

## **Выявление актуальных тем исследований по вопросам энергетической политики и безопасности в публикациях издательств Elsevier и MDPI за 2021–2023 гг.**

### **Часть 2. Кластеризация публикаций. Автоматическое экстрактивное (извлекающее) реферирование**

**Б.Н. Чигарев**

Институт проблем нефти и газа РАН, г. Москва, Россия  
E-mail: [bchigarev@ipng.ru](mailto:bchigarev@ipng.ru)

**Аннотация.** Данная статья посвящена выявлению актуальных тем исследований в публикациях за 2021–2023 гг. издательств Elsevier и MDPI по вопросам энергетической политики и безопасности на основе кластеризации публикаций.

Показано, что именные группы, извлеченные из заголовков и аннотаций, могут быть использованы для кластеризации публикаций, позволяющей определить актуальные темы исследований. Извлечение именных групп осуществлялось с использованием программного пакета Sifaka. Строилась таблица из 5000 наиболее часто встречаемых именных групп, термины которых оставались в текстах заголовков и аннотаций, остальные термины удалялись. Подготовленные таким образом тексты использовались в процессе кластеризации библиометрических записей с использованием демоверсии программы Carrot2 и алгоритма Lingo3G. Тексты аннотаций для каждого кластера подвергались автоматическому экстрактивному реферированию с использованием программы sumy с алгоритмом Text-rank и списка стоп-слов stopwords-en-ISO.txt. Продемонстрирована возможность использования технологии экстрактивного реферирования для описания доминирующей тематики отдельных кластеров публикаций.

Приведен список доминирующих тематик в публикациях по вопросам энергетической политики и безопасности в публикациях издательств Elsevier и MDPI за 2021–2023 гг.

**Ключевые слова:** энергетическая политика, темы исследований, библиометрический анализ, Sifaka, Carrot2, кластеризация публикаций, экстрактивное реферирование.

**Для цитирования:** Чигарев Б.Н. Выявление актуальных тем исследований по вопросам энергетической политики и безопасности в публикациях издательств Elsevier и MDPI за 2021–2023 гг. Часть 2. Кластеризация публикаций. Автоматическое экстрактивное (извлекающее) реферирование // Актуальные проблемы нефти и газа. 2023. Вып. 2(41). С. 84–107. <https://doi.org/10.29222/ipng.2078-5712.2023-41.art6>

#### **Введение**

Часть 1 данной статьи была посвящена выявлению актуальных тем исследований по вопросам энергетической политики и безопасности в публикациях издательств Elsevier и MDPI за 2021–2023 гг., основывалась на использовании методов кластеризации терминов и

реферирования тем с помощью ChatGPT. Цели публикации части 2:

– Выявление актуальных тем исследований в публикациях за 2021–2023 гг. издательств Elsevier и MDPI по вопросам энергетической политики и безопасности на основе кластеризации публикаций.

– Демонстрация возможности использования именных групп извлеченных из текстов заголовков и аннотаций для кластеризации публикаций как основы выявления актуальных тем исследований.

– Демонстрация возможности использования технологии экстрактивного реферирования<sup>1</sup> для описания доминирующих тем кластера публикаций.

### **Актуальность кластеризации публикаций**

Кластеризация публикаций – это процесс объединения похожих публикаций в группы на основе их содержания, методологии или других характеристик. Кластеризация публикаций на основе их содержания позволяет исследователям увидеть, какие темы изучаются чаще всего, а какие области остаются без внимания.

Кластеризация публикаций может помочь исследователям более эффективно выявлять актуальные исследования. Группируя публикации на основе их сходства, исследователи могут быстро определить наиболее актуальные исследования и не тратить время на просеивание неактуальных.

Кластеризация публикаций может способствовать сотрудничеству и обмену знаниями между исследователями. Выявляя других исследователей, работающих над схожими темами, исследователи могут более эффективно сотрудничать и обмениваться знаниями и ресурсами.

В целом, кластеризация публикаций является важным инструментом

для исследователей, позволяющим понять тематику большого числа публикаций, определить актуальные исследования, а также облегчить сотрудничество и обмен знаниями. Примерами работ, раскрывающими значимость методов кластеризации документов, могут служить публикации [1–3].

