000 conferences, more than 12 million patents, more than 100,000 monographs, more than 42,000 trademarks and more than 400,000 technical standards.

Figure 1.4 shows how Price distribution works. A relatively small number of journals publish an absolute majority of relevant scientific articles, which in their turn get the majority of citations.

Only 3,000 journals publish 80% of articles and 92% of what is cited.

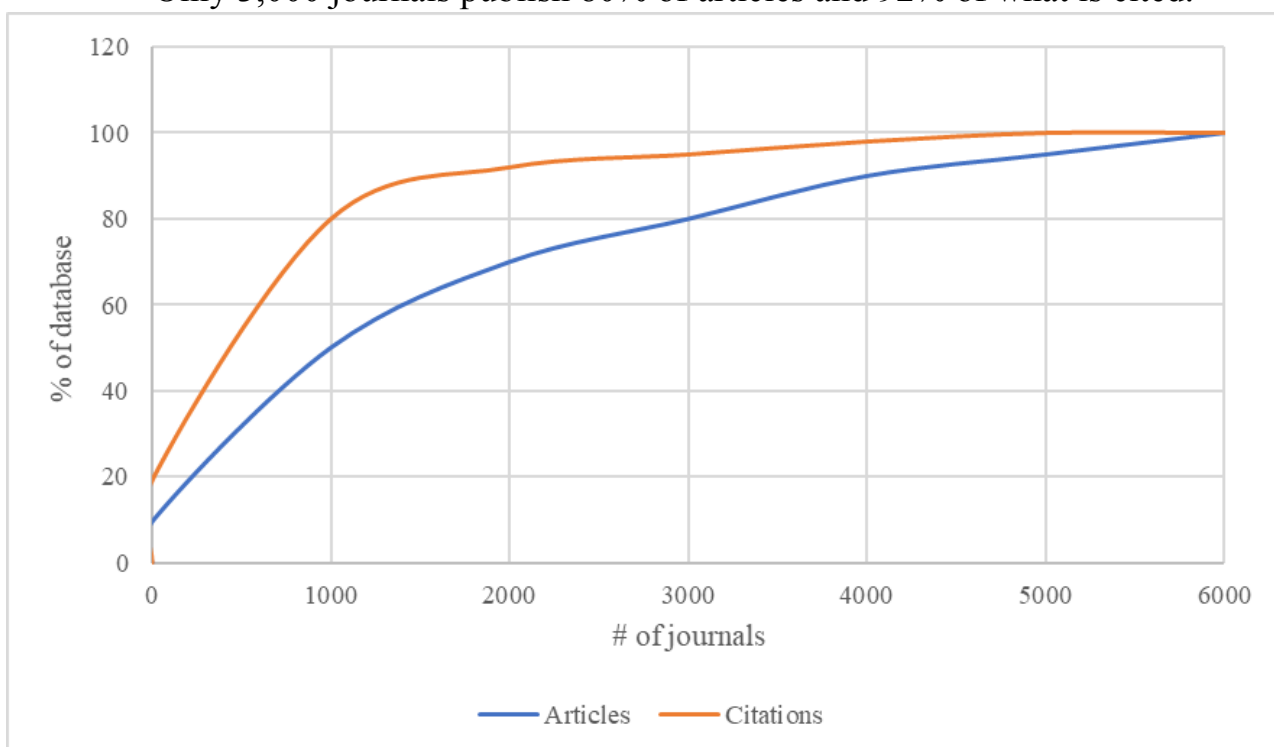


Figure 1.4. Illustration of the uneven distribution of articles and their citations by scientific journals Web of Science CC, 2003

A separate type of publications are those in scientific blogs on the Internet and **Open Access** journals (open access journals are electronic journals providing full access to published articles for all users of the Internet). The first Open Access journals were those of the Public Library of Science (PLoS) [1].

Open access makes relevant a "quality control" of such publications: whether they meet the standards of publication ethics and whether the data can be trusted. Until recently, **Jeffrey Bill**, a librarian at the University of Colorado, wrote and updated "Bill's List". Inclusion of a journal in this list upon the results of an analysis of its citations (including self-citation, mutual citation of journals issued by one publisher) and / or any reference to violations of review rules or incorrect information about the editorial board was often one of the reasons for its elimination from the Journal Citation Report. In 2013, 66 journals [1] were removed from the Journal Citation Report for manipulating citations. At present, many societies are engaged in identifying abusive publishers. For example, Cabell Scholarly Analytics regularly publishes "black" and "white" lists of publications.

Currently, the number of scientific publications is so big that a scientist reading, on average, more than 200 articles per year, seems unable to read even major publications in their narrow subject area, let alone tracking publications in the related fields, which is crucial for interdisciplinary research [16].

## 1.2. Publications Databases as Search and Analysis Tools

Abstract Journals were the first to facilitate searching for information, and then in the late 19th century the first indices appeared - those of legal documents Shepard's Citations and Index Medicus.

A breakthrough in the technology of searching and analyzing scientific information came from *Dr. Eugene Garfield's* brilliant idea of using scientific references as a means of scientific search. This idea was implemented in the form of the **Science Citation Index (SCI)**. In 1960, Garfield founded the *Institute of Scientific Information (ISI)* where he started creating databases of machine-readable indices of scientific publications with an ability to make a record of citations.

Abstracts database is a bibliographic database containing bibliographic records, including annotations, abstracts or any other indicators of the document content. Its creation made it easier to conduct large-scale research on citation analysis.

Library databases can be classified as:

- bibliographic records only;
- databases of abstracts containing abstracts of publications and some additional information about the article;
- full-text databases containing, as the name suggests, full texts of the articles included in the database;
- mixed bases.

In 1964, Dr. Eugene Garfield launched the **Science Citation Index (SCI)**. The **Social Sciences Citation Index (SSCI)** was introduced in 1973 and the **Arts & Humanities Citation Index (AHCI)** in 1978. These databases, which are currently the foundation of the **Web of Science Core Collection (WoS CC)**, index all types of publications in scientific journals selected according to strictly defined rules. The selection rules are posted on the official website of **Clarivate** [4]. It is editors with the relevant qualifications who provide this selection; journals are added and removed from the Web of Science Core Collection database throughout the year.

Since 1990, the **Conference Proceeding Citation Index (CPCI)** has been compiling conference index databases. Since January 2014, this data has been available on the Web of Science platform [18]. The selection of materials is provided in accordance with the rules published on the official website of Clarivate [5].

In 2010, **Book Citation Index (BKCI)** started to be indexed. Requirements for indices of books are available on the website of Clarivate [6].

In 2015, the **Emerging Sources Citation Index (ESCI)** was added to the Web of Science Core Collection [18]. This section keeps record of journals that have submitted applications for indexing in the Web of Science Core Collection. After a few years of monitoring these journals, a positive decision may be made to include them into one of the "senior" journal bases (SCIE, SSCI, AHCI) or to stop indexing them in the Core Collection. Figure 1.5 shows the selection scheme for the Web of Science Core Collection [1].

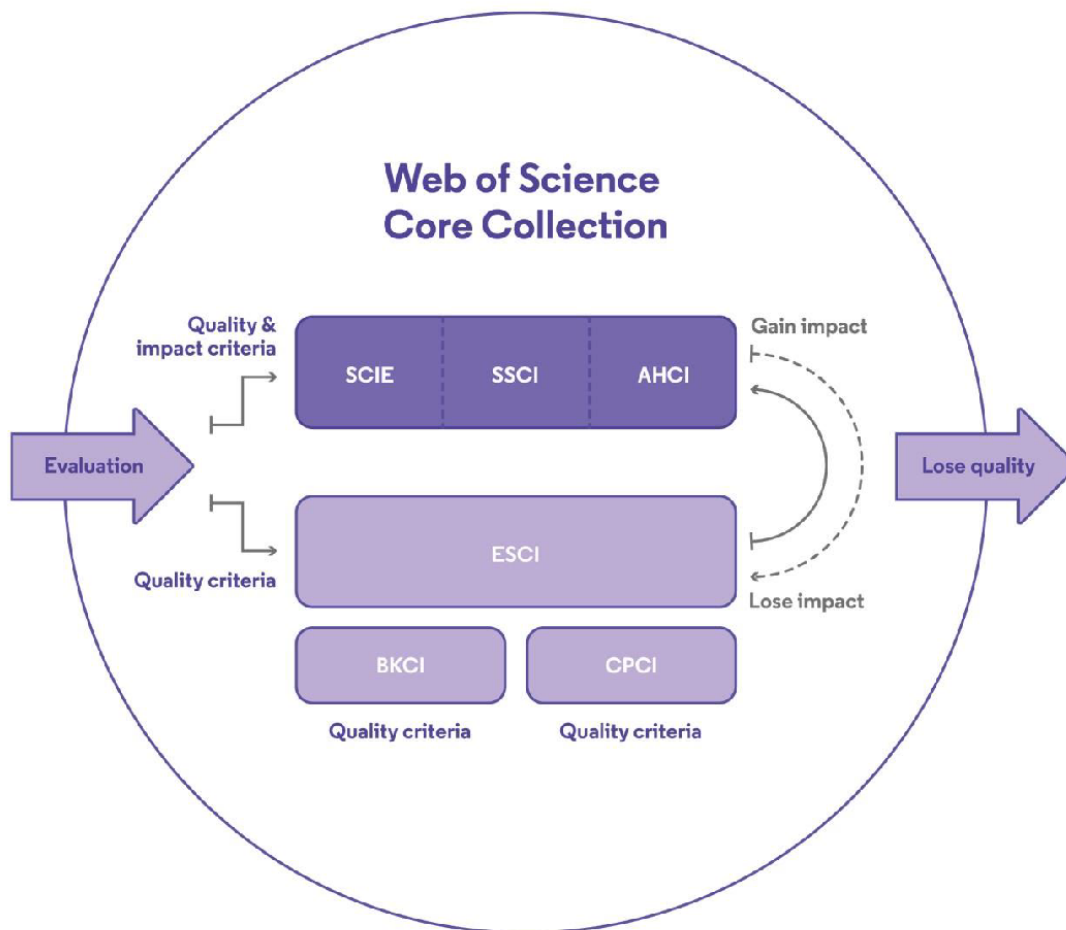


Figure 1.5. Selection of journals in the Web of Science Core Collection [1]

**Science Citation Index Expanded (SCIE)** contains journals on natural, technical, medical sciences (physics, chemistry, mathematics, biology, computer science and computer engineering, medicine, and others) archived since 1900 [1];

**Social Science Citation Index (SSCI)** contains journals on social sciences (economics, management, sociology, law, political science, history, and others) archived since 1900 [1];

**Arts & Humanities Citation Index (A&HCI)** contains journals on humanities (history, literary studies, art history, religious studies, and others) archived since 1975 [1];

***Emerging Sources Citation Index (ESCI)*** contains journals of the entire spectrum of scientific disciplines that are not included in the above indexes, which are called "senior" indexes, but evaluated by Clarivate for inclusion in them, the so-called "reception" WoS CC, containing lesser-known publications, archived since 2005 [1];

***Conference Proceeding Citation Index (CPCI)*** contains proceedings of conferences devoted to natural, technical and medical sciences, archived since 1990 [1];

***Conference Proceeding Citation Index - Social Science & Humanities*** contains proceedings of conferences devoted to social and humanitarian sciences, archived since 1990 [1];

***Book Citation Index – Science*** contains books on natural, technical, medical sciences, archived since 2005 [1];

***Book Citation Index – Social Sciences & Humanities*** contains books on social sciences and humanities, archived since 2005 [1].

