

# Why IEEE Xplore Matters for Research Trend Analysis in the Energy Sector

B.N. Chigarev\*

Oil and Gas Research Institute of Russian Academy of Sciences, Moscow, Russia.

**Abstract** — The paper aims to briefly compare and analyze the results of queries to IEEE Xplore and the leading abstract databases Scopus and Web of Science to identify research trends. Some errors were revealed in the Author Keywords in Web of Science. Therefore, a more detailed analysis that involved comparing various types of key terms was made only for IEEE Xplore and Scopus platforms. The study employed IEEE Access journal metadata as indexed on both platforms. Sample matching for IEEE Xplore and Scopus was achieved by comparing DOI. The IEEE Xplore metadata contains more key term types, which provides an advantage in analyzing research trends. Using NSPEC Controlled Terms from expert-compiled vocabulary provides more stable data, which gives an advantage when considering the change of terms over time. Apriori, an algorithm for finding association rules, was used to compare the co-occurrence of the terms for a more detailed description of sample subjects on both platforms. VOSviewer was used to analyze trends in scientific research based on IEEE Xplore data. The 2011-2021 ten-year period was divided into two sub-intervals for comparing the occurrence of Author Keywords, IEEE Terms, and NSPEC Controlled Terms. Bibliometric data of the IEEE conference proceedings was used to illustrate the importance of context in estimating the growth rate of publishing activity on a topic of interest.

**Index Terms:** bibliometric analysis, IEEE Xplore, INSPEC Controlled Terms, keywords co-occurrence, research trends, Scopus.

## I. INTRODUCTION AND OBJECTIVES

The increasingly sophisticated and competitive landscape of scientific works demands an in-depth analysis of research trends for decision-making in developing an innovation development strategy.

This topic is very diverse and is well represented in scientific publications that address various aspects relevant for identifying research trends. For example, the research in [1] relied on bibliometric methods to map intellectual structures and research trends. Data was collected from citations and co-citations found in Science Citation Index Expanded, Social Science Citation Index, and Arts and Humanities Citation Index. The application of bibliometric analysis allowed the authors to identify research trends related to innovative entrepreneurship. Multicriteria decision analysis (MCDA) approaches to incorporating social criteria and evaluating participatory mechanisms in the decision-making process for renewable energy projects are discussed in [2]. The authors expect that in the future, developing countries with a high potential for energy production from renewable sources will face problems in assessing the potential social implications of the decisions made. According to [3], science plays a significant part in decision-making on the sustainable development of renewable energy. The study applies a textual analysis approach to 2 533 Scopus-indexed metadata published from 1990 to 2016, based on a Latent Dirichlet Allocation Topic Model. The models created include up to 1 100 topics. The most developed ones are energy storage, photonic materials, nanomaterials, and biofuels. The establishment of sustainable energy systems will require future research to focus not only on technical energy infrastructure but also on related economic, environmental, and political issues. The analysis presented in [4] aims to identify the major trends in research on artificial intelligence (AI) in business. The authors conduct a bibliometric analysis of Web of Science and Scopus data. They identify 11 clusters and the most common terms used in AI research whose analysis has shown a growing scientific interest in synergies between AI and business. Identifying research trends helps make

---

\* Corresponding author.  
E-mail: [bchigarev@ipng.ru](mailto:bchigarev@ipng.ru)

<http://dx.doi.org/10.38028/esr.2021.03.0005>

Received September 22, 2021. Revised October 11, 2021.

Accepted November 03, 2021. Available online November 28, 2021.

This is an open access article under a Creative Commons Attribution-NonCommercial 4.0 International License.

© 2021 ESI SB RAS and authors. All rights reserved.

decisions on the selection of prospective research topics. Citation and publication delays constrain such analysis. Therefore, the authors of [5] use an approach called predicting the frequency of author-specified keywords to identify research trends. A long short-term memory neural network (LSTM) is used for the analysis. It is noted that the feature characterizing the potential for community development is especially significant in the long-term prediction.

The energy transition to low-carbon power sources requires significant development of the power grid infrastructure and optimization of its operation.

Grid infrastructure topics are well represented in the Scopus abstracts database. For example, the query TITLE-ABS-KEY ("grid infrastructure") yields 3 157 papers (as of September 2021).

Therefore, only a brief list of publications that reveal relevant issues of this topic will be given.

Key aspects affecting the integration of microgrids in the broader context of energy transformation are presented in [6]. In contrast to other decentralized energy systems, microgrids interact with centralized grid infrastructure. The authors' analysis shows that California's path to microgrids is mostly driven by legislative and regulatory pressures toward clean energy and symbiotic relationships between regime influencers and the microgrid niche. The authors of [7] analyze the spread of solar energy in Portugal, both nationally and locally. They note that the energy transition must be implemented in a multiscale, multilateral, and intersectoral perspective. Since solar power plants require access to land and electric grids, the establishment of the solar energy infrastructure involves interaction with local communities. In [8], the authors present a macroeconomic assessment of planned investments in power grid infrastructure in Germany. Investments in power grid infrastructure are mainly aimed at achieving environmental and energy policy goals. Using a statistical analysis, the authors show how the multiplier effect of grid investments impacts macroeconomic outcomes: production, added value, employment, and tax revenues. The net multiplier effect on production volume is positive, whereas the other effects are negative. Research related to smart grids as forerunners of the Energy Internet, which should connect producers and consumers of electricity with renewable energy sources and storage units, is discussed in [9]. The paper presents a systematic review of the literature related to the current state of the Energy Internet. The authors found that although the infrastructure, technology, and system design are reasonably ready for the transition to the Energy Internet, the major obstacles are defined by regulations.

These tasks are classical for the experts of the Institute of Electrical and Electronics Engineers (IEEE). Research trends in this area are reasonable to justify the use of the metadata of the IEEE Xplore platform, which is currently insufficiently used in the bibliometric analysis.

Each abstract database has its strengths and weaknesses to be considered in the bibliometric analysis to identify the trends in scientific research. There is extensive literature dedicated to comparative analysis of capabilities of abstract databases, as well as the errors and issues that arise in them.

