Article

Technological Sovereignty as a Current Energy Security Challenge. Preliminary Analysis

Boris Chigarev [*]

Article

Technological Sovereignty as a Current Energy Security Challenge. Preliminary Analysis

Boris Chigarev

Oil and Gas Research Institute, Russian Academy of Sciences, Moscow, Russia; bchigarev@ipng.ru

Abstract: Energy security is often interpreted as independence from fossil fuels, but a one-sided approach can lead to dependence on high-value-added technologies. The development of artificial intelligence, which requires high energy consumption, chips and servers, is shifting competition in manufacturing and services from energy security to technological sovereignty. With the development of technology, sovereignty has shifted from military independence to freedom from economic coercion by other states and large corporations. The aim of this study was to identify suitable tools for analyzing abstract texts from tens of thousands of bibliometric records and pre-assessing relevant topics related to the energy sector to effectively analyze trends in technological sovereignty issues. In this paper 10 thousand bibliometric records for the year 2024, sorted by relevance and exported from the open abstract database Scilit on the query: "energy AND technology" in [Title, Abstract, Keyword], Content Type: JOURNAL-ARTICLE, English. Filters were applied on the "Subject" category most related to technology: Power Systems & Electric Vehicles, Energy Systems & Technologies, Electrical Energy Management, and Nuclear Technology & Instrumentation. The main theme of the bibliometric data analyzed was renewable energy. Twelve clusters were identified based on keywords, of which three were closest to the topic for which this research was funded: *hydrogen*, *heat energy storage* and *greenhouse gas emissions*. These clusters reflect keywords derived from both Yake! and PatternRank. The Yake! program outperforms PatternRank in terms of run time and representation of found keywords in abstract texts. The feasibility of using *AnyAscii* for text preprocessing is demonstrated. Using artificial intelligence to create text based on key phrases speeds up text processing, but the need for manual editing remains. The study showed that there is a need to expand data sources, e.g. using OnePetro for oil and gas topics, IEEE Xplore for energy systems issues, Semantic Scholar to evaluate the role of AI in the energy sector.

Keywords: technological sovereignty; energy security; bibliometric data; text analysis; key phrases

1. Introduction

Relevance. With the development of technology, sovereignty has shifted from military independence to freedom from economic coercion by other states and large corporations. States must ensure control over critical technologies, access to them from independent countries, and guaranteed and long-term access to monopoly suppliers from countries such as the United States or China [1].

It should be noted that the new is the well-forgotten old, as illustrated by an article from 1983 [2] with an evocative title "Technological Sovereignty: Forgotten Factor in the 'Hi-Tech' RAZZAMATAZZ". The publication's main claim is that technological sovereignty refers to the freedom to select, generate and utilize commercially necessary technologies for industrial innovation, as opposed to technological self-sufficiency, which implies possession of all necessary technologies. **Note:** according to Cambridge dictionary [1] RAZZAMATAZZ is noisy and noticeable activity, intended to attract attention.

The study of technological sovereignty in Russia and the EU countries, according to a systematic review of scientific literature and expert interviews [3], has been significantly intensified over the last

---

[1] https://dictionary.cambridge.org/dictionary/english/razzmatazz

three years, with the focus on developing an adequate definition of it to understand and address the challenges of technological sovereignty. (Kapoguzov E.A., Pakhalov A.M. (2024). Technological sovereignty: Conceptual approaches and perceptions by the Russian academic experts. Journal of the New Economic Association, 3 (64), 244–250 (in Russian))

The increasing role of digitalization in the economy, and thus in economic security, makes it relevant to achieve data sovereignty — the control and management of data within one's own jurisdiction and operational sovereignty — the independent operation and maintenance of critical digital infrastructure and services.

The article [4] argues that Big Tech firms have not only challenged traditional sovereignty but also established a complex symbiotic relationship due to their technical advantages.

The study [5] aims to provide a clear understanding of data sovereignty in the context of new data-driven technologies, addressing the challenges faced by various stakeholders in retaining control over their data.

The paper [6] explores the concept of digital sovereignty through embedded 'situated practices' of political and economic projects that aim to create autonomous digital infrastructures in a hyper-connected world.

The relevance of a comprehensive study of technological sovereignty as a current energy security challenge is also due to the high degree of influence of politics in promoting the interests of specific countries, so that [7] notes that at the UN Climate Change Conference in November 2021, 34 countries, including Germany, pledged to end international financing for fossil fuel extraction by 2022. But Senegal's President Macky Sall criticized the commitment, saying it would be a "death blow" to the African economy. Senegal, which was on the verge of exporting gas, sees its reserves as crucial to economic growth. However, in the wake of the European energy crisis, German Chancellor Olaf Scholz has announced plans to partner with Senegal on gas exports.

When 34 countries, many of them technological and financial leaders, announce their intention to end international funding for fossil fuel extraction by 2022, research funding and scientific publications are bound to be affected. And the Glasgow conference is not unique.

Motivation. Energy security can include independence from fossil fuels. This is more characteristic of European countries, for example. But an emphasis on renewable energy sources can lead to a different kind of dependence and thus security: dependence on high-value-added technologies and components. Not all countries, especially developing countries, are self-sufficient in renewable energy technologies, the efficient operation of which depends not only on the renewable energy generators themselves but also on a developed infrastructure, including capacity balancing, energy storage, charging stations, and so on. In other words, independence from fossil energy sources may turn into technological dependence. The development of AI implies the development of data centers, the operation of which requires not only high energy consumption but also a large number of servers (or microprocessors) and software. As AI is seen as the next stage in improving the competitiveness of manufacturing and services, developing countries may become dependent on countries with advanced technologies. In other words, national security issues will shift from energy security to technological security.

Energy security is still important, it's just that countries with advanced science and technology will have better leverage to manipulate developing countries, for example by blocking high-tech equipment or access to artificial intelligence-based services.

An example of the growing interdependence of energy and AI technologies is reflected in the report[2] and the appearance of the journal *Energy and AI* at Elsevier[3] in 2020.

From a broader perspective, the convergence of competencies in energy&technology&R&D can be expected to become the next hot topic of economic security.

---

[2] https://www.iea.org/reports/energy-and-ai — IEA (2025), Energy and AI, IEA, Paris

[3] https://www.scimagojr.com/journalsearch.php?q=2101047378&tip=sid

The relevance of the topic under discussion may be indicated by the following reports: IEA, Energy Technology Perspectives 2020 dataset[4], and World Energy Investment 2024 Datafile[5].

A short list of key statements with examples of works revealing their content on technological issues of energy security topic

The technological dimension of energy security involves examining how modern technology can ensure reliable and sustainable energy supplies, protect infrastructure and maintain energy independence. The purpose of the study [8] is to examine the relationship between technological progress and energy security in the context of global uncertainty, financial development, globalization and infrastructure development of newly industrialized countries.

Energy efficiency and energy saving technologies such as improved insulation, LED lighting and efficient technologies reduce energy demand, increase energy security. The study [9] considers energy saving as the most important aspect of energy security of the economy, compares energy intensity and energy efficiency of Russia with other countries. It highlights the significant potential for energy saving in the industrial sector and identifies barriers to its realization, including the lack of specialists with relevant education.

Nuclear technology is a stable source of energy with low carbon emissions. The review [10] assesses the economic, climate and environmental viability of global nuclear power, emphasizing its importance to global energy security and climate change mitigation. The paper [11] explores the potential of nuclear power in the energy transition, highlighting its significant contribution to the global electricity mix, particularly in developed countries, and its minimal greenhouse gas emissions.

A resilient energy infrastructure, including redundant transmission lines, microgrids and distributed energy resources, ensures a stable energy supply. The paper [12] presents a modern method for securing critical infrastructure in energy transmission networks, integrating cryptographic mechanisms with biometric data to enhance cyber threat protection.

Renewable energy technologies, including solar, wind, hydroelectric and geothermal, play an important role in energy security but require energy storage solutions. The paper [13] discusses the importance of transitioning to renewable energy sources as a means of mitigating climate change and ensuring long-term energy security. The necessity of sustainable supply chains and strategies to reduce dependence on foreign suppliers is emphasized, addressing challenges such as geopolitical tensions, trade restrictions, and natural disasters. The review [14] examines the potential of hydrogen as a clean energy carrier, emphasizing its role in replacing fossil fuels and recent advances in hydrogen production technologies. Due to the growing economy, India is experiencing an increasing demand for energy. The government is regulating the use of fossil fuels and promoting renewable energy sources such as geothermal. However, geothermal energy in India is yet to be explored. The study [15] demonstrates the potential of geothermal energy in the Indian subcontinent and its applications.

Digitally dependent energy infrastructure requires robust cybersecurity measures such as encryption, firewalls and intrusion detection systems. The study [16] examined the use of artificial intelligence to enhance the security of critical energy infrastructure, resulting in a 98% increase in threat detection and a 70% reduction in incident response time.

Advanced technologies such as horizontal drilling and hydraulic fracturing have increased the availability of natural gas and oil, and carbon capture and storage (CCS) technologies are helping to reduce the environmental impact of fossil fuel use. China has drilled its first ultra-deep scientific exploration well, reaching a depth of over 10,000 meters, at CNPC's Take-1 well in the Fuman oil field. The five-year plan aims to develop the field to produce 35.7 million barrels of oil per year by 2025 [17]. Fossil fuels play a critical role in ensuring national energy security. A comprehensive

---

[4] https://www.iea.org/data-and-statistics/data-product/energy-technology-perspectives-2020-dataset — IEA, Paris

[5] https://www.iea.org/data-and-statistics/data-product/world-energy-investment-2024-datafile — IEA, Paris

analysis of China's future fossil fuel demand is essential, especially in the context of political stability, economic normalization, carbon emission reduction and enhanced energy security [18].

Energy storage systems, including lithium-ion, solid-state and next-generation batteries, hydro storage, flywheels and superconducting magnets, are critical to addressing energy security issues but face technical challenges. The commissioning of the second unit of the Astravets NPP in Belarus in 2023 will increase the need for controllability and security of the Belarusian power system. Energy storage systems can help to balance load curves, and the paper [19] evaluates the performance of lithium-ion energy storage systems. The study [20] proposes a power allocation approach for multi-location energy storage systems based on security regions. It analyzes hourly loads and average intraday loads to determine charging and discharging strategy trends. Cost-effective configurations can be achieved by considering energy conversion devices and equipment prices.

On the need for preliminary work on selecting tools for analyzing scientific texts and identifying keywords for collecting publications on the multifaceted topic under consideration

Energy security, which encompasses various aspects such as political-economic, technological, resource, legal and environmental, means the uninterrupted availability of energy sources at an affordable price[6].

Technological aspects include the development of energy consumption, distribution, transmission and production technologies that contribute to the efficient use of energy [21].

The study examines the technological aspect of energy security, emphasizing that the lack of technological sovereignty can significantly worsen the overall energy security situation. Russia, which has significant natural resources, can solve economic and legal issues independently, while technological solutions, including in the energy sector, are largely dependent on imports. Large markets such as China, India and the EU are oriented towards renewable energy sources due to the large volume of imports, while Russia benefits from developing renewable energy sources to save on domestic consumption and free up resources for export or deeper processing.

In order to effectively analyze trends in technological challenges to energy security, it is very important to select appropriate methods and tools to track the texts of publications in these areas.

In a number of open access abstract databases, the Author Keywords field in exported bibliometric records is poorly populated. In addition, in OnePetro, for example, keywords are available in individual publication records, but only the title and abstract fields are available when the data are exported. The second problem is that author keywords are not always available in the abstract texts. Even if we take several thousand quality records from ScienceDirect that contain author keywords, no more than 50% of all author keywords will be found in the text of all abstracts. Therefore, author keywords alone may not be sufficient to expand the search and find relevant publications. It is necessary to extract keywords from the text of the abstracts. Creating a list of keywords for a given topic is necessary not only for collecting relevant literature, but also for text mining.

To confirm the above, let us cite the paper [22], which shows that on average 56.7% of the author's keywords appear in the abstract and body of the article.

The analysis carried out by Babaii, E., & Taase, Y. [23] showed that 46% of the key words specified by the author were present in the text of the peer-reviewed articles.

Another example is the average percentage of keywords in the text of abstracts: Cyberleninka → 42.41%; MathNet → 43.64% [24].

In [25] the study compares a selection of keywords proposed by the author with words from a generalized controlled vocabulary for articles. The effectiveness of the controlled vocabulary for keyword generation is shown. The value of the controlled vocabulary for keyword selection can be demonstrated using bibliometric data from the IEEE Xplore platform of the IEEE Terms field.

