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

B. N. Chigarev[*]

* Corresponding author.
E-mail: bchigarev@ipng.ru

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

Abstract — **The article is aimed at a brief comparison and analysis of the results of queries to IEEE Xplore and the leading abstract databases Scopus and Web of Science to identify research trends. Some errors in the Author Keywords in Web of Science have been revealed. Therefore, a more detailed analysis was conducted by comparing different types of key terms for IEEE Xplore and Scopus platforms only. I used IEEE Access journal metadata as indexed on both platforms. The sample match for IEEE Xplore and Scopus was achieved by comparing DOI. The IEEE Xplore metadata contains more types of key terms, which provides an advantage in analyzing research trends. Using NSPEC Controlled Terms from expert-compiled vocabulary provides a more stable data, which gives an advantage when considering the change of terms over time. Apriori, an algorithm for finding association rules, has been used to compare co-occurrence of 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. Using the IEEE conference proceedings bibliometric data, I illustrated the importance of context in estimating the rate of growth 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 [1-5].

Systematic bibliometric analysis of metadata of scientific publications and conference proceedings reveals major trends in R&D. Traditionally, the abstract databases Scopus and Web of Science (WoS) are used for this purpose, but 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.

The energy transition to low-carbon power sources requires significant development of the power grid infrastructure and optimization of its operation [6-9]. These tasks are classical for the experts of the Institute of Electrical and Electronics Engineers (IEEE). Research trends in this area are reasonable to identify using the metadata of the IEEE Xplore platform, which is currently insufficiently used in bibliometric analysis.

The aim of this article was to highlight some of the IEEE Xplore features, which, along with its openness, may provide additional benefits compared to the closed 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 occur frequently in them. This approach makes it possible to assess research trends by the occurrence of key terms describing published documents. The terms can be author's keywords, documents text mining terms, or experts-controlled terms from a subject vocabulary [10-12].

The article is devoted only to this aspect of bibliometric analysis.

To clarify the underestimation of IEEE Xplore as a source of publication metadata, several comparisons were made for queries containing the basic terms bibliometrics OR scientometrics and the names of leading abstract databases. I got the following results for queries in Scopus:

- TITLE-ABS-KEY ((bibliometric* OR scientometric*) AND "ieee xplore») →11 results
- TITLE-ABS-KEY ((bibliometric* OR scientometric*) AND "scopus») →3797 results

1

- TITLE-ABS-KEY ((bibliometric* OR scientometric*) AND ("WoS" OR "web of science")) →5919 document results

And for Web of Science Core Collection:

- (bibliometric* OR scientometric*) AND "ieee xplore" (Topic) →10 results
- (bibliometric* OR scientometric*) AND "scopus" (Topic) →3026 results
- (bibliometric* OR scientometric*) AND ("WoS" OR "web of science") (Topic) →4840 results

I found no results for ("All Metadata": ((bibliometric* OR "All Metadata": scientometric*) AND "All Metadata":"ieee xplore»)) on IEEE Xplore platform.

IEEE Xplore platform provides a comprehensive list of metadata for publications, which enables a comprehensive bibliometric analysis[1].

IEEE Xplore platform metadata list, you can use to analyze the topics of published materials:

- Abstract
- Author Keywords
- 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[2].

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 (3564); SMPTE (3022); SAE (2942); River Publishers (2351); BIAI (1517). No giants such as Elsevier and Springer Nature are in the list, but dominate the publications of the IEEE itself, the platform is focused 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 large number of conference materials metadata compared to journal articles and standards-related documents.

For 2 582 653 papers in 2011-2020 we got: Conferences (1 992 101); Journals (482 568); Magazines (66 809); Books (25 937); Early Access Articles (9611); Standards (5297); Courses (330).

Conference proceedings reflect industry interests more than peer-reviewed publications; for example, the major Publication Topics for IEEE Xplore in 2011-2020 are: learning (artificial intelligence) (103 944); feature extraction (67 483); optimization (64 359); neural nets (46 323); Internet (43 459); cloud computing (40 371); mobile robots (36 370); image classification (36 029); control system synthesis (35 968); medical image processing (35 231); wireless sensor networks (33 215); power grids (32 965) - distinct engineering challenges.

