Preprints.org

Article

Not peer-reviewed version

IEEE Terms Analysis of 2019–2024 IEEE Xplore Data on the Topic of Energy Systems

Boris Chigarev *

Posted Date: 20 May 2024

doi: 10.20944/preprints202405.1282.v1

Keywords: IEEE Terms; IEEE Xplore; energy systems; VOSviewer; Scimago Graphica; fpgrowth

Article

IEEE Terms Analysis of 2019–2024 IEEE Xplore Data on the Topic of Energy Systems

Boris Chigarev

Oil and Gas Research Institute of the Russian Academy of Sciences (OGRI RAS), Moscow, Russia.

Abstract: Relevance of the subject. Keyword selection is critical to finding relevant literature and requires justification for querying abstract databases and other sources to ensure accuracy and completeness. The topic of Energy Systems is of great importance due to the introduction of renewable energy sources, modern control methods, and complex energy systems during the energy transition. **The aim of this study.** The goal of this paper was to identify keywords that would be useful to subject matter experts when collecting literature on a particular topic, based on entries in the IEEE Terms field. **Data.** Bibliometric data were exported from the IEEE Xplore platform in the following order: for 2019-2023, 2000 records sorted by relevance, for 2024, 1680 records current as of April 11, 2024. **Analytical methods applied and software used.** VOSviewer, Scimago Graphica implementing the Clauset-Newman-Moore algorithm and agglomerative hierarchical clustering method implemented in Multidendrograms. **Results.** The main issues of the Energy Systems topic are presented in tabular and graphical formats. The fpgrowth utility offers flexible data preparation options, which makes it worthwhile to conduct a separate study to analyze the score of co-occurrence terms given by its algorithm.

Keywords: IEEE Terms; IEEE Xplore; energy systems; VOSviewer; Scimago Graphica; fpgrowth

---

I. Introduction

A. Relevance of the Subject

The preparation of literature reviews [1] and systematic reviews [2,3] requires, at the first stage, justification of the choice of query to abstract databases and other sources of information, which provides an acceptable level of accuracy and completeness of the collected sources for the disclosure of the analyzed topic.

The query is constructed from keywords linked by logical operators and filters embedded in a specific abstract database. Therefore, the justification of the choice of keywords plays an essential role in the search for literature relevant to a given topic.

The above points out the importance of giving reasons for the choice of keywords. The significance of the subject of energy systems is due to the fact that during the energy transition, the incorporation of renewable energy sources and energy storage devices, the use of modern control and optimization methods significantly complicates the structure of energy systems [4,5]. This makes the analysis of scientific publications on this topic relevant.

B. The Aim of this Study

The purpose of this paper was to identify combinations of keywords that would be useful to subject matter experts in gathering literature on a topic of interest within the above-mentioned theme. Relevant keywords were identified based on entries in the IEEE Terms field.

C. Literature Review

A search via the open access abstract databases Scilit.net and Sciencedirect.com for the query 'Energy System" AND "IEEE Term' in the [Title, Abstract, Keyword] fields yielded no publications matching the query.

A query for "All Metadata" "Energy System" AND "All Metadata" "IEEE Term" to the ieeexplore.ieee.org platform also yielded no results as articles.

Consequently, there is no direct equivalent to our work on these platforms.

On the request: Title, abstract, keywords: "Energy System" AND Keywords to the platform sciencedirect.com received — Review articles (11) and Research articles (20). Some of them, which are closest to our task, have been considered in this brief literature review. Brief comments on them are given.

The purpose of this paper [6] was to review and explore the application of Artificial Intelligence (AI) and Machine Learning (ML) in the energy domain using the VOSviewer program. A thorough analysis was performed on the 2000 most recent and most cited articles found using keywords related to energy, AI and MO. A visualization of their co-occurrence was performed. Difference from our work: keyword co-occurrence was evaluated using VOSviewer program only. Other topics were considered, not energy systems.

The paper [7] applied natural language processing using TextRank and TF-IDF algorithms to extract keywords from energy research project descriptions, which has not been done before for the EnArgus database. It was shown that TF-IDF gives better results for keyword extraction. The use of the EnArgus database is rarely encountered in bibliometric studies, it is worthwhile to further explore the possibilities of this database for analyzing energy systems topics.

The study [8] critically analyzes the use of machine learning in photovoltaic and solar energy research using the PRIZMA[1] approach to the Scopus database, including publication trends and bibliometric analysis. PRIZMA approach plays an essential role in writing systematic reviews, but really this article only uses the capabilities of VOSviewer software and Scopus analytics.

The systematic literature review [9] utilized six databases and a set of carefully selected keywords derived from a preliminary analysis of the energy community concept and its many variations. Databases used: Web of Science, Science Direct, SciELO, DOJA, IEEEX and ACM. The value of this paper for our study is that it highlights the importance of standardizing the terminology used. Therefore, we use IEEE terms that have undergone expert editing.

The paper [10] encompasses a complete overview and classification of thermal energy storage technologies used in the housing environment, while considering trends and prospects of prior and current studies. In this paper, we are interested in Tables 1–3, which show the queries used for different categories and the keywords used to group terms related to different topics.

Table 1. Occurrence of the top 20 IEEE Terms in journals with the highest number of publications in the studied bibliometric data sample.

Journal/IEEE Term	Batteries	Computational	Costs	Energy consumption	Energy management	Energy storage	Generators	Internet of Things	Load modeling	Microgrids	Optimization	Power system stability	Real-time systems	Renewable energy	Resource management	Switches	Task analysis	Uncertainty	Voltage control	Wireless
CSEE Journal of Power and Energy Systems	26	6	27	5	12	20	16	2	28	20	36	23	12	45	10	10	1	31	20	1
IEEE Access	438	113	281	310	207	175	174	142	252	278	512	235	143	472	191	95	131	186	213	168
IEEE Internet of Things Journal	46	46	28	129	18	5	2	339	12	6	118	5	27	12	122	8	121	9	2	116
IEEE Sensors Journal	10	4	3	15		1	2	14	2		12		4		5	1	6		1	23
IEEE Systems Journal	45	18	37	19	26	23	25	9	51	33	59	28	31	34	27	17	14	28	33	24

[1] https://www.prisma-statement.org/ — Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA)

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2024 doi:10.20944/preprints202405.1282.v1