---

[6] https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2024/Apr/IRENA_Geopolitics_transition_energy_security_2024.pdf

Evaluations of the effectiveness of keyword extraction from text are always quite subjective, so they should not be overemphasized. For example, a very informative paper [26] compares the extracted keywords based on the similarity between the extracted keywords and the original keywords. However, as noted above, the original keywords themselves appear in the abstracts no more than 50% of the time, even for large lists. In the same paper, there is a very revealing Table 5, which shows that for 1393 datasets across 13 topics/categories, the average time in seconds to identify 10 keywords for each method was: BERT (1007), YAKE (43.28), RAKE (4.02), TEXTRANK (173.1), CHATGPT (2463), HYBRID (1007 + 121 = 1128). That is, for CHATGPT compared to YAKE, the ratio is 2463/43.28=56.9. For large texts, the difference in identifying keywords in a minute or an hour makes a difference, especially if such a procedure is to be applied regularly.

The paper [27] provides a fairly detailed comparison of a number of keyphrase extraction algorithms, including those that use the Multilingual Text-to-Text Transformer (mT5). They conclude that this approach outperforms the Yake algorithm! However, the following should be noted: a) the phrases generated by Yake! contain terms that should be categorized as stop words, e.g. "впервые, общий, особенный, явленный (vpervye, obshhij, osobennyj, yavlenny`j)". (Example 2, Table 3). These terms are not included in the list of Russian stop words in the Yake! package. This list can be easily edited, in fact, this is one of the advantages of the methods of keyword extraction without the use of large language models - they are quite easy to modify/adapt to the needs of the current task. The texts in the examples are very short. The Yake! algorithm focuses on local statistics of keyword candidates, for very short texts with a small variety of terms it is difficult to talk about statistical estimates. When using LLM, the statistical estimates are represented in this external model.

A note on evaluating the significance of terms in the algorithm The Yake! The authors rely on the assumption of Machado et al. [28] who state that "the higher the number of different terms that co-occur with the candidate term on both sides, the less significant the term will be".

An attempt has been made to use an algorithm that utilizes T5 Transformer published on Github[7].

The program works well when the text consists of a few abstracts, but even 300 abstracts are processed in 36 minutes. It may be advisable to first cluster the texts and then extract key phrases for each cluster of texts. For 10000 abstracts the expected time may be more than 2 hours.

Given this brief analysis, the aim of this study was to select appropriate tools for efficient text analysis of abstracts from several tens of thousands of bibliometric records and for a preliminary evaluation of relevant topics related to technologies used in the energy sector.

2. Materials and Methods

In this paper, we used 10 thousand bibliometric records for the year 2024, sorted by relevance and exported from the open abstract database Scilit on the query: "energy AND technology" in [Title, Abstract, Keyword], Content Type: JOURNAL-ARTICLE, English. This query returned 70.6K such records, current as of 03/25/2025. We also applied filters by "Subject" category, which may be most related to Technologies: Power Systems & Electric Vehicles, Energy Systems & Technologies, Electrical Energy Management, and Nuclear Technology & Instrumentation. I.e. an attempt was made to distinguish publications related to such sections as Sustainability Studies, Telecommunications, Economics. The filters were applied due to the fact that the Scilit platform limits the export of bibliometric records to 10 thousand for one request, so it was necessary to select 10 thousand of the most relevant records out of 70.6K records. After applying filters, 14K records remained in the sample and 10K most relevant records were exported from the database.

At the first stage of work on the analysis of technological aspects of energy security, the main attention was paid to the selection of the most reliable methods of working with text fields of bibliometric records. The bibliometric records themselves, collected on a fairly broad query, were

---

[7] https://github.com/Shivanandroy/KeyPhraseTransformer — KeyPhraseTransformer is built on T5 Transformer architecture

considered as the first iteration to identify dominant themes described by keywords (keyphrases, terms in the context of this article).

The records were exported in RIS format (which has the most complete representation of text fields). Files in RIS format were merged and converted to TSV format. Tabular data are the most convenient for implementing samples and are well imported by programs such as VOSviewer, Bibliometrix, and others. Tables need only to be brought to some standardized form, in this case Scopus, by changing the names of column headings and changing the separators between terms. A peculiarity of the bibliometric records of the Scilit platform is a fairly good filling of the abstract field. In our sample of 10,000 records, the abstract field was filled in for 9921 records. The best performance can be found either in paid databases or in publishers' databases, e.g., ScienceDirect from Elsevier or IEEE Xplore from IEEE.

The Scilit platform provides system-generated keywords, not the author's keywords. This column contained 923 empty fields and is not used in this paper. One of the main tasks of this study was to select a suitable method for identifying keywords (phrases) from abstract texts. This task is relevant to the planned general work because the export of records from databases such as OnePetro contains only the abstract field, although individual records contain keyword fields, but these values are not exported when the database is freely available. Another example is RSS records from thematic sites, where the analog of the abstract can be the *description* field, which can be used to determine the subject of publications.

Text fields can contain terms in different spellings, different quotes, dashes, and even hieroglyphs or Cyrillic characters. Therefore, a selection of the most convenient, well-supported program solving this issue was made. The choice was made in favor of *AnyAscii[8]:* Unicode to ASCII transliteration. A variant implemented in the Rust language was used. The changes made by *AnyAscii* (and others utilities used in this work) were checked using the WinMerge[9] program, which allows text files to be compared and comparison reports to be generated. The analysis showed that the vast majority of abstract texts remained unchanged. The changes included replacing the short dash with a hyphen, making quotation marks uniform, and transliteration of Cyrillic and hieroglyphics. That is, while saving the main text, it was easier to eliminate problems that may cause failures in the subsequent analysis of texts, for example, replacement/deletion of characters that are not processed by the program.

Text pre-processing also included removing bracketed text, html tags, TeX formula entries, and removing quotation marks. Some characters were detected only after applying some text processing utilities, e.g. '|' or '\' can affect the execution of regular expressions.

The text cleaning described above is also important for the subsequent dictionary lemmatization, which can be seen as a search and replace spelling of numerous terms. In this case, a list of more than 200 thousand strings collected from Github and other sources and constantly updated with new entries, e.g., deepfakes → deepfake, was used.

Alternatively, one could use well-established packages such as Language Processing spaCy; they are good for frequently used actions, but using simple utilities makes it easier to check each step of text conversion, and dictionary completion is easier to implement. Also, specialized utilities can run faster than more general-purpose packages that perform more checks and additional actions that are not needed in a particular case.

This study presents a comparison of only two methods for identifying keywords/phrases from abstract texts: KeyphraseVectorizers PatternRank [29] and Yake! [30]. Other methods were also tested, such as KeyPhraseTransformer, but they had to be abandoned at this stage, e.g., due to the execution time on texts with 10 thousand abstracts or the complexity of the parameter selection. I.e., the test checks went well, but errors were generated on the sample of texts used. This does not exclude

---

[8] https://github.com/anyascii/anyascii — Unicode to ASCII transliteration

[9] https://github.com/WinMerge/winmerge

the use of other approaches to keyword identification in the future. Other fast methods, such as PKE, were not used because the literature indicates the advantage of PatternRank and Yake!.

In this work the KeyphraseVectorizers_PatternRank variant with the parameter keyphrase_vectorizers → KeyphraseCountVectorizer was used. Pre-prepared texts of abstracts were used.

The convenience of PatternRank is that it is easy to get five keywords for each abstract (or 5 keywords for all abstracts). But there is no explicit way to pass a parameter for the number of keywords using the examples given by the developer. In Yake! there is such a possibility. So, to somehow compare the generated words for all abstracts, we used the condition that 514 PatternRank terms occur 10 or more times, then 514 can be used as a parameter for Yake! and thus make a comparison of the same number of terms.

In the Python example given by the developer, Yake! is used for large text at once, but there is nothing stopping us from using it for each abstract. For this purpose, we used an implementation of Yake! written in Rust.

Hyperfine[10] was used to estimate program execution time.

The keywords obtained for each abstract by both methods were used by VOSviewer [31] to cluster and plot terms based on their co-occurrence. To normalize terms, e.g. singular and plural, the keywords were subjected to dictionary lemmatization.

For clarity, the same VOSviewer was used for each cluster when constructing the term co-occurrence networks. The attached materials contain files in JSON format, allowing to view the obtained graphs on the service app.vosviewer.com.

The resulting keywords can be used to describe the research topics most commonly identified in bibliometric data. The services playground.allenai.org + quillbot.com/summarize were used to demonstrate the power of large linguistic models to generate texts describing research topics using keywords. Manual editing of the results obtained at each stage was carried out. Examples are given for each cluster obtained using the keywords Yake!

Note: Other experiments were also conducted, including aggregation of abstract texts using the GSDMM algorithm and using the KeyPhraseTransformer package, but they did not yield significantly different keyword representations. A detailed comparison of many methods was not within the scope of this study, so in order not to overload the paper they are not given here.

3. Results and Discussion

Before running the PatternRank and Yake! packages, the abstract texts were cleaned up. First, the program *anyascii* was used, which performed transliteration for non-Latin texts (Russian, Chinese), put different types of quotes into the same form, which made them easier to delete later, replaced short dashes with hyphens, etc. HTML markup tags were removed. Removed formula text (usually TeX) and backslashes (to avoid misinterpretation in regular expressions). Note: you can use \ \, but in our case it would be redundant. Explanations and abbreviations in parentheses have been removed. This preprocessing of the text not only normalized the writing of the abstract texts, but also allowed the software packages used to work without generating errors when parsing the texts.

The text itself was not changed by more than 90%. The conversion results were checked using the WinMerge program. Of course, it would be possible to find a suitable Python package that would do most of the text preparation on the fly, but when doing research, it is more important to perform detailed checks of the results obtained than to reduce the time of the job execution.

After compiling the lists of keywords (phrases) obtained with the PatternRank and Yake! packages, their texts were subjected to dictionary lemmatization. Such a reduction of term spelling variants plays an essential role in the construction of a co-occurrence network. Alternatively, the thesaurus_terms files of VOSviewer can be used, but dictionary lemmatization works similarly and is a more universal tool, applicable not only to VOSviewer.

---

[10] https://github.com/sharkdp/hyperfine — A command-line benchmarking tool

3.1. Compiling Keyword Lists from Text of Abstracts Using Yake

Reasons for choosing this method of compiling keyword lists from abstract texts:

- YAKE! is an unsupervised keyword extraction method, meaning it does not require any labeled training data or prior knowledge about the text [32].
- YAKE! does not depend on dictionaries, thesauri, or any external linguistic resources [33].
- The method is lightweight and computationally efficient, making it suitable for real-time or large-scale processing of abstracts. In our case, the "YAKE!" task was completed in less than a minute.
- YAKE! can identify both single-word and multi-word keyphrases.
- YAKE! takes into account the context of words and phrases in the text, which allows you to identify meaningful keywords rather than frequently occurring words that may not reflect the meaning of the text.
- YAKE! is implemented both in Python and Rust.

yake-rust offers better performance, but for sufficiently long texts (e.g. several hundred abstracts) it generates an error, which disappears if you shorten the text length. So, for long texts (tens of thousands of abstracts) you should use the Python version of Yake!

3.2. Compiling Keyword Lists from Text of Abstracts Using KeyphraseVectorizers PatternRank

For the sake of brevity, we will use PatternRank instead of KeyphraseVectorizer PatternRank for the remainder of the text.

Reasons for choosing this method of compiling keyword lists from abstract texts:

- The method relies on a set of the most well-known text processing software packages: texts are annotated with part-of-speech tags using spaCy (star 31.3K on GitHub), and KeyBERT (star 3.8K on GitHub) to extract key phrases.
- According to the authors' assertion PatternRank: "texts are annotated with spaCy part-of-speech tags", "Extract grammatically accurate keyphases based on their part-of-speech tags", "The advantage of using KeyphraseVectorizers in addition to KeyBERT is that it allows users to get grammatically correct keyphrases instead of simple n-grams of pre-defined lengths".
- Tests[11] conducted by the authors of this package show the best results in comparison with KeyBERT, YAKE (the fastest keyphrase extraction), and SingleRank. TF-IDF, YAKE, RAKE - statistical methods, SingleRank, TextRank - based on graphs, and KeyBERT - deep learning method. Note: the package pytextrank[12] was also tested. It has a high rating on GitHub (star 2.2K) [34], which gives good results on texts of several hundred abstracts, but in our case, ~10 thousand abstracts, it worked extremely slowly, the work was interrupted after more than two hours of waiting.
- The KeyphraseVectorizers PatternRank package is arranged so that it picks up keyphrases for each of the abstracts. This is useful for later use in programs such as VOSviewer. The results can be interpreted as index keywords.
- On a computer with 8-core AMD Ryzen 7 5700G; 32 GB RAM, KeyphraseVectorizer PatternRank processing of the collected abstract texts was completed in 20-25 minutes.