The study presented in [10] examines 3 073 351 citations found by Web of Science, Scopus, Google Scholar, Microsoft Academic, Dimensions and Open Citations Index of CrossRef of 2 515 English-language highly cited papers (from 252 subject categories) published in 2006. The authors conclude that in terms of coverage Microsoft Academic and Dimensions are good alternatives to Scopus and WoS in many subject categories. However, it is worth noting that the metadata structure of these sources differs significantly, especially in terms of the keywords offered by the systems and the classification of subject categories. Authors of [11] claim that the original purpose of scientific publications was to provide a global exchange of scientific results, ideas, and discussions among the academic community to achieve better scientific results. Nowadays, many of the most crucial decisions on industrial and economic growth priorities, allocation of financial resources, educational policies, creation of opportunities for collaboration, acquisition of status, employment of academic staff, and others rely on the evaluation of scientific results and research quality approximated as publication impact has become the most significant criterion. The authors aim to provide all potential users with a comprehensive description of the two main bibliographic databases – Web of Science and Scopus. The variety of publications devoted to comparing individual abstract databases is enormous; one can find a suitable comparison for the most famous databases. For example, the authors of [12] have found that Google Scholar indexed the most recent papers indexed in WoS, and now they can be found through Google Scholar. The ratio of quantity and quality of citations, threats to WoS, and weaknesses of Google Scholar are discussed. Some publications compare abstract databases by specific indicators. For example, in [13], a comparative analysis of journal coverage is made for three databases (Web of Science, Scopus, and Dimensions) to understand and visualize their differences. The analysis employed the most recent lists of major journals from the three databases. Findings indicate that the databases differ significantly in journal coverage with Web of Science being the most selective and Dimensions – the most comprehensive. Comparison of the data presented in abstract databases indicates that the specific direction of research and authors' affiliation are also important. In [14], the authors compare three resources (Web of Science, Google Scholar, and Scopus) to determine the resource with the most representative coverage of citations of South African environmental research. The study has found that Web of Science extracts most citation results, followed by Google Scholar and Scopus. WoS shows the best results in terms of overall coverage of journal samples and also

extracts the largest number of unique articles. A multiple-copy study shows that WoS and Scopus find no duplicates, whereas Google Scholar finds a few of them. Scopus provides the fewest inconsistencies in terms of content verification compared to the other two citation resources.

When conducting bibliometric research, it is essential to understand what errors and inaccuracies a researcher may encounter when using the metadata of leading abstract databases. A wide range of studies is devoted to this issue. We will cite some of them, which reveal this problem to the greatest extent. In [15], the authors focus on a systematic analysis of duplicate entries in Scopus, and [16] presents empirical analysis and classification of database errors in Scopus and Web of Science. The study in [17] analyzes the so-called “phantom citation” (i.e., articles about which the WoS reports that they are citing an article when, in fact, they are not). An analysis of citations (and article references) in two English-language and two non-English-language sources shows that phantom citations and other indexing errors are about twice as common in non-English-language articles. These and other errors affect about 1% of citations in the WoS database. This factor influences the calculation of h-indices or other indicators of research impact. Another aspect of citation problems [18] is missing citations, i.e., the lack of links between the cited article and the corresponding citing article. This study is based on an extensive sample of scientific articles concerned with engineering and manufacturing and focuses on the old data in Scopus and WoS databases. The main results of this study are as follows: 1) both databases are slowly correcting old missed citations, and 2) a small fraction of initially corrected citations may suddenly disappear from the databases over time.

The developers of the free software VOSviewer, widely used in bibliometric research, made a significant contribution to an analysis of issues with the metadata of leading abstract databases. In [19], they present a large-scale comparison of five interdisciplinary sources of bibliographic data (Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic) for 2008-2017. Scopus was pairwise compared with each of the other data sources. The authors emphasize the importance of combining comprehensive coverage of scientific literature with a flexible set of filters for its selection.

Both the citation rate and the impact factor of a journal are significant to assess individual articles and their authors. Authors of [20], using computer modeling, show that under certain conditions, the impact factor is a more accurate indicator of the value of articles, whereas, under other conditions, the number of citations received by an article is a more precise indicator of its value than the impact factor, i.e., it is crucial to critically discuss research assessment criteria. This statement is especially significant for new publications whose citations have not yet been formed in abstract databases.

Systematic bibliometric analysis of metadata of

scientific publications and conference proceedings reveals major R&D trends. Traditionally, abstract databases - Scopus and Web of Science (WoS) - are used for this purpose, however, there are a growing number of specialized platforms that allow collecting information for such analysis, such as OnePetro, IEEE Xplore, and Semantic Scholar.

Specialized abstract databases may better reflect the opinions of experts in the field than general databases.

This paper aims to highlight some of the IEEE Xplore features, which, along with its openness, may provide additional benefits compared to closed access databases Scopus and WoS.

#### *On the underestimation of IEEE Xplore as a source of bibliometric metadata*

Document topics are most commonly defined by a set of terms that describe the subject area well and frequently occur in them. This approach makes it possible to assess research trends by the occurrence of key terms describing published documents. Terms can be the author's keywords, documents text mining terms, or experts-controlled terms from a subject vocabulary [21-23].

The paper focuses only on this aspect of bibliometric analysis.

To clarify the underestimation of IEEE Xplore as a source of publication metadata, several comparisons are made for the queries containing the basic terms bibliometrics OR scientometrics and the names of leading abstract databases. The following results are obtained

For queries in Scopus:

- TITLE-ABS-KEY ((bibliometric\* OR scientometric\*) AND “ieee xplore”) → 11 results;
- TITLE-ABS-KEY ((bibliometric\* OR scientometric\*) AND “scopus”) → 3 797 results;
- TITLE-ABS-KEY ((bibliometric\* OR scientometric\*) AND (“WoS” OR “web of science”)) → 5 919 document results.

For queries in Web of Science Core Collection:

- (bibliometric\* OR scientometric\*) AND “ieee xplore” (Topic) → 10 results;
- (bibliometric\* OR scientometric\*) AND “scopus” (Topic) → 3 026 results;
- (bibliometric\* OR scientometric\*) AND (“WoS” OR “web of science”) (Topic) → 4 840 results.

No results were found for “All Metadata:” bibliometric\* OR “All Metadata:” scientometric\* AND “All Metadata:” “ieee xplore” on the IEEE Xplore platform.

IEEE Xplore platform provides a comprehensive list of metadata for publications, which enables a comprehensive bibliometric analysis ( <https://ieeexplore.ieee.org/Xplorehelp/searching-ieee-xplore/advanced-search>.)

The list of IEEE Xplore platform metadata can be used to analyze the topics of published materials:

- Abstract;
- Author Keywords;

Table 1. Examples of mismatches between Author Keywords on the Web of Science platform and the keywords in the publications themselves [24-26].

Author Keywords by WoS	Correct Keywords	DOI of article
<u>Big Data</u> ; <u>Data analysis</u> ; <u>Tools</u> ; <u>Social networking (online)</u> ; <u>Computer languages</u> ; <u>Companies</u> ; <u>Big data analytics</u> ; <u>Data analytics</u> ; <u>Deep learning</u> ; <u>Machine learning</u>	<u>Big data analytics</u> ; <u>data analytics</u> ; <u>deep learning</u> ; <u>machine learning</u>	<a href="https://doi.org/10.1109/ACCESS.2019.2923270">10.1109/ACCESS.2019.2923270</a>
<u>Text categorization</u> ; <u>Semantics</u> ; <u>Feature extraction</u> ; <u>Natural language processing</u> ; <u>Bit error rate</u> ; <u>Task analysis</u> ; <u>Neural networks</u> ; <u>Text classification</u> ; <u>Text representations</u> ; <u>Label embedding</u>	<u>label embedding</u> ; <u>Text classification</u> ; <u>text representations</u>	<a href="https://doi.org/10.1109/ACCESS.2019.2954985">10.1109/ACCESS.2019.2954985</a>
<u>Task analysis</u> ; <u>Rehabilitation robotics</u> ; <u>Lighting</u> ; <u>Clutter</u> ; <u>Computer vision</u> ; <u>Training</u> ; <u>Machine intelligence</u> ; <u>Robotic vision systems</u>	<u>Machine intelligence</u> ; <u>robotic vision systems</u>	<a href="https://doi.org/10.1109/ACCESS.2019.2955480">10.1109/ACCESS.2019.2955480</a>

Table 2. Top 25 key terms according to Scopus for 1 250 records.