IEEE Transactions on Applied Superconductivity	9	1	3		1	12	7		4	3	9	7		3		5			11	1
IEEE Transactions on Communications	13	3	5	13	1			10	1		47		1	1	45	5	15	2		50
IEEE Transactions on Green Communications and Networking	23	9	8	38		1		26	2	2	50		4	12	52	5	27	1		47
IEEE Transactions on Industrial Electronics	71	10	9	14	24	17	24	2	11	26	31	19	13	12	3	41	3	17	64	6
IEEE Transactions on Industrial Informatics	49	11	24	27	35	21	16	12	44	50	72	23	23	31	13	9	18	48	18	12
IEEE Transactions on Industry Applications	116	11	77	9	27	43	55	2	58	84	67	53	27	73	8	38	5	37	106	3
IEEE Transactions on Intelligent Transportation Systems	29	4	18	34	12	7	2	6	7	4	39	1	12	6	10	4	7	8	5	10
IEEE Transactions on Power Electronics	73	2	6	3	12	31	12	1	6	23	11	25	6	13		65	2	1	75	9
IEEE Transactions on Power Systems	41	38	64	6	22	68	82		79	33	79	108	42	95	18	7	2	92	39	
IEEE Transactions on Smart Grid	95	36	78	17	70	59	55	2	90	116	135	44	49	73	19	7	7	101	70	2
IEEE Transactions on Sustainable Energy	71	19	49	4	21	65	30	1	47	52	87	46	28	78	12	7	2	72	51	
IEEE Transactions on Transportation Electrification	54	4	10	15	27	23	3		9	6	29	7	15	11	1	8	3	5	16	
IEEE Transactions on Vehicular Technology	68	18	12	67	28	5	8	20	5	3	99	2	18	7	93	10	52	7	9	93
IEEE Transactions on Wireless Communications	11	11	1	18	1	1		13			62		1	3	55	5	23	2		107
Journal of Modern Power Systems and Clean Energy	26	6	30	2	12	19	19		38	21	32	28	22	39	6	8	2	39	21	

Table 2. Occurrence of major IEEE Terms by years.

IEEEterm/Year	2019	2020	2021	2022	2023	2024
Batteries	293	307	297	269	282	148
Computational modeling	91	88	79	112	97	95
Costs			54	279	367	242
Energy consumption	210	227	187	190	176	124
Energy management	130	120	102	92	139	71
Energy storage	127	151	137	108	132	61
Generators	116	120	112	115	101	80
Internet of Things	111	96	111	127	142	123
Load modeling	126	159	175	154	139	90
Microgrids	164	161	181	142	164	92

Optimization	360	361	327	335	362	257
Power system stability	107	136	142	149	184	106
Real-time systems	116	113	115	125	97	96
Renewable energy sources	118	162	170	267	321	173
Resource management	176	132	140	162	152	127
Switches	77	84	104	105	102	105
Task analysis	85	110	98	136	142	152
Uncertainty	105	122	165	123	168	119
Voltage control	148	128	177	189	193	130
Wireless communication	195	142	159	150	158	117

Table 3. Occurrence of major IEEE Terms for authors with the largest number of publications.

Author/ IEEE Term	Batteries	Computational	Costs	Energy	Energy	Energy storage	Generators	Internet of Things	Load modeling	Microgrids	Optimization	Power system	Real-time systems	Renewable energy	Resource	Switches	Task analysis	Uncertainty	Voltage control	Wireless
J. Wang	18	7	15	23	12	6	4	17	16	11	41	11	8	24	18	11	20	16	13	19
X. Li	23	6	10	15	6	8	7	15	14	12	30	14	12	10	26	10	17	15	18	18
H. Li	15	5	3	10	4	4	8	4	9	6	17	13	6	11	10	7	9	8	8	12
H. Zhang	12	10	16	11	12	8	5	8	14	12	42	10	8	14	19	10	16	22	17	19
Y. Wang	35	14	19	20	14	21	11	13	24	21	42	19	18	18	17	15	12	13	23	19
Z. Li	19		9	9	5	7	7	6	12	17	27	5	8	21	16	6	10	19	9	9
X. Liu	16	7	7	11	2	6	4	8	8	13	27	12	10	12	12	9	10	11	14	11
Y. Li	22	11	19	19	21	17	12	13	15	15	41	19	15	29	29	11	15	15	15	20
Y. Chen	20	15	12	17	10	8	2	10	7	4	26	3	9	7	12	10	13	7	7	11
J. Liu	11	11	14	11	6	14	4	6	13	9	16	7	6	15	11	12	13	12	14	15
H. Wang	26	7	11	22	9	13	9	12	14	8	37	14	8	15	14	6	17	15	12	10
Y. Xu	14	9	15	6	5	13	13	6	15	19	35	13	8	17	16	5	8	27	9	11
Z. Wang	13	9	10	24	9	12	10	3	10	7	28	12	3	16	20	10	13	10	17	12
X. Wang	18	8	17	15	5	11	9	12	18	9	43	14	3	22	24	6	12	18	10	19
J. Li	19	4	21	22	13	9	3	9	16	7	44	6	8	24	18	6	15	11	8	13
Y. Liu	29	11	11	25	17	17	10	8	22	14	41	24	16	16	16	19	22	17	19	18
Y. Zhang	29	16	24	34	15	16	11	23	17	10	66	15	10	33	30	9	26	21	18	27
X. Zhang	31	11	21	21	4	14	4	8	18	13	29	14	10	26	14	13	13	21	20	15
S. Wang	16	6	7	8	6	8	8	6	11	8	18	12	8	17	5	5	4	7	13	7
J. Zhang	14	6	12	15	3	8	7	7	10	6	34	8	5	13	13	7	10	11	9	16

This review may be useful with respect to the PRISMA approach, the use of EnArgus, SciELO, DOJA, and ACM databases, approaches to term standardization, and the use of groups of terms for different topics. However, in compiling the literature review, it was not possible to find works directly related to the study of the topic "Energy Systems" using IEEE Terms.

II. Materials and Methods

A. Data

Bibliometric data were exported from the IEEE Xplore platform in the following order: for 2019-2023 by 2000 records sorted by relevance, for 2024 - 1680 records current as of April 11, 2024.

By querying the database: ("Document Title": energy system) OR ("Abstract": energy system) and applying the filters: "Single Year" as well as "Journals" the records → 11680 were retrieved.

A query of the database was conducted: ‘Document Title’: energy system OR ‘Abstract’: energy system. The filters "Single Year" and "Journals" were then applied, resulting in the retrieval of 11,680 records.

Only records with filled-in fields, specifically the Digital Object Identifier (DOI) and the Institute of Electrical and Electronics Engineers (IEEE) Terms, were utilized. The presence of the DOI allows for the identification of the publication and facilitates its retrieval via the Internet. In this study, we analyzed data from the IEEE Terms field; therefore, records with an empty IEEE Terms field are not of interest.