When conducting bibliometric research, it is advisable to choose the relevance of the topic according to the materials of conferences or patent studies, and by the analysis of peer-reviewed articles to assess the scientific validity of the chosen topic. It is important that the choice of the goal and methods of its achievement should not be based on a closed set of data, it is better that these sets have overlap, but do not coincide.

II. ANALYSIS AND RESULTS

A.   Comparing keywords of IEEE Xplore and Scopus platforms

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

While comparing the expressiveness of keywords in different platforms, I came across the fact that in the Web of Science system there are many errors in the Author Keywords field, and the Keywords Plus field contains few terms. This issue requires additional and more comprehensive study and is beyond the scope of this article, so Table 1 provides only a few examples to illustrate this problem.

Table 1. Examples of mismatches between Author Keywords on the Web of Science platform
and the keywords in the publications themselves [13-15]

Author Keywords by WoS	Correct Keywords	DOI of article
Big Data; Data analysis; Tools; Social networking (online); Computer languages; Companies; Big data analytics; data analytics; deep learning; machine learning	Big data analytics; data analytics; deep learning; machine learning	10.1109/ACCESS. 2019.2923270
Text categorization; Semantics; Feature extraction; Natural language processing; Bit error rate; Task analysis; Neural networks; Text classification; text representations; label embedding	label embedding; Text classification; text representations	10.1109/ACCESS. 2019.2954985

---

[1] https://ieeexplore.ieee.org/Xplorehelp/searching-ieee-xplore/advanced-search

[2] https://ieeexplore.ieee.org/Xplorehelp/searching-ieee-xplore/command-search#summary-of-data-fields

Task analysis; Rehabilitation robotics; Lighting; Clutter; Computer vision; Training; Machine intelligence; robotic vision systems	Machine intelligence; robotic vision systems	10.1109/ACCESS. 2019.2955480

The examples are taken from 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.

I have not encountered such a problem on the IEEE Xplore and Scopus platforms, so in the further I use these systems.

In order to compare key terms in different systems, a set of publications indexed in both systems must be established. The IEEE Access journal, which is indexed in all the above systems, has a large number of publications and fits the bill, for example, according to Scopus 18 073 publications in 2020.

IEEE Xplore and Scopus allow the export of 2000 metadata for a single query, which is enough for a qualitative comparison. To select 2000 articles out of 18000, they were sorted by citation on each platform, and then the first 2000 bibliometric metadata were exported. The citation rate of articles is determined based on platform's own data, so there is a difference in the lists of articles in the 2000 most cited for each platform. Articles with the same DOI were sampled to resolve this issue. There were 1250 such articles. For comparison, the intersection of the 2000 most cited articles in the journal in 2020 between the systems Web of Science and IEEE Xplore was 1207, which compares with 1250 and indicates the consistency of the results.

It should be noted that for a sample of 1250 records, there is no discrepancy between the Author Keywords in both systems, so the following two tables list them once. Tables 2 and 3 each show the 25 most common key terms: Author Keywords and Index Keywords[3] for Scopus; IEEE Terms and INSPEC Controlled Terms for IEEE Xplore. N in the tables denotes the occurrence of the term.

Table 2: Top 25 key terms according to Scopus for 1250 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
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

The general topics of the terms presented in the table can be described as: deep learning; machine learning; blockchain; convolutional neural network; internet of things. You can use table data to generate new queries for further literature gathering.

Table 3: Top 25 key terms according to IEEE Xplore for 1250 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	optimisation	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

---

[3] https://service.elsevier.com/app/answers/detail/a_id/21730/supporthub/scopus/

3

cloud computing	57	particle swarm optimisation	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

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 are no less expressive than Index Keywords reflect the subject of publications, so that 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.

When studying trends in topics of scientific publications assessed by frequency of occurrence (or co-occurrence) of terms, the controlled dictionary yields more stable results, since index terms differ wider in bibliometrics metadata at different time periods. In turn, Author Keywords, being the most subjective, better reflect the current state of the topics and it is expedient to use them to identify emerging trends in the topics of publications. The IEEE Xplore platform provides both capabilities. A detailed analysis of these statements is beyond the scope of this article 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 Apriori, an algorithm for finding associative rules.