Author Keywords	N Index Keywords	N
deep learning	112 deep learning	181
machine learning	67 learning systems	156
blockchain	61 internet of things	114
convolutional neural network	37 convolutional neural networks	92
internet of things	37 network security	84
security	36 5g mobile communication systems	75
iot	32 blockchain	69
5g	30 classification (of information)	64
edge computing	25 convolution	64
covid-19	23 deep neural networks	64
artificial intelligence	21 energy utilization	62
optimization	21 feature extraction	62
feature selection	19 energy efficiency	61
image encryption	18 forecasting	61
feature extraction	17 surveys	61
particle swarm optimization	17 learning algorithms	60
smart grid	16 particle swarm optimization (pso)	56
classification	15 cryptography	54
cloud computing	14 machine learning	51
energy efficiency	14 digital storage	47
data privacy	13 electric power transmission networks	47
anomaly detection	12 long short-term memory	47
energy management	12 network architecture	46
intrusion detection	12 support vector machines	46
lstm	12 internet of things (iot)	45

- Document Title;
- Index Terms;
- INSPEC Controlled Terms;
- INSPEC Non-controlled Terms;
- Standard Dictionary Terms;
- Standards ICS Terms.

INSPEC Controlled Terms, Keywords from the INSPEC expert-edited dictionary are of particular interest (<https://ieeexplore.ieee.org/Xplorehelp/searching-ieee-xplore/command-search#summary-of-data-fields>.)

The list of publishers whose publications are indexed in Xplore is the second feature of this platform: IEEE (2 477 765); OUP (39 031); IET (21 473); MIT Press (11 958); VDE (10 124); Wiley (3 564); SMPTE (3 022); SAE (2 942); River Publishers (2 351); BIAI (1 517). No giants such as Elsevier and Springer Nature are on the list, but the publications of the IEEE itself dominate. The

platform focuses on industry interests. When analyzing research trends, it is important to understand the priorities of the IEEE community.

The feature of IEEE Xplore is the high number of conference materials metadata compared to journal articles and standards-related documents.

Out of 2 582 653 papers in 2011-2020, there were those of Conferences (1 992 101), Journals (482 568), Magazines (66 809), Books (25 937), Early Access Articles (9 611), standards (5 297), and courses (330).

Conference proceedings reflect industry interests more than peer-reviewed publications. For example, the major Publication Topics for IEEE Xplore in 2011-2020 were learning (artificial intelligence) (103 944), feature extraction (67 483), optimization (64 359), neural nets (46 323), the Internet (43 459), cloud computing (40 371), mobile robots (36 370), image classification

Table 3. Top 25 key terms according to IEEE Xplore for 1 250 records.

IEEE Terms	N	INSPEC Controlled Terms	N
feature extraction	209	learning (artificial intelligence)	307
optimization	143	feature extraction	155
machine learning	114	internet of things	122
task analysis	92	optimization	116
mathematical model	80	convolutional neural nets	111
training	80	neural nets	81
computational modeling	75	cryptography	78
internet of things	68	pattern classification	72
deep learning	67	diseases	64
predictive models	63	power engineering computing	62
wireless communication	63	5g mobile communication	59
data models	62	image classification	58
5g mobile communication	60	cloud computing	56
cloud computing	57	particle swarm optimization	54
computer architecture	57	mobile computing	53
heuristic algorithms	55	power grids	52
support vector machines	54	probability	52
security	53	recurrent neural nets	51
neural networks	50	security of data	48
blockchain	47	support vector machines	48
encryption	46	data privacy	47
monitoring	46	search problems	47
reliability	46	internet	46
sensors	44	medical image processing	46
protocols	43	distributed power generation	45

(36 029), control system synthesis (35 968), medical image processing (35 231), wireless sensor networks (33 215), and power grids (32 965), which are distinct engineering challenges.

In bibliometric research, it is advisable to choose the topic relevance according to the materials of conferences or patent studies and analyze peer-reviewed articles to assess the scientific validity of the topic. The choice of the goal and methods of its achievement must not rest on a closed set of data. These sets should overlap but not coincide.

## II. ANALYSIS AND RESULTS

### A. Comparison of keywords of IEEE Xplore and Scopus platforms.

Let us briefly explain why Scopus but not Web of Science was chosen for comparison.

The comparison of the expressiveness of keywords in different platforms has revealed that the Web of Science system contains many errors in the Author Keywords field, and the Keywords Plus field has few terms. This issue

requires additional, more comprehensive study and is beyond the scope of this paper. Therefore, Table 1 provides only a few examples to illustrate it.

The examples are taken from the IEEE Access journal, which provides access to the full text, making it easy to compare the author's keywords in the system and in the article. Last accessed July 10, 2021.

Such an issue, however, has not been encountered on the IEEE Xplore and Scopus platforms, which is why these systems are used further in the study.

To compare the key terms in different systems, one must establish a set of publications indexed in both systems. The IEEE Access journal, which is indexed in all the above systems, has a whole host of publications and fits the bill, for example, according to Scopus 18 073 publications in 2020.

IEEE Xplore and Scopus allow the export of 2 000 metadata for a single query, which is enough for a qualitative comparison. To select 2 000 articles out of 18 000, they are sorted by citation on each platform, and then the first 2 000 pieces of bibliometric metadata are exported. Citation rate of the articles is determined based on the platform's data,

Table 4. The 25 most commonly co-occurring key terms in the Author Keywords and Index Keywords fields of Scopus metadata records

Author Keywords	% Index Keywords	%
machine_learning*deep_learning	7.32 learning_systems*deep_learning	25.93
convolutional_neural_network*deep_learning	6.50 convolutional_neural_networks*deep_learning	15.74
feature_extraction*deep_learning	3.25 deep_neural_networks*deep_learning	12.65
artificial_intelligence*machine_learning	2.85 convolutional_neural_networks*learning_systems	10.80
covid-19*deep_learning	2.85 convolution*deep_learning	10.19
intrusion_detection*deep_learning	2.44 convolution*convolutional_neural_networks	9.57
cnn*deep_learning	2.44 long_short-term_memory*deep_learning	8.95
artificial_intelligence*deep_learning	2.44 learning_algorithms*learning_systems	8.64
classification*deep_learning	2.44 support_vector_machines*learning_systems	8.33
lstm*deep_learning	2.44 forecasting*learning_systems	8.02
security*machine_learning	2.03 convolutional_neural_networks*learning_systems*deep_learning	8.02
anomaly_detection*deep_learning	2.03 convolution*convolutional_neural_networks*deep_learning	7.41
data_analytics*machine_learning	2.03 deep_neural_networks*convolutional_neural_networks	7.41
cnn*lstm	2.03 classification*learning_systems	7.10
covid-19*machine_learning	2.03 deep_neural_networks*learning_systems	7.10
q-learning*reinforcement_learning	1.63 learning_algorithms*deep_learning	7.10
natural_language_processing*deep_learning	1.63 convolution*learning_systems	6.79
neural_network*deep_learning	1.63 feature_extraction*learning_systems	6.48
internet_of_things*machine_learning	1.63 classification*deep_learning	6.48
pandemic*covid-19	1.63 deep_neural_networks*convolutional_neural_networks*deep_learning	6.48
sentiment_analysis*deep_learning	1.63 deep_neural_networks*learning_systems*deep_learning	6.17
cnn*lstm*deep_learning	1.63 convolution*deep_neural_networks	5.86
artificial_intelligence*machine_learning*deep_learning	1.63 decision_trees*learning_systems	5.56
image_classification*deep_learning	1.22 reinforcement_learning*deep_learning	5.56
attention_mechanism*deep_learning	1.22 network_security*learning_systems	5.56