Of the 11680 entries, 105 do not contain the terms IEEE (11571) and 11566 contain the terms IEEE and DOI. In this paper, 11566 entries are used. Since programs such as VOSviewer cannot directly import data in IEEE Xplore format, we renamed the fields as data field names from Scopus and corrected the delimiters between the terms.

B. Analytical Methods Applied and Software Used

The general characteristics of the IEEE Terms field data are presented in the form of pivot tables. The 20 most frequently occurring IEEE Terms and, respectively, for example, the 20 journals with the largest number of publications were selected for their construction. The SELECT, COUNT() with GROUP BY and PIVOT operators were used to construct the summary tables.

The clustering of EEE Terms was performed using the following programs: VOSviewer [11] which implements the Leiden algorithm [12] for unweighted graphs and Scimago Graphica [13] which implements the Clauset-Newman-Moore algorithm [14], applicable to both weighted and unweighted graphs.

The Agglomerative Hierarchical Clustering method, implemented in Multidendrograms, was employed to construct the dendrogram of IEEE Terms [15].

The co-occurrence of IEEE Terms was evaluated using the utility fpgrowth, as implemented by Christian Borgelt in Ref. [16].

III. Results and Discussion

A. General Characteristics of Field ‘IEEE Terms’ Data

In order to determine which IEEE Terms are most appropriate for searching in certain journals, 20 journals with the largest number of publications on the topic of Energy Systems were selected and the occurrence of IEEE Terms in their publications related to this topic was determined. The results presented in Table 1 facilitate the selection of terms when querying these journals.

IEEE Access is the journal with the highest number of publications per year. This determined the high occurrence of all 20 key terms. As expected, the IEEE Internet of Things Journal has a high frequency of occurrence of the term Internet of Things. IEEE Transactions on Industry Applications has a strong focus on Batteries and Voltage control, and IEEE Transactions on Power Systems has a strong focus on Power system stability.

The evolution of the subject matter of the publications over time was evaluated by examining the change in the most frequently occurring terms presented in Table 2.

The number of publications for 2024 is smaller than for previous years due to the fact that the exported data are current as of April 2024.

The publications predominantly address the topics of optimization, batteries, and renewable energy sources. The costs topic has exhibited the highest growth over time.

The same methodology employed for journals was utilized to evaluate the subject matter of individual authors' publications (Table 3).

The table indicates that publications on the topic are significantly dominated by authors with Chinese surnames.

To support this conclusion, the most common affiliations of authors whose publications contain terms from the top 20 were identified. Such a check is important because an author with a Chinese surname may not work in China. The results are shown in Table 4.

Table 4. Occurrence of top IEEE terms among authors with top 20 affiliations.

Author Affiliations/ IEEE Term	Batteries	Computational	Costs	Energy consumption	Energy management	Energy storage	Generators	Internet of Things	Load modeling	Microgrids
School of Electrical … China	1		13		2	16	14	4	15	9
Department of Electrical … China	39	16	23	19	5	28	14	1	9	10
School of Electrical … China	40	16	40	6	18	6	15		57	20
College of Energy … China	7	10	16		7	1	2	6	12	5
School of Electric … China	14	4	6		7	6	16		25	12
School of Electrical … China	78	9	42	38	23	28	1		20	26
School of Electrical … Singapore	45	4	18	1	9	29	14		19	51
School of Electrical … China	31		12	4	31	15	6		9	8
State Key Laboratory … China	7	9	34	7	16	5	34		20	5
School of Electrical … China	7	1	6	2	6	6	11	4	5	11
College of Electrical … China	5	15	18	4	3	15	7		30	15
Department of Electrical …India	55		7		11	2	6	4	6	34
College of Electrical … China	14		8	1	10	11	12		11	17
Department of Energy …Denmark	30	5	3	3	9	21	8	3	14	43
College of Electrical … China	28	2	15	2	4	32	34		29	14
State Key Laboratory …China	15	6	11		1	9	15		12	18

Author Affiliations/ IEEE Term	Optimization	Power system stability	Real-time systems	Renewable energy	Resource management	Switches	Task analysis	Uncertainty	Voltage control	Wireless
School of Electrical … China	21	32	7	16	3	6		18	16	4
Department of Electrical … China	24	23	6	38		20	1	26	30	
School of Electrical … China	64	39	12	42	5	17		66	18	
College of Energy … China	30	15	19	24	4	1		38	2	
School of Electric … China	11	14	2	13		13		11	1	
School of Electrical … China	20	19	20	22	11	24		20	35	
School of Electrical … Singapore	19	15	10	14	9	3		22	34	4
School of Electrical … China	23	14	14	12					21	
State Key Laboratory … China	31	45	4	50	8	19	8	31	6	
School of Electrical … China	15	25		7	4	19		22	24	5
College of Electrical … China	14	6	11	19	9	4	1	23	7	3
Department of Electrical …India	2	9	3		7	24		1	69	1
College of Electrical … China	12	31	2	36	1	20		9	24	1
Department of Energy …Denmark	16	21	5	22		11		13	31	2
College of Electrical … China	30	41	12	55	6	17	3	67	40	
State Key Laboratory …China	25		6	5		4		16	12	

National Renewable …USA	19	5	13		14	12	2		33		31	15	8	22	5			3	21	3
National Mobile …China	1		1	19				4			49	5			34		11			25
State Key Laboratory … China	8	19	3	30				29	3		13		3		25	4	29	1		35
School of Electrical … Australia	18	4	29	6	4	23	14	3	9	6	22	5	15	20	12		7	6	14	5

List of complete names of the top 20 affiliations, presented in the same sequence as in Table 4:

- School of Electrical Engineering and Automation, Wuhan University, Wuhan, China
- Department of Electrical Engineering, Tsinghua University, Beijing, China
- School of Electrical Engineering, Southeast University, Nanjing, China
- College of Energy and Electrical Engineering, Hohai University, Nanjing, China
- School of Electric Power Engineering, South China University of Technology, Guangzhou, China
- School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China
- School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
- School of Electrical Engineering, Beijing Jiaotong University, Beijing, China
- State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing, China
- School of Electrical and Information Engineering, Tianjin University, Tianjin, China
- College of Electrical Engineering, Sichuan University, Chengdu, China
- Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India
- College of Electrical and Information Engineering, Hunan University, Changsha, China
- Department of Energy Technology, Aalborg University, Aalborg, Denmark
- College of Electrical Engineering, Zhejiang University, Hangzhou, China
- State Key Laboratory of Advanced Electromagnetic Engineering and Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, China
- National Renewable Energy Laboratory, Golden, CO, USA
- National Mobile Communications Research Laboratory, Southeast University, Nanjing, China

The dominance of affiliations with Chinese institutions is evident.