3.3. Some Characteristics of Keywords/Keyphrases

514 keyphrases generated by PatternRank occur 10 or more times.

---

[11] https://towardsdatascience.com/unsupervised-keyphrase-extraction-with-patternrank-28ec3ca737f0/

— Unsupervised Keyphrase Extraction with PatternRank

[12] https://github.com/DerwenAI/pytextrank

The 514 keyphrases contain 1201 single terms (for comparison: 1436 single terms from 514 keywords generated by Yake!). There are 276 unique one-syllable terms without lemmatization and 224 unique one-syllable terms with lemmatization.

All records contain 49604 non-unique keywords, 24584 unique keywords after lemmatization compared to 26157 before lemmatization. For comparison with Yake! 49685, 33217 and 31846 — i.e. diversity is higher.

The importance of lemmatization can be illustrated by an example: in the list of keywords, there are 88 results for "renewable energy technologies" and 50 results for "renewable energy technology".

Difference between before and after lemmatization: 26157-24584=1573. It is difficult to manually compile 1573 entries in the thesaurus_terms file before using VOSviewer.

The difference between the unique keywords extracted from the annotation texts is 31846 Yake! vs. 24584 PatternRank, but their occurrence in the annotation texts themselves is high: 452004 Yake! and 259553 in PatternRank. The Yake! terms occur significantly more often in the texts. Therefore, it is more appropriate to use Yake! terms for our task, especially since they are easier/faster to obtain.

We can also make such a comparison - translate multi-word keywords into single-word keywords and count the unique number of such single-word terms for both cases: 5803 (PatternRank) vs. 7360 (Yake!). Their intersection, innerJoin, equals 3423 terms. And in this case, Yake! yields a greater variety for search terms in abstract texts.

3.4. Keyword Co-Occurrence Networks Generated by Yake! Algorithm

The keyword co-occurrence networks were constructed using the VOSviewer program. The network was constructed for all records and keyword clusters were defined. Each cluster was then represented by a separate plot. Next, examples of text generated by playground.allenai.org using the selected keyword combination were reported. The generated text was subjected to abstract summarization using the quillbot.com/summarize service. At each stage, the generated text was manually edited.

The choice of playground.allenai.org is due to the ability to view documents from the training data that have exact text matches with the model response using the OLMoTrace feature. The model did not have direct access to these documents when generating the response. The documents are retrieved after the response is generated. This service allows individuals to experiment with and understand the behavior of various AI models, including OLMo, developed by the Allen Institute for AI (Ai2)[13] .

The choice of quillbot.com/summarize is based on personal testing of existing abstractive summarization services. QuillBot not only summarizes, but also rewords the text, which can be particularly useful. In addition, quillbot.com/summarize displays the keywords used for summarization, which is important for understanding the summarization process. The quillbot.com service offers other text conversion features such as translation, grammar checker and AI humanizer. The latter is useful if you have difficulty expressing a particular part of the text in a more readable way. More information about text summarization can be found in [35,36].

On the need for manual editing: On the one hand, AI-generated texts can be verbose, with repetitions of key phrases (phraseological diversity is more characteristic of human-written text), on the other hand, it allows to reduce authorial bias in text formation. Generalization of the text can lead to the exclusion of semantically significant phrases. That said, editing generalized texts, such as reviews, is a faster process than writing the text itself. Specialized texts are better written independently. The importance of phraseological diversity is discussed in [37].

3.5. Clusters of the Keyword Co-Occurrence Network Identified by Yake-Rust

This section uses the keywords obtained by the yake-rust program for each of the abstracts of 10 thousand bibliometric records (79 of them did not contain annotation texts).

---

[13] https://allenai.org/about

Figure 1 shows 12 clusters of key terms identified by the yake-rust program. The keywords co-occurrence network was built by VOSviewer with the following parameters: total number of key terms → 31832, of which 861 occurred 5 or more times. 500 terms with the highest total link strength were used to build the network. With the minimum number of 20 terms in a cluster, 12 clusters were obtained.
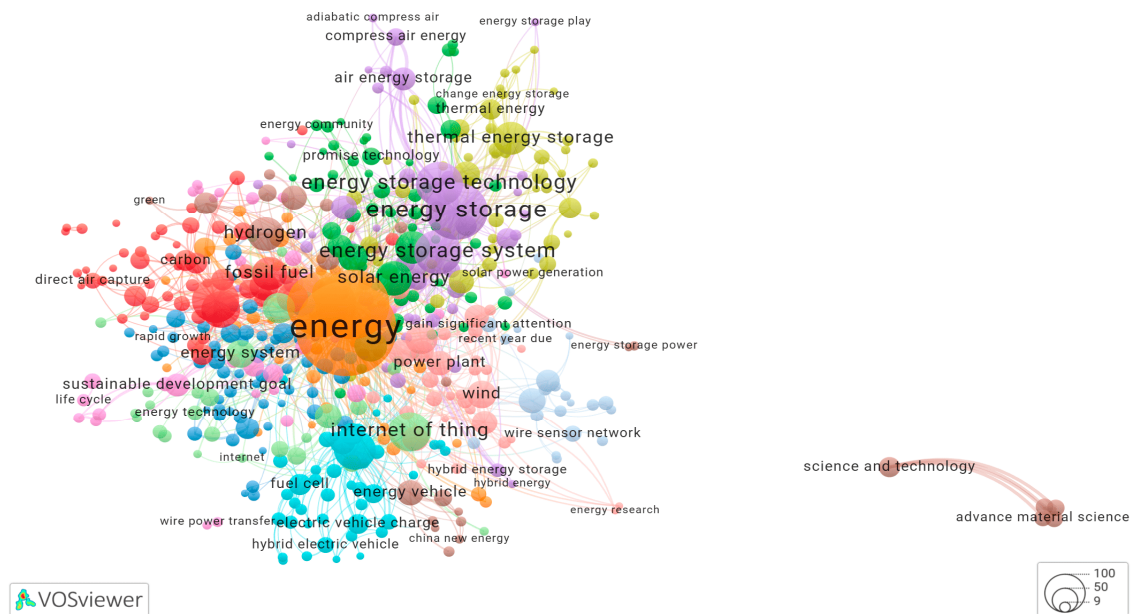

Figure 1. Co-occurrence network of keywords obtained by yake-rust. 12 clusters.

The most common terms in all clusters: energy (821), renewable energy source (443), renewable energy (385), energy storage (310), energy storage system (181), energy storage technology (179), greenhouse gas emission (179), internet of thing (145), electric vehicle (140), climate change (121), solar energy (109), hydrogen (105), fossil fuel (100), thermal energy storage (99), solar (98), energy source (89), recent year (87), energy consumption (84), power (84), carbon emission (79), energy system (70), renewable energy technology (70), system (70), wind (70), reduce carbon emission (69), power system (61), renewable energy system (60), phase change material (59), energy harvest (58), energy vehicle (58).

As can be seen from the list above, the topic of renewable energy sources dominates. A characteristic feature of the extracted keywords is the frequent repetition of related key phrases, e.g. energy storage, energy storage system, energy storage technology, thermal energy storage. In future research, it is better to use two-word terms to describe broad topics and three-word terms for more specific topics. Further research on this statement is recommended. One-word terms are too general, e.g., energy, wind, system.

To explore each cluster's network in more detail, each of the 12 clusters was depicted on its own graph. A query was made to the playground.allenai.org service to identify one of the most frequent topic in each cluster. The response received was summarized both manually and using the features of the quillbot.com/summarize service to make the text more concise. The purpose of these actions was to analyze some possibilities of artificial intelligence services to accelerate the process of text abstracting as one of the key functions of analytical reviews preparation.

3.5.1. Cluster-1 of Network Shown in Figure 1

Figure 2 shows the graph of the first cluster, the entire keyword network. It was created based on the data exported from VOSviewer after creating the network shown in Figure 1.
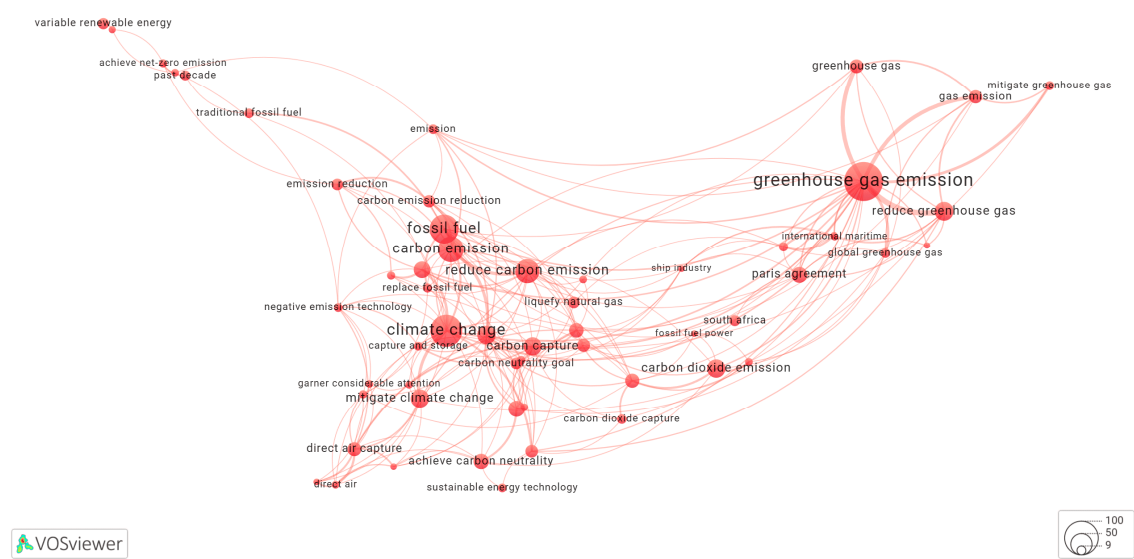
Figure 2. Network of keywords co-occurrence in the first cluster.

The dominant terms in this cluster: greenhouse gas emission, climate change, fossil fuel, carbon emission, reduce carbon emission, reduce greenhouse gas. The negative context of the term *fossil fuels* is emphasized. The theme of maintaining the dominant position of *fossil fuels* in electricity and heat generation is not visible in the figure.

The result of applying the procedures *playground.allenai.org + quillbot.com/summarize + manual editing* to the term *greenhouse gas emission* with 179 occurrences:

"Emission monitoring technology, including satellite imaging, atmospheric monitoring stations, and emission inventories.

Emission reduction technologies involve improving energy efficiency, adopting renewable energy sources, and using carbon capture and storage (CCS) technologies to capture and store CO2 emissions at their source.

Mitigation technologies, including carbon sequestration through reforestation, afforestation, and soil carbon sequestration, and direct air capture systems, aim to remove CO2 from the atmosphere.

Technologies like biofuels, hydrogen fuel cells, and electric vehicles are crucial in reducing GHG emissions in the transportation sector.