hence the difference in the lists of articles in the 2 000 most cited ones for each platform. Articles with the same DOI are sampled to resolve this issue. There are 1 250 such articles. For comparison, in 2020, the intersection of 2 000 most cited journal articles between the Web of Science and IEEE Xplore systems was 1 207, which compares with 1 250 and indicates the consistency of the results.

It is worth noting that for a sample of 1 250 records, there is no discrepancy between the Author Keywords in both systems. For this reason, the following two Tables list them once. Tables 2 and 3, each, show the 25 most common key terms: Author Keywords and Index Keywords ([https://service.elsevier.com/app/answers/detail/a\\_id/21730/supporthub/scopus/](https://service.elsevier.com/app/answers/detail/a_id/21730/supporthub/scopus/)) for Scopus; IEEE Terms and INSPEC Controlled Terms for IEEE Xplore. N in the Tables denotes the occurrence of the term.

The general topics of the terms presented in the Table can be described as deep learning, machine learning, blockchain, convolutional neural network, and the Internet of things. Data from the Table can be used to generate new queries for further collection of literature.

The terms: feature extraction and distributed power generation, power grids, data privacy are more pronounced in IEEE Xplore metadata than in Scopus, but in general, the coverage of topics in both cases is close in nature.

IEEE Xplore data is in the public domain, Author Keywords on this platform and in Scopus coincide, and INSPEC Controlled Terms reflect the subject of publications

no less expressively than Index Keywords, thus the features of IEEE Xplore are attractive for bibliometric analysis to detect research trends. An additional advantage is that experts in a narrower subject area moderate the INSPEC Controlled Terms vocabulary, and therefore, it better reflects engineering topics.

The study on the trends in topics of scientific publications assessed by frequency of occurrence (or co-occurrence) of terms indicates that the controlled dictionary yields more stable results since index terms differ wider in bibliometrics metadata at different periods. In turn, Author Keywords, being the most subjective, better reflect the current state of the topics, and it is advisable to use them to identify emerging trends in publication topics. The IEEE Xplore platform provides both capabilities. A detailed analysis of these statements is beyond the scope of this paper and deserves a separate study.

*B. Assessment of the co-occurrence of key terms based on the Apriori algorithm.*

The interrelationship of key terms can describe a topic in more detail than a set of individual terms. One method of solving this problem is the Apriori algorithm designed to find associative rules.

This section used the key terms: Author KW, Index KW, IEEE Terms, and INSPEC Terms (abbreviated from INSPEC Controlled Terms).

The set of terms that occur together was reduced by

Table 5. The 25 most commonly co-occurring key terms in the IEEE Terms and INSPEC Terms fields of IEEE Xplore metadata records.

IEEE Terms	% INSPEC Terms	%
feature_extraction*machine_learning	21.5 feature_extraction*learning-artificial_intelligence	29.36
deep_learning*feature_extraction	20 convolutional_neural_nets*learning-artificial_intelligence	22.94
training*machine_learning	9.5 neural_nets*learning-artificial_intelligence	17.43
support_vector_machines*machine_learning	9 pattern_classification*learning-artificial_intelligence	16.51
predictive_models*machine_learning	8 convolutional_neural_nets*feature_extraction	13.76
support_vector_machines*feature_extraction	8 convolutional_neural_nets*feature_extraction*learning-artificial_intelligence	13.46
data_models*machine_learning	7.5 image_classification*learning-artificial_intelligence	12.54
optimization*machine_learning	6.5 recurrent_neural_nets*learning-artificial_intelligence	10.09
machine_learning_algorithms*feature_extraction	6.5 diseases*learning-artificial_intelligence	9.17
neural_networks*machine_learning	6 image_classification*feature_extraction	8.56
task_analysis*feature_extraction	6 support_vector_machines*learning-artificial_intelligence	8.26
training*feature_extraction	6 medical_image_processing*learning-artificial_intelligence	8.26
task_analysis*deep_learning	5.5 image_classification*feature_extraction*learning-artificial_intelligence	8.26
training*deep_learning	5.5 object_detection*learning-artificial_intelligence	7.65
prediction_algorithms*machine_learning	5 internet_of_things*learning-artificial_intelligence	7.65
computational_modeling*machine_learning	5 optimisation*learning-artificial_intelligence	7.34
task_analysis*machine_learning	5 medical_image_processing*image_classification	6.73
machine_learning_algorithms*machine_learning	5 power_engineering_computing*learning-artificial_intelligence	6.42
diseases*machine_learning	4.5 medical_image_processing*image_classification*learning-artificial_intelligence	6.42
neural_networks*feature_extraction	4 neural_nets*feature_extraction	6.42
predictive_models*data_models	4 pattern_classification*feature_extraction	6.12
support_vector_machines*deep_learning	4 image_segmentation*learning-artificial_intelligence	5.81
support_vector_machines*feature_extraction*machine_learning	4 image_classification*convolutional_neural_nets*learning-artificial_intelligence	5.81
computer_architecture*machine_learning	3.5 image_classification*convolutional_neural_nets	5.81
sentiment_analysis*feature_extraction	3.5 pattern_classification*feature_extraction*learning-artificial_intelligence	5.81

imposing additional constraints, which involved sampling rows from Scopus and IEEE Xplore metadata with the word “learning.” Tables 2 and 3 show the following terms with the word “learning:” deep learning, machine learning, learning systems, learning (artificial intelligence), learning algorithms, which evidences the relevance of such a restriction on sampling.

With this constraint applied to 1 250 Scopus data records will yield:

- 246 rows containing the learning string — Author Keywords;
- 324 rows containing the learning string — Index Keywords.
- And with this constraint applied to 1 250 IEEE Xplore data records will yield:
- 200 — IEEE Terms;
- 327 — INSPEC Terms.

Values in those two lists are comparable in order of magnitude.

Preparing the data for the Apriori algorithm involved standard actions: lowercasing strings, combining terms with different endings, removing unwanted characters, and combining words in a term into a single string by replacing

spaces with underscores.