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

To reduce the set of terms occurring together, I imposed additional constraints by sampling rows from Scopus and IEEE Xplore metadata containing the word "learning". Referring to tables 2 and 3, you will find the following terms containing the word "learning": deep learning, machine learning, learning systems, learning (artificial intelligence), learning algorithms, which indicates the relevance of such a restriction on sampling.

Applying this constraint to 1250 Scopus data records will yield:
- Author Keywords — 246 rows containing the learning string
- Index Keywords — 324 rows containing the learning string

And applying this constraint to 1250 IEEE Xplore data records:
- IEEE Terms — 200
- INSPEC Terms — 327

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

Preparing the data for the apriori algorithm involved both 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: % is the percentage of this key term group in the total list of term groups that passed the 1% threshold. The spaces between terms have been replaced with * for clearer viewing.

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

4

	%		%
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

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, respectively for IEEE Terms and INSPEC Terms.

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

According to the author of this article, 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

The term feature_extraction, frequently appeared in this list with different co-terms, indicate the importance of data dimensionality reduction in pattern recognition and time series tasks, etc.

It is of interest to analyze in more detail 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 term feature_extraction in bibliometric metadata of IEEE Xplore platform in 2011-2021

Sampling on query: ("Publication Topics":"feature extraction") OR ("IEEE Terms":"feature extraction"); Filters Applied: 2011 – 2020 gives 136 983 results, of which:

- Conferences (113 268)
- Journals (21 944)
- Early Access Articles (1058)
- Magazines (637)
- Books (73)
- Standards (2)
- Courses (1)

Main Publication Topics:

- feature extraction (12 951)
- learning (artificial intelligence) (7859)
- image classification (4466)
- convolutional neural net (2835)
- object detection (2457)
- neural net (2366)
- image segmentation (2314)
- image representation (2037)
- support vector machine (1859)
- pattern classification (1746)
- medical image processing (1713)
- geophysical image processing (1666)
- medical signal processing (1444)
- video signal processing (1272)
- computer vision (1243)
- signal classification (1149)
- image matching (1077)
- remote sensing (1065)
- image colour analysis (1058)
- disease (1030)
- 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, colour analysis, texture and resolution for solving tasks: medical image, geophysical image, medical signal processing, remote sensing and face recognition.

For comparison, let's 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:

- Conference Paper (44 846)
- Article (43 566)
- Review (854)
- Book Chapter (766)
- Editorial (40)

- Book (37)
- Letter (37)
- Short Survey (17)
- Note (16)

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

E.   Using VOSviewer to briefly analyze research trends on the topic "feature extraction"

VOSviewer [16 17], a software tool for constructing and visualizing bibliometric networks, is widely used in bibliometric analysis. For example, in the abstract database Scopus on the query: TITLE-ABS-KEY (VOSviewer) we get 1.437 results, and in the database WoS, the query: VOSviewer (Topic) - 1.086 results.

In the context of this article, it is useful to show the fundamental possibility of using this program to identify research trends in the data Author Keywords, IEEE Terms and INSPEC Terms of IEEE Xplore platform. The tasks of detailed analysis of research trends on the topic "feature detection" in this article are not posed.

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 more detailed analysis, it is useful to track changes in the composition of key terms in individual clusters formed by VOSviewer.

The sampling of bibliometric metadata for this section was done as follows: a query ("IEEE Terms": "feature extraction") OR ("Publication Topics": "feature extraction") for each year of the interval: 2011-2021, if the number of publications meeting the request per year did not exceed 2000, all metadata was downloaded, if the number of publications exceeded 2000, only the metadata of the first 2000 most cited journal articles was exported (last year not complete, data as of 15-07-2021). Metadata was summed for two-time intervals, 2011-2017 and 2018-2021, yielding a close number of records in each sub-sample, 7522 and 8000 entries, respectively.

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

Keyword 2011-2017	N	Keyword 2018-2021	N
feature extraction	480	deep learning	1045
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

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, while 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 shift from feature extraction applications (face recognition, remote sensing, synthetic aperture radar, biometrics, fault diagnosis) to big data algorithms: deep learning, convolutional neural network.