### **Материалы и методы**

*Библиометрические данные* по вопросам энергетической политики и безопасности за 2021–2023 гг. были собраны из двух источников: ScienceDirect издательства Elsevier и сайта mdpi.com издательства MDPI. Данные актуальны на 14.03.2023. Т. е. использовались те же самые исходные данные, что и в первой части публикации.

Итоговое число записей, которое использовалось в работе, составило 2612 строки.

*Методика предварительной обработки текстов* заголовков и аннотаций включала следующие стадии:

- объединение текстов заголовков и аннотаций в один абзац;
- удаление текстов, заключенных в скобки, например, Internet of Things (IoT) – Internet of Things, methane (CH<sub>4</sub>) – methane;
- замена наиболее часто встречаемых терминов во множественном числе на единственное число;
- удаление кавычек;
- замена или удаление символов, не соответствующих кодировке utf-8.

<sup>1</sup> [https://ru.wikipedia.org/wiki/Автоматическое\\_реферирование](https://ru.wikipedia.org/wiki/Автоматическое_реферирование) – Автоматическое реферирование, аннотирование или суммаризация – создание краткой версии текстового документа с помощью компьютерной программы.

Каждый абзац текста записывался как отдельный файл и индексировался программой `sifakaBuildIndex.jar`<sup>2</sup>. Индекс включал в себя и именные группы (noun phrases). Используя `sifakaTextMiner.jar` из программного пакета Sifaka, строилась таблица из 5000 наиболее часто встречаемых именных групп.

В файле, содержащем тексты заголовков и аннотаций и прошедшим предварительную обработку, оставались только термины из таблицы 5000 наиболее часто встречаемых именных групп. Преимущество такого подхода, по сравнению с исключением из текста стоп-слов, заключается в том, что выбор списка стоп-слов достаточно субъективен, а в данном случае в тексте остаются лишь термины, входящие в список наиболее часто встречаемых именных групп, а другие термины автоматически исключаются.

Полученные вышеизложенным методом 2612 текстовых записей добавлялись как отдельная колонка к общей таблице библиометрических записей. Кластеризация библиометрических записей проводилась по текстам сформированного автором столбца. Предложенный подход позволяет сохранить значения полей, по которым не проводилась кластеризация, что позволяет, например, использовать поле DOI для определения статей, вошедших в тот или иной кластер.

Учитывая, что данное исследование проводилось как проверка концепции (proof of concept) – т. е. демонстрации того, что предлагаемый автором подход для выявления актуальных тем исследований из библиометрических данных осуществим,

для кластеризации записей использовалась демоверсия системы Carrot2 с алгоритмом Lingo3G<sup>3</sup>, находящаяся в открытом доступе, простая в работе, позволяющая экспортировать полученные результаты и предоставляющая высококачественное графическое представление данных. Как демоверсия, данная система имеет ограничение на объем обрабатываемых данных. Используемая предварительная обработка текстов позволила уложиться в существующие ограничения.

Экспортируемые из Carrot2 результаты кластеризации записей использовались для дальнейшего анализа. Тексты аннотаций для каждого кластера подвергались автоматическому экстрактивному реферированию. Учитывая ограничения на размер статьи, в результатах исследования представлены лишь три примера полученных текстов рефератов, раскрывающих их тематику для кластеров с наибольшим числом публикаций.

## Результаты и обсуждения

### Результаты извлечения именных групп с использованием Sifaka

Индексация 2612 текстов записей заголовков и аннотаций производилась при следующих параметрах: удаление стоп-слов и стемминг Кроветца не использовались, так как данные процедуры препятствуют выделению именных групп в тексте. В индекс включались именные группы (Noun Phrases) и биграммы (Bigrams). В дальнейшей работе использовались 5000 наиболее часто встречаемых именных групп, примеры которых представлены в табл. 1.

<sup>2</sup> <https://www.lemurproject.org/sifaka.php> – Sifaka – приложение для анализа текста, построенное на базе поисковой системы с открытым исходным кодом.