Energy efficiency technologies, such as LED lighting, energy management systems, and insulation improvements"

3.5.2. Cluster-2 of Network Shown in Figure 1

Figure 3 shows the graph of the second cluster, the entire keyword network.
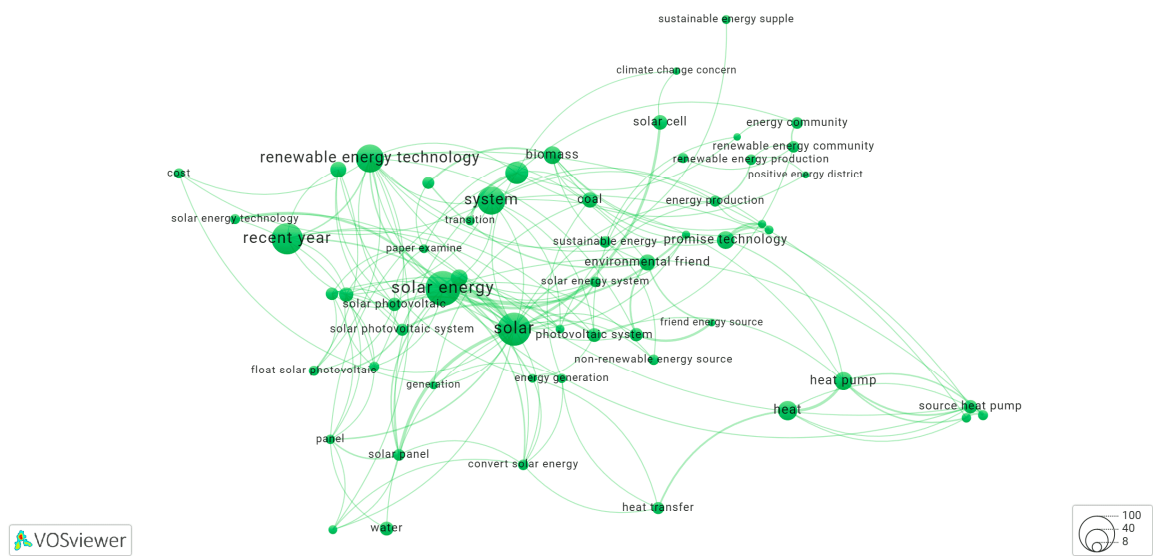
Figure 3. Network of keywords co-occurrence in the second cluster.

The dominant terms in this cluster: solar energy, solar, recent year, renewable energy technology, system, energy transition, heat, heat pump. The topic is relevant, but consideration of environmental issues on panel waste disposal and extraction of materials for their production is not presented in the figure.

The result of applying the procedures *playground.allenai.org + quillbot.com/summarize + manual editing* to the term *solar energy technology* (compiled from *solar energy* → 156; *renewable energy technology* → 98) :

"Photovoltaic cells and solar panels are widespread solar energy technologies, with advances such as biphasic and perovskite cells increasing their efficiency.

Concentrated solar power uses mirrors or lenses to concentrate sunlight, heat fluids, produce steam, and drive turbines to generate electricity, often with thermal energy storage.

Solar thermal collectors heat water or other liquids to provide hot water to residential, commercial, and industrial facilities.

Passive solar design harnesses the sun's energy without mechanical systems, optimizing building orientation, using materials with high thermal mass, and shading to optimize solar utilization.

Solar energy storage, primarily lithium-ion batteries, play a critical role in storing excess energy for use at night, with research exploring more efficient options such as flow batteries and solid-state batteries.

Smart grid integration: smart meters, smart appliances and grid management software enable efficient distribution and utilization of solar energy."

3.5.3. Cluster-3 of Network Shown in Figure 1

Figure 4 shows the graph of the third cluster, the entire keyword network.
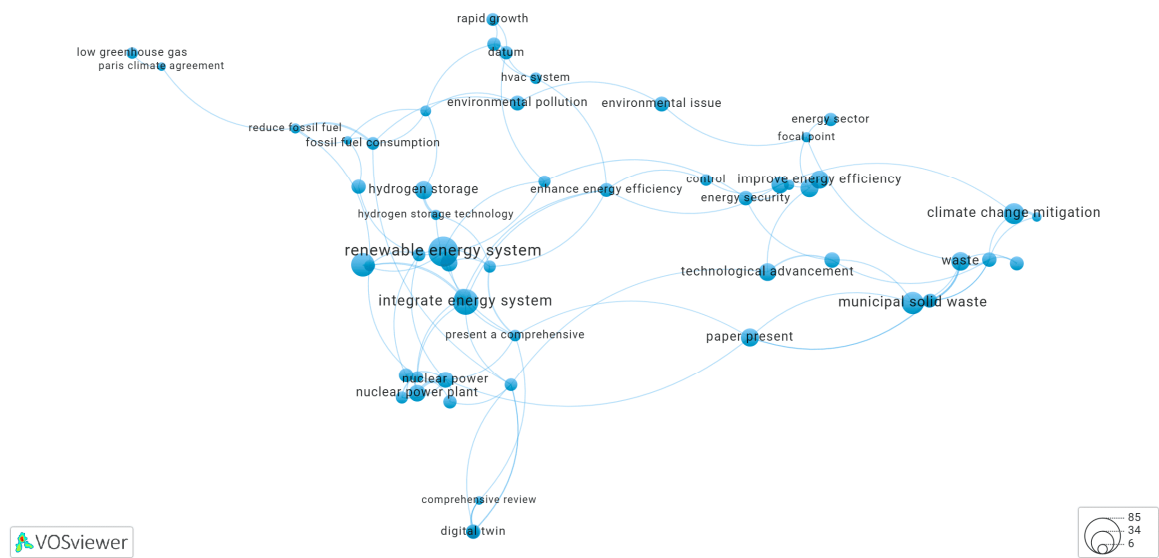
Figure 4. Network of keywords co-occurrence in the third cluster.

The dominant terms in this cluster: renewable energy system, integrate energy system, global warm, municipal solid waste, climate change mitigation, improve energy efficiency, waste, hydrogen storage. The topic of *hydrogen storage* in the context of an *integrated renewable energy system* is really interesting.

The result of applying the procedures playground.allenai.org + quillbot.com/summarize + manual editing to the term integrate energy system technology (compiled from renewable energy system → 60; integrate energy system → 44) :

"The integration of renewables requires equalizing their intermittent nature by forecasting their generation and balancing them with other energy resources.

Energy storage technologies like batteries, pumped hydro, and thermal storage are integrated to manage renewable energy variability, ensure grid stability, and support peak power savings.

Smart grid technology uses digital technology to improve the efficiency and reliability of energy distribution, enabling real-time energy management and demand response programs.

Demand Response and Energy Management Systems enable utilities and customers to adjust energy consumption during peak demand periods, thereby reducing the electric grid's load.

High voltage DC transmission lines are used to transport large amounts of electricity over long distances. They facilitate the connection of remote renewable energy sources to demand centers.

Microgrids are small autonomous networks that utilize distributed energy resources, energy storage, and intelligent control systems to optimize energy consumption and enhance resilience during grid outages."

3.5.4. Cluster-4 of Network Shown in Figure 1

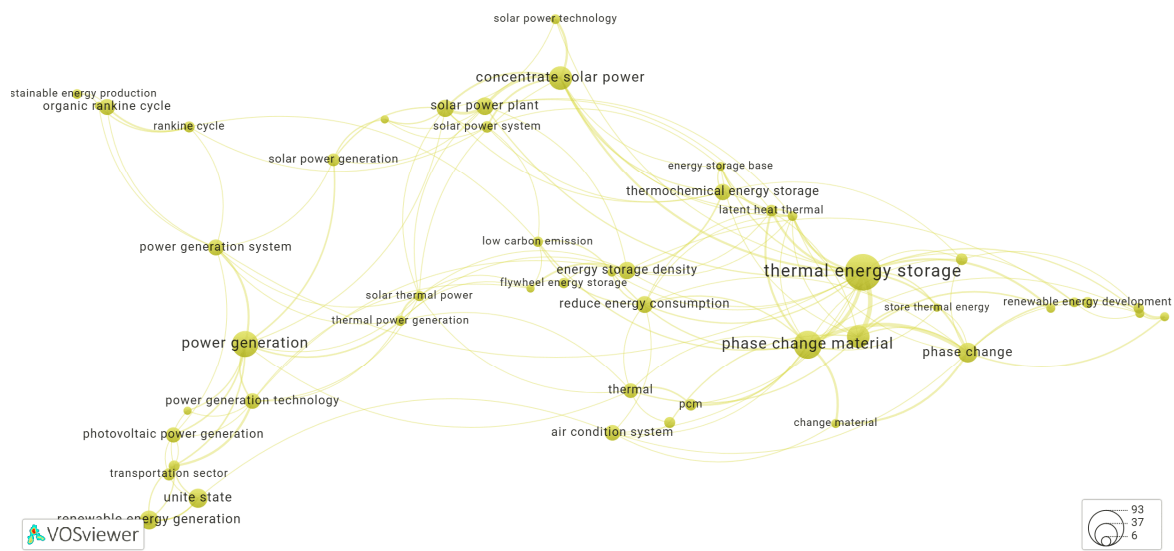Figure 5 shows the graph of the fourth cluster, the entire keyword network.

Figure 5. Network of keywords co-occurrence in the fourth cluster.

The dominant terms in this cluster: thermal energy storage, phase change material, power generation, concentrate solar power, thermal energy, phase change, unite state, renewable energy generation. To conserve energy, temperature equalization is necessary, e.g. for daily variations in sunlight and air heating.

The result of applying the procedures playground.allenai.org + quillbot.com/summarize + manual editing to the term thermal energy storage bade on phase change material (compiled from thermal energy storage → 99; phase change material → 59): "Materials with phase transitions absorb and release heat energy, allowing it to be stored and released at nearly constant temperature.
PCMs can be organic or inorganic; organic PCMs consist of paraffins, fatty acids and polymers, inorganic PCMs include salt hydrates and metals.

PCMs offer advantages in thermal energy storage, including high energy density, temperature control, flexibility and reduced environmental impact. They can maintain constant temperatures during phase change, meet specific application requirements and improve heating and cooling efficiency.

PCMs face problems of thermal conductivity, stability and cost. Solutions include additives with high thermal conductivity or developing systems with good heat exchanger surfaces. Cost remains a barrier to widespread adoption."

3.5.5. Cluster-5 of Network Shown in Figure 1

Figure 6 shows the graph of the fifth cluster, the entire keyword network.
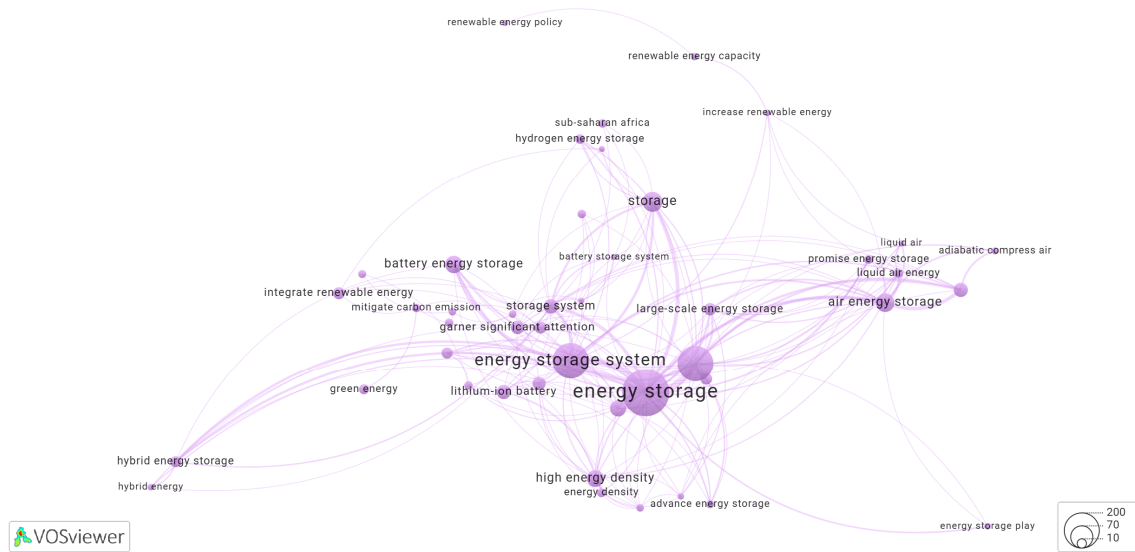
Figure 6. Network of keywords co-occurrence in the fifth cluster.

The dominant terms in this cluster: energy storage, energy storage system, energy storage technology, storage, air energy storage, battery energy storage, high energy density, energy storage device, storage system. The issue of energy storage is very relevant, especially in distributed systems, due to the uneven nature of renewable energy generation. It is useful to consider *energy systems* issues in a broader context [38–40].

The result of applying the procedures playground.allenai.org + quillbot.com/summarize + manual editing to the term energy storage system technology (compiled from energy storage → 310; energy storage system → 181; energy storage technology → 179): "Lithium-ion batteries are widely used in energy storage systems due to their high energy density and long service life. Lead-acid batteries are still used in stationary applications due to their lower cost. Flow batteries, in which the chemical components are dissolved in a liquid separated by a membrane, are suitable for grid-scale energy storage. Sodium sulfur batteries are high-temperature batteries. Solid electrolyte technology has the potential to provide higher energy density and faster charging.

Pumped hydro storage involves pumping excess energy from a lower reservoir to an upper reservoir, and then using turbines to generate electricity when needed.