### C. Results of applying the Apriori algorithm to the formed samples.

The 25 most frequent term groups for Scopus records for Author Keywords and Index Keywords, respectively, are listed in Table 4. The designation in Tables 4 and 5 are: % is the percentage of this key term group in the overall list of term groups that passed the 1% threshold, symbol \* is used to replace the spaces between terms for more convenient viewing.

Table 4 shows that in the first 25 groups of key terms, the joint occurrence of two terms prevails. The joint occurrence of three terms is not very informative: cnn\*lstm\*deep\_learning and artificial\_intelligence\*machine\_learning\*deep\_learning. The application domain for deep\_learning is most often found as feature\_extraction, which corresponds to the general theme of the bibliometric metadata set used.

Table 5 presents the 25 most common groups of terms for records from the IEEE Xplore platform for IEEE Terms and INSPEC Terms, respectively.

The findings of this paper suggest that the advantage of

groups of terms used in IEEE Xplore compared with the terms in Table 4 for Scopus is that they more capaciously describe the subject area due to a combination of terms describing methods and their object of application, for example:

- feature\_extraction\*machine\_learning;
- deep\_learning\*feature\_extraction;
- data\_models\*machine\_learning;
- task\_analysis\*feature\_extraction;
- training\*feature\_extraction;
- diseases\*machine\_learning;
- feature\_extraction\*learning-artificial\_intelligence;
- image\_classification\*learning-artificial\_intelligence;
- image\_classification\*feature\_extraction;
- pattern\_classification\*feature\_extraction.

Term “feature\_extraction,” which frequently appears on the list with different co-terms, indicates the significance of the data dimensionality reduction in pattern recognition and time-series problems, and others.

It is of interest to make an in-depth analysis of the context in which the term feature\_extraction appears in publications indexed by IEEE Xplore and how this context changes over time.

#### *D. Analysis of the context for the term “feature\_extraction” in bibliometric metadata of IEEE Xplore platform in 2011-2021.*

Sampling in the query (“Publication Topics:” “feature\_extraction”) OR (“IEEE Terms:” “feature\_extraction”), with the filters 2011–2020, gives 136 983 results, of which:

- 113 268 – Conferences;
- 21 944 – Journals;
- 1 058 – Early Access Articles;
- 637 – Magazines;
- 73 – Books;
- 2 – Standards;
- 1 – Courses.

Main Publication Topics are:

- feature\_extraction (12 951);
- learning (artificial intelligence) (7 859);
- image\_classification (4 466);
- convolutional\_neural\_net (2 835);
- object\_detection (2 457);
- neural\_net (2 366);
- image\_segmentation (2 314);
- image\_representation (2 037);
- support\_vector\_machine (1 859);
- pattern\_classification (1 746);
- medical\_image\_processing (1 713);
- geophysical\_image\_processing (1 666);
- medical\_signal\_processing (1 444);
- video\_signal\_processing (1 272);
- computer\_vision (1 243);
- signal\_classification (1 149);
- image\_matching (1 077);
- remote\_sensing (1 065);

- image\_color\_analysis (1 058);
- disease (1 030);
- regression\_analysis (987);
- face\_recognition (932);
- image\_texture (911);
- image\_fusion (880);
- image\_resolution (877).

These topics can be summarized as follows: feature extraction by convolutional neural nets, support vector machines and regression analysis for image classification, segmentation, representation, matching, color analysis, texture and resolution for solving the problems of medical image, geophysical image, medical signal processing, remote sensing, and face recognition.

For comparison, let us show the results of the query AUTHKEY (“feature\_extraction”) OR INDEXTERMS (“feature\_extraction”) AND PUBYEAR > 2010 to the Scopus database, which provides metadata to 90 283 documents, of which:

- 44 846 – Conference Paper;
- 43 566 – Article;
- 854 – Review;
- 766 – Book Chapter;
- 40 – Editorial;
- 37 – Book;
- 37 – Letter;
- 17 – Short Survey;
- 16 – Note.

Thus, there is significantly more conference material on this request in IEEE Xplore than in Scopus over the same period.

#### *E. VOSviewer for a brief analysis of research trends for the “feature\_extraction” topic.*

VOSviewer [27, 28], a software tool for constructing and visualizing bibliometric networks, is widely used in the bibliometric analysis. For example, in the Scopus database, to the query TITLE-ABS-KEY (VOSviewer), we obtain 1 437 results, and in the WoS database, to the query VOSviewer (Topic) – 1 086 results.

In the context of this paper, it is instrumental to feature the primary possibility of using this program to identify research trends in the data of Author Keywords, IEEE Terms, and INSPEC Terms of IEEE Xplore platform. The paper does not set the objectives to provide a detailed analysis of research trends for the topic “feature\_extraction.”

The easiest way to assess the possibility of using VOSviewer to analyze trends in scientific research according to IEEE Xplore data is to break the 10-year interval into two sub-intervals and compare the occurrence of Author Keywords, IEEE Terms and INSPEC Terms in them. For a more detailed analysis, it is sensible to track changes in the composition of the key terms in individual clusters formed by VOSviewer.

A sampling of bibliometric metadata for this section was made as follows. The query (“IEEE Terms:”



Table 6. Comparison of occurrence of Author Keywords for two time intervals. N is the occurrence of the term in the sample.

Keyword 2011-2017	N	Keyword 2018-2021	N
feature extraction	480	deep learning	1 045
classification	303	convolutional neural network	942
feature selection	200	feature extraction	435
deep learning	187	machine learning	350
machine learning	185	classification	242
face recognition	130	feature selection	209
pattern recognition	128	fault diagnosis	174
support vector machine	119	object detection	148
remote sensing	116	transfer learning	147
image classification	113	cnn	132
object detection	102	feature fusion	131
segmentation	99	attention mechanism	123
sparse representation	99	image classification	113
biometrics	98	remote sensing	109
synthetic aperture radar	90	person re-identification	94
computer vision	81	computer vision	89
image segmentation	71	action recognition	87
support vector machines	66	semantic segmentation	79
dimensionality reduction	65	generative adversarial network	78
object recognition	64	deep convolutional neural network	73
action recognition	63	support vector machine	70
change detection	63	deep neural network	69
image retrieval	63	pattern recognition	69
fault diagnosis	58	face recognition	66
image processing	58	image segmentation	66

Table 7. Comparison of the occurrence of IEEE Terms for two periods.

IEEE Terms 2011-2017	N	IEEE Terms 2018-2021	N
feature extraction	5 935	feature extraction	7 050
training	1 284	training	1 665
visualization	898	task analysis	1 411
support vector machines	701	visualization	832
vectors	660	convolution	757
image segmentation	626	deep learning	698
robustness	611	semantics	696
image color analysis	606	image segmentation	679
shape	511	three-dimensional displays	664
accuracy	496	support vector machines	637
computational modeling	453	data mining	560
cameras	451	machine learning	560
kernel	448	neural networks	548
histograms	424	cameras	492
data mining	419	computational modeling	473
remote sensing	388	data models	428
detectors	372	image color analysis	416
estimation	369	kernel	414
algorithm design and analysis	356	correlation	402
hidden markov models	355	object detection	398
semantics	355	remote sensing	379
three-dimensional displays	355	convolutional neural network	374
image edge detection	347	sensors	370
correlation	336	shape	365
databases	335	robustness	329

Table 8. Comparison of the occurrence of INSPEC Terms for two time intervals.