<sup>3</sup> <https://search.carrotsearch.com/#/workbench> – Приложение для кластеризации с использованием алгоритма Lingo3G.

Таблица 1

**40 наиболее часто встречаемых терминов (именных групп)**

Термин	N <sub>d</sub>	N <sub>t</sub>	Термин	N <sub>d</sub>	N <sub>t</sub>
energy policy	793	1000	energy technologies	78	108
energy security	565	824	power plants	73	102
energy consumption	267	479	energy development	71	102
energy sources	331	461	electricity consumption	53	100
energy system	268	394	electricity generation	81	98
energy transition	234	386	energy market	66	94
energy efficiency	196	331	wind energy	48	90
climate change	218	260	energy mix	75	90
co <sub>2</sub> emissions	132	244	power generation	68	89
energy sector	162	225	carbon neutrality	63	88
carbon emissions	121	190	energy generation	65	81
case study	156	170	energy intensity	40	78
energy resources	117	161	eu countries	46	77
energy poverty	59	157	electricity demand	55	73
energy demand	118	157	ghg emissions	37	72
greenhouse gas emissions	127	135	electricity price	45	72
energy use	81	125	development goals	69	72
energy production	95	116	hydrogen production	35	72
power system	77	111	electricity market	46	65
energy supply	84	109	policy implications	62	64

Сокращения: N<sub>d</sub> – число записей, в которых встречаются указанные термины, N<sub>t</sub> – общая встречаемость термина.

Из табл. 1 видно, что представленные термины хорошо описывают тематику энергетической политики и безопасности и демонстрируют понятную семантику словосочетаний.

В части 1 статьи показано, что число одинаковых терминов в списках из 5000 терминов, выбранных по используемой в этой части исследования процедуре, и авторскими ключевыми словами составляет 755. Сравнение авторских ключевых слов и полученных терминов в данной части исследований составило 842 из общего числа 5000. Численно результаты различаются мало. Сравнение 5000 терминов, полученных в первой и второй частях исследований, дает 938 общих значений, т. е. даже выше, чем

в сравнении с авторскими ключевыми словами.

#### **Результаты подготовки столбца записей, основанной на выявленных именных группах в текстах заголовков и аннотаций**

Основной целью данного раздела являлось формирование столбца записей, по которым в дальнейшем осуществлялась их кластеризация.

Процедура формирования такого столбца заключалась в том, что в абзацах текстов, состоящих из заголовков и аннотаций каждой публикации, оставлялись только термины, включенные в список 5000 наиболее часто встречаемых именных групп.

Примеры таких записей представлены в табл. 2.