***Index Chemical and Current Chemical Reactions*** are the two databases on chemical compounds and reactions included in WoS CC [1].

Materials are selected according to rigid rules regarding the authenticity of publications, layout, professional expertise of editors and other requirements. The main data sources for the Web of Science Core Collection are carefully selected journals, conference proceedings and books [7]. The structure of these databases of abstracts makes it possible to analyze and search by article titles, their authors, keywords, author affiliations, journal and conference titles, subject categories and research areas [16].

For all the indexed articles, including indexing of reference lists, citations are accounted for automatically. Based on these data, journal indicators are calculated, it becomes possible to analyze citation indicators of individual authors, institutions or countries for any set of publications for any period. Currently, the Web of Science Core Collection indexes 148 Russian editions in SCIE, 3 SSCI journals and 10 A&HCI editions [18].

### 1.3. Databases Hosted on the Web of Science Platform

In addition to the Web of Science Core Collection database, the Web of Science platform [1] hosts other databases (indices) of scientific citation, which are available by subscription of institutions.

The pie chart in Figure 1.6 represents databases hosted on the Web of Science platform [1].

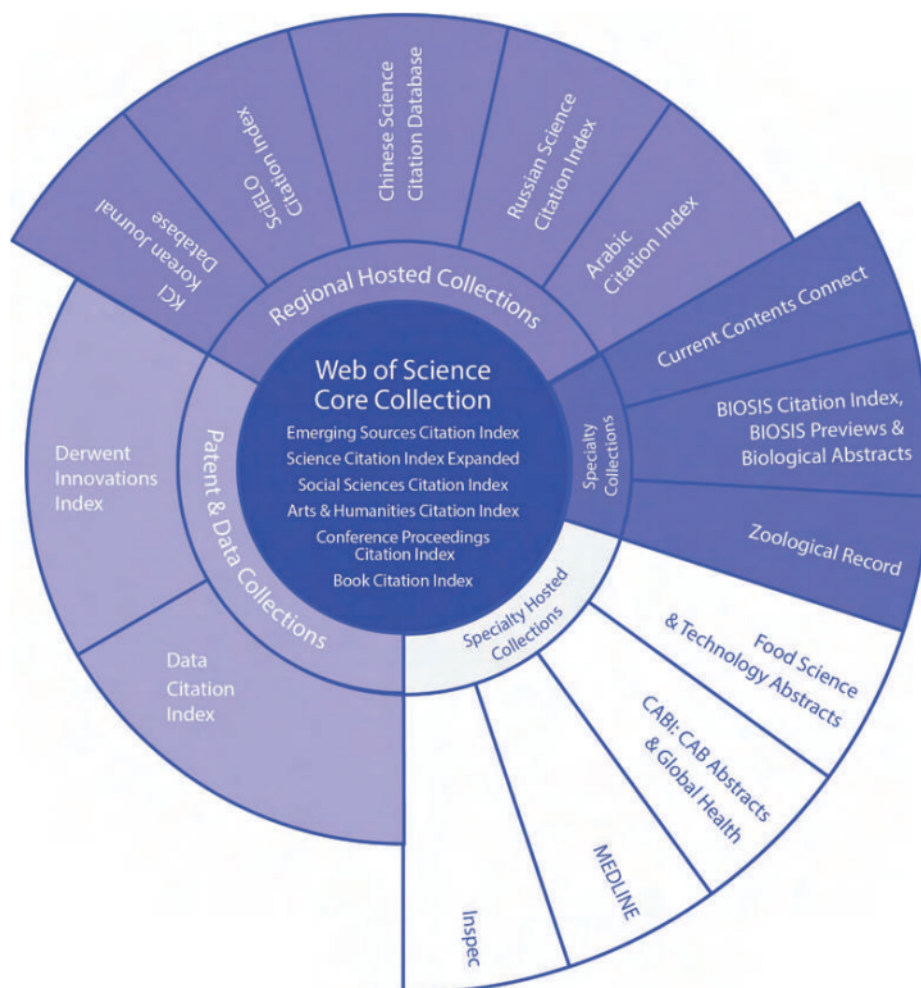


Figure 1.6. Web of Science databases [8]

Descriptions of all databases (grounds for content selection, composition, any specific features) are available on the official website of Clarivate Web of Science [8].

**Derwent Innovation Index** – this database includes more than 80 million patent descriptions from 59 patent agencies around the world, since 1963 [9]. A distinctive feature of this patent database is special annotations and its own Derwent classification, in addition to the international patent classification code. Moreover, inclusion in the database results from a combination of patents registered with various patent agencies. Hyperlinks in the patent description lead to the cited and quoting patents.



**Data Citation Index** is a database of 400 open data repositories since 2001 [10]. In a number of journals, a mandatory requirement for publication is the data being available for the general audience. A description of such an article includes an icon "Associated data" (Associated data) and you can proceed to the repository using the link provided in the article. The repository description specifies the articles where this open data has been used.

Since 2009, the Web of Science platform [18] has been hosting regional scientific citation indices indexing the highly relevant sources of scientific publications in their regions and countries:

Index of Chinese scientific publications *Chinese Science Citation Database* – 2009.

Index of Spanish and Portuguese scientific publications *SciELO Citation Index* – 2013.

Index of Korean scientific publications *KCI Korean Journal Index* – 2014.

Index of Russian scientific publications RSCI - *Russian Science Citation Index* [18] – 2015.

Index of Arab scientific publications *Arab Citation Index* – 2020.

## 1.4. Russian Science Citation Index

In 2015, Clarivate together with the Scientific Electronic Library eLIBRARY.RU developed the **Russian Science Citation Index (RSCI)** on the Web of Science platform [18]. In this index of Russian scientific publications, RSCI indexes scientific journals from Russia and CIS countries selected by multi-stage bibliometric and expert evaluation from among those indexed in the **Russian Index Science Citation (RISC)** [11]. Initially, about 600 scientific journals from the RSCI were selected in RISC, the index is regularly reviewed, thus, by the end of 2020, 792 journals were included in RSCI.

Figure 1.7 shows a comparison of the distribution of journals in RSCI by subject areas with the same distribution in the RISC. The RISC includes more social science journals and interdisciplinary journals than the RSCI.

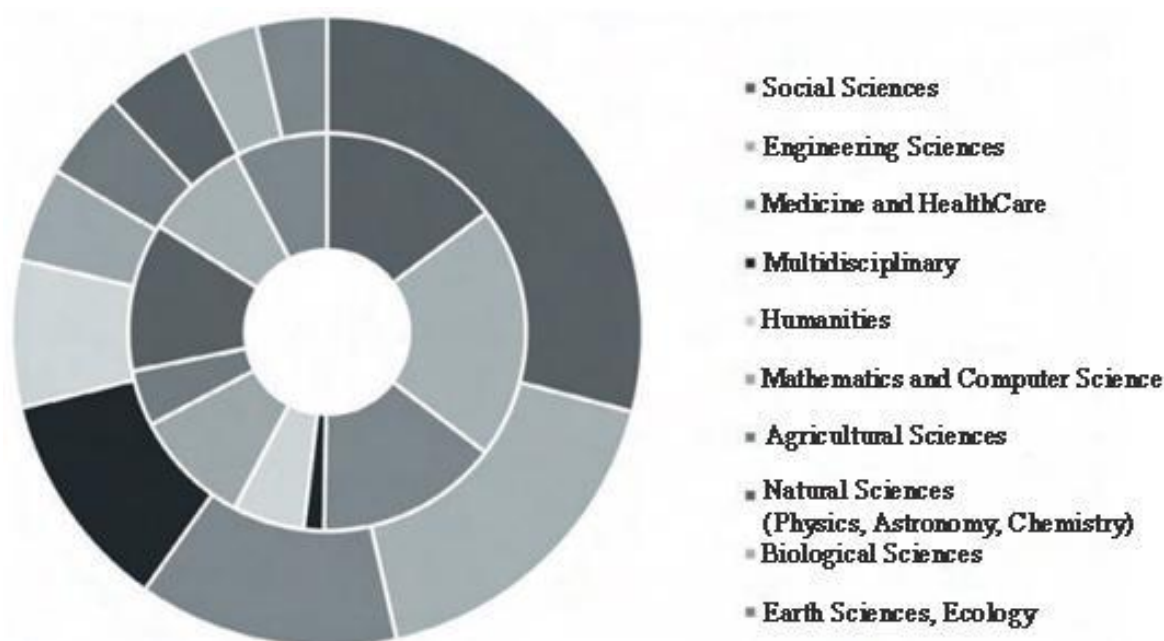


Figure 1.7. Distribution by subject area of the RISC (outer or outer circle) and RSCI (inner circle) [18]

Figure 1.8 shows the analysis of the number of publications in Russian journals belonging to various groups of journals, and their citations. The analysis shows that 23% of the total number of publications indexed in the RISC are the WoS and/or Scopus and RSCI, which make up the core of the RISC. On the other hand, these publications accumulated 83% of citations.