INSPEC Terms 2011-2017	N	INSPEC Terms 2018-2021	N
feature extraction	3 991	feature extraction	7 044
image classification	1 388	learning-artificial intelligence	3 450
learning-artificial intelligence	1 309	image classification	2 024
geophysical image processing	757	convolutional neural nets	1 567
support vector machines	687	object detection	1 178
image segmentation	620	neural nets	998
object detection	589	image segmentation	994
image representation	588	image representation	972
medical image processing	586	support vector machines	754
medical signal processing	498	pattern classification	680
neural nets	457	medical image processing	625
image matching	453	geophysical image processing	602
remote sensing	442	computer vision	572
video signal processing	409	medical signal processing	543
face recognition	384	video signal processing	519
image texture	352	signal classification	498
pattern classification	348	image fusion	451
signal classification	346	image color analysis	445
regression analysis	319	remote sensing	421
statistical analysis	309	image matching	418
hyperspectral imaging	295	diseases	407
image color analysis	290	image motion analysis	394
computer vision	289	fault diagnosis	387
pattern clustering	284	recurrent neural nets	375
synthetic aperture radar	278	image texture	370

Table 9. Top 10 terms for each of the 6 clusters shown in Figure 1.

Label (red)	cluster	N	Label (turquoise)	cluster	N	Label (blue)	cluster	N
neural nets	1	1 455	medical signal processing	2	1 041	feature extraction	3	11 035
support vector machines	1	1 441	electroencephalography	2	510	learning-artificial intelligence	3	4 759
fault diagnosis	1	544	neurophysiology	2	437	pattern classification	3	1 028
principal component analysis	1	460	cameras	2	384	video signal processing	3	928
wavelet transforms	1	381	traffic engineering computing	2	345	face recognition	3	732
entropy	1	251	medical disorders	2	315	regression analysis	3	667
condition monitoring	1	246	pose estimation	2	307	pattern clustering	3	542
time series	1	233	image sensors	2	257	statistical analysis	3	515
mechanical engineering computing	1	230	electrocardiography	2	255	optimization	3	491
power engineering computing	1	219	stereo image processing	2	244	graph theory	3	488
Label (yellow)	cluster	N	Label (violet)	cluster	N	Label (green)	cluster	N
signal classification	4	844	image classification	5	3 412	medical image processing	6	1 211
recurrent neural nets	4	407	object detection	5	1 767	diseases	6	660
radar imaging	4	406	image segmentation	5	1 614	cancer	6	330
probability	4	399	convolutional neural nets	5	1 567	biomedical mri	6	257
gaussian processes	4	338	image representation	5	1 560	biomedical optical imaging	6	256
matrix algebra	4	323	geophysical image processing	5	1 359	brain	6	248
bayes methods	4	264	image matching	5	871	computerized tomography	6	197
gesture recognition	4	197	remote sensing	5	863	eye	6	172
speech recognition	4	191	computer vision	5	861	tumors	6	159
hidden markov models	4	184	image color analysis	5	735	patient diagnosis	6	150



“feature extraction”) OR (“Publication Topics:” “feature extraction”) is made for each year of the interval 2011-2021. If the number of publications meeting the request per year did not exceed 2 000, all metadata was downloaded, and if the number of publications exceeded 2 000, only the metadata of the first 2 000 most cited journal articles was exported (last year was not complete, data as of 15-07-2021). Metadata was summed for two intervals, 2011-2017 and 2018-2021, yielding a close number of records in each sub-sample, 7 522 and 8 000 entries, respectively.

The subject for both periods is similar – feature extraction for image analysis.

Author Keywords in 2018-2021 are more related to deep learning and neural networks, whereas, in 2011-2017, the focus is on feature selection and classification, i.e., closer to the main query (feature extraction). It can be assumed that over time, the authors' interests have shifted from feature extraction applications (face recognition, remote sensing, synthetic aperture radar, biometrics, fault diagnosis) to big data algorithms: deep learning, convolutional neural networks.

Tables 7 and 8 were built similarly to Table 6 but only for IEEE Terms and INSPEC Controlled Terms.

In IEEE Terms, the “feature extraction” themes are expressed in all periods, which is due to the request itself. However, whereas previously, the publications had emphasized classic problems, for example, visualization, support vector machines, vectors, image segmentation, image color analysis, remote sensing, hidden Markov models, and image edge detection; the subsequent periods, as in the case of Author Keywords, saw more modern, big data-related topics, including convolution, deep learning, semantics, data mining, machine learning, neural networks, and three-dimensional displays. There is a significant increase in the interest in the field of application algorithms: “three-dimensional displays.”

Overall, there is a good consistency in results for Author Keywords and IEEE Terms. Therefore, it is advisable to combine them in bibliometric analysis.

INSPEC Controlled Terms are chosen from an export-controlled dictionary. Therefore, the overall set of terms for different time intervals is more stable. This factor may give an advantage in using the INSPEC Controlled Terms when considering in detail the change in dominant terms in individual years compared to Author Keywords.

The second feature of INSPEC Controlled Terms is the more frequent appearance of terms describing the applied fields of research, e.g., geophysical image processing, medical image processing, medical signal processing, video signal processing, fault diagnosis, diseases. This fact is essential, for example, when collecting materials on the specific methods of data analysis applied in a given area of research. The IEEE Xplore platform provides such a possibility.

The INSPEC Controlled Terms dictionary is periodically updated by experts and can be used to analyze

emerging trends in research. This is a separate task for bibliometric analysis. However, even the simple fact that the term “recurrent neural nets” in the above data occurs only among INSPEC Controlled Terms indicates their importance for research trend analysis.

VOSviewer allows creating a general picture (landscape) of research and thematic clustering based on the co-occurrence of key terms.

In this paper, the VOSviewer is used only as applied to the INSPEC Controlled Terms for the entire 2011-2021 timeframe. The choice of INSPEC Controlled Terms is due to their control by INSPEC experts. Expert assessments are the most expensive and difficult to rank data. Thus, the export-controlled dictionaries, the level of peer review of scientific articles, the rating of journals and organizations, and the citation rate of papers are crucial in analyzing research trends because they indirectly reflect expert opinion.

Fig. 1 presents the results of term network and co-occurrence-based clustering for INSPEC Controlled Terms for all metadata by the query (“IEEE Terms:” “feature extraction”) OR (“Publication Topics:” “feature extraction”) for 2011-2021. By removing the records without INSPEC Controlled Terms, we get 14 840 lines to analyze.

The total number of INSPEC Controlled Terms for this sample was 3 086, of which 1 216 occurred more than five times. Out of these terms, 1 000 with the highest overall level of links were used to construct a network of terms.