B. IEEE Terms Clustering Results Obtained Using VOSviewer Program

This section presents the results of IEEE Terms clustering obtained using the default parameters of VOSviewer. It is found that the sample used contained 3180 unique IEEE Terms, and 1425 of them appeared 5 or more times. 9 clusters were obtained.

The list of the most frequently occurring terms and their total link strength is presented in Table 5. For most of the terms, the total link strength is often larger the more times the term occurs, but not always, e.g., the 'term real-time systems' occurs 659 times and has a total link strength of 3731, and the term 'energy management' — 651 and 3964 respectively. Terms with high total link strength are good candidates for multi-term queries.

Table 5. Most common IEEE terms and their total link strength.

IEEE Term	occurrences	total link strength

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2024 doi:10.20944/preprints202405.1282.v1

optimization	1994	11429
batteries	1594	9065
renewable energy sources	1198	7132
energy consumption	1109	6257
voltage control	957	5538
costs	936	5786
wireless communication	918	5223
microgrids	898	5249
resource management	883	5168
load modeling	842	4888
power system stability	818	4744
uncertainty	799	4652
task analysis	721	4135
energy storage	710	4159
internet of things	708	4073
real-time systems	659	3731
energy management	651	3964
generators	643	3723
switches	576	3192
computational modeling	561	3241
mathematical model	545	2953
wireless sensor networks	510	2880
sensors	495	2679
energy harvesting	476	2684
reliability	476	2689
power systems	457	2646
mathematical models	448	2608
topology	434	2462
reactive power	425	2453
servers	409	2412

The results presented in this table can be considered as a description of the dominant topic in all publications whose bibliometric data we exported from the IEEE Xplore platform.

The IEEE Terms clustering obtained with the VOSviewer program is shown in Figure 1.

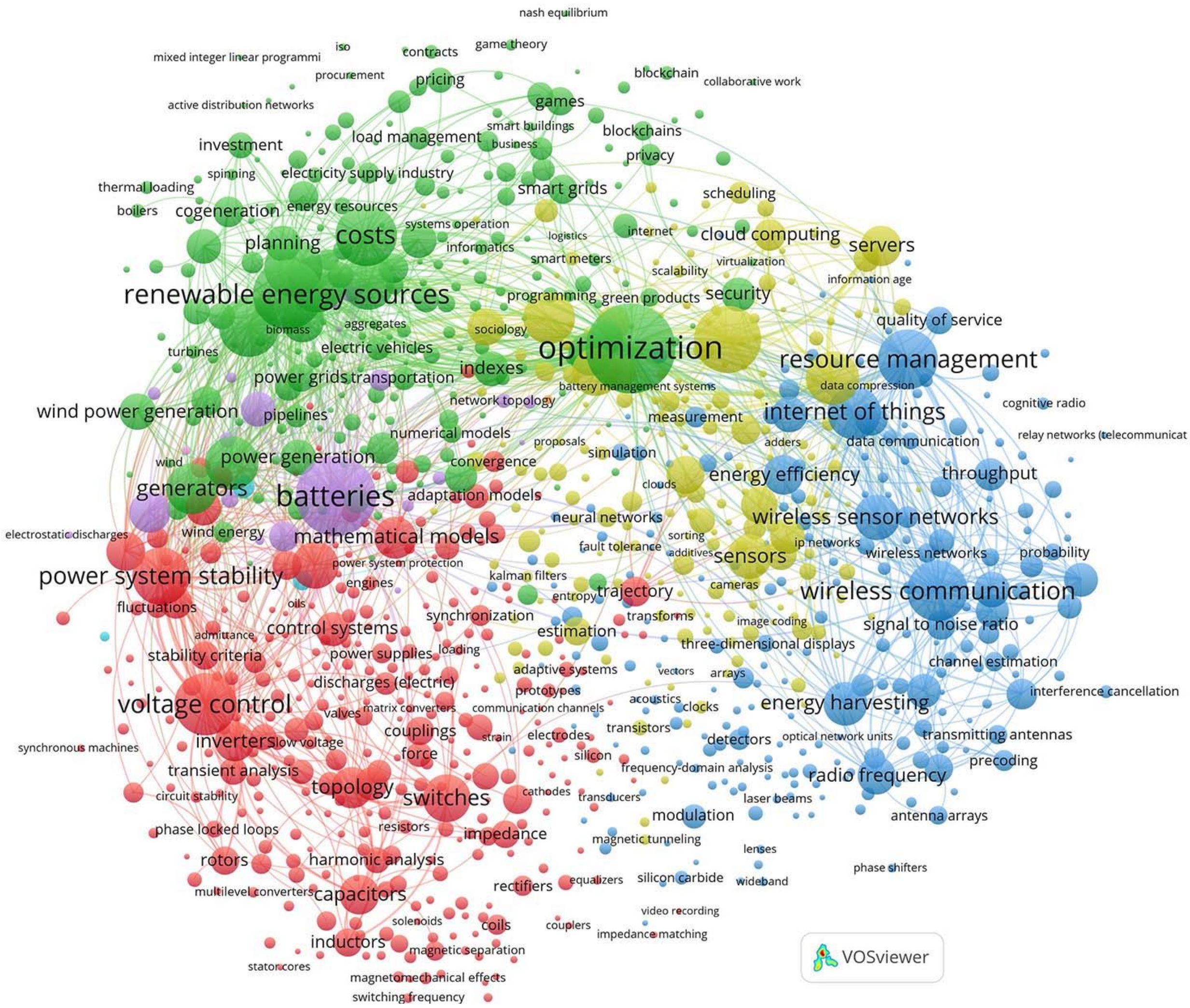

Figure 1. 9 clusters of co-occurrence of IEEE Terms, obtained with the VOSviewer program.

Given the parameters, clusters 7,8,9 contain one term, cluster 6 contains three, and cluster 5 contains 20.

Four clusters with 311,234,232,197 terms have highest representation.

Cluster 1 (red): voltage control (957), power system stability (818), switches (576), mathematical model (545), mathematical models (448), topology (434), capacitors (398), inverters (391), frequency control (374), power system dynamics (314), control systems (287), voltage (279), wind turbines (261), trajectory (221), stability analysis (218).

Cluster 2 (green): optimization (1994), renewable energy sources (1198), costs (936), microgrids (898), load modeling (842), uncertainty (799), energy storage (710), energy management (651), generators (643), reliability (476), power systems (457), wind power generation (331), stochastic processes (324), planning (320), power generation (317).

Cluster 3 (blue): wireless communication (918), resource management (883), internet of things (708), wireless sensor networks (510), energy harvesting (476), energy efficiency (387), radio frequency (363), relays (325), delays (322), receivers (290), noma (287), interference (278), protocols (268), throughput (266), array signal processing (242).