This ratio of publications to citations is close to that used by Dr. Eugene Garfield to form the Science Citation Index.

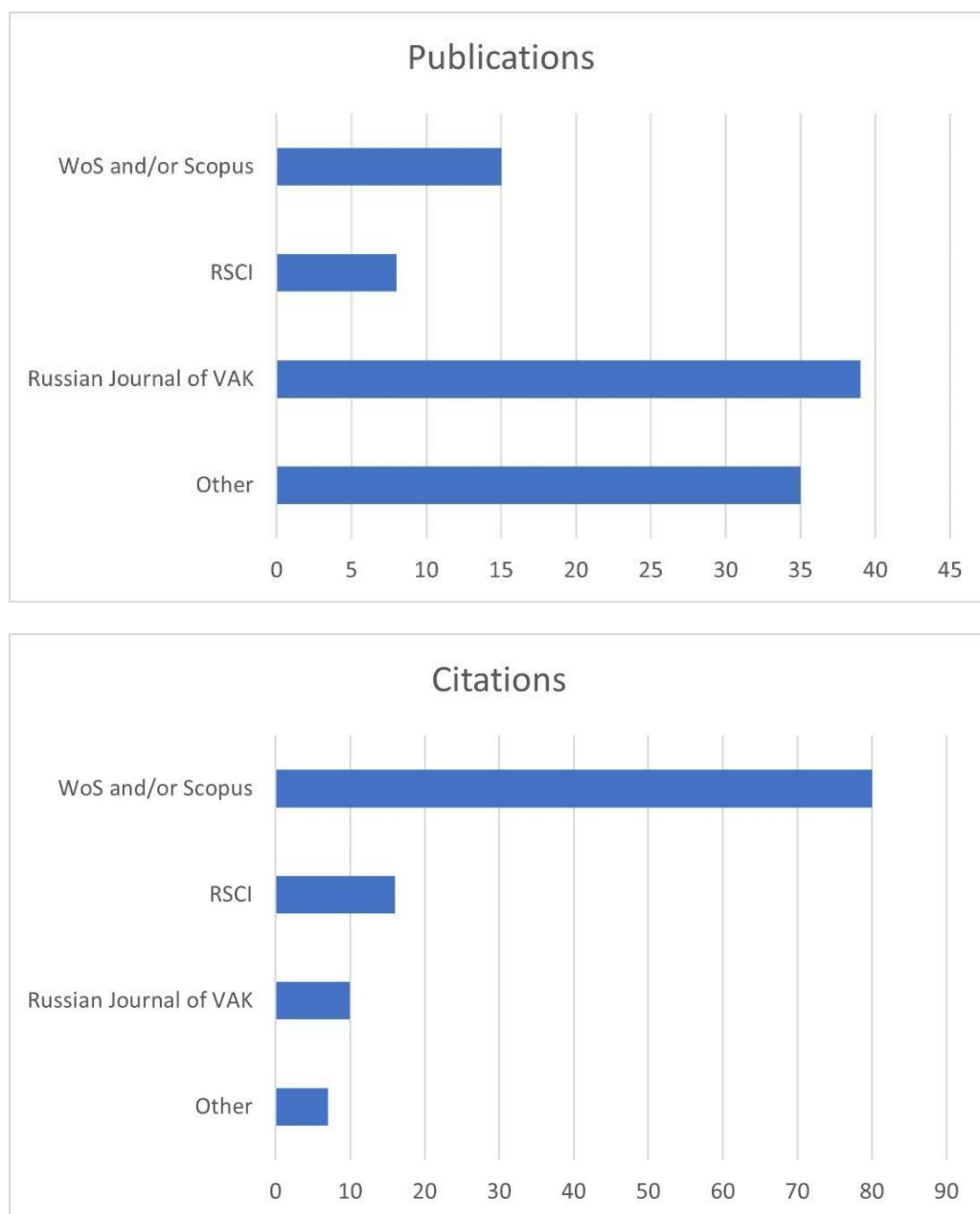


Figure 1.8. Proportion of number of publications and citations of Russian authors in different groups of journals [18]

It should be noted that the list of RSCI metadata is much broader than that of RSCI metadata. For example, **eLibrary** does not contain any information about the English titles of articles, names and affiliations of authors, article summaries in English. However, the metadata of publications are regularly updated retrospectively.

Currently, the RSCI database has been internationally recognized, which is confirmed by the query statistics for 2016-2018 shown in Figure 1.9.

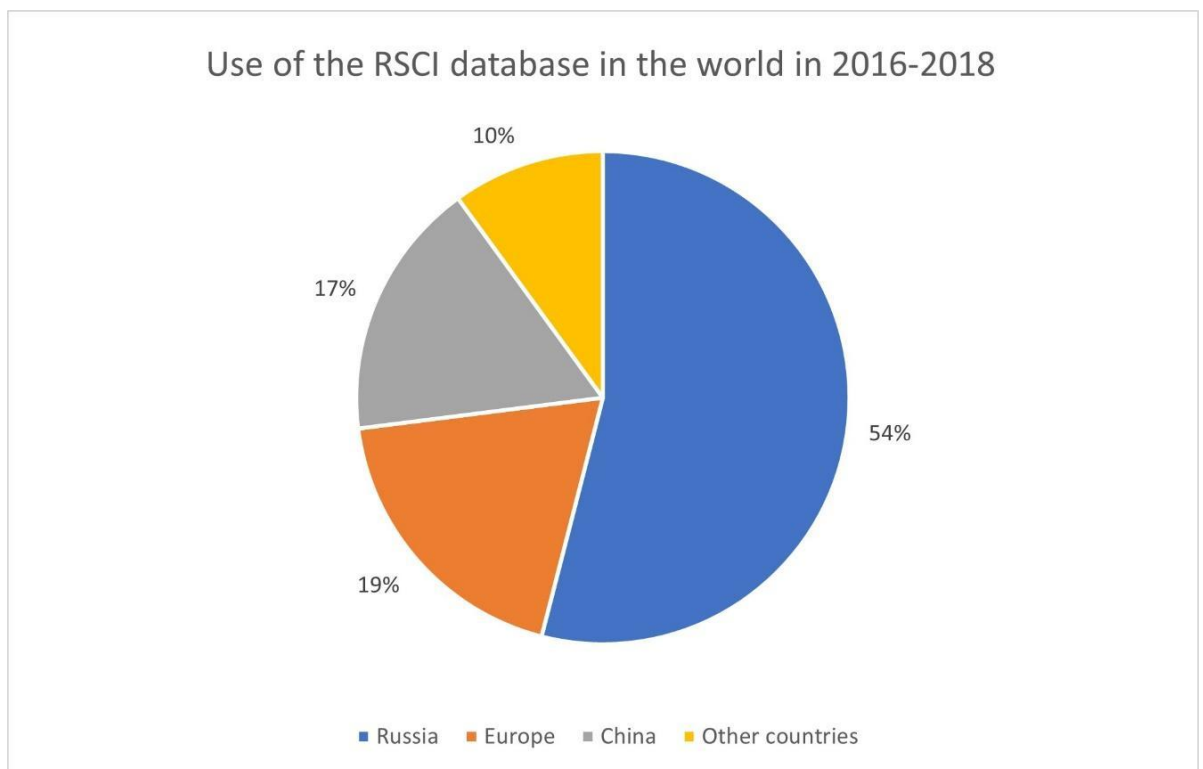


Figure 1.9. Statistics of RSCI usage on the Web of Science platform

Combining all databases on one platform has a number of advantages, for example, when searching the entire platform, the user can also get a huge amount of helpful information:

- years of the most active research on the subject of interest;
- a list of countries and institutions conducting such research;
- a list of the most active researchers;
- a list of journals publishing the research results;
- a list of journals that publish the most cited papers on the topic;
- a list of funds supporting research;
- a list of conferences presenting the relevant findings.

This helps to better plan the research strategy, collaboration and publication of the results.

## 1.5. Test Questions

1. Name the main function of a scientific journal.
2. When did the first scientific journals appear?
3. Name the number of scientific journals that are currently being issued.
4. What is an open access journal?
5. What is "Bill's List"?
6. Which journals are included in the Science Citation Index (SCI)?
7. Which journals are included in the Social Sciences Citation Index (SSCI)?
8. Which journals are included in the Arts & Humanities Citation Index (AHCI)?
9. Which databases are included in the Web of Science Core Collection?
10. Which journals are included in the Emerging Sources Citation Index (ESCI)?
11. Which journal index is older - ESCI or SSCI?
12. What metadata can be used to search the Web of Science Core Collection database?
13. List the databases hosted on the Web of Science platform.
14. What information does the Derwent Innovation Index database include?
15. What information is contained in the Data Citation Index database?
16. List the regional scientific citation indices included in the Web of Science platform.
17. What is the difference between the Russian Science Citation Index (RSCI) and Russian Index Science Citation (RISC) databases?
18. Which database includes more scientific journals - RSCI or RISC?
19. What is the difference between RSCI and RISC metadata?
20. List the groups of scientific journals included in the core of the RSCI.
21. What information can a user find on the Web of Science platform?

## **CHAPTER 2. INDICATORS IN BIBLIOMETRY**

Bibliometric indicators refer to the citation of scientific publications. Bibliometric indicators are included in scientific citation databases, or can be calculated using such databases [21].

Bibliometrics is application of mathematical and statistical methods to scientific publications [12, 13]. Bibliometric assessment is based on the assumption that the authors report a vast majority of their scientific discoveries and research results [21] in international scientific publications where they can be read and then quoted by other scientists. The number of citations of a journal article can be considered a reflection of the impact of this article on the scientific community.

Applied bibliometrics analyzes the number of scientific articles, citation of these articles and the relationship between articles, authors and scientific areas.