If there is no limit on the number of terms in the cluster, we obtain 8 of them, which is a lot for the primary analysis. With the minimum number of terms in the cluster of 40 to 90, there are 6 clusters, with the most common ones shown in Table 9. The wide range of values (40-90) indicates the stability of the resulting clusters. This parameter is useful to adjust the number of clusters to be formed depending on the study objectives.

In VOSviewer, clusters are ordered by the number of unique terms but not by the total number of terms. Therefore, the central term of the “feature extraction” sample is included in the third cluster.

Express analysis of research trends employed VOSviewer's ability to display the change over time (Overlay in terms of VOSviewer) in the occurrence of terms used in the network. The graph of overlay over time is presented in Fig 1.

“Object detection” and “learning-artificial intelligence” are the most frequently used terms in recent times, but they are rather general in nature.

For a more detailed analysis of particular emerging research trends, it is more interesting to choose several specific terms, such as “fault diagnosis” and “condition monitoring” from the red cluster.

**Note:** In this paper, the terms are used as they appear on the IEEE Xplore platform. For example, “Conferences” means conference proceedings, “Publication Topics”

corresponds to the dictionary of INSPEC Controlled Terms.

Next, the data meeting the query (“Publication Topics:” “fault diagnosis”) AND (“Publication Topics:” “feature extraction”) was used. In 2011–2021, IEEE Xplore indexed 2 042 documents that match this request, including 1 477 in Conferences and 563 in Journals. In 2011, only 65 papers were posted, including 64 in Conferences and 1 in Journals, whereas 2020 saw already 498 papers, with 296 in Conferences and 202 in Journals.

It follows from this data that in the context of the general topic of “feature extraction,” in 2011, the “fault diagnosis” issue was mainly raised at conferences, and only one journal article was indexed, whereas, in 2020, there were already 202 articles and their number became commensurate with the number of conference proceedings. This situation confirms the well-known fact that it is easier to detect the emerging trends in conference proceedings than in scientific publications.

Similar dynamics are observed for the term “condition monitoring.” The query (“Publication Topics:” “state monitoring”) AND (“Publication Topics:” “feature extraction”) for 2021–2021 found 892 documents, of which 639 in Conferences and 252 in Journals.

- 2011 → 40 in Conferences and 3 in Journals;
- 2020 → 136 in Conferences and 79 in Journals.

The terms “fault diagnosis” and “condition monitoring” are included in the same cluster, as in Fig. 1. This fact is consistent with the results of the above two queries. The distribution of publications by “Publication Topic” for them is shown in Table 10.

To show that context matters, the data from queries that include “fault diagnosis” and “condition monitoring,” but without the context of “feature extraction,” was used.

In 2011–2021, 23 795 documents related to the query (“Publication Topics:” “fault diagnosis”) were indexed, including 19 171 in Conferences and 4 516 in Journals.

- 2011 → Conferences (1 470) and Journals (144), all of → 1 624
- 2020 → Conferences (2 311) and Journals (990), all of → 3 316

In 2011–2021, 10 942 documents related to the query (“Publication Topics:” “condition monitoring”) were indexed, including 8 865 in Conferences and 1 991 in Journals.

- 2011 → Conferences (792) and Journals (63), all of → 862
- 2020 → Conferences (1 099) and Journals (450), all of → 1 559

It follows from the above data that in the broader context, the decade-long increase in the interest in the terms “fault diagnosis” and “condition monitoring” in all publications is about two times, which is significantly less than in the context of “feature extraction.”

The decrease in growth is due to a slight increase in the number of conference proceedings. For scientific

publications, the gain is more significant.

Thus, the conclusion can be made that for the largely common problems of “fault diagnosis” and “condition monitoring,” the growth of interest in them is due to the application of more advanced analytical methods for solving them, which require the procedure of “feature extraction.”

### III. CONCLUSION

Bibliometric analysis has shown that the IEEE Explore platform is an undervalued resource, despite its some advantages over well-known Scopus and WoS abstract databases, the main of which are:

- open access to the platform;
- a wide variety of key terms allowing a more detailed study of research trends;
- citation rate of publications is assessed within a specialized database, i.e., the opinion of experts in a particular subject area dominates.

The WoS system contains some inconsistencies between the Author Keywords in the database and the Author Keywords in the full texts of publications, making it difficult to use them when analyzing the topics of publications by keywords.

The comparability of the topics identified by the key terms of publications indexed in IEEE Xplore and Scopus is shown. At the same time, the controlled vocabulary, when used to identify research topics and trends from metadata of samples that satisfy queries, has the following advantages:

- the stability of the controlled vocabulary gives a better ability to compare key terms in the samples at different time intervals;
- co-occurrence of such terms better describes the topics of publications because it provides more balance between the terms defining methods of analysis and research objects.

A significant feature of IEEE Xplore is the large host of indexed conference proceedings, which helps identify emerging trends in research in an earlier stage.

The reasonableness of using the Apriori algorithm to identify multiple co-occurrences of terms to describe topics of indexed publications is demonstrated.

The possibility of using VOSviewer to build a landscape of scientific research and identify trends in topics is shown. Officially, VOSviewer does not support exporting data from IEEE Xplore, but it is easy to pre-process data to use this great program.

This study was not intended to explore in detail all the features of the IEEE Xplore platform for bibliometric analysis and identification of research trends, as the objective behind it was to attract the attention of specialists from the energy sector to the capabilities of this platform and encourage its wider use in their work.

## ACKNOWLEDGMENT

This paper was written within the framework of the state assignment (topic “Fundamental basis of innovative technologies of the oil and gas industry (fundamental, search and applied research),” No AAAA-A19-119013190038-2).