Cluster 4 (khaki): energy consumption (1109), task analysis (721), real-time systems (659), computational modeling (561), sensors (495), servers (409), monitoring (360), predictive models (348), power demand (329), computer architecture (300), heuristic algorithms (297), data models (278), cloud computing (257), hardware (246), training (219).

Cluster 5 (violet): batteries (1594), reactive power (425), state of charge (299), hybrid power systems (231), degradation (126), predictive control (75), decentralized control (71), marine vehicles

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2024 doi:10.20944/preprints202405.1282.v1

(62), network topology (57), aging (48), communication networks (44), automatic generation control (42), energy loss (40), us department of defense (30), uninterruptible power systems (21).

C. IEEE Terms clustering results obtained using the Scimago Graphica program

VOSviewer is a high quality, self-sufficient program for bibliometric analysis, but this is its limitation, the clustering algorithm, data pre-processing and graphing are predefined. If, for example, the researcher needs to work with weighted graphs, or apply a different algorithm for clustering selected terms, or make lists of key terms according to a given subject dictionary, then it will be necessary to implement a more flexible approach to data analysis, allowing at each step to implement different approaches to work with data.

VOSviewer allows you to import Pajek NET format data and use the 'thesaurus_terms' file to replace the spelling of terms, this overcomes a number of limitations listed above, but it is still appropriate to use this program for its intended purpose.

Another approach could be to use more versatile programs such as Scimago Graphica programs or utilities — fpgrowth, sed, grep will allow you to build a pip to implement more flexible data analysis.

Figures 2 and 3 show the result of this approach: the co-occurrence of IEEE Terms was determined using fpgrowth, the terms themselves were selected based on their co-occurrence, and the sed utility and plain text editor were used to get the correct format for Scimago Graphica.

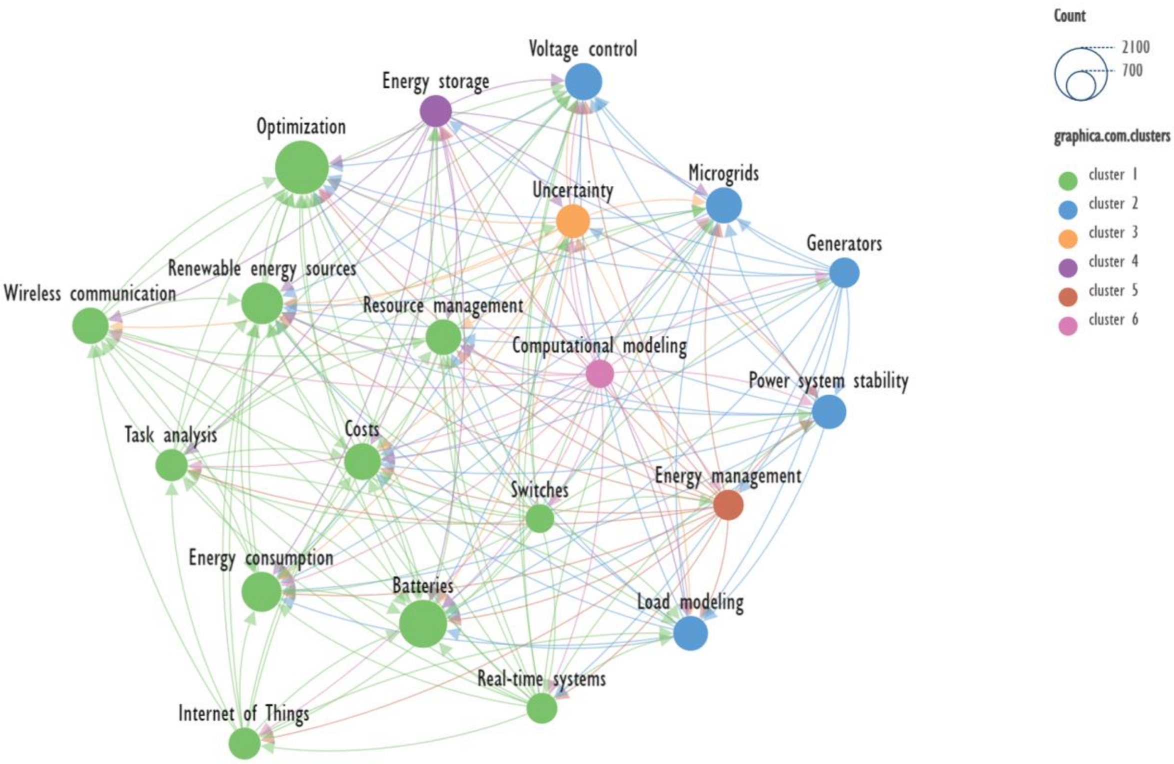

Figure 2. Clustering of the co-occurrence of the 20 most frequent IEEE terms, obtained using the Scimago Graphica program for unweighted edges. **Note:** Figures 2 and 3 show **undirected graphs**, for clarity we added arrows on the edges — with terms of which clusters the considered term is more related. In addition, the possibility of using this program to display directed graphs is shown. Actually, the main purpose of this paper is to demonstrate the capabilities of the used programs.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2024 doi:10.20944/preprints202405.1282.v1