Таблица 2

**Примеры абзацев текстов и строк, составленных из терминов именных групп**

Заголовок статьи	Строка, содержащая термины из 5000 именных групп
"My Electricity" Program Effectiveness Supporting the Development of PV Installation in Poland	electricity program effectiveness development pv installation poland studies show legitimacy renewable energy mechanisms problems allocation funds paper situation funds photovoltaic micro-installations poland electricity program article results analyses inequalities use funds electricity program inequalities intensity support terms maximization electricity production electricity program pv micro-installations average power pv capacity financial resources electricity program funds allocation surpluses average income small households potential project selection criteria analysis criteria income households
Can Michigan's Upper Peninsula Achieve Justice in Transitioning to 100% Renewable Electricity? Survey of Public Perceptions in Sociotechnical Change	upper justice renewable electricity survey public perceptions change cost energy western upper northern state united states situation wup residents electricity bills interest renewable electricity region questions factors wup residents transition re support re transition wup research factors residents re transition wup research public perceptions residents' survey regression results show residents' initiative consumption probability support wind energy development re transition support high wup trend project preferences results research roadmap future community-engaged planning re region
The Impact of the Government Policy on the Energy Efficient Gap: The Evidence from Ukraine	impact government policy energy gap evidence paper check impact investment energy efficiency gap findings bibliometric analysis growth research interests core energy efficiency gap central quality institutions increase green investments energy sector relationships investment synergy effects energy efficiency gaps national economy analysis times series world data bank eurostat period methods unit root test data test vec-modelling cointegration analysis findings energy efficiency gaps year increase green energy investments year stability public perception increase public perception points stability points recovery speed energy sector policy energy efficiency gaps government level public perception stability
Addressing the Urban Heat Islands Effect: A Cross-Country Assessment of the Role of Green Infrastructure	urban heat islands effect cross-country assessment green infrastructure urban heat islands effect urban areas temperature increases local heat waves intensity reductions levels life quality paper reviews green infrastructure intensity impact human comfort comprehensive review literature paper reports analysis case studies set cities 13 countries geographical regions climate zones results common green infrastructure urban areas conditions impacts addition study impacts urban challenges city implications paper complex heat waves climate change urban green resources benefits impacts paper need city planners effects new building projects
Decomposition Analysis of the Evolution of the Local Energy System as a Tool to Assess the Effect of Local Actions: Methodology and Example of Malmö, Sweden	decomposition analysis evolution local energy system tool effect local actions methodology paper use decomposition analysis effect local energy-related actions climate change mitigation policy evaluation planning local level assessment impact local actions challenge perspective change factors local energy-related greenhouse gas emissions local authorities methodology decomposition model changes local energy system causes/effects methodology effect local actions case study
Modelling Bidding Behaviour on German Photovoltaic Auctions	modelling bidding behaviour photovoltaic auctions article renewable energy support allocation types auctions framework auction theory rules photovoltaic feed-in auctions work bidding strategies cost electricity methodology set rules price nash-equilibrium bidding strategy function monotone functions monotone functions framework auction theory modelling results pv support allocation auction system cost pricing rule participants valuation auction nash-equilibrium cost perspective application pricing rule policy decision

Подчеркнем, что в правой колонке представлен не полный текст, а перечень слов, по которым осуществлялось установление близости текстов в процессе их кластеризации с использованием демоверсии программы Carrot2.

В Прил. 1 дан авторский перевод текстов из этой таблицы, не претендующий на профессиональный перевод.

В табл. 2 подобраны примеры, содержащие небольшое число терминов в правой колонке. Для того, чтобы грубо оценить объем текстов 2612 записей заголовков, аннотаций и терминов из именных групп, приведем их размер в килобайтах: 274 КБ, 3763 КБ и 1717 КБ, соответственно. Таким образом, применяемый автором метод позволил сократить объем текста в поле, по которому проводилась кластеризация, в 2,35 раза по сравнению с суммой полей заголовков и аннотаций. Это не только позволило проводить кластеризацию по более релевантным терминам, но и использовать демоверсию программы Carrot2 и алгоритма Lingo3G, имеющие ограничения на объем текста, по которому проводится кластеризация.

#### **Результаты кластеризации 2612 библиометрических записей с использованием алгоритма Lingo3G**

Преимущество кластеризации библиометрических записей (фактически публикаций) для выявления актуальных задач или тем в рамках заданного направления исследований, по сравнению с кластеризацией ключевых терминов, заключается в том, что конечным источником информации для составления,

например, аналитического обзора являются именно публикации, а не ключевые термины. По терминам, описывающим актуальную задачу или тему, все равно необходимо собирать публикации для составления обзора.