## REFERENCES

- [1] J. J. M. Ferreira, C. I. Fernandes, and S. Kraus, “Entrepreneurship research: mapping intellectual structures and research trends,” *Rev Manag Sci*, vol. 13, no. 1, pp. 181–205, Feb. 2019, DOI: 10.1007/s11846-017-0242-3.
- [2] R. A. Estévez, V. Espinoza, R. D. Ponce Oliva, F. Vásquez-Lavín, and S. Gelcich, “Multi-criteria decision analysis for renewable energies: research trends, gaps and the challenge of improving participation,” *Sustainability*, vol. 13, no. 6, p. 3515, Mar. 2021, DOI: 10.3390/su13063515.
- [3] M. W. Bickel, “Reflecting trends in the academic landscape of sustainable energy using probabilistic topic modeling,” *Energ Sustain Soc*, vol. 9, no. 1, p. 49, Dec. 2019, DOI: 10.1186/s13705-019-0226-z.
- [4] J. L. Ruiz-Real, J. Uribe-Toril, J. A. Torres, and J. De Pablo, “Artificial intelligence in business and economics research: trends and future,” *Journal of Business Economics and Management*, vol. 22, no. 1, pp. 98–117, Oct. 2020, DOI: 10.3846/jbem.2020.13641.
- [5] W. Lu, S. Huang, J. Yang, Y. Bu, Q. Cheng, and Y. Huang, “Detecting research topic trends by author-defined keyword frequency,” *Information Processing & Management*, vol. 58, no. 4, p. 102594, Jul. 2021, DOI: 10.1016/j.ipm.2021.102594.
- [6] W. Ajaz and D. Bernell, “California’s adoption of microgrids: A tale of symbiotic regimes and energy transitions,” *Renewable and Sustainable Energy Reviews*, vol. 138, p. 110568, Mar. 2021, DOI: 10.1016/j.rser.2020.110568.
- [7] L. Silva and S. Sareen, “Solar photovoltaic energy infrastructures, land use and sociocultural context in Portugal,” *Local Environment*, vol. 26, no. 3, pp. 347–363, Mar. 2021, DOI: 10.1080/13549839.2020.1837091.
- [8] L. Schreiner and R. Madlener, “A pathway to green growth? Macroeconomic impacts of power grid infrastructure investments in Germany,” *Energy Policy*, vol. 156, p. 112289, Sep. 2021, DOI: 10.1016/j.enpol.2021.112289.
- [9] A. Joseph and P. Balachandra, “Smart grid to energy internet: a systematic review of transitioning electricity systems,” *IEEE Access*, vol. 8, pp. 215787–215805, 2020, DOI: 10.1109/ACCESS.2020.3041031.
- [10] A. Martín-Martín, M. Thelwall, E. Orduna-Malea, and E. Delgado López-Cózar, “Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations’ COCI: a multidisciplinary comparison of coverage via citations,” *Scientometrics*, vol. 126, no. 1, pp. 871–906, Jan. 2021, doi: 10.1007/s11192-020-03690-4.
- [11] R. Prancutė, “Web of Science (WoS) and Scopus: the titans of bibliographic information in today’s academic world,” *Publications*, vol. 9, no. 1, p. 12, Mar. 2021, DOI: 10.3390/publications9010012.
- [12] J. C. F. de Winter, A. A. Zadpoor, and D. Dodou, “The expansion of Google Scholar versus Web of Science: a longitudinal study,” *Scientometrics*, vol. 98, no. 2, pp. 1547–1565, Feb. 2014, DOI: 10.1007/s11192-013-1089-2.
- [13] V. K. Singh, P. Singh, M. Karmakar, J. Leta, and P. Mayr, “The journal coverage of Web of Science, Scopus, and Dimensions: A comparative analysis,” *Scientometrics*, vol. 126, no. 6, pp. 5113–5142, Jun. 2021, DOI: 10.1007/s11192-021-03948-5.
- [14] L. S. Adriaanse and C. Rensleigh, “Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison,” *The Electronic Library*, vol. 31, no. 6, pp. 727–744, Nov. 2013, DOI: 10.1108/EL-12-2011-0174.
- [15] J.-C. Valderrama-Zurián, R. Aguilar-Moya, D. Melero-Fuentes, and R. Aleixandre-Benavent, “A systematic analysis of duplicate records in Scopus,” *Journal of Informetrics*, vol. 9, no. 3, pp. 570–576, Jul. 2015, DOI: 10.1016/j.joi.2015.05.002.
- [16] F. Franceschini, D. Maisano, and L. Mastrogiacomio, “Empirical analysis and classification of database errors in Scopus and Web of Science,” *Journal of Informetrics*, vol. 10, no. 4, pp. 933–953, Nov. 2016, DOI: 10.1016/j.joi.2016.07.003.
- [17] M. A. García-Pérez, “Strange attractors in the Web of Science database,” *Journal of Informetrics*, vol. 5, no. 1, pp. 214–218, Jan. 2011, DOI: 10.1016/j.joi.2010.07.006.
- [18] F. Franceschini, D. Maisano, and L. Mastrogiacomio, “Do Scopus and WoS correct ‘old’ omitted citations?,” *Scientometrics*, vol. 107, no. 2, pp. 321–335, May 2016, DOI: 10.1007/s11192-016-1867-8.
- [19] M. Visser, N. J. van Eck, and L. Waltman, “Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic,” *Quantitative Science Studies*, vol. 2, no. 1, pp. 20–41, Apr. 2021, DOI: 10.1162/qss\_a\_00112.
- [20] L. Waltman and V. A. Traag, “Use of the journal impact factor for assessing individual articles: Statistically flawed or not?,” *F1000Research*, vol. 9, p. 366, Mar. 2021, DOI: 10.12688/f1000research.23418.2.
- [21] Balili, U. Lee, A. Segev, J. Kim, and M. Ko, “TermBall: tracking and predicting evolution types of research topics by using knowledge structures in scholarly big data,” *IEEE Access*, vol. 8, pp. 108514–108529, 2020, DOI: 10.1109/ACCESS.2020.3000948.
- [22] J. C. Valderrama-Zurián, C. García-Zorita, S. Marugán-Lázaro, and E. Sanz-Casado, “Comparison of MeSH terms and KeyWords Plus terms for more accurate classification in medical research fields.

A case study in cannabis research,” *Information Processing & Management*, vol. 58, no. 5, p. 102658, Sep. 2021, DOI: 10.1016/j.ipm.2021.102658.

- [23] S. Lozano, L. Calzada-Infante, B. Adenso-Díaz, and S. García, “Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature,” *Scientometrics*, vol. 120, no. 2, pp. 609–629, Aug. 2019, DOI: 10.1007/s11192-019-03132-w.
- [24] F. Amalina et al., “Blending big data analytics: review on challenges and a recent study,” *IEEE Access*, vol. 8, pp. 3629–3645, 2020, DOI: 10.1109/ACCESS.2019.2923270.
- [25] Y. Dong, P. Liu, Z. Zhu, Q. Wang, and Q. Zhang, “A fusion model-based label embedding and Self-Interaction Attention for text classification,” *IEEE Access*, vol. 8, pp. 30548–30559, 2020, DOI: 10.1109/ACCESS.2019.2954985.
- [26] F. Feng, R. H. M. Chan, X. Shi, Y. Zhang, and Q. She, “Challenges in task incremental learning for assistive robotics,” *IEEE Access*, vol. 8, pp. 3434–3441, 2020, DOI: 10.1109/ACCESS.2019.2955480.
- [27] N. J. van Eck and L. Waltman, “Software survey: VOSviewer, a computer program for bibliometric mapping,” *Scientometrics*, vol. 84, no. 2, pp. 523–538, Aug. 2010, DOI: 10.1007/s11192-009-0146-3.
- [28] F. Rizzi, N. J. van Eck, and M. Frey, “The production of scientific knowledge on renewable energies: Worldwide trends, dynamics and challenges and implications for management,” *Renewable Energy*, vol. 62, pp. 657–671, Feb. 2014, DOI: 10.1016/j.renene.2013.08.030.



**Boris N. Chigarev** is a leading engineer for scientific and technical information at the Oil and Gas Research Institute of the Russian Academy of Sciences. He received the degree of Ph.D. in chemical physics, including combustion and explosion physics, from V.I. Kurchatov Institute of Atomic Energy in 1989. His research interests are bibliometric analysis and trends in energy research and development.