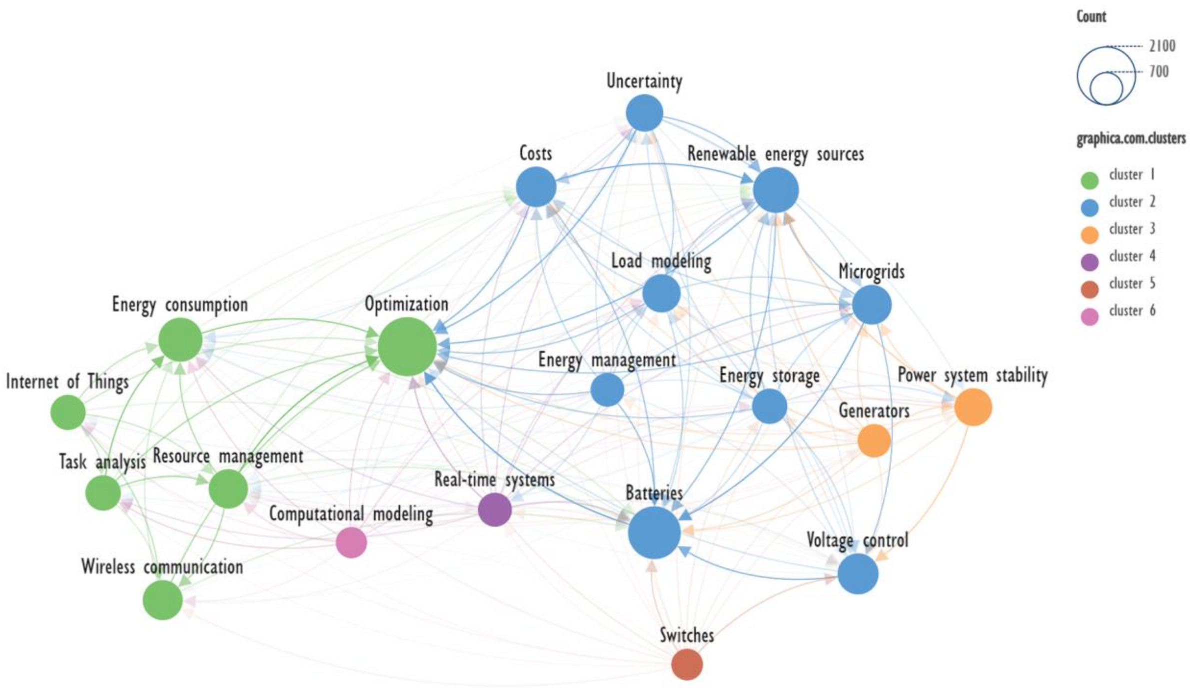

Figure 3. Clustering of the co-occurrence of the 20 most frequent IEEE Terms obtained using the Scimago Graphica program for weighted edges.

Using a weighted graph significantly redistributes nodes between clusters. This is most noticeable by moving the terms 'Batteries', 'Energy Storage', and 'Energy Management' to the blue cluster.

Another strength of Scimago Graphica is the ability to use different Layout algorithms. For example, in the previous two cases the LinLog method was used, in Figure 4 the use of Degre Top-Bottom is shown. Other Layout methods did not give a good representation of the data in this graph.

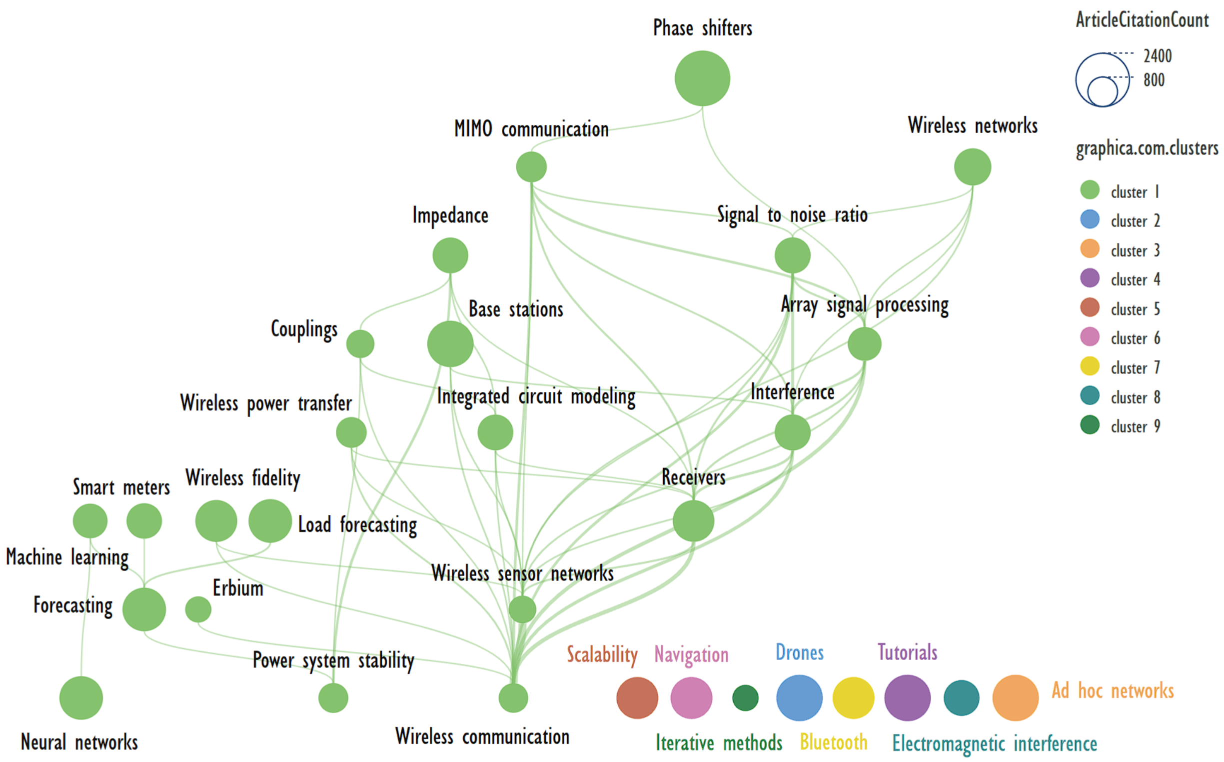

Figure 4. Clustering of the co-occurrence of 30 IEEE Terms taken from the most cited articles, obtained using the Scimago Graphica weighted relationships program. The weights of the terms were estimated by the average citations of the articles in the sample in which they occur.

D. IEEE Terms clustering results obtained using Multidendrograms program

Hierarchical clustering of terms is the most established method, which allows easy and clear interpretation of the obtained results. The algorithms implementing this approach are well studied and optimized. For large ones, this approach is not feasible [17], but the task of selecting the right list of terms for which the subject matter expert will make a query, for example, when collecting materials for writing a systematic review, will consist of sequentially applying constraints on their selection and their total number will not be large. In this case, the use of a hierarchical clustering of the terms is appropriate. As in the results described in the previous section, the results of hierarchical clustering depend on the estimation of the weight of the co-occurrence of terms. Examples of dendrograms for weighted and unweighted graphs are shown in Figures 5 and 6.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2024 doi:10.20944/preprints202405.1282.v1