### **2.1. Bibliometric Tools. Analytical add-ins of the Web of Science Core Collection Database**

The Web of Science Core Collection [1] database refers to "bibliometric databases" or to "databases of (scientific) citation" (citation database, citation index).

#### **2.1.1. Principles of Bibliometric Databases Arrangement**

Scientific journals provide the main content of bibliometric databases. Noteworthy, the databases, most often, do not contain the full text of the journal articles. They provide the following information on each article to the user:

- Bibliographic information about the article (publisher's imprint: author(s) name(s), article title, journal title, year of publication, volume, issue, pages) [17].
- Abstract of the article — if it is available in the authentic text of the publication, citation bases do not make abstracts for the papers where they are initially absent [17].
- Keywords; sometimes these are two sets of keywords — those attributed to the article in the authentic text of the publication ("author's keywords"); and keywords "assigned" by the citation database based on its internal thesaurus of keywords and automated algorithms [17] (they analyze the titles of other sources cited in the article).
- The subject (category) attributed to the article and the type of publication.
- Institutions where authors work, the so-called "affiliations", with their postal addresses and, sometimes, e-mail addresses of authors.

- ***The list of references cited in the article*** — this feature is the key to the database and makes the latter "bibliometric".
- Various secondary fields: the journal's ISSN number, the language of the authentic paper, some information about grant support, the name and address of the publisher, any other relevant information. [1].

Bibliometric databases may include, in addition to journals, books and conference proceedings.

Each journal has ***a thematic heading***; it is assigned to a scientific discipline (it can be assigned to more than one of them). The thematic heading of the article is determined by the thematic heading of the journal [21].

Each publication is assigned ***the document type***: scientific research article (Article), scientific review (Review), editor's note (Editorial), Letter (Letter), Book Review (Book Review), or other.

Based on the institutional addresses indicated by the authors, the so-called affiliations, the article is considered a publication of the university/laboratory/organization/country. Moreover, the author can specify several employers.

There are several ways to account for the articles written by multiple authors.

- The simplest of them is ***whole (or total) counting***, which assumes that each of the co-authors is credited with one article. Whether the author wrote the article on their own or in collaboration with other colleagues — it is still one (whole) publication in his statistics. The same is true for institutions and countries [20].
- Besides, there are several ways of ***fractional counting***. If there are  $n$  authors in the article, then it is assumed that everyone has written  $1/n$  articles, and the total publication activity of the author is made up of such shares for all their works. The same is true for institutions.

## 2.1.2. Analytical Tools: Journal Citation Reports, Essential Science Indicators and InCites

To collect citation statistics, the Web of Science Core Collection [1] uses the existing analytical tools: **Journal Citation Reports**, **Essential Science Indicators** [1] and **InCites databases**.

These analytical superstructures collect and aggregate data from the Web of Science Core Collection [1] on various "information units": journals, countries, institutions, universities, scientists.

**Journal Citation Reports (JCR)** is a database of bibliometric indicators of journals in general. It publishes data on the number of articles published in the journal, the number of links received by the journal, the chronological distribution of links made/received, and the impact factor of the journal [21].

**Essential Science Indicators (ESI)** is a database of bibliometric indicators of authors, journals, institutions, countries. It publishes the data on the number of articles published by the author/journal/institutions/country and their citation (for journals, the information in ESI is less detailed than in JCR). It contains a section on highly cited articles (included in the 1% of the most cited) and promising scientific "fronts" determined by a special bibliometric procedure [21].

**InCites Benchmarking & Analytics (InCites)** is an analytical tool that provides a detailed and deep analysis of bibliometric indicators of institutions, individual scientists, and countries. It uses normalization of citations by the areas of science and by the journals [21].

In JCR and ESI, indicators are calculated only for natural sciences and social sciences journals, that is, for SCIE and SSCI databases. InCites accounts for all journal, book and conference resources of WoS CC.



## 2.2. Impact Indicators: Indicators of Article Impact

The most common class of citation indicators includes those that estimate the number of references received by one article included in a certain set of publications. These can be articles from a particular journal or a paper by one author [21], a team, an institution, from a country. Such indicators of academic influence are called impact indicators.

The interval during which the articles evaluated have been published is called a "**publication window**" [21]. For example, it can include articles written in some scientific institutions for the five-year period of 2016-2020.

When calculating the impact indicator, it is important to consider the time interval during which the articles referring to the one evaluated were published. For example, some articles published in 2020 made some refer to the article published within the publication window of 2016-2020. This interval is called the "**citation window**"[21].

When calculating the impact indicator of the institution, we consider the average number of references made in the papers of the year 2020 to articles published in 2016-2020 by the scientists of a scientific institution. The average number of references means that the number of references registered is divided by the number of articles in the publication window (the number of articles published by scientists of a scientific institution in 2016-2020).

### 2.2.1. Journal Impact Factor

The most famous impact indicator is the journal impact factor, the **Journal Impact Factor (IF)**[1].

When calculating the impact factor of a journal, *a two-year publication window* and *a one-year citation window* are used. It changes from year to year, so it is calculated for a specific (reporting) year. For the year  $Y$ , the impact factor of the journal is equal to the ratio of the number of all references received in the year  $Y$  by the articles of this journal published in the years  $Y-1$  and  $Y-2$  to the number of these articles (that is, to the number of journal articles published in years  $Y-1$  and  $Y-2$ ) [21]. Thus, the publication window is two years [ $Y-1$ ;  $Y-2$ ] (the average citation of this set of journal articles is evaluated), the citation window is one year [ $Y$ ] (it is the references made in the year  $Y$  are taken into account).

The impact factor refers to the average number of references received in the reporting year by journal articles published during the previous two years [20].

For a visual representation, we introduce some formulae. Let the variable  $P_x$  be the number of articles published in the evaluated journal in the year  $X$ ,  $C_{y \rightarrow x}$  is the number of references received in the year  $Y$  by the articles of this journal published in the year  $X$  [21]. Then the formula for the impact factor IF of the journal in 2021 will have the form

$$IF = \frac{C_{2021 \rightarrow 2020} + C_{2021 \rightarrow 2019}}{P_{2020} + P_{2019}}.$$

Figure 2.1 shows a screenshot of the JCR database for the field of acoustics, which are ordered in the descending order of the impact factor. In addition, many publishers post the current values of the impact factors of their journals on their web pages accessible to the audience [20].

	Full Journal Title	Total Cites	Journal Impact Factor
1	ULTRASONICS SONOCHEMISTRY	17,314	7.279
2	ULTRASOUND IN OBSTETRICS & GYNECOLOGY	12,336	5.595
3	ULTRASCHALL IN DER MEDIZIN	2,238	4.613
4	IEEE-ACM Transactions on Audio Speech and Language Processing	3,110	3.531
5	JOURNAL OF SOUND AND VIBRATION	36,167	3.123
6	IEEE TRANSACTIONS ON ULTRASONICS FERROELECTRICS AND FREQUENCY CONTROL	11,266	2.989
7	JOURNAL OF VIBRATION AND CONTROL	5,075	2.865
8	ULTRASONICS	7,026	2.598

Figure 2.1. – Acoustic journals, in descending order of impact factor. Screenshot of JCR database interface (Clarivate)

The data on impact factors of journals are published annually in the JCR database. The data for each year appear in June-July of the following year. The impact factor of a journal for 2021 can be checked only in the summer of 2022 [1].

It is the journal “CA-A Cancer Journal for Clinicians” [1] that has the maximum impact factor of 2020, IF=223.7. This journal has many references, although it does not publish a large number of articles. The journal “Nature Reviews Materials” is three times behind this publication, IF=74.4.

There are only a few leading journals and many journals with a low impact factor. As a result, the median distribution of impact factors of natural sciences journals included in SCIE is 1.91. This means that a half of all the 9153 journals have an impact below 1.91, and half of them have an impact above that figure. For SSCI social sciences journals, the median is even lower, only 1.40.

## 2.2.2. Five-year Impact Factor of the Journal. Immediacy Index

On the other hand, in a number of scientific disciplines, especially in social sciences, the professional community does not have time to fully absorb new knowledge in such a short period as two years, and it is advisable to use an indicator with a wider publication window. Therefore, the values of **the five-year impact factor** of the journals are also published in the JCR database [21].

For the year  $Y$ , the five-year impact factor of the journal is equal to the ratio of the number of all references received in the year  $Y$  by its articles published in the years  $Y-1$  to  $Y-5$  to the number of these articles (that is, to the number of journal articles published in the years  $Y-1$  to  $Y-5$ ). Thus, the impact factor refers to the average number of references made in the reporting year to journal articles published during the previous five years [21].

For a visual representation, the formula for the impact factor of the journal in 2021 will have the form

$$IF_5 = \frac{C_{2021 \rightarrow 2020} + C_{2021 \rightarrow 2019} + C_{2021 \rightarrow 2018} + C_{2021 \rightarrow 2017} + C_{2021 \rightarrow 2016}}{P_{2020} + P_{2019} + P_{2018} + P_{2017} + P_{2016}}.$$

The maximum five-year impact factor for SCIE journals is 177.3, the median is 1.99. It is higher than the median two-year impact of 1.91, which demonstrates that the five-year indicator more amply covers citations received by the journal.

When calculating the two-year and five-year impact factors, references made to journal articles published in the reporting year are not taken into account. If only the indicators described above are taken into account, then the references of the reporting year "disappear" for the journal.

The indicator that records references of "the same year" is also published in the JCR database and is called the **Immediacy Index (II)** [1]. This index shows how quickly the scientific world reacts to articles and uses them. The Immediacy index assumes a one year publication window and a one-year citation window, and they coincide - this is the reporting year  $Y$ . To calculate the index, the number of all references received in the year  $Y$  by the journal articles published in the same year  $Y$  is divided by the number of these articles [21]

$$II = \frac{C_{2021 \rightarrow 2021}}{P_{2021}}.$$

The journal "CA-A Cancer Journal for Clinicians" [1] had the maximum II in JCR-2018 equal to 52.6. The median for SCIE journals is low – 0.49. For almost every journal, the Immediacy Index is the lowest indicator of all the impact factors considered.