Compressed air energy storage systems store compressed air in underground caverns or tanks, which is heated when needed and sent to turbines to generate electricity.

Thermal energy storage in the form of molten salt. Ice storage involves the production of ice during off-peak hours in energy-intensive cooling systems. Long-term storage of thermal energy, e.g. in boreholes.

Flywheels store energy. Their fast response time makes them suitable for frequency control and short-term energy storage.

Flywheel energy storage systems are suitable for frequency control and short-term energy storage due to their fast response time.

Hydrogen, produced through water electrolysis, can be utilized to store excess renewable energy, which can be utilized in fuel cells or burned for electricity generation.

Superconducting magnetic energy storage in coils, providing fast response time and power system stability, but their application is limited by high cost and technical problems."

3.5.6. Cluster-6 of Network Shown in Figure 1

Figure 7 shows the graph of the sixth cluster, the entire keyword network.

Figure 7. Network of keywords co-occurrence in the sixth cluster.

The dominant terms in this cluster: electric vehicle, artificial intelligence, energy management, fuel cell, electric vehicle charge, energy management system, alternative energy source, hybrid electric vehicle. On this figure, the term *artificial intelligence* appears in the context of *electric vehicles and energy management*. Nowadays, the topic of *artificial intelligence* is more and more related to data centers and their energy consumption, and this issue requires a separate study.

The result of applying the procedures playground.allenai.org + quillbot.com/summarize + manual editing to the term electric vehicle charge technology (compiled from electric vehicle → 140; electric vehicle charge → 36): DC fast charging technology for rapid charging speed, allowing up to 80% of the battery capacity to be charged in 30 minutes to 1 hour.

Standard connector matching technologies for EV charging including J1772 in North America, Mennekes in Europe, and CHAdeMO and Combined Charging System for fast DC charging.

Wireless charging, a technology enabling electric vehicles to charge without a physical plug, is currently less efficient and slower than wired charging.

Developing a robust charging infrastructure for electric vehicles that includes physical charging stations, grid modernization, and smart energy management systems.

3.5.7. Cluster-7 of Network Shown in Figure 1

Figure 8 shows the graph of the seventh cluster, the entire keyword network.
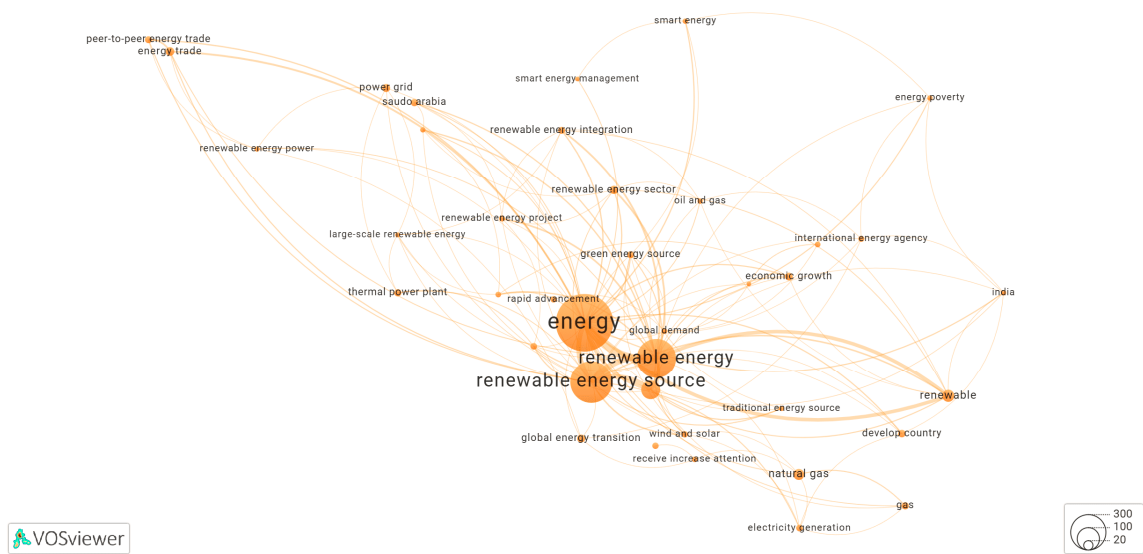
Figure 8. Network of keywords co-occurrence in the seventh cluster.

The dominant terms in this cluster: energy, renewable energy source, renewable energy, energy source, renewable, natural gas, energy trade, renewable energy sector. If labeling this cluster with the key term *energy*, one can see the dominance of topics related to renewable energy sources, publications containing the terms *natural gas*, *oil and gas* are very poorly represented.

The result of applying the procedures *playground.allenai.org + quillbot.com/summarize + manual editing* to the term *renewable energy source technologies* (compiled from *renewable energy source* → 443): Solar energy uses photovoltaic cells to convert sunlight into electricity, while solar thermal systems heat fluids to generate steam for turbines. Concentrated Solar Power (CSP) technology focuses sunlight onto a small area for electricity generation.

Wind turbines convert wind energy into mechanical power, generating electricity. Offshore wind farms utilize stronger, more consistent winds over the sea.

Hydropower utilizes water power to generate electricity using dams, and pumped storage facilities.

Biomass energy is generated by burning organic materials like wood, agricultural waste, or energy crops, using advanced technologies like gasification, pyrolysis, and anaerobic digestion.

Geothermal energy uses the Earth's internal heat to generate electricity by drilling holes in geothermal reservoirs, bringing hot water or steam to the surface, which drives turbines and heat pumps to heat and cool buildings.

Tidal energy used to generate electricity. Ocean thermal energy conversion that uses the temperature difference between warm surface seawater and cold deep seawater to generate electricity.

Hydrogen energy can be produced from renewable sources, such as solar or wind power, through electrolysis, which can be stored and used to generate electricity.

Biofuels like ethanol and biodiesel, derived from biomass or recycled lubricants, can be utilized in vehicles or as fuel additives to petroleum-based fuels.

3.5.8. Cluster-8 of Network Shown in Figure 1

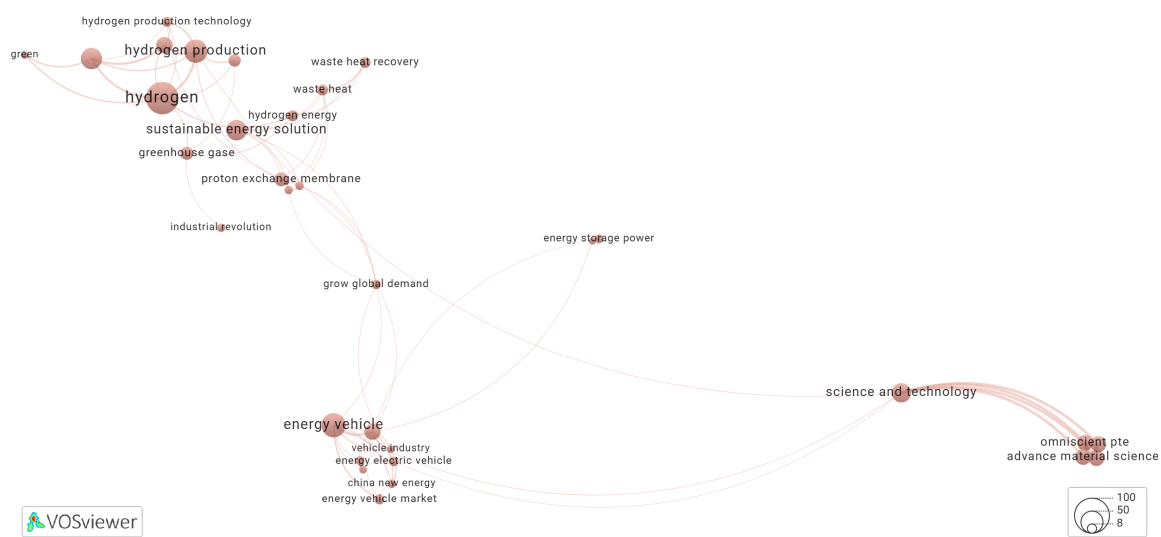Figure 9 shows the graph of the eighth cluster, the entire keyword network.

Figure 9. Network of keywords co-occurrence in the eighth cluster.

The dominant terms in this cluster: hydrogen, energy vehicle, hydrogen production, green hydrogen, sustainable energy solution, science and technology, green hydrogen production, energy vehicle industry. In my opinion, it is advisable to focus on *hydrogen* and *advanced materials science*. Of particular interest is the topic of *Critical Raw Materials for Energy Transition*, which is underrepresented in the bibliometric dataset surveyed. An example of work that exposes this topic is the review [41], which shows that the energy transition has increased the demand, extraction and supply of critical materials. Critical materials concepts should prioritize long-term sustainability over politics. The availability of critical materials poses serious risks to the long-term sustainability of economies.

The result of applying the procedures playground.allenai.org + quillbot.com/summarize + manual editing to the term hydrogen production technology (compiled from hydrogen → 105; hydrogen production technology → 10): Steam reforming of methane is the most common industrial method of hydrogen production, involving a high-temperature reaction with steam to produce hydrogen and carbon monoxide. This endothermic two-step process results in significant carbon emissions as a by-product.

Autothermal reforming, a process involving the self-sustaining oxidation of methane using oxygen and steam, has the potential to enhance energy efficiency.

Gasification of coal at high temperatures using steam and oxygen to produce a syngas mixture containing hydrogen and carbon monoxide which can be further processed by shear conversion to produce additional hydrogen and CO_2.

The process of water electrolysis, utilizing renewable energy sources, produces "green" hydrogen without releasing greenhouse gases, using three main types of electrolyzers: alkaline, proton exchange membrane, and solid oxide.

Microorganisms can produce hydrogen through fermentation or direct biophotolysis in algae, but these methods are in research and development stages and are not commercially viable on a large scale.

Photobiological production using algae and cyanobacteria to produce hydrogen requires further research to overcome challenges such as low efficiency and scalability.

Thermochemical cycles, including solar thermochemical water splitting, utilize heat to convert water into hydrogen and oxygen, a potential future technology that utilizes concentrated solar energy.

3.5.9. Cluster-9 of Network Shown in Figure 1

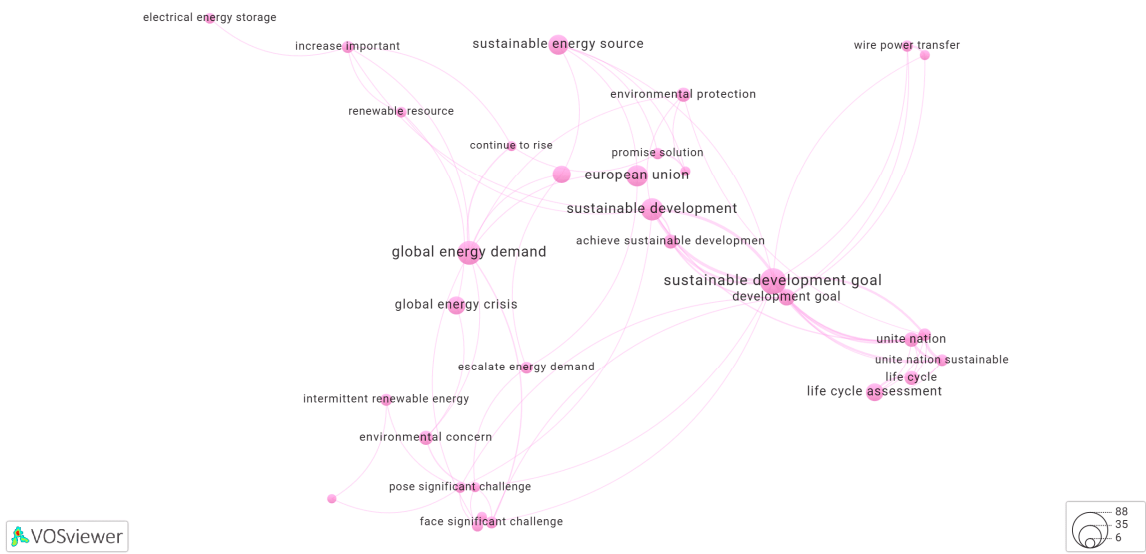Figure 10 shows the graph of the ninth cluster, the entire keyword network.

Figure 10. Network of keywords co-occurrence in the ninth cluster.

The dominant terms in this cluster: sustainable development goal, global energy demand, sustainable development, European union, sustainable energy source, energy demand, global energy crisis, life cycle assessment. This cluster includes publications least related to technology, but rather to political economy.