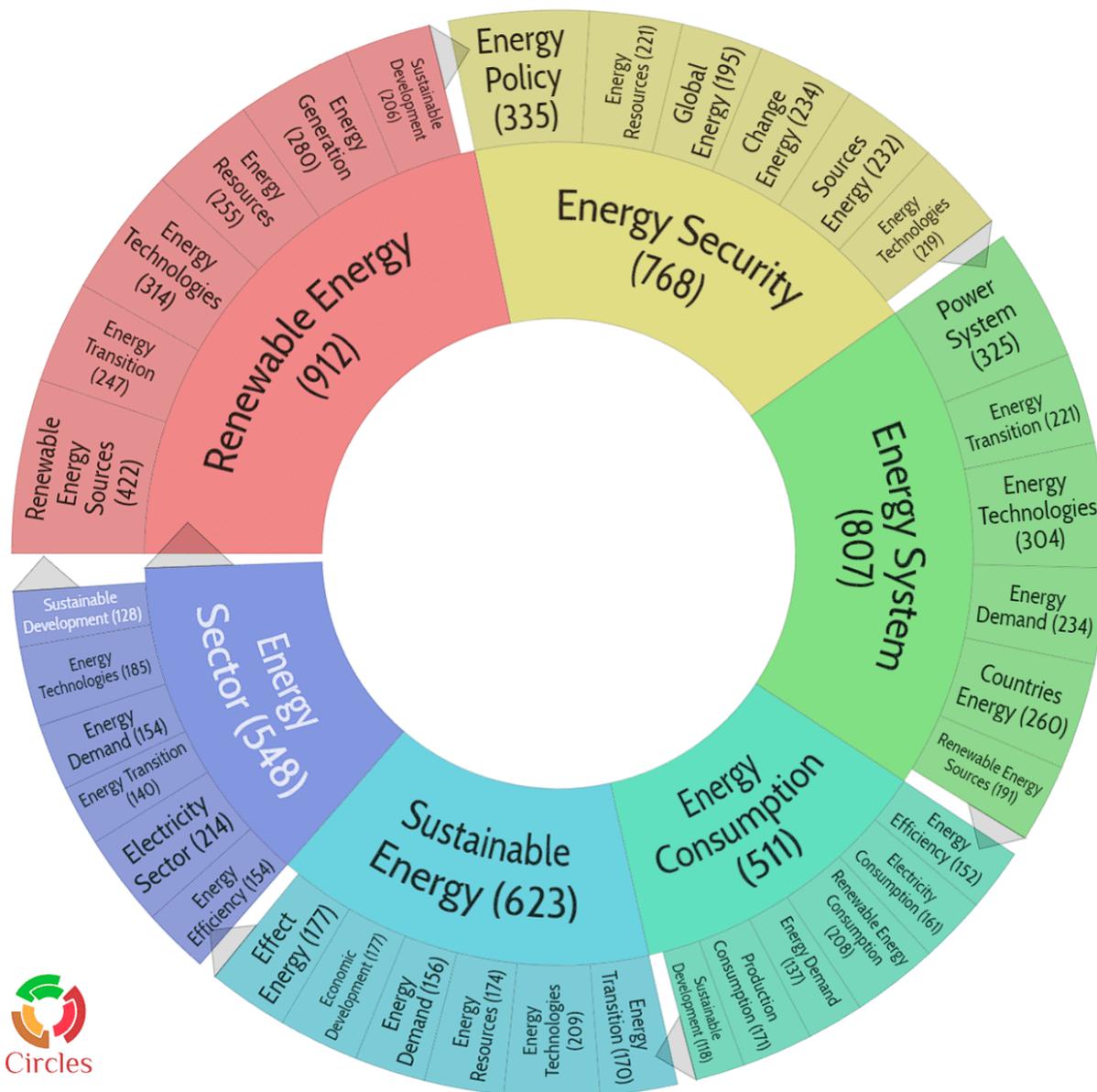
Саму кластеризацию записей можно проводить, используя различные поля или их комбинацию. Например, поля авторов публикации или названий журналов, в которых статья опубликована, могут быть использованы совместно с текстами полей заглавий статей и аннотаций. Библиометрическое сопряжение (Bibliometric coupling) можно рассматривать как частный случай кластеризации публикаций, мерой близости которых выступает пристатейный список литературы.

Общая картина кластеризации библиометрических записей по мере близости терминов, составленных из именных групп, выявленных из полей заголовков и аннотаций, представлена на рис. 1.

Параметры программы, отличные от параметров по умолчанию: minClusterSize – 0,2, minLabelWords – 2:

minClusterSize – определяет минимально допустимый размер кластера по отношению к размеру родительского кластера. Например, значение 0,2 означает, что кластеры не должны содержать менее 20% документов родительского кластера (всех документов в случае кластеров верхнего уровня);

minLabelWords – определяет минимальную длину метки в словах. Метки, состоящие из меньшего количества слов, не будут генерироваться.



**Рис. 1.** Тематика кластеров первого и второго уровня, составленная с использованием демоверсии программы Carrot2 и алгоритма Lingo3G.  
 На графике представлены 6 из 8 кластеров с наибольшим числом публикаций

Представленная на рис. 1 диаграмма хорошо отражает как основную тематику энергетической политики и безопасности, так и подкатегории основных тем.