Recently, there has been a worldwide trend towards the growth of impact indicators. The reason for this is a question that requires some special research. For example, the average size of the list of references in scientific articles has been gradually increasing. An important practical conclusion follows from this trend: if the impact factor of a journal has grown, this does not necessarily mean that the journal has become better. In this case, there is only one adequate approach - to compare the changes of indicators in a journal with the general background of the journals on the same academic discipline.

The solution is to use the same n-publication window and citation window. In this case, the average number of references received over the last full n-years by those articles that were published during the same last n years is considered (the average per article).

Two indicators are normally used to assess the level of journals self-citation. The numerator of both indicators is the number of references received by the journal from articles published in i.e. the number of "self-citations". The denominator of the first indicator is the number of all references received by the journal. This indicator is called the self-citation coefficient I. It shows the proportion of the links it received from itself in the whole body of references received by the journal. The denominator of the second indicator is equal to the number of all references made by the journal [17]. It is called the self-citation coefficient II and shows the proportion of references leading to itself in the whole body of citations made by the journal [1].

## 2.3. Indicators of the Relative Influence of Articles: Relative, Normalized Impact Indicators

The central fact to be considered when evaluating the effectiveness of scientists or institutions, as well as when evaluating the quality of journals, is the strong dependence of absolute bibliometric indicators [17] on the area of science of the author/institution or the journal publication. .

No specific features of the area of science are taken into account in the absolute values of impact indicators. Therefore, relative bibliometric indicators are used to meet the following tasks:

- Comparison of research units from various scientific areas.
- Comprehensive assessment of institutions conducting research in several areas.

### 2.3.1. Relative Impact Factor

The purpose of *relative impact indicators* is to evaluate the scientific activity of research units compared to other similar subjects from the same area of science. This comparison is "relative to competitors". For example, you may have to compare the activities of a physician and a biologist, or to get an integral indicator of an institution whose employees publish articles in various fields [17].

Consider a relative metric for a journal. The simplest of them is the ratio of the impact factor of the publication to the average impact factor of the discipline assigned to the journal. On the other hand, the discipline average is not normally calculated as the ratio of the sum of all the impact factors of the journals of this discipline to the number of such journals, but somewhat differently, as the so-called **Aggregate Impact Factor** [17] is used.

A *discipline* refers to a collection of articles published in all journals assigned to a given discipline. A discipline is considered as a single "*meta-journal*" for which the *traditional impact factor* is calculated: the ratio of the number of all references received in the year  $Y$  by the articles of this discipline published in the years  $Y-1$  and  $Y-2$  to the number of articles of this discipline published in the years  $Y-1$  and  $Y-2$ . This indicator for every discipline as a whole is also published in JCR [17]. It is the "**aggregated**" ("**collected**") **Impact Factor** of the discipline that is taken as the denominator; the impact of each journal is divided by this value to get a relative impact factor [1]. Again, it is different from the arithmetic mean of all journal impact factors in the discipline.

In general, we can say that if the relative impact factor of the journal is more than one, it means that on average the articles in the journal considered are cited more often than an average article published in this discipline, and vice versa [17].

A similar indicator – *the relative citation of a scientist* is equal to the ratio of the average citation of their articles (the average number of references per article) to the average citation of the articles of a certain reference group, against which we consider the activity of a scientist [17].

The main question when calculating relative indicators is what to choose as a *reference group*. For example, if we select the articles of the university/faculty where the scientist works, then we will evaluate the effectiveness of this scientist's activities relative to their colleagues working at this university/faculty. However, it should be noted that if a university publishes in several fields, then the comparison will not be adequate: if a scientist publishes in a highly cited area, for example, in medicine, then they will have an undeserved advantage over their colleagues [17], if they do so in a lower cited area, for example, in history, then they will be at a disadvantage.

If we select all the articles of the journal in which the scientist is published as a *reference group*, then we will have an accurate assessment of how much their papers are higher or lower in citation than the average level of this journal [17].

The choice of a reference group depends on the tasks of bibliometric research an analyst faces.

### 2.3.2. Rank Indicators

There is another approach used when comparing interdisciplinary journals. You do not have to compare the absolute values of bibliometric indicators (impact factor, five-year impact factor, influence index), we can only compare the ranking positions publications take in their disciplines. This approach is referred to as **rank method** [1].

The rank method consists in dividing the resulting ordered list rating into  $n$  equal parts and determining in which of these parts the journal falls. For example, if  $n$  is assumed to be *four*, then we introduce **quartiles ( $n=4$ )** — the journals of the first quartile  $Q1$  (falling into the top quarter of the list by the parameter under consideration), the second quartile  $Q2$ , the third quartile  $Q3$ , the fourth quartile  $Q4$ . The system of ranking indicators supposes that the journals in the first quartile are higher than those in the second quartile (not only in the discipline considered, but also in any other discipline). On the other hand, journals that fall into the same quartile in different disciplines are considered [17] the same in terms of scientific level within this method.

The most difficult issue when using the rank method is choosing the indicator  $n$ . This indicator sets the resolution of the rank indicator, the greater  $n$ , the higher the resolution. However, journals that appear to be equal when choosing  $n=4$  (falling into the same quartile) can get a different level at  $n=5$  "*quintile*", and so on. The limiting case of the rank method is when the entire series with the index  $n$  consists of  $n$  journals.

One of the advantages of the ranking method is that it warrants against strong asymmetry of citation distributions, when a small number of journals/authors/institutions receive a significant number of references and create such averages for their group that not all other participants [17] can achieve them.



### 2.3.3. High Citation

Another approach, radically different from the ones discussed above, is to deal with the "extremes" observed in it — those articles that have received an abnormally small or abnormally large number of references, rather than to calculate the average citation of the entire array of articles. Therefore, we will call the indicators based on these methods those of **High Citation**, meaning "extreme", the maximum and the minimum limit of possible citation.

The minimum side is obvious, these are the publications that have not received a single reference, on the maximum side there are the so-called **Highly Cited Papers** [17].

A distinctive feature of the "extreme" approach is that only the proportion of articles that fall into two "marginal" categories is determined, and the number of references [17] other articles actually receive is not taken into account.

A Highly Cited Paper, within the framework of the ESI, is a publication that falls into the 1% of the most cited papers in the world among those published in the given scientific [17] area and in the given year.

The proportion of scientific papers that have never been cited determines the "*non-citation coefficient*". Obviously, if an article does not have any references now, this does not mean that it will not be cited in the future. Therefore, for an adequate analysis of the "*non-citation coefficient*", a time interval is allocated, the value of which is fixed for the publications of any year of publication, for example, 5 years prior, so that older publications do not have an advantage over those recently published.

### 2.3.4. Citation Profiles

A *generalization of High Citation* is the study of the proportion of articles of any scientific unit that fall into each **percentile ( $n=100$ )** (the value that a given random variable does not exceed with a fixed probability given as a percentage) by citation among similar articles.

Let us take one article and all the numerous "similar" ones — as usual, this means articles on the same subject, of the same year of publication and the same type. Let us arrange the entire array in the ascending order of the number of references received by the articles, after which we divide this ordered list, for example, into 10 intervals equal in the number of articles. This division is often used, each interval, in this case, being referred to as a **decile ( $n= 10$ )**.

The method essentially consists in taking all the publications of the author / institution / country / region. Of course, it is preferable to confine yourself to those published for a certain period. We have to find the decile into which each article falls and get the total distribution for the area under study: how many of its articles fall into the decile of 0-10%, how many in 10-20%, ..., how many in the most cited decile of 90-100%. For each article, the decile will be determined by its own set of "similar" ones (with the same characteristics of the topic, year of publication and type), however the results — the deciles obtained — can be combined across the entire array of publications, and this will be quite appropriate [1].

The author's citation profile is a histogram of the distribution of the proportion of articles published by the author by citation deciles. The histogram shows how often the author publishes “strong” or “weak” articles. This is an ample image of the bibliometric effectiveness of the scientist's publications.

### 2.3.5. “Weighted” Indicators

The following two indicators are called "*weighted*" because they take into account the citations received by the journal with different weights — depending on how "credible" is the source from which the citation is obtained [17]. This credibility, in turn, depends on the citation of the source of the reference itself. A reference from the Highly Cited journal Nature is valued more than a reference from a much less known source.

Specialists from the laboratory of Karl Bergstrom (George Washington University) proposed the weighted journal indicator, called the **Eigenfactor** (Own factor), in 2007.

The mechanism by which weighted indicators can be illustrated is the "distribution of prestige" to journals using an iterative procedure. First, each magazine is assigned the same initial "prestige" equal to one. At the first stage of the iteration, each journal equally "divides" its prestige unit among all references. He "sends" references to other journals, thus transferring a share of his prestige. The fewer references are made in the journal, the greater the "weight" of each of them. Thus, the "prestige" received by each journal from other journals is calculated [21]. Further iterations are repeated until a quasi-stable state is reached, when each further iteration step practically does not change the value of the "prestige" received by each of the journals.

The **Article Influence**, which characterizes an average article in the journal, is the journal's own factor normalized by the number of articles in the issue.

Both indicators have been published in JCR since 2007. The Article Influence index levels out the difference in the citation activity in various scientific areas, since the "prestige" of the journal is distributed equally [17] among all references originating from it.

## 2.4. Scientific Journal Indicators: SJR and SNIP

In order to avoid the shortcomings of the impact factors of journals and to be able to compare the level of journals from different scientific areas, some complementary techniques of evaluating the quality of journals [21] were developed: **Scimago Journal Rank (SJR)** and **Source Normalized Impact per Paper (SNIP)** [1].

**SJR**, like the weighted journal indicator **Eigenfactor** (Own factor), takes into account the value of references; a reference from a more prestigious journal is accepted with more weight than a reference from a less prestigious one [20]. Moreover, the self-citations of the journal in this type of calculation are accepted only at 33% [1].

At the first iteration, each journal is assigned a rank equal to one, and then this rank, proportionally to the number of citations received, is redistributed between the journals whose articles have been cited. After a certain number of such reallocations of ranks, the highest ranking journal that has received the maximum number of citations gets the maximum value of the indicator. In addition, one citation from a prestigious journal appears to be more valuable than a citation from a less [17] prestigious journal, which only distributes citations, but does not receive them in return.