The dominant terms in this cluster: sustainable development goal → 48; global energy demand → 40; sustainable development → 35; European union → 34. The cluster is not about technology.

3.5.10. Cluster-10 of Network Shown in Figure 1

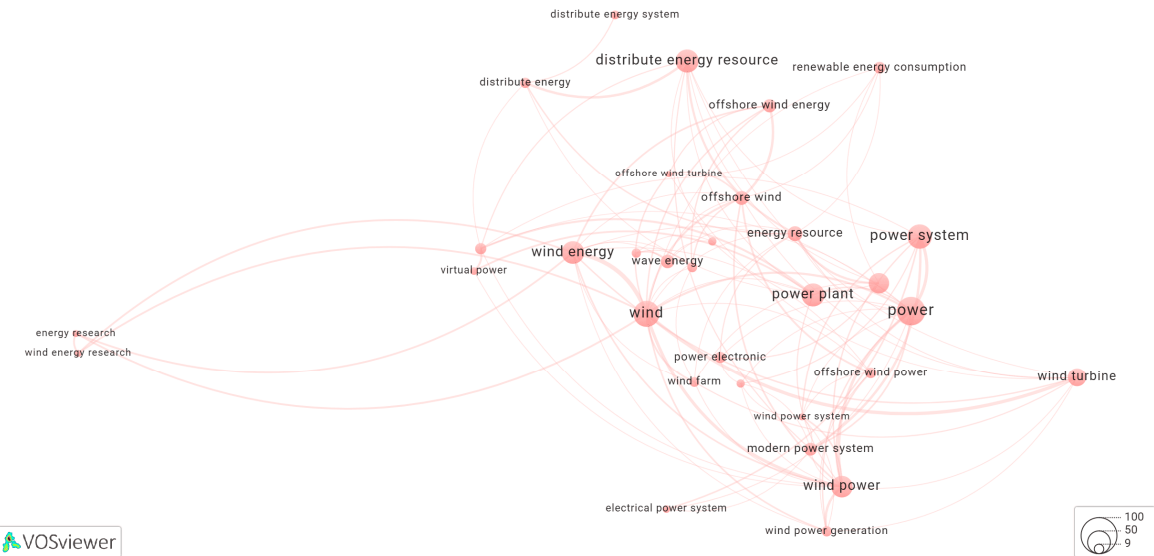Figure 11 shows the graph of the tenth cluster, the entire keyword network.

Figure 11. Network of keywords co-occurrence in the tenth cluster.

The dominant terms in this cluster: power, wind, power system, distribute energy resource, power plant, wind energy, wind power, renewable energy resource. Keywords describe the topic of energy generation, but renewable energy dominates.

The result of applying the procedures *playground.allenai.org + quillbot.com/summarize + manual editing* to the term *wind power system technologies* (compiled from *power → 84; wind → 70; power system → 61*): The two main types of wind turbines are horizontal axis turbines and vertical axis turbines, which capture wind from any direction without orientation.

Direct-drive generators use a gearless mechanism to generate electricity, which reduces wear and tear and improves efficiency, but increases initial cost. Gear-driven generators use a gearbox to increase turbine shaft speed and are more affordable but require more frequent maintenance.

Yaw control systems steer the nacelle into the wind to maximize power, and angle control systems adjust blade angles to maintain optimum angle of attack and control turbine speed during high winds or stops for maintenance.

Steel Towers are a traditional choice that balances cost, weight, and durability. Concrete Towers offer better corrosion resistance and longer life spans but are heavier and may be more expensive to install. Composite Towers, made from fiberglass, are lighter and easier to transport but may cost more.

Power electronics, such as inverters, are used to change the frequency and voltage of the generated electricity to meet the requirements of the power grid. And DC/AC converters.

Tools to measure wind speed, direction and other wind characteristics to determine the best location for installing wind turbines.

3.5.11. Cluster-11 of Network Shown in Figure 1

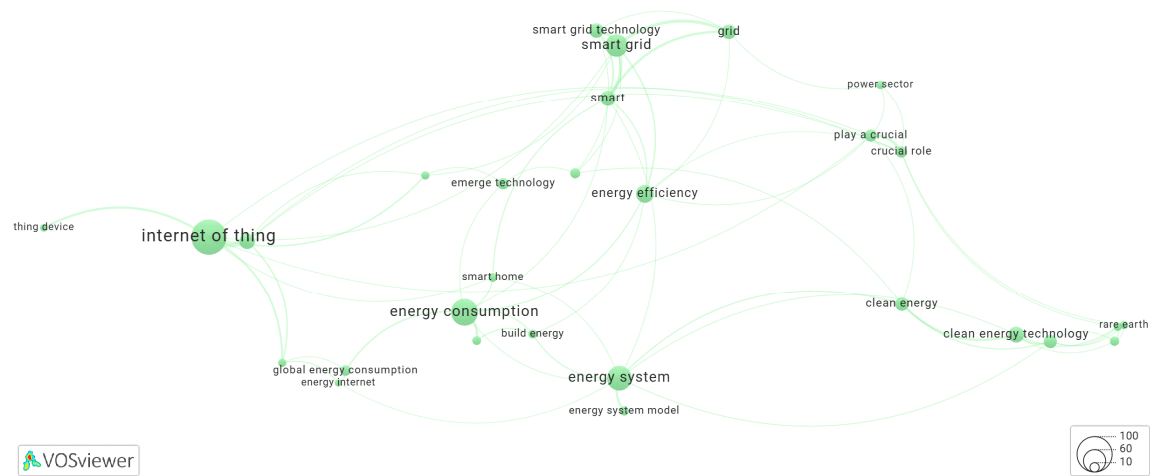Figure 12 shows the graph of the eleventh cluster, the entire keyword network.

Figure 12. Network of keywords co-occurrence in the eleventh cluster.

The dominant terms in this cluster: internet of thing, energy consumption, energy system, smart grid, energy efficiency, clean energy technology, smart, thing technology. The topic can be assigned to *Energy Systems*, including *Internet of Things* and *Smart Grid*. The note for Cluster 5 also applies here.

The result of applying the procedures *playground.allenai.org + quillbot.com/summarize + manual editing* to the term *internet of thing technologies*: Sensors and transducers include temperature, humidity, accelerometers, gyroscopes, and pressure sensors.

Microcontrollers such as Arduino or Raspberry Pi are compact, inexpensive, and have built-in I/O capabilities. Microprocessors provide processing power to analyze data.

Communication protocols like Wi-Fi, Bluetooth, Zigbee, Z-Wave, LoRaWAN, and NB-IoT vary in range, power consumption, bandwidth, and complexity.

Cloud services like AWS IoT, Microsoft Azure IoT Hub, and Google Cloud IoT offer platforms for managing IoT devices and data.

IoT gateways act as intermediaries between IoT devices and the cloud, processing and aggregating sensor data, handling communication protocols, and performing basic analytics before sending data to the cloud, and connecting devices without direct internet access.

Software for device management, data processing and user interaction. Middleware provides the necessary software and services to enable communication and interoperability between disparate systems.

User interfaces and applications that allow users to interact with the IoT system, view or control data.

IoT systems generate large amounts of data that need to be analyzed (including using machine learning algorithms) to identify patterns, predict future events, and make evidence-based decisions. To prevent data leakage and unauthorized access to IoT systems, robust security measures such as encryption, secure boot, firmware updates, and network security protocols are required.

Many IoT devices are battery-powered, so energy efficiency and battery life are critical. Technologies such as energy harvesting, solar charging, and low-power modes are often used to maximize battery life.

Edge computing technologies, which are implemented on IoT devices or gateways, process data closer to the source, reducing latency and bandwidth utilization.

3.5.12. Cluster-12 of Network Shown in Figure 1

Figure 13 shows the graph of the twelfth cluster, the entire keyword network.
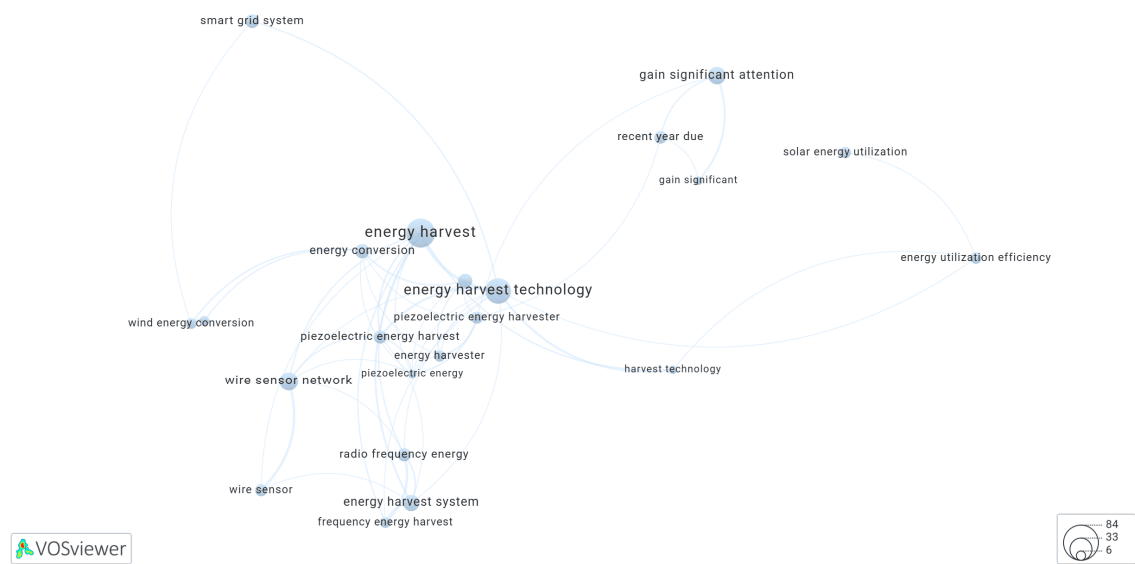

Figure 13. Network of keywords co-occurrence in the twelfth cluster.

The dominant terms in this cluster: energy harvest, energy harvest technology, wire sensor network, gain significant attention, energy harvest system, energy conversion, energy conversion technology, smart grid system, piezoelectric energy harvest, radio frequency energy. The topic of *energy harvesting technology* is interesting in this case in the context of the autonomous operation of sensors and the energy sources for them: *piezoelectric energy harvest and radio frequency energy*.

The result of applying the procedures playground.allenai.org + quillbot.com/summarize + manual editing to the term energy harvest technology (compiled from energy harvest → 58; energy harvest technology → 45) : Thermoelectric energy harvesting utilizes the Seebeck effect to generate electricity through temperature differences, such as between the human body and the surrounding air or hot equipment and a cooler environment.

Piezoelectric energy harvesting is suitable for applications such as footstep generators in pedestrian areas, motion sensors, or capturing the vibration energy of machines.

Photovoltaic energy harvesting is a method that uses flexible solar cells to harvest energy from ambient light for small devices.

Radio frequency energy harvesting uses radio waves emitted by television, radio broadcasting, mobile networks, or Wi-Fi signals for low-power devices.

Harvesting biochemical energy, such as glucose in the human body, to generate electricity for medical implants and wearable devices.

Using microturbines and piezoelectric systems to generate hydroelectric power on a small scale from fluid streams, such as water in pipes or streams.

Energy harvesting by wind microturbines in remote areas or as supplementary energy sources. Triboelectric nanogenerators generate electricity by rubbing two materials against each other; they can harvest energy from human movement, vibrations, and even water waves.

Electromagnetic induction generates electricity by moving a conductor through a magnetic field, and is used in applications involving shaking and rotating machinery.

Note: The use of artificial intelligence to generate text based on key phrases reduces bias in the selection of the main tasks to be solved in the topic under consideration. The generated text contains many repetitions of key terms, so it is appropriate to apply abstractive text summarization to it (reducing it by about half for short texts). However, summarization can lead to the deletion of some key terms, which will distort the informativeness of the text. Therefore, subsequent manual editing cannot be avoided. In my opinion, the creation of efficient text editors that combine all three stages of text processing may be in demand. In general, the use of AI in such a context speeds up text processing, but the need for manual editing remains.

3.6. Clusters of the Keyword Co-Occurrence Network Identified by KeyphraseVectorizers PatternRank

This section uses the keywords obtained by the yake-rust PatternRank for each of the abstracts of 10 thousand bibliometric records.