Пример экспортируемых из демоверсии программы Carrot2 данных представлены в табл. 3.

Данный пример показывает удобство работы с подобными данными: колонки

Cluster Level 1 и Cluster Level 2 указывают на тематику, к которой относится статья, по DOI можно найти не только библиометрические данные, но, нередко, и полные тексты статьи. По заголовку публикации можно более детально оценить тематику рассматриваемой в таблице записи.

Таблица 3

**Пример экспортируемых данных для кластера «возобновляемая энергетика»  
(Renewable Energy) и кластера второго уровня «источники возобновляемой энергии»  
(Renewable Energy Sources)**

Title	DOI	Cluster Level 1	Cluster Level 2
Green Technological Development and Deployment in the Association of Southeast Asian Economies (ASEAN) – At Crossroads or Roundabout?	10.3390/su13020758	Renewable Energy	Renewable Energy Sources
Expert Views on the Future Development of Biogas Business Branch in Germany, The Netherlands, and Finland until 2030	10.3390/su13031148	Renewable Energy	Renewable Energy Sources
Green Energy in Central and Eastern European (CEE) Countries: New Challenges on the Path to Sustainable Development	10.3390/en14040884	Renewable Energy	Renewable Energy Sources
An Analysis of Renewable Energy Usage by Mobile Data Network Operators	10.3390/su13041886	Renewable Energy	Renewable Energy Sources

Учитывая, что данная работа выполнялась в формате подтверждения концепции (proof of concept) о рациональности использования кластеризации публикаций на основе использования именных групп, выявленных в текстах заголовков и аннотаций, а не в виде всестороннего исследования по выявлению актуальных тем по вопросам энергетической политики и безопасности в публикациях издательств Elsevier и MDPI, далее приведены лишь примеры выявления тем для трех кластеров и, соответственно, кластеров второго уровня на основе автоматического реферирования входящих в них текстов.

Демонстрация примеров организована следующим образом:

– Осуществлялась выборка публикаций, относящихся к трем кластерам второго уровня: источники возобновляемой энергии (renewable energy sources), энергетическая политика (energy policy) и энергетическая система (power system). Это подкластеры с наибольшим числом публикаций, входящие в кластеры также с

наибольшим числом публикаций (renewable energy), (energy security) и (energy system).

– Тексты аннотаций для каждого кластера подвергались автоматическому экстрактивному реферированию с использованием программы sumy<sup>4</sup> с алгоритмом Text-rank и списка стоп-слов stopwords-en-ISO.txt<sup>5</sup>. Использовалось ограничение на суммирование до 7 предложений.

– Полученные 7 предложений рассматривались как высокорейтинговые, по ним находились публикации, аннотации которых включали данные предложения. Соответственно, найденные публикации отражали тематику рассматриваемого кластера второго уровня.

– Для удобства чтения 7 предложений переводились на русский, а исходные тексты размещались в Прил. 2.

<sup>4</sup> <https://github.com/miso-belica/sumy> – Модуль для автоматического реферирования/суммирования текстовых документов.

<sup>5</sup> <https://github.com/stopwords-iso/stopwords-en> – Коллекция стоп-слов английского языка.

Примечание: в проводимом исследовании при использовании программы *sumu* использовались и другие входящие в нее алгоритмы (*edmundson*, *Isa*, *Lex-rank*), равно как и разное ограничение на число предложений в итоговом реферате. Полученные результаты существенно различались, что вообще характерно для библиометрических исследований и глубокого анализа текстов. Метафорически это можно назвать как субъективный взгляд различных алгоритмов на анализируемые тексты. Но подобное разнообразие имеет и положительную сторону – позволяет снизить предвзятость выбора. Детальное рассмотрение данной проблемы выходит за рамки данной статьи.