To obtain the final value of the **SJR** of the journal, a citation window of 3 years is set and the number of papers published in the journal normalizes the resulting «prestige». Thus, the **SJR** indicator takes into account the level of citing journals and is independent of the field of scientific research.

**SNIP** takes into account the citation potential of the journal.

To calculate the citation potential of the journal under study, the individual environment of the journal is determined. For that, all journals that have cited the articles of a given journal over the past 10 years are selected, and the average length of the list of references in the citing articles is calculated. Then, all references that are not within the database from which the indicators are calculated, are excluded from the lists. This makes it possible to neglect the differences in the traditions of citation in different fields of knowledge [21].

The standard three-year impact factor is divided by the citation potential, and a rating indicator of the journal independent of the field of knowledge is obtained. The crucial factor for calculating **SNIP** is the completeness of the database.

**SJR** and **SNIP** are calculated using the Scopus database. The main indicator of the journal level in **Scopus** is the **CiteScore** indicator; it is calculated as a classic impact factor with a citation window of 3 years.

The **RISC** calculates more than 50 different indicators for journals, which makes possible an ample evaluation of a journal [1].

The scientist's awareness of various bibliometric indicators helps them to develop an optimal publication strategy and choose the journals that are most suitable for their prestige and scientific profile.

## 2.5. Test Questions

1. What are bibliometric indicators?
2. What information about the article do databases contain?
3. Can bibliometric databases include books?
4. Can an author specify multiple affiliations?
5. What is the main function of the following databases: Journal Citation Reports, Essential Science Indicators and InCites?
6. What indicators are collected in the Journal Citation Reports (JCR) database?
7. What data are published in Essential Science Indicators (ESI)?
8. What are the features of the InCites Benchmarking & Analytics (InCites) database?
9. What are impact indicators?
10. What specifies the impact factor of the journal?
11. What is the publication window and the citation window? For which journal indicators are these intervals used?
12. In which database and when are the data on the impact factors of the journal published for every further year?
13. How is the five-year impact factor of the journal calculated?
14. What is the ranking method? What is a journal of the first quartile? How many quartiles are there in total?
15. How is the Scimago Journal Rank (SJR) indicator determined? What does SJR take into account?

## CHAPTER 3. ASSESSMENT OF THE EFFECTIVENESS OF SCIENTIFIC ACTIVITY

Back in the 19th century, mathematician Charles Babbage proposed using data on the number of scientific articles to assess the fame of a scientist. However, this method of evaluation did not take any account of the quality of scientific publications.

In 1867, the first Catalog of Scientific Papers appeared which made possible analyzing statistics on publications. As early as then, bibliometric indicators began to be used all over the world to evaluate scientific activity, distribute grants, and make decisions about employment.

### 3.1. Tendencies of Developing the Methods of Assessment of Scientific Activity

Currently, improper use of simple citation indicators and the number of publications only causes distrust of bibliometric on the part of the scientific community and may result in some errors in evaluating activities of individual scientists and institutions. Even the most adequate use of *bibliometric indicators* cannot replace *expert assessment*.

Citation analysis helps to investigate scientific communication, identify the existing and emerging research areas – the "hot trends". However, when evaluating scientific research, it is important that the concept of "*impact*" is not identical to the concept of the "*quality*" of scientific research [1].

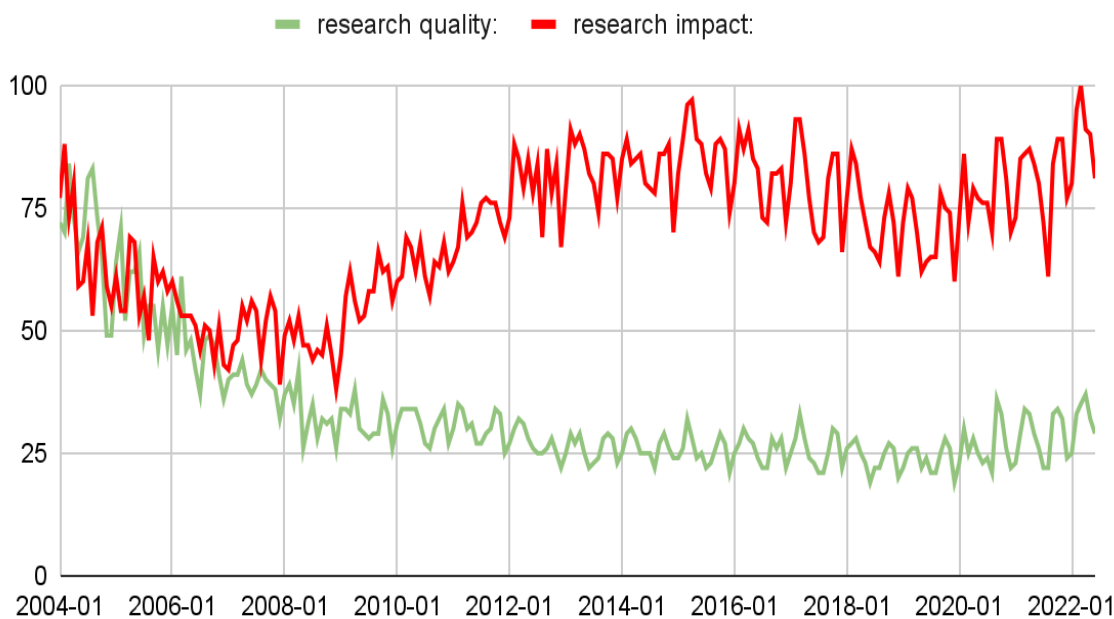


Figure 3.1. Comparison of queries in Google about the quality and quantity of research

The comparison of Google search queries on the quality and quantity of research shown in Figure 3.1 shows that over the past twenty years, the number of queries on the "influence" of an article has increased and the number of queries on the "quality" of research has decreased.

When choosing a journal for publication, scientists should be guided primarily by the journal's thematic focus rather than its "influence", and scientific administrators - by the content of the articles rather than the quantitative indicators of the scientists' publication activity.

The most adequate assessment of scientific activity requires combinations of bibliometric indicators and expert assessments in different proportions (as indicated in Table 1).

Table 1. Approximate ratio of methods for assessing scientific activity

Unit of assessment	Expert assessment	Bibliometric indicators
Country	20%	80%
Organization	40%	60%
Research group	60%	40%
Scientist	80%	20%

Back in 2012, the San Francisco Declaration on the Evaluation of Scientific Activity was adopted, calling, among other things, for avoiding using the value of the **Impact Factor Journal** to assess the *quality of articles*. In order to obtain an assessment adequate to the goal, it is necessary to involve experts who, with their opinion, can compensate for the imperfection of formal indicators and will give an assessment taking into account the characteristics of the evaluated scientists, teams and institutions. Scientometrics assessment thus acts as a supplement to expert opinion. A similar appeal to the scientific community was made in Russia in 2019 [1].

Professional expertise is to be skillfully combined with the competent use of objective Scientometrics indicators at different levels of evaluation of a scientist / institution / country, and in different scientific fields, and this combination can and has to be used in different proportions.



## 3.2. Methods for Assessing the Effectiveness of Scientists

### 3.2.1. Hirsch Index

In 2005, when selecting candidates for the position of Professor of Physics at the University of California, San Diego, *Jorge Hirsch* proposed introducing a new indicator. With the help of this indicator, J. Hirsch tried to give a comprehensive assessment of the effectiveness of a scientist by taking into account both the number of publications and their citation (quantity and quality). The evaluation method was named after the author's surname - the **Hirsch Index** (*h-index*).

An example of calculating the *h-index* is shown in Figure 3.2. Each point on this graph corresponds to one publication. The serial number of the publication in the list ordered by citation is plotted along the axis *X* (the first articles received the most references). The citation of publications is plotted along the axis *Y* – how many references the article with the corresponding number received. The first publications in the order of *h*, that is, the most cited ones, make up the **Hirsch core**.

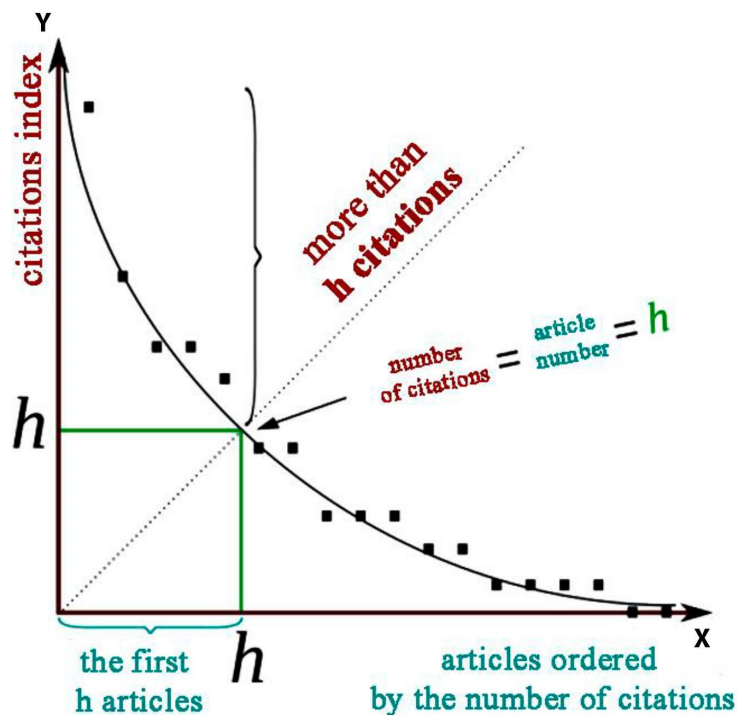


Figure 3.2. Calculation of **Hirsch** Author Index

According to the definition, the *Hirsch Index* of an array of publications of a scientist will be equal to *h* if there are *h* publications from this array, each of

which have received at least  $h$  citations, and each of the other articles received no more than  $h$  citations [21].

If we connect the points to obtain a graph  $y=f(x)$ , then the intersection point of this graph with the bisector  $y=x$  will correspond to the *h-index*.