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

ТАБЛИЦА 7. Comparison of the occurrence of IEEE Terms for two time periods.

IEEE Terms 2011-2017	N	IEEE Terms 2018-2021	N
feature extraction	5935	feature extraction	7050
training	1284	training	1665

	N		N
visualization	898	task analysis	1411
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

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

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

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	3991	feature extraction	7044
image classification	1388	learning-artificial intelligence	3450
learning-artificial intelligence	1309	image classification	2024
geophysical image processing	757	convolutional neural nets	1567
support vector machines	687	object detection	1178
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 colour analysis	445
regression analysis	319	remote sensing	421
statistical analysis	309	image matching	418
hyperspectral imaging	295	diseases	407
image colour 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

INSPEC Controlled Terms are chosen from an expert-controlled dictionary, so the overall set of terms for the different time intervals is more stable, this may give an advantage in using the INSPEC Controlled Terms when considering in detail the change of 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 application of specific methods of data analysis in a given area of research. IEEE Xplore platform provides such a possibility.

The INSPEC Controlled Terms dictionary is periodically updated by experts, which can be used to analyze emerging trends in research. This is a separate task for bibliometric analysis, but 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 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 controlled by INSPEC experts. Expert assessments are the most expensive and difficult to rank data, so expert-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 query ("IEEE Terms": "feature extraction") OR ("Publication Topics": "feature extraction") for 2011-2021. By removing the records that do not have INSPEC Controlled Terms, we get 14840 lines to analyze.

The total number of INSPEC Controlled Terms for this sample was 3086, of which 1216 occur more than 5 times. From these terms, the 1000 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 get 8 clusters, which is a lot for primary analysis. With a minimum number of terms in a cluster of 40 to 90 we get 6 clusters. The most common terms of which are shown in Table 9. The wide range of values: 40-90 indicates the stability of the resulting clusters. This parameter is useful for adjusting the number of clusters to be formed depending on the objectives of the study.
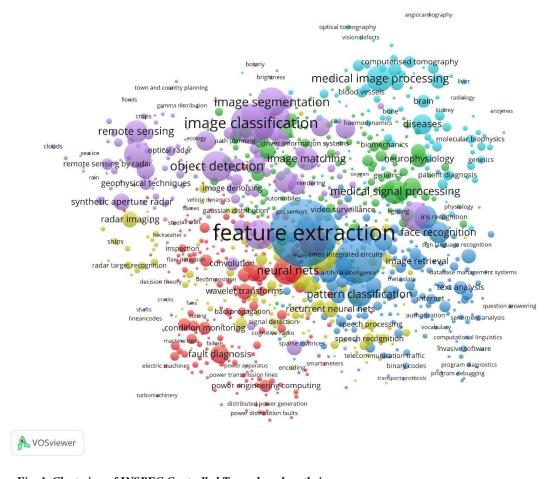

Fig. 1. Clustering of INSPEC Controlled Terms based on their co-occurrence.

TABLE 9. Tor 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	1455	medical signal processing	2	1041	feature extraction	3	11035
support vector machines	1	1441	electroencephalography	2	510	learning-artificial intelligence	3	4759
fault diagnosis	1	544	neurophysiology	2	437	pattern classification	3	1028

9

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	optimisation	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	3412	medical image processing	6	1211
recurrent neural nets	4	407	object detection	5	1767	diseases	6	660
radar imaging	4	406	image segmentation	5	1614	cancer	6	330
probability	4	399	convolutional neural nets	5	1567	biomedical mri	6	257
gaussian processes	4	338	image representation	5	1560	biomedical optical imaging	6	256
matrix algebra	4	323	geophysical image processing	5	1359	brain	6	248
bayes methods	4	264	image matching	5	871	computerised tomography	6	197
gesture recognition	4	197	remote sensing	5	863	eye	6	172
speech recognition	4	191	computer vision	5	861	tumours	6	159
hidden markov models	4	184	image colour analysis	5	735	patient diagnosis	6	150