*Примечание к переводу текстов:* алгоритм *Text-rank* отдает предпочтение длинным предложениям. *Text-rank*, в сравнении с алгоритмом *Lex-rank*, выбирает предложения примерно в 2,5 раза длиннее (для нашего набора текстов аннотаций). Длинные предложения не только неудобно читать, но их крайне неудобно переводить. Если сделать «нормальный» перевод, состоящий из нескольких предложений и включающий синонимы повторяющихся слов, то станет непонятно, почему алгоритм *Text-rank*, примененный к английскому тексту, высоко ранжировал именно эти предложения. Длинные предложения, содержащие повторяющиеся термины, являются хорошими кандидатами для *Text-rank*. Поэтому, при переводе приходилось идти на компромисс – стараться не отдаляться от перевода, близкого к дословному, но при этом сохранить

минимальную их читаемость на русском языке. По английским текстам выбранных предложений находилась содержащая их библиометрическая запись (фактически DOI), что и вносилось в список литературы, который, согласно излагаемой процедуре, хорошо описывает выбранную тематику.

#### **Перевод результатов автоматического реферирования для подкластера «источники возобновляемой энергии»**

Рассмотрены: производство первичной энергии из возобновляемых источников, доля энергии из возобновляемых источников в конечном потреблении энергии, структура производства энергии из возобновляемых источников и доля энергии из возобновляемых источников в энергии, используемой транспортным сектором [4].

В качестве цели исследования рассматривался энергетический переход на основе разработки стратегии внедрения политики возобновляемой энергетики во всей энергетической системе с использованием всех возобновляемых источников энергии для прогнозирования будущих энергетических потребностей и потенциала снижения выбросов углерода [5].

Согласно полученным результатам, наиболее значимыми макроуровневыми рисками, с которыми сталкивается генерация возобновляемой энергии в Иране, являются «инфляция и колебания цен», «коррупция», «сложные процедуры лицензирования», «исследования и разработки и потенциал отечественных технологий», «санкции», «колебания обменного курса» и «права собственности и договорные риски» [6].

В существующих исследованиях анализировались региональная диспетчеризация возобновляемых источников энергии и планирование мощностей, однако, в них не полностью изучено влияние технических и экономических характеристик систем хранения энергии на интеграцию возобновляемых источников энергии и энергетический переход, а важность систем хранения энергии для энергетического перехода в настоящее время игнорируется [7].

Кроме того, было сказано, что системы хранения энергии, создание микросетей, сочетание солнечной, ветровой энергии и систем хранения энергии, а также политика в области возобновляемых источников энергии являются одними из способов преодоления экономических и энергетических последствий использования энергии ветра [8].

Были использованы шесть основных критериев (социальный, экономический, качество энергии, политический, технический, экологический) и шестнадцать подкритериев, а также разработана модель оценки для установления приоритетов наиболее подходящих возобновляемых источников энергии для производства электроэнергии в развивающихся странах, в которой рассматриваются четыре основных ресурса (солнечная энергия, энергия ветра, гидроэнергия и энергия биомассы) и их потенциал в производстве электроэнергии [9].

Для перехода на возобновляемые источники энергии очень важны соотношение спроса и предложения возобновляемой энергии, отношение потребления энергии к валовому внутреннему продукту (ВВП), эластичность производства энергии, эластичность

потребления энергии, эффективность преобразования энергии и инвестиции в НИОКР через «зеленое» финансирование [10].

#### **Перевод результатов автоматического реферирования для подкластера «энергетическая политика»**

Ассоциация государств Юго-Восточной Азии (АСЕАН) сталкивается с огромными проблемами в отношении будущего энергетического ландшафта и того, как энергетический переход будет охватывать новую архитектуру, включая умную политику и технологии для обеспечения доступности энергии для энергетической безопасности и устойчивости [11].

Энергетический ландшафт АСЕАН в перспективе будет зависеть от текущих действий, политики и инвестиций, направленных на преобразование энергетической системы, основанной на ископаемом топливе, в более чистую энергетическую систему, однако все решения и меры энергетической политики, которые будут приняты в ходе энергетического перехода, должны быть взвешены с учетом потенциально более высоких затрат на энергию, проблем доступности и рисков энергетической безопасности [11].