If a scientist has written 200 articles, but each of them received only 10 references, then the Hirsch Index of this scientist will be equal to 10. On the other hand, if a scientist has written only 10 articles, each of which has received 200 references, his *Hirsch Index* will still be 10.

To achieve a high value of the *Hirsch Index*, scientists have to write many papers that receive many citations. The *Hirsch Index* tries to reflect a balanced assessment of both the publication activity of a scientist and the citation [21] of their works.

In his work, Hirsch studied the publications of American physicists and found that the *Hirsch Index* of the order of 10-12 corresponds to a permanent research position at a large university, 15-20 to membership in the American Physical Society, and the Index of 45 and higher to membership in the National Academy of Sciences of the USA [1].

It should be noted that the *Hirsch Index* reflects the level of performance of a scientist quite adequately; it depends on the field of knowledge and other factors as well as on all scientometric indicators [21]. Its author himself [14] recognizes all the limitations of the *Hirsch Index*.

Limitations or disadvantages of the *Hirsch Index*:

- The *Hirsch Index* is a "lifetime" indicator of the academic age of a scientist: this results in the inequality of young and distinguished specialists (the *Hirsch Index* is lower for scientists with a lower "academic age"). A scientist may fail to have published anything for many years, but the *h-index* will not be lower than when they were at the peak of their career.
- The *Hirsch Index* does not depend on the disciplinary field; it is not a normalized indicator for the areas of science: it is possible to compare the effectiveness of scientists only within the same scientific area. Figure 3.3 shows the results of a comparison of the *Hirsch Indices* for scientists from various fields.
- The *Hirsch Index* is an integer, i.e. it has a low resolution.
- The *Hirsch Index* does not take into account the actual number of references received by articles above and below the  $h$  point: only the position of this point is important. Imagine two scientists have published 100 articles each and each article of the first one has received 10 references, and only 10 articles of the second one received 10 references, and the remaining 90 articles were not cited at all. In

this case, with an obvious disparity in the performance of these scientists, they will have the same values of the *Hirsch Index* = 10.

- The *Hirsch Index* does not consider the contribution to the overall citation of the most highly cited publications: the total citation of publications included in the *Hirsch core* does not affect the h-index.
- The *Hirsch Index* does not take into account the number of co-authors in publications and the type of paper.

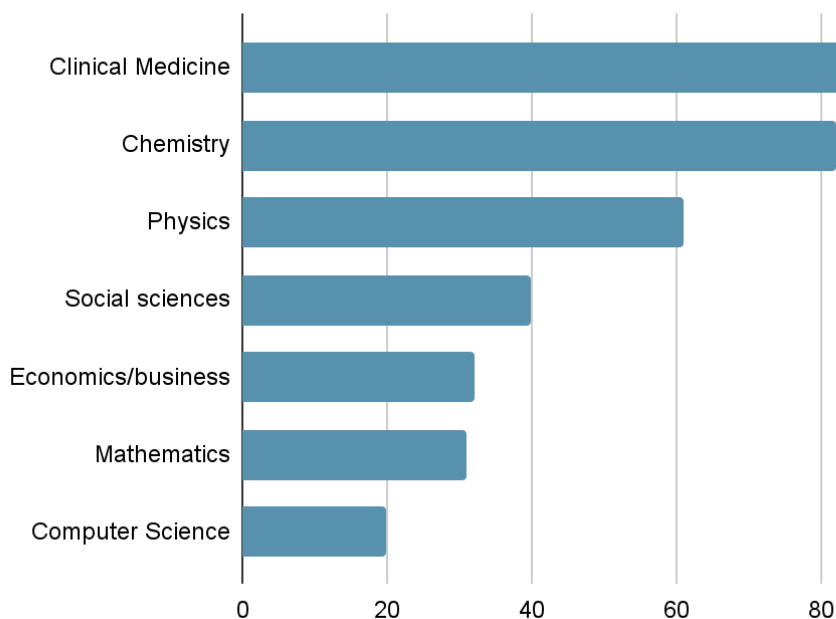


Figure 3.3. Average Hirsch Index for different areas

To compensate for the shortcomings, more than a hundred different variations of the *Hirsch Index* have already been suggested. For example, one of the options proposed by *Jorge Hirsch* himself takes into account extreme positions (the so-called outliers). This is referred to as the  $\alpha$ -index, the ratio of the number of all citations of the author to the square of the *Hirsch Index*:  $\alpha = \frac{N_{\text{обш}}}{h^2}$ .

Other variations take into account, for example, the duration of a scientist's career. To account for the citation of articles included in the *Hirsch core*, it is also proposed to additionally calculate the average number of references to articles included [21] in the *Hirsch core*, or to determine the median of the number of citations  $h$  of articles included in the *Hirsch core* of the author's publications. To account for the number of co-authors in publications, it is proposed to take into account the result of dividing the *h-index* by the average number of authors in articles that are included in the *Hirsch core* of publications [1].

Based on the *h-index*, a **hierarchical Hirsch Index** is created - the *I-index*. The hierarchical index evaluates the publication effectiveness at different levels: journal, institution, university, country. *I-index* shows that a scientific institution has an *index I*, if at least *I* scientists of the organization have a *Hirsch Index* of at least *I*. Similarly, at the next level, it is possible to rank all scientific institutions of the country [21]. The *Hirsch Index* has been so successful because a number of science administrators considered it a simple and universal tool for evaluating scientists and institutions. It is very tempting to be able to evaluate the effectiveness of scientific activity with one number. In fact, a deeper analysis has shown that the *h-index* is not a panacea, but only one of the bibliometric indicators that must be used in combination with other metrics for an effective and adequate assessment of scientific activity.

### 3.2.2. Beamplots: a New Standard for Citation Statistics

In 2021, Clarivate proposed a new method for evaluating the effectiveness of scientific activity – **beamplots** (span diagrams). Beamplots visualize the effectiveness of an author based on the citations of their publications over a long time interval. An example of beamplots is shown in Figure 3.4.

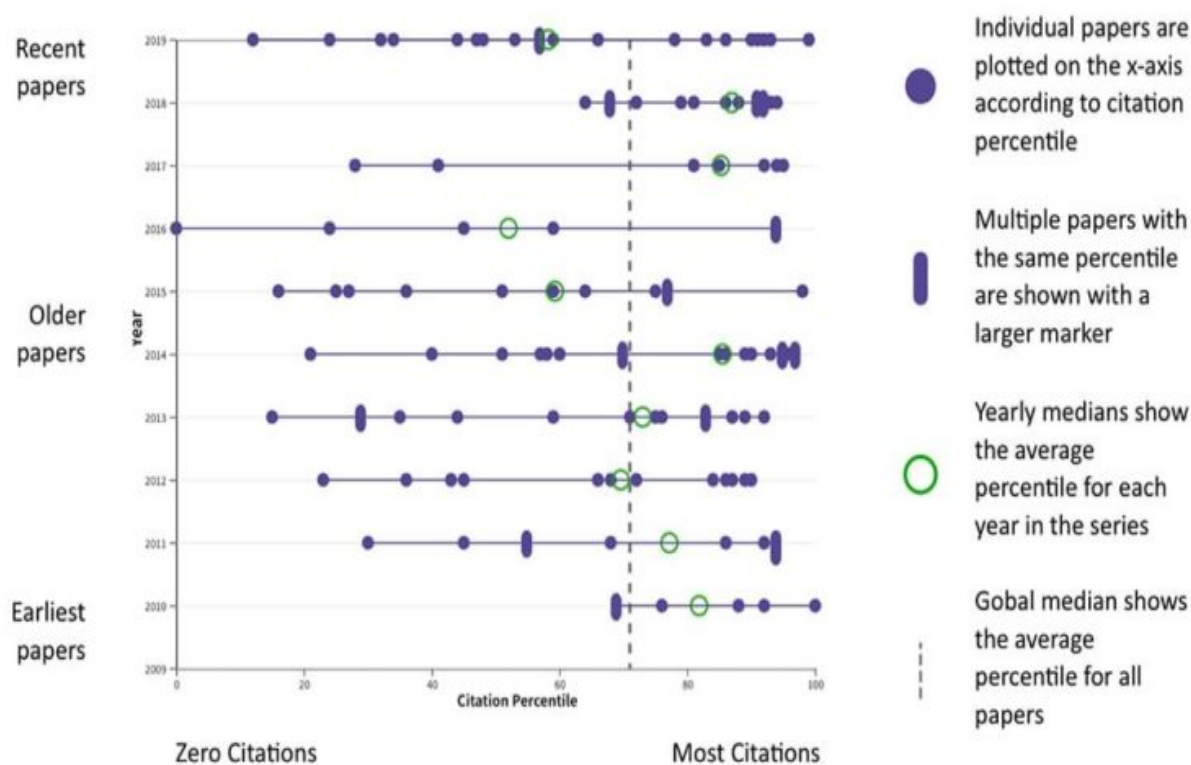


Figure 3.4. **Beamplots** by the author

Beamplots in WoS CC are based on *percentiles*.

A percentile is a value that a given random variable does not exceed with a fixed probability given as a percentage.

All publications are sorted by citation, if the publications have all the three properties:

- belong to the same type,
- belong to the same subject area,
- were published in the same year.

The *zero percentile* is zero citations. The *hundredth percentile* is the maximum citation among all publications of the same type, the same year, the same subject area.

For each narrow scientific, a specific type of paper and a specific year, the maximum and minimum number of citations is determined. The number of citations of a specific publication related to the same field and published in the same year is determined according to this interval.

For example, if the maximum number of citations of a publication in a particular scientific field in one year is 100, and the minimum is zero, then a publication with 80 citations will fall into the 80 percentile. Normalization within narrow scientific fields allows, with the introduction of certain assumptions, to compare different scientific fields.

A set of timelines is built, each of which represents a time interval (usually one year). At each time interval, the author's publications are plotted in accordance with the percentiles scored. The more the dot is to the right; the better the publications are cited. For each year, the arithmetic mean is determined, and then the arithmetic mean for the entire career of a scientist is plotted (dotted line).