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

For express analysis of research trends, I used 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 represented in Fig 1.
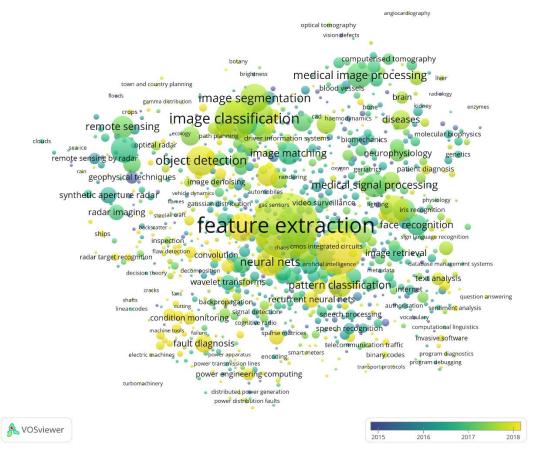

Fig.2. Trends in term occurrence over time for 2011-2021.

10

"Object detection" and "learning-artificial intelligence" – are some of 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 some 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, in the system "Conferences" means conference proceedings, "Publication Topics" corresponds to the dictionary of INSPEC Controlled Terms.

Next, data meeting the query were used: ("Publication Topics":"fault diagnosis") AND ("Publication Topics":"feature extraction"). In 2011-2021, IEEE Xplore indexed 2042 documents that match this request, of which Conferences (1477) and Journals (563). In 2011, only 65 papers were posted, Conferences (64) and Journals (1), while in 2020 there are already 498 papers → Conferences (296) and Journals (202).

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

A similar dynamic is 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: Conferences (639) and Journals (252).

- 2011→Conferences (40) and Journals (3)
- 2020→Conferences (136) and Journals (79)

The terms "fault diagnosis" and "condition monitoring" are included in the same cluster, as in Fig. X., this fact is consistent with the results of the above two queries, the data on the distribution of publications by "Publication Topics" for which are shown in Table 10.

TABLE 10. Top 5 "Publication Topics" for queries containing the terms "fault diagnosis" and "condition monitoring", respectively

Publication Topics	N	Publication Topics	N
fault diagnosis	2042	condition monitoring	892
feature extraction	2042	feature extraction	892
mechanical engineering computing	637	fault diagnosis	514
learning-artificial intelligence	554	mechanical engineering computing	327
condition monitoring	514	learning-artificial intelligence	234

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

23 795 documents were indexed in 20211-2021 related to the query: ("Publication Topics": "fault diagnosis"), of which: Conferences (19 171) and Journals (4516).

- 2011→Conferences (1470) and Journals (144), all of →1624
- 2020→Conferences (2311) and Journals (990), all of →3316

10 942 documents were indexed in 20211-2021 related to the query: ("Publication Topics":" condition monitoring"), of which: Conferences (8865) and Journals (1991).

- 2011→ Conferences (792) and Journals (63), all of →862
- 2020→ Conferences (1099) and Journals (450), all of → 1559

It follows from the above data that in the broader context, the decade-long increase in 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, we can conclude that for the fairly common tasks of "fault diagnosis" and "condition monitoring", the growth of interest in them is due to the application of more modern analytical methods for their solution, which require the procedure of "feature extraction".

III. CONCLUSION

Bibliometric analysis showed that the IEEE Explore platform is an undervalued resource, despite a number of its advantages over well-known Scopus and WoS abstract databases, the main ones:

- open access to the platform
- a wide variety of key terms, allowing a more detailed study of research trends

- the assessment of the citation rate of publications is carried out within a specialized database, i.e., the opinion of experts in a particular subject area dominates.

The WoS system contains a number of 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, using controlled vocabulary 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 samples at different time intervals
- the joint occurrence of such terms better describes the topics of publications because more balanced includes terms that describe methods of analysis and research objects.

A significant feature of IEEE Xplore is a large number of indexed conference proceedings, which allow the identification of emerging trends in research at an earlier stage.

I show the reasonableness of using a priori algorithm to identify multiple co-occurrences of terms to describe topics of indexed publications.

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

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

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*, p. 112289, May 2021, doi: 10.1016/j.enpol.2021.112289.

[9]. 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]. 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.

[11]. 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.

[12]. 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.

[13]. 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.

[14]. 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.

[15]. 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.

[16]. 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.

[17]. 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 Oil and Gas Research Institute of 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 research trends in energy sector R&D. https://orcid.org/0000-0001-9903-2800