Текущая энергетическая политика Индонезии, которая опирается на дешевое ископаемое топливо и сосредоточена на двух из трех аспектов энергетической трилеммы, а именно энергетической безопасности и энергетической справедливости, может препятствовать ее усилиям по увеличению доли возобновляемых источников энергии [12].

Тематический анализ наиболее цитируемых публикаций выявил четыре взаимосвязанные темы устойчивого энергетического развития: доступ к недорогим современным энергетическим услугам, устойчивое энергоснабжение, устойчивое использование энергии и энергетическая безопасность [13].

Технологический прогресс и энергетические инновации могут помочь повысить энергетическую безопасность Китая путем диверсификации источников/путей энергоснабжения и повышения энергоэффективности, что важно для развития чистой и устойчивой энергетики [14].

Анализ показывает, что хотя возобновляемые источники энергии могут быть перспективным решением для повышения энергетической безопасности, текущий энергетический кризис показал, что нынешний масштаб инвестиций в возобновляемые источники энергии не готов удовлетворить текущие потребности в энергии, что вынуждает страны вновь вводить дополнительные источники углеводородного топлива, такие как уголь, затормаживая глобальный энергетический переход [15].

Межпровинциальный экспорт энергии оказывает перевернутое U-образное воздействие на доступность местной энергии, и поскольку структура экспорта энергии положительно влияет на доступность местной энергии, увеличение доли электроэнергии в общем экспорте энергии на 1% может привести к улучшению доступности местной энергии на 0,05% [16].

Примечание: первые два предложения взяты из одной аннотации, такое не часто встречается, можно предположить, что это

обусловлено частым появлением термина ASEAN в текстах аннотаций статей, посвященных энергетической политике.

#### **Перевод результатов автоматического реферирования для субкластера «энергетические системы»**

Новизна данного исследования заключается в том, что в нем рассматривается количество и мощность венгерских НМКЕ<sup>6</sup> в районах различных электроэнергетических компаний за определенный период времени, чтобы изучить возможный способ их эффективной интеграции в систему электроснабжения путем определения номинальной мощности накопителей энергии и энергоемкости предлагаемых систем хранения энергии [17].

Страны АСЕАН предприняли дальновидные шаги по увеличению доли возобновляемых источников энергии в обычных энергосетях, не препятствуя их текущему развитию. В данном исследовании представлены: политика, потенциал, прогресс и проблемы в области солнечной энергетики, существующие в странах АСЕАН; эти страны обладают огромным потенциалом солнечной энергии, поскольку расположены вблизи экватора, поэтому им следует сосредоточиться на использовании солнечной энергии для получения электричества и тепла, но для удовлетворения пикового спроса и обеспечения надежности возобновляемой энергии критически важным вопросом может стать разработка передовых систем хранения энергии [18].

---

<sup>6</sup> НМКЕ – малые электростанции для домашних хозяйств.

В данной статье рассматриваются государственная политика и цели, направленные на развитие энергетического потенциала, достижения в области хранения энергии и проблемы внедрения систем возобновляемой энергии в странах АСЕАН; в ней также подробно описаны рекомендуемые политические меры для ускорения использования возобновляемой энергии населением и снижения углеродного следа в течение следующего десятилетия путем внедрения передовых систем хранения энергии [18].

Начиная с первых примитивных ветряных турбин в конце 1800-х годов, мировых войн, энергетического кризиса 1970-х годов и десятилетий растущей обеспокоенности состоянием окружающей среды, этот исторический отчет описывает, как приоритеты политики в Дании постепенно смещались от внимания к диверсификации энергетики, энергоэффективности и энергетической независимости к устойчивому развитию и возобновляемым источникам энергии, при этом ветроэнергетика и ветроэнергетические технологии возглавили список приоритетов [19].

Хотя существующие исследования анализируют региональную диспетчеризацию возобновляемой энергии и планирование мощностей, они не в полной мере изучают влияние технических и экономических характеристик систем хранения энергии на интеграцию возобновляемой энергии и энергетический переход [7].

Авторы статьи отметили, что системы хранения энергии, создание микросетей,

сочетание солнечной, ветровой энергии и систем хранения энергии, а также политика в области возобновляемых источников энергии являются одними из способов борьбы