#### Advantages of Beamplots:

- You can see how frequently the author's publications are cited.
- It is possible not only to observe the dynamics of publication activity, but also the dynamics of citations.

Beamplots allow visually comparing authors from different scientific fields and whose numbers of publications differ significantly. Visualization makes it easier to reveal the dynamics of publication activity. Constructing Beamplots is already available in many scientific databases, for example, in the Web of Science Core Collection [1].

### 3.3. Altmetrics: Evaluation Methods Other than Citation Analysis

With the development of Internet technologies, some new ways of evaluating the effectiveness of scientific activity, other than bibliometric methods, have appeared. These methods of evaluating publications are based on taking into account the real attention to publications, rather than not on the number of citations in databases.

The opportunity to process a large number of information sources has made it possible to assess the actual usefulness of a particular publication. This is especially true in relation to those areas of science in which the citation indicators are poorly correlated with the actual reflection of the publication value [19].

In 2010, the *Manifesto of altmetrics* was adopted [15], in order to provide scientists with tools for rapid analysis of the popularity of scientific literature. **Altmetrics** are new methods of Scientometrics that evaluate the results of scientific activity based on the presence, mention and use of publications on the Internet rather than on the number of citations. *Altmetrics* takes into account downloading articles in citation managers, mentions in social networks and scientific blogs, discussion of articles by the readers on journal websites [1] and other similar indicators.

One of the main advantages of *altmetrics* is that it reduces the communication cycle. Thus, for example, evaluating a publication by the *Hirsch Index*, needs a certain period of time to fix the number of citations (relative to the year of publication), whereas using *altmetrics*, it is possible to assess the publication importance of the article weeks or even days after the publication. The high speed of *altmetrics* data presentation allows a scientist to quickly receive the information about the publications recommended by their colleagues, share the articles read, and set up notifications about new arrivals.

*Altmetrics* is becoming a powerful tool of public expertise of publications and is already implemented as a complementary alternative expertise in open access journals.

Unlike the *Impact Factor* of the journal, *altmetrics* reflect the value of the article itself outside the closed academic community and do not depend on the journal's prestige indicators. *Altmetrics systems* may appear to be more reliable indicators for journals than conventionally calculated *Impact Factors*, which may be subject to manipulation.

### 3.4. Test Questions

1. Are the concepts of "impact" and "quality" of scientific research identical?
2. Why is it that the Impact Factor of the journal cannot be used to evaluate the quality of articles?
3. What combination of indicators is to be used for a proper assessment of scientific activity?
4. When, who and why suggested using the Hirsch Index?
5. Give a definition of the Hirsch Index.
6. Under what conditions will a scientist have a high Hirsch Index?
7. List the disadvantages of the Hirsch Index.
8. Give a definition of the  $\alpha$ -Hirsch index. What does it take into account?
9. Which publications form the core of Hirsch?
10. Why is the Hirsch Index so popular when evaluating the effectiveness of scientific activity?
11. What are Beamplots?
12. Describe the advantages of Beamplots.
13. Do Beamplots allow comparing the effectiveness of scientific activity of authors from different scientific fields?
14. What are altmetrics?
15. Explain the advantage of the altmetrics system over databases of citations.



## CONCLUSION

Scientometrics is a dynamically developing field in which new sources of scientific data are constantly emerging, as well as new tools and approaches to evaluate the effectiveness of scientific activity.

The list of indicators available in Clarivate's tools and even more so, of those existing in bibliometric, is much broader than the main bibliometric and journal indicators discussed in this tutorial.

There are no ideal bibliometric indicators. When evaluating the effectiveness of scientific activity of a scientist, an institution or a whole country, it is necessary to use an integrated approach: bibliometric research is to be complemented by professional expert assessments, as well as by altmetrics, a powerful public examination of publications on the Internet.

Using bibliometric only, it is impossible to make decisions in the management of science – in the distribution of grants and funding, implementation of personnel policy, opening / closing of laboratories, programs, directions. The role of bibliometric is to provide additional information about the "scientific units" considered, to assist decision-making. Scientometrics communities use the concept of "informed decision", "decision with full information", and bibliometric analysis plays an important role in the implementation of this concept.

## ABBREVIATIONS

**A&HCI** — Arts & Humanities Citation Index [1] (Index of cited literature in the humanities and arts).

**BKCI-S** — Book Citation Index Science [1].

**CPCI-S** — Conference Proceedings Citation Index Science [1].

**DOI** — Digital Object Identifiers.

**ESI** — Essential Science Indicators [1].

**ESCI** — Emerging Sources Citation Index [1].

**IF** — Impact Factor [1].

**ISI** — Institute for Scientific Information [1].

**ISSI** — International Society for Scientometrics and Informetrics [1].

**JCR** — Journal Citation Reports [1].

**JIF** — Journal Impact Factor [1].

**JNCI** — Journal Normalized Citation Impact [1].

**NCI** — Normalized Citation Impact.

**R&D** — Research and Development.

**RSCI** — Russian Science Citation Index.

**SCI** — Science Citation Index [1].

**SCIE** — Science Citation Index Expanded [1] (Index of cited literature in science and technology).

**SJR** — Scimago Journal Rank [1].

**SNIP** — Source Normalized Impact per Paper [1] (citation normalized by reference sources).

**SSCI** — Social Sciences Citation Index [1] (Index of cited social science literature).

**TRL** — Technology readiness level (Levels of technology readiness to practice).

**WoS CC** — Web of Science Core Collection [1].

**WoS** — Web of Science [1] (Clarivate information platform with its main information products).

## LIST OF SOURCES USED

1. Handbook on Scientometrics: Science and Technology Development Indicators, Second edition. 2021. 358 p. DOI 10.15826/B978-5-7996-3154-3 (second edition).
2. Akopov A. I. Scientific journals: a review of scientific developments and an attempt at typological differentiation against the background of socio-economic and professional problems // Scientific and Cultural Journal. – 2007. – № 12 (157).
3. Price D., Beaver D. Cooperation in the "Invisible College" // Communication in Modern Science: S. translations. - M.: Progress, 1976. - P. 335-350.
4. Web of Science Journal Evaluation Process and Selection Criteria. n. d. – <https://clarivate.com/webofsciencegroup/journal-evaluation-process-and-selection-criteria/>.
5. Web of Science Conference Proceedings Selection Process. n. d. Accessed April 18, 2020. – <https://clarivate.com/webofsciencegroup/essays/web-science-conference-proceedings-selection-process/>.
6. The Selection Process for the Book Citation Index in Web of Science. n. d. Accessed April 19, 2020. – <https://clarivate.com/webofsciencegroup/essays/selectionprocess-book-citation-index-web-science/>.
7. Editorial Selection Process. Web of Science Core Collection. n. d. Accessed April 9, 2020. – <https://clarivate.com/webofsciencegroup/solutions/editorial/>.
8. Web of Science Platform: Data. n. d. Accessed April 9, 2020. – <https://clarivate.com/webofsciencegroup/solutions/web-of-science/>.
9. Derwent Innovations Index on Web of Science. n. d. Accessed April 9, 2020. – <https://clarivate.com/webofsciencegroup/solutions/webofscience-derwent-innovation-index/>.
10. Data Citation Index. n. d. Accessed April 9, 2020. – <https://clarivate.com/webofsciencegroup/solutions/webofscience-data-citation-index/>.
11. Regulations of the Bibliographic Database. Russian Science Citation Index. – 2018. – [https://www.elibrary.ru/projects/rsci/reglament\\_RSCI.pdf](https://www.elibrary.ru/projects/rsci/reglament_RSCI.pdf).
12. Bredihin S. V., Kuznetsov A. Yu. Methods of bibliometric and the market of electronic scientific periodicals. – Novosibirsk, Moskva: IVMiMG SO RAS, 2012. – 248 p.
13. Bredihin S. V., Kuznetsov A. Yu., Shcherbakova N. G. Analysis of citation in bibliometric. – Novosibirsk, Moskva: IVMiMG SO RAS, NEIKON, 2013. – 344 p.
14. Conroy G. What is Wrong with the H-Index, According to Its Inventor // Nature Index. – 2020. – <https://www.natureindex.com/news-blog/whats-wrong-with-the-hindex-according-to-its-inventor>.
15. Priem J., Taraborelli D., Groth P., Neylon C., Altmetrics: A manifesto. 2010. Oct. 26. – <http://altmetrics.org/manifesto>.

16. Moskaleva O.V. Research Publications as a Means of Communication, Analysis and Assessment of Research Activity // Handbook on Scientometrics: Science and Technology Development Indicators. - 2014. - P. 110-156. - DOI 10.15826/B978-5-7996-1352-5.0006
17. Pislyakov V.V. Bibliometric Indicators in the Thomson Reuters Information Resources // Handbook on Scientometrics: Science and Technology Development Indicators. - 2014. - P. 75-106. - DOI 10.15826/B978-5-7996-1352-5.0005
18. Moskaleva O.V. Research Publications as a Means of Communication // Handbook on Scientometrics: Science and Technology Development Indicators, Second edition. - 2021. - P. 140-176. - DOI 10.15826/B978-5-7996-3154-3.007
19. Moskaleva O.V. Research Publications as a Means of Analysis and Assessment of Research Activity // Handbook on Scientometrics: Science and Technology Development Indicators, Second edition. - 2021. - P. 221-278. - DOI 10.15826/B978-5-7996-3154-3.009
20. Pislyakov V.V. Bibliometric Indicators in the Clarivate Databases // Handbook on Scientometrics: Science and Technology Development Indicators, Second edition. - 2021. - P. 177-220. - DOI 10.15826/B978-5-7996-3154-3.008
21. Handbook on Scientometrics: Science and Technology Development Indicators. 2014. 250 p. DOI 10.15826/B978-5-7996-1352-5.0000 (first edition).

**Marusina Maria Yakovlevna**

**Scientific Index Search**

**Training manual**

In the author's edition

Editorial-Publishing Department of ITMO University

Head of EPD

Signed for printing

Order № 4802

Circulation

Printed on a risograph

N.F. Gusarova

**Editorial-Publishing Department**

**ITMO University**

197101, St. Petersburg, Kronverksky pr., 49, Letter A