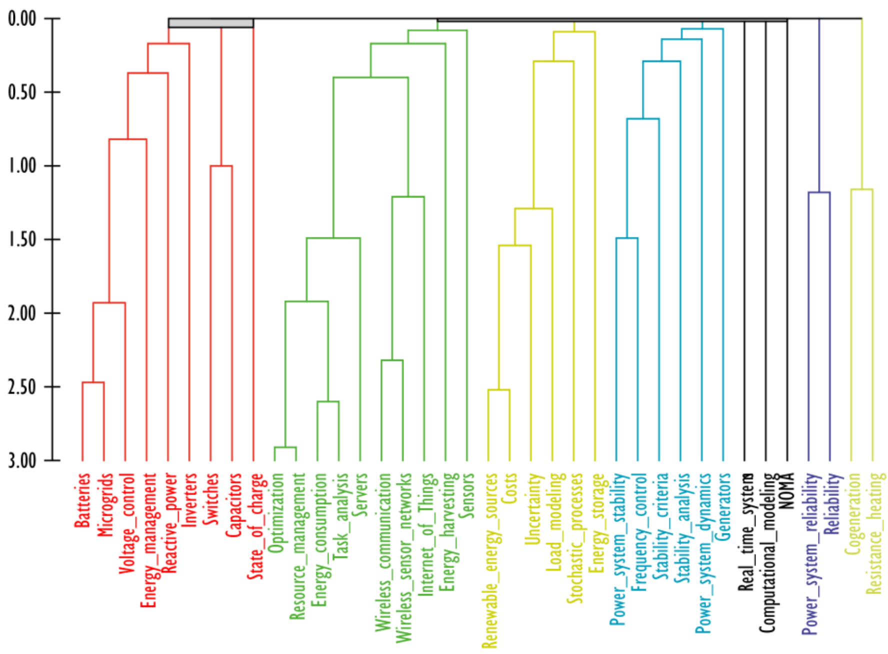

Figure 5. Dendrogram of IEEE Terms obtained by applying fpgrowth -s1m2n2 algorithm and Clustering algorithm with parameters → Beta Flexible → Weighted.

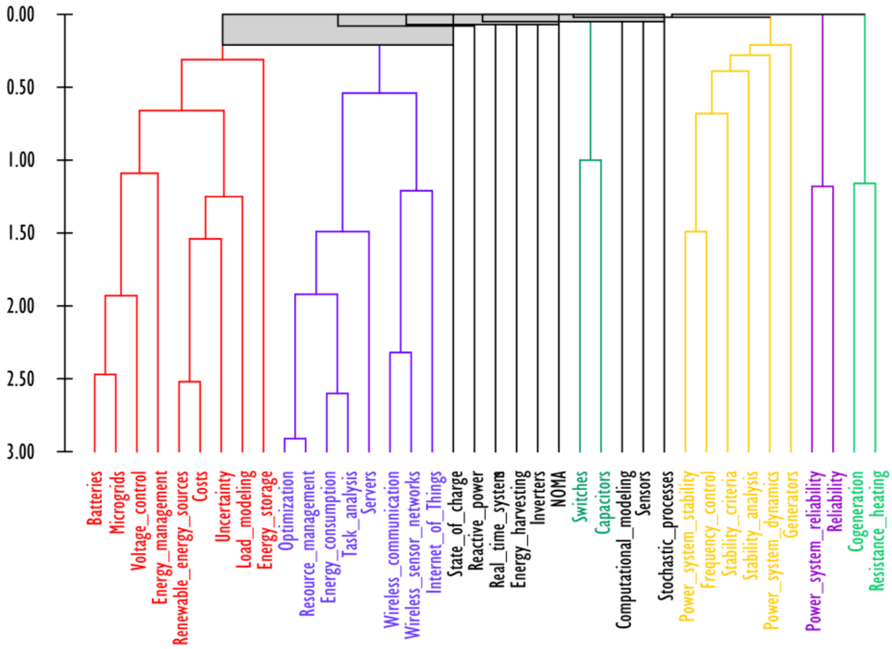

Figure 6. Dendrogram of IEEE Terms obtained by applying fpgrowth -s1m2n2 algorithm and Clustering algorithm with parameters → Beta Flexible → Unweighted.

The dendrogram was constructed using the program MultiDendrograms-5.2.1 with the following parameters [15]: Tour of measure → Similarity; Precision → 2; Clustering algorithm → Beta Flexible → Weighted.

Hierarchical clustering is the most easily interpretable results, but like any clustering is sensitive to the parameters chosen, e.g. if you change Clustering algorithm → Beta Flexible → Weighted to Clustering algorithm → Beta Flexible → unweighted

It is noteworthy that the changes in clustering affect the terms with the lowest similarity. When the similarity measure is changed: single linkage, complete linkage, and arithmetic linkage significantly change the nature of the dendrogram. Therefore, any sequence of mining text should be considered as a clue to a choice, not as something that is the only one that is correct.

E. IEEE term co-occurrence estimates obtained using the fpgrowth utility

The simplest way of estimating the co-occurrence of terms can be obtained using Apriori class algorithms.

Examples of descriptions of possible topics of publications in three IEEE Terms obtained using the fpgrowth utility. The numbers on the right side of the lists are measures of co-occurrence of terms given by the fpgrowth utility.

The three most occurring topics that contain the term **Voltage_control** are as follows:

- Voltage_control → Renewable_energy_sources → Batteries → 0.2247
- Voltage_control → Optimization → Batteries → 0.164204
- Voltage_control → Optimization → Renewable_energy_sources → 0.0950652

The three most occurring topics that contain the term **Costs** are as follows:

- Costs → Optimization → Renewable_energy_sources → 0.656814
- Costs → Batteries → Renewable_energy_sources → 0.535822
- Costs → Batteries → Optimization → 0.406188

The three most occurring topics that contain the term **Microgrids** are as follows:

- Microgrids → Voltage_control → Batteries → 0.760522
- Microgrids → Renewable_energy_sources → Batteries → 0.544465
- Microgrids → Optimization → Batteries → 0.492611

The three most occurring topics that contain the term **Resource_management** are as follows:

- Resource_management → Wireless_communication → Optimization → 0.760522
- Resource_management → Energy_consumption → Optimization → 0.604961
- Resource_management → Energy_consumption → Wireless_communication → 0.30248

The three most occurring topics that contain the term **Load_modeling** are as follows:

- Load_modeling → Costs → **Renewable_energy_sources** → 0.432115
- Load_modeling → Costs → Optimization → 0.328407
- Load_modeling → Renewable_energy_sources → Optimization → 0.311123

The three most occurring topics that contain the term **Power_system_stability** are as follows:

- Power_system_stability → Microgrids → Voltage_control → 0.319765
- Power_system_stability → Voltage_control → Renewable_energy_sources → 0.250627
- Power_system_stability → Microgrids → Renewable_energy_sources → 0.2247

The three most occurring topics that contain the term **Uncertainty** are as follows:

- Uncertainty → Costs → Renewable_energy_sources → 0.561749
- Uncertainty → Costs → Optimization → 0.501253
- Uncertainty → Load_modeling → Optimization → 0.432115

The three most occurring topics that contain the term **Task_analysis** are as follows:

- Task_analysis → Resource_management → Energy_consumption → 0.916083
- Task_analysis → Optimization → Energy_consumption → 0.881514
- Task_analysis → Optimization → Resource_management → 0.75188

The three most occurring topics that contain the term **Energy_storage** are as follows:

- Energy_storage → Costs → Renewable_energy_sources → 0.371619
- Energy_storage → Optimization → Renewable_energy_sources → 0.362976
- Energy_storage → Microgrids → Renewable_energy_sources → 0.319765

The three most occurring topics that contain the term **Real-time_systems** are as follows:

- Real-time_systems → Batteries → Optimization → 0.33705
- Real-time_systems → Uncertainty → Optimization → 0.285196
- Real-time_systems → Task_analysis → Energy_consumption → 0.259269

The three most occurring topics that contain the term **Energy_management** are as follows:

- Energy_management → Optimization → Batteries → 0.570391
- Energy_management → Microgrids → Optimization → 0.509895
- Energy_management → Microgrids → Batteries → 0.458042

The three most occurring topics that contain the term **Generators** are as follows:

- Generators → Costs → Renewable_energy_sources → 0.276553
- Generators → Microgrids → Batteries → 0.259269
- Generators → Optimization → Renewable_energy_sources → 0.241984

The three most occurring topics that contain the term **Computational_modeling** are as follows:

- Computational_modeling → Energy_consumption → Task_analysis → 0.466684
- Computational_modeling → Resource_management → Task_analysis → 0.371619
- Computational_modeling → Optimization → Task_analysis → 0.319765

The fpgrowth utility offers flexible data preparation options, which makes it worthwhile to conduct a separate study to analyze the score of co-occurrence terms given by its algorithm.

IV. Conclusions

Various methods are presented for using IEEE Terms to define keywords for queries to collect publications for writing literature reviews and systematic reviews.

The main issues of the Energy Systems topic are presented in tabular and graphical formats.

The feasibility of constructing pivot tables for the comprehensive evaluation of analyzed bibliometric data exported from the abstract database is demonstrated.

The straightforward approach to analyze key terms based on their co-occurrence is to use the VOSviewer program and Apriori class algorithms.

For a more detailed analysis of the co-occurrence of terms, it is recommended to utilize programs Scimago Graphica and Multidendrograms, with preliminary preparation of a sample of bibliometric data and selection of an appropriate clustering method and its parameters.

V. Possible applications of the findings

The findings of this study can be utilized as a framework for developing queries to reference databases when gathering materials for the compilation of literature and systematic reviews.

Acknowledgment: This work was funded by the Ministry of Science and Higher Education of the Russian Federation, State Assignment No. 122022800270-0.

References

1. S. Kraus et al., 'Literature reviews as independent studies: guidelines for academic practice', *Rev Manag Sci*, vol. 16, no. 8, pp. 2577–2595, Nov. 2022, doi: 10.1007/s11846-022-00588-8.
2. K. S. Khan, R. Kunz, J. Kleijnen, and G. Antes, 'Five Steps to Conducting a Systematic Review', *J R Soc Med*, vol. 96, no. 3, pp. 118–121, Mar. 2003, doi: 10.1177/014107680309600304.
3. P. Siddaway, A. M. Wood, and L. V. Hedges, 'How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses', *Annu. Rev. Psychol.*, vol. 70, no. 1, pp. 747–770, Jan. 2019, doi: 10.1146/annurev-psych-010418-102803.
4. S. Bolwig et al., 'Review of modelling energy transitions pathways with application to energy system flexibility', *Renewable and Sustainable Energy Reviews*, vol. 101, pp. 440–452, Mar. 2019, doi: 10.1016/j.rser.2018.11.019.
5. Deniz and M. S. Çeliktaş, 'Analysis of Energy Transition Pertaining to the Future Energy Systems', in *Handbook of Smart Energy Systems*, M. Fathi, E. Zio, and P. M. Pardalos, Eds., Cham: Springer International Publishing, 2023, pp. 1535–1555. doi: 10.1007/978-3-030-97940-9_92.
6. Entezari, A. Aslani, R. Zahedi, and Y. Noorollahi, 'Artificial intelligence and machine learning in energy systems: A bibliographic perspective', *Energy Strategy Reviews*, vol. 45, p. 101017, Jan. 2023, doi: 10.1016/j.esr.2022.101017.

 doi:10.20944/preprints202405.1282.v1

7. J. Richarz, S. Wegewitz, S. Henn, and D. Müller, 'Graph-based research field analysis by the use of natural language processing: An overview of German energy research', *Technological Forecasting and Social Change*, vol. 186, p. 122139, Jan. 2023, doi: 10.1016/j.techfore.2022.122139.
8. Zaidi, 'A bibliometric analysis of machine learning techniques in photovoltaic cells and solar energy (2014–2022)', *Energy Reports*, vol. 11, pp. 2768–2779, Jun. 2024, doi: 10.1016/j.egyr.2024.02.036.
9. D. De São José, P. Faria, and Z. Vale, 'Smart energy community: A systematic review with metanalysis', *Energy Strategy Reviews*, vol. 36, p. 100678, Jul. 2021, doi: 10.1016/j.esr.2021.100678.
10. E. Borri, G. Zsembinszki, and L. F. Cabeza, 'Recent developments of thermal energy storage applications in the built environment: A bibliometric analysis and systematic review', *Applied Thermal Engineering*, vol. 189, p. 116666, May 2021, doi: 10.1016/j.applthermaleng.2021.116666.
11. 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.
12. V. A. Traag, L. Waltman, and N. J. Van Eck, 'From Louvain to Leiden: guaranteeing well-connected communities', *Sci Rep*, vol. 9, no. 1, p. 5233, Mar. 2019, doi: 10.1038/s41598-019-41695-z.
13. Y. Hassan-Montero, F. De-Moya-Anegón, and V. P. Guerrero-Bote, 'SCImago Graphica: a new tool for exploring and visually communicating data', *EPI*, p. e310502, Sep. 2022, doi: 10.3145/epi.2022.sep.02.
14. Clauset, M. E. J. Newman, and C. Moore, 'Finding community structure in very large networks', *Phys. Rev. E*, vol. 70, no. 6, p. 066111, Dec. 2004, doi: 10.1103/PhysRevE.70.066111.
15. Fernández and S. Gómez, 'Solving Non-Uniqueness in Agglomerative Hierarchical Clustering Using Multidendrograms', *J Classif*, vol. 25, no. 1, pp. 43–65, Jun. 2008, doi: 10.1007/s00357-008-9004-x.
16. Borgelt, 'An implementation of the FP-growth algorithm', in *Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations*, Chicago Illinois: ACM, Aug. 2005, pp. 1–5. doi: 10.1145/1133905.1133907.
17. F. Nielsen, 'Hierarchical Clustering', in *Introduction to HPC with MPI for Data Science*, Cham: Springer International Publishing, 2016, pp. 195–211. doi: 10.1007/978-3-319-21903-5_8.