Figure 14 shows 11 clusters of key terms identified by the PatternRank program. The keywords co-occurrence network was built by VOSviewer with the following parameters: total number of key terms → 24584, of which 1200 occurred 5 or more times. 500 terms with the highest total link strength were used to build the network. With the minimum number of 20 terms in a cluster, 11 clusters were obtained.
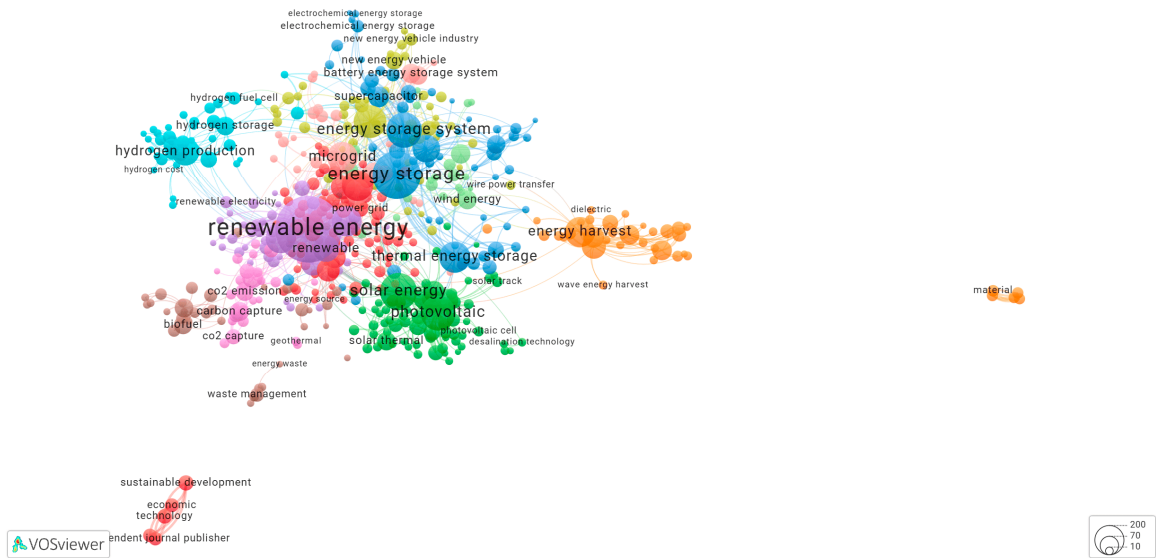
Figure 14. Co-occurrence network of keywords obtained by PatternRank. 11 clusters.

3.6.1. Cluster-3 of Network Shown in Figure 14

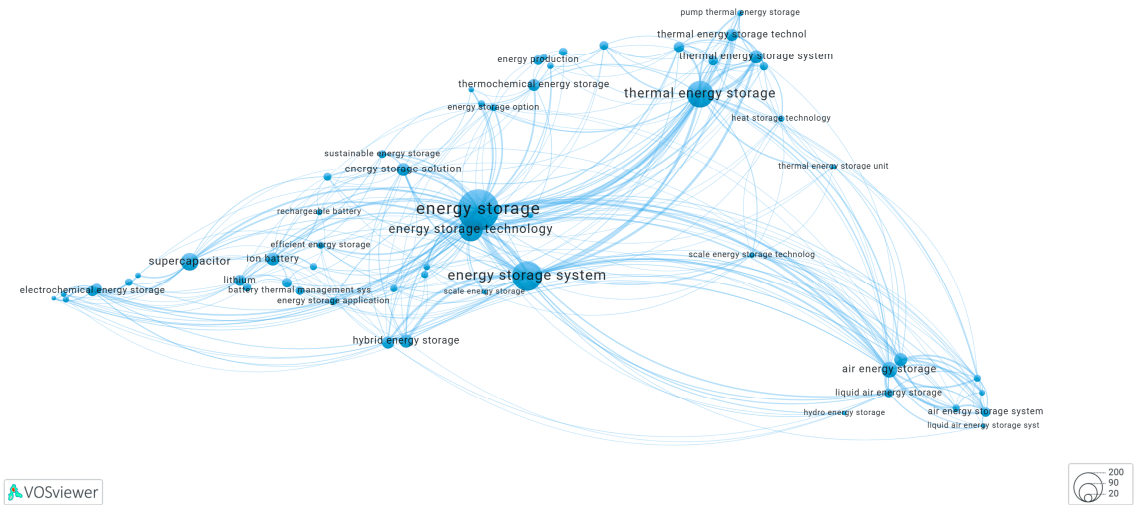Figure 15 shows the graph of the third cluster of Pattern Rank's keyword network.

Figure 15. Network of keywords co-occurrence in the third cluster. Keywords obtained by PatternRank.

The dominant terms in this cluster: energy storage, energy storage system, thermal energy storage, energy storage technology, supercapacitor, air energy storage, thermal energy storage system, energy storage solution, hybrid energy storage, compress air energy storage, ion battery, hybrid energy storage system, thermal energy storage technology, thermochemical energy storage.

Different aspects of the *energy storage* theme are reflected very consistently. The closest is cluster 5 from the previous section (Figure 6).

3.6.2. Cluster-6 of Network Shown in Figure 14

Figure 16 shows the graph of the sixth cluster of Pattern Rank's keyword network.

Figure 16. Network of keywords co-occurrence in the sixth cluster. Keywords obtained by PatternRank.

The dominant terms in this cluster: hydrogen production, hydrogen storage, hydrogen, green hydrogen production, water electrolysis, hydrogen energy, hydrogen economy, hydrogen energy storage, hydrogen technology, hydrogen fuel cell, hydrogen production technology, hydrogen generation.

Different aspects of the *hydrogen production* topic are reflected more consistently compared to cluster 8 of the previous section (Figure 9).

3.6.3. Cluster-9 of Network Shown in Figure 14

Figure 17 shows the graph of the ninth cluster of Pattern Rank's keyword network.

Figure 17. Network of keywords co-occurrence in the ninth cluster. Keywords obtained by PatternRank.

The dominant terms in this cluster: carbon emission, carbon capture, co2 emission, co2 capture, greenhouse gas emission, emission, carbon capture technology, biogas, co2, carbon emission reduction.

Compared with cluster 1 of the previous section (Figure 2), different aspects of the *carbon emission* topic are reflected more consistently.

3.7. A Few Preliminary Thoughts on the Results Obtained in This Paper

The choice of graphs for comparison from the two sections is due to the fact that the issues of *hydrogen*, *thermal energy storage* and *greenhouse gas emissions* presented in the clusters obtained from both of their derivation options are closest to the theme on which this study was conducted (State Assignment No. FMME-2025-0012).

Preliminary, subjective opinion: terms of clusters formed from keywords generated by PatternRank are more similar in topic (more homogeneous due to use of LLM), on the other hand, keywords from Yake! give more links to terms from related topics.

In the bibliometric data for the query "energy AND technology", oil and gas topics are not presented separately. The most common term that can be attributed to this topic is fossil fuels. This can be explained by the fact that the issues of renewable energy technologies have been actively promoted in political decisions for a long time and, as a consequence, actively financed. Therefore, when promoting technological issues of energy security related to oil and gas, it is necessary to take into account such competition from renewable energy sources. For renewable energy systems, energy storage, balancing generation and consumption are among the key challenges. Hydrogen, terminal energy storage and greenhouse gas emissions are examples of overlapping interests that should not be underestimated in oil and gas projects. That is, it is appropriate to consider competitive opportunities for oil and gas projects in the areas of hydrogen, terminal energy storage, and greenhouse gas emissions. Examples: Use of spent wells as a source and storage of thermal energy. Burying greenhouse gases or using them to displace hydrocarbons from reservoirs. Developing technologies to produce hydrogen and soot instead of hydrogen and carbon monoxide. I.e. for energy security issues it is important to assess the competitiveness of the oil and gas sector not only in terms of economic indicators, but also in terms of issues that are declared as benefits in the RES, e.g. carbon is valuable in materials science.

For a more detailed disclosure of the topic "Technological Sovereignty as a Current Energy Security Challenge" on the basis of bibliometric analysis, it is necessary to expand the data sources, e.g. to use OnePetro for oil and gas topics, IEEE Xplore for energy systems issues, Semantic Scholar to evaluate the role of AI in the energy sector, etc.

Both ways of defining keywords suffer from the appearance of terms that are similar in meaning but different in detail, e.g. energy, renewable energy, renewable energy technology. One-word terms greatly expand the search results, while three-word terms narrow them. The choice is determined by the task at hand, and the selection problem can be solved using regular expressions, e.g. by the number of spaces in key terms.

4. Conclusions

The prevalence of topics related to renewable energy in the texts of abstracts of bibliometric data collected on the query "energy AND technology" in the Scilit database for the year 2024 is shown.

Twelve clusters were identified based on keywords from the Yake! program, three of which are closest to the topic for which the study was funded. The objectives represented in these clusters are as follows: *hydrogen*, *heat energy storage* and *greenhouse gas emissions* are reflected by keywords derived from both Yake! and PatternRank. Clusters formed from keywords derived from PatternRank appear more homogeneous in terms of topic, while keywords from Yake! provide more links to terms from related topics.

Artificial intelligence can generate text based on key phrases, reducing task selection bias. However, when summarizing the text, it may remove key terms, which distorts the informative nature of the text. Therefore, manual editing is necessary. The use of AI speeds up text processing, but the need for manual editing remains

The study shows that it is necessary to expand data sources, e.g. use OnePetro for oil and gas topics, IEEE Xplore for energy systems issues, Semantic Scholar to evaluate the role of AI in the energy sector, etc.

The advantage of the Yake! program over PatternRank in terms of both execution time and representation of the obtained keywords in the abstract texts is shown. Both ways of defining keywords suffer from the emergence of terms that are similar in meaning but different in detail, e.g. energy, renewable energy, renewable energy technology.

The feasibility of using *AnyAscii* for text preprocessing is demonstrated.

On the Need for Further Research

The above analysis showed the predominance of the topic of renewable energy in the texts of the studied bibliometric records. At the same time, the issues of technological sovereignty are not addressed. For example, in all the bibliometric records analyzed in this paper, the terms *technological independence* and *technological sovereignty*, the relevance of which is important especially in the context of growing competition in digital technologies and AI, did not appear once.

Here are just a few references that reveal the relevance of the topic of technological sovereignty: "The debate on technological sovereignty revolves around a state-owned digital infrastructure aimed at creating interoperable, non-exclusive services. It includes full national ownership of the entire technology stack, including data centers, and modest attempts to provide technology choices."[14] . Resolution of the Government of the Russian Federation of April 15, 2023, No. 603 "On Approval of Priority Areas of Technological Sovereignty Projects..."[15] .

At the same time, no technology can be competitive without stable and affordable energy, which is the reason for the attention to the topic of Technological Sovereignty as a Current Energy Security Challenge.

Although the topic of technological sovereignty has been widely discussed in recent years, academic publications are still scarce. For example, query: *"technological independence"* in ScienceDirect, Research articles [Title, abstract, keywords]: returned 9 results in all years. And "Technology Sovereignty" — 1 result in all years [42].

A search for "Technological Sovereignty" on onepetro.org returned 2 results [43,44].

References

1. H. Hauser, 'Technology, Sovereignty and Realpolitik', in *Consensus or Conflict?*, H. Wang and A. Michie, Eds., Singapore: Springer Nature Singapore, 2021, pp. 233–242. doi: 10.1007/978-981-16-5391-9_21.
2. P. Grant, 'TECHNOLOGICAL SOVEREIGNTY: FORGOTTEN FACTOR IN THE "HI-TECH" RAZZAMATAZZ', *Prometheus*, vol. 1, no. 2, 1983, doi: 10.1080/08109028308628930.
3. Е. А. Капогузов and А. М. Пахалов, 'Технологический суверенитет: концептуальные подходы и восприятие российскими академическими экспертами: Technological sovereignty: Conceptual approaches and perceptions by the Russian academic experts', *Journal of the New Economic Association*, no. 3(64), pp. 244–250, Oct. 2024, doi: 10.31737/22212264_2024_3_244-250.

---

[14] https://ecdpm.org/work/tech-sovereignty-and-new-eu-foreign-economic-policy

[15] http://static.government.ru/media/acts/files/1202304170025.pdf

4. H. Gu, 'Data, Big Tech, and the New Concept of Sovereignty', *J OF CHIN POLIT SCI*, vol. 29, no. 4, pp. 591–612, Dec. 2024, doi: 10.1007/s11366-023-09855-1.

5. P. Hummel, M. Braun, M. Tretter, and P. Dabrock, 'Data sovereignty: A review', *Big Data & Society*, vol. 8, no. 1, p. 2053951720982012, Jan. 2021, doi: 10.1177/2053951720982012.

6. F. Musiani, '*Infrastructuring* digital sovereignty: a research agenda for an infrastructure-based sociology of digital self-determination practices', *Information, Communication & Society*, vol. 25, no. 6, pp. 785–800, Apr. 2022, doi: 10.1080/1369118X.2022.2049850.

7. L. Van Vliet, J. Herzog-Hawelka, and C. McDonnell, 'Neo-colonialism and leaving fossil fuels underground: a discourse analysis of the potential German-Senegalese gas partnership', *Energy Research & Social Science*, vol. 125, p. 104121, Jul. 2025, doi: 10.1016/j.erss.2025.104121.

8. J. Wang, S. Ghosh, O. A. Olayinka, B. Doğan, M. I. Shah, and K. Zhong, 'Achieving energy security amidst the world uncertainty in newly industrialized economies: The role of technological advancement', *Energy*, vol. 261, p. 125265, Dec. 2022, doi: 10.1016/j.energy.2022.125265.

9. Moscow Aviation Institute (National Research University) and R. S. Golov, 'Energy saving as a factor of energy security: personnel support', *Scient. Work. Fr. Econom. Soc. Rus.*, vol. 249, no. 5, pp. 126–145, 2024, doi: 10.38197/2072-2060-2024-249-5-126-145.

10. R. Prăvălie and G. Bandoc, 'Nuclear energy: Between global electricity demand, worldwide decarbonisation imperativeness, and planetary environmental implications', *Journal of Environmental Management*, vol. 209, pp. 81–92, Mar. 2018, doi: 10.1016/j.jenvman.2017.12.043.

11. M. Asif, B. Solomon, and C. Adulugba, 'Prospects of Nuclear Power in a Sustainable Energy Transition', *Arab J Sci Eng*, vol. 50, no. 5, pp. 3467–3477, Mar. 2025, doi: 10.1007/s13369-024-09557-7.

12. A. Manowska, M. Boros, M. W. Hassan, A. Bluszcz, and K. Tobór-Osadnik, 'A Modern Approach to Securing Critical Infrastructure in Energy Transmission Networks: Integration of Cryptographic Mechanisms and Biometric Data', *Electronics*, vol. 13, no. 14, p. 2849, Jul. 2024, doi: 10.3390/electronics13142849.

13. Yewande Mariam Ogunsuji, Olamide Raimat Amosu, Divya Choubey, Bibitayo Ebunlomo Abikoye, Praveen Kumar, and Stanley Chidozie Umeorah, 'Sourcing renewable energy components: building resilient supply chains, reducing dependence on foreign suppliers, and enhancing energy security', *World J. Adv. Res. Rev.*, vol. 23, no. 2, pp. 251–262, Aug. 2024, doi: 10.30574/wjarr.2024.23.2.2297.

14. S. O. Akpasi, I. M. Smarte Anekwe, E. K. Tetteh, U. O. Amune, S. I. Mustapha, and S. L. Kiambi, 'Hydrogen as a clean energy carrier: advancements, challenges, and its role in a sustainable energy future', *Clean Energy*, vol. 9, no. 1, pp. 52–88, Jan. 2025, doi: 10.1093/ce/zkae112.

15. S. Ganguly and U. Bhan, 'Scope of Geothermal Energy in Indian Energy Security', in *Energy Storage and Conservation*, A. K. Sahu, B. C. Meikap, and V. K. Kudapa, Eds., Singapore: Springer Nature Singapore, 2023, pp. 15–19. doi: 10.1007/978-981-99-2870-5_3.

16. J. Govea, W. Gaibor-Naranjo, and W. Villegas-Ch, 'Transforming Cybersecurity into Critical Energy Infrastructure: A Study on the Effectiveness of Artificial Intelligence', *Systems*, vol. 12, no. 5, p. 165, May 2024, doi: 10.3390/systems12050165.

17. P. D. Szymczak, 'CNPC, Sinopec Drill Ultra Deep in Search of Energy Security', *Journal of Petroleum Technology*, vol. 75, no. 07, pp. 20–25, Jul. 2023, doi: 10.2118/0723-0020-JPT.

18. Y. Huang *et al.*, 'Forecast of Fossil Fuel Demand Based On Low Carbon Emissions from the Perspective of Energy Security', *Chem Technol Fuels Oils*, vol. 58, no. 6, pp. 1075–1082, Jan. 2023, doi: 10.1007/s10553-023-01490-z.

19. K. M.A., N. N.L., and N. A.N., 'The Use of Energy Storage to Improve Controllability and Security of the Belarusian Power System', *Energy Systems Research*, vol. 6, no. 3(23), pp. 28–35, Oct. 2023, doi: 10.25729/esr.2023.03.0003.

20. Z. Yan, Y. Zhang, and J. Yu, 'Allocative approach to multiple energy storage capacity for integrated energy systems based on security region in buildings', *Journal of Energy Storage*, vol. 84, p. 110951, Apr. 2024, doi: 10.1016/j.est.2024.110951.

21. Z. Yang, C. Hao, S. Shao, Z. Chen, and L. Yang, 'Appropriate technology and energy security: From the perspective of biased technological change', *Technological Forecasting and Social Change*, vol. 177, p. 121530, Apr. 2022, doi: 10.1016/j.techfore.2022.121530.

22. W. Lu, Z. Liu, Y. Huang, Y. Bu, X. Li, and Q. Cheng, 'How do authors select keywords? A preliminary study of author keyword selection behavior', *Journal of Informetrics*, vol. 14, no. 4, p. 101066, Nov. 2020, doi: 10.1016/j.joi.2020.101066.

23. Esmat Babaii, Yoones Taase, 'Author-assigned Keywords in Research Articles: Where Do They Come from?' *Iranian Journal of Applied Linguistics (IJAL)*, vol. 16, no. 2, pp. 1-19, 2013 url: https://ijal.khu.ac.ir/article-1-1786-fa.pdf

24. Morozov D.A., Glazkova A.V., Tyutyulnikov M.A., Iomdin B.L. Keyphrase Generation for Abstracts of the Russian-Language Scientific Articles. *NSU Vestnik. Series: Linguistics and Intercultural Communication*. vol. 21, no 1, pp. 54-66, 2023, (In Russ.) doi: 10.25205/1818-7935-2023-21-1-54-66.

25. A. Roy and S. Ghosh, 'Freedom Versus Standard in Article Keyword Generation: An Empirical Study', *Journal of Library Metadata*, vol. 24, no. 4, pp. 291–305, Oct. 2024, doi: 10.1080/19386389.2024.2395198.

26. U. Ahmed, C. Alexopoulos, M. Piangerelli, and A. Polini, 'BRYT: Automated keyword extraction for open datasets', *Intelligent Systems with Applications*, vol. 23, p. 200421, Sep. 2024, doi: 10.1016/j.iswa.2024.200421.

27. A. V. Glazkova, D. A. Morozov, M. S. Vorobeva, and A. Stupnikov, 'Keyphrase generation for the Russian-language scientific texts using mT5', *Model. anal. inf. sist.*, vol. 30, no. 4, pp. 418–428, Dec. 2023, doi: 10.18255/1818-1015-2023-4-418-428.

28. D. Machado, T. Barbosa, S. Pais, B. Martins, and G. Dias, 'Universal Mobile Information Retrieval', in *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, vol. 5615, C. Stephanidis, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 345–354. doi: 10.1007/978-3-642-02710-9_38.

29. T. Schopf, S. Klimek, and F. Matthes, 'PatternRank: Leveraging Pretrained Language Models and Part of Speech for Unsupervised Keyphrase Extraction':, in *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Valletta, Malta: SCITEPRESS - Science and Technology Publications, 2022, pp. 243–248. doi: 10.5220/0011546600003335.

30. R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, and A. Jatowt, 'YAKE! Keyword extraction from single documents using multiple local features', *Information Sciences*, vol. 509, pp. 257–289, Jan. 2020, doi: 10.1016/j.ins.2019.09.013.

31. 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.

32. R. Campos, V. Mangaravite, A. Pasquali, A. M. Jorge, C. Nunes, and A. Jatowt, 'A Text Feature Based Automatic Keyword Extraction Method for Single Documents', in *Advances in Information Retrieval*, vol. 10772, G. Pasi, B. Piwowarski, L. Azzopardi, and A. Hanbury, Eds., Cham: Springer International Publishing, 2018, pp. 684–691. doi: 10.1007/978-3-319-76941-7_63.

33. R. Campos, V. Mangaravite, A. Pasquali, A. M. Jorge, C. Nunes, and A. Jatowt, 'YAKE! Collection-Independent Automatic Keyword Extractor', in *Advances in Information Retrieval*, vol. 10772, G. Pasi, B. Piwowarski, L. Azzopardi, and A. Hanbury, Eds., Cham: Springer International Publishing, 2018, pp. 806–810. doi: 10.1007/978-3-319-76941-7_80.

34. P. Nathan, *DerwenAI/pytextrank: v3.1.1 release on PyPi*. (Mar. 25, 2021). Zenodo. doi: 10.5281/ZENODO.4637885.

35. M. Luo, B. Xue, and B. Niu, 'A comprehensive survey for automatic text summarization: Techniques, approaches and perspectives', *Neurocomputing*, vol. 603, p. 128280, Oct. 2024, doi: 10.1016/j.neucom.2024.128280.

36. A. P. Widyassari *et al.*, 'Review of automatic text summarization techniques & methods', *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1029–1046, Apr. 2022, doi: 10.1016/j.jksuci.2020.05.006.

37. R. Hu, J. Wu, and X. Lu, 'Word-Combination-Based Measures of Phraseological Diversity, Sophistication, and Complexity and Their Relationship to Second Language Chinese Proficiency and Writing Quality', *Language Learning*, vol. 72, no. 4, pp. 1128–1169, Dec. 2022, doi: 10.1111/lang.12511.

38. Y. D. Severina, V. A. Shakirov, and L. N. Takaishvili, 'Modeling the Development of Energy Systems of Remote Areas in the Context of the Energy Transition', *Energy Systems Research*, vol. 7, no. 4(28), pp. 38–48, Dec. 2024, doi: 10.25729/esr.2024.04.0005.

39. D. I.S., 'Methods for Analyzing and Increasing Cyber Resilience of Smart Energy System Facilities', *Energy Systems Research*, vol. 6, no. 3(23), pp. 75–81, Oct. 2023, doi: 10.25729/esr.2023.03.0007.

40. Chigarev, B.N., 'A Brief Analysis of Topics of the IEEE Conference on Energy Internet and Energy System Integration in 2017–2021', *Energy Systems Research*, vol. 6, no. 3(23), pp. 36–49, Oct. 2023, doi: 10.25729/esr.2023.03.0004.

41. V. Lundaev, A. A. Solomon, T. Le, A. Lohrmann, and C. Breyer, 'Review of critical materials for the energy transition, an analysis of global resources and production databases and the state of material circularity', *Minerals Engineering*, vol. 203, p. 108282, Nov. 2023, doi: 10.1016/j.mineng.2023.108282.

42. J. Edler, K. Blind, H. Kroll, and T. Schubert, 'Technology sovereignty as an emerging frame for innovation policy. Defining rationales, ends and means', *Research Policy*, vol. 52, no. 6, p. 104765, Jul. 2023, doi: 10.1016/j.respol.2023.104765.

43. Technology Development Center for Fuel and Energy Complex under the Ministry of Energy of the Russian Federation, RF, Moscow, The Diplomatic Academy of the Ministry of Foreign Affairs of Russia, RF, Moscow, O. V. Zhdaneev, K. N. Frolov, and Technology Development Center for Fuel and Energy Complex under the Ministry of Energy of the Russian Federation, RF, Moscow, 'Scientific and technological priorities of the fuel and energy complex of the Russian Federation until 2050', *OIJ*, no. 10, pp. 6–13, 2023, doi: 10.24887/0028-2448-2023-10-6-13.

44. State University of Management RF, Moscow, V. Ya. Afanasyev, D. A. Suslov, State University of Management RF, Moscow, S. V. Chuev, and State University of Management RF, Moscow, 'Soviet experience in the development of economic and industrial potential under sanctions', *OIJ*, no. 12, pp. 156–160, 2022, doi: 10.24887/0028-2448-2022-12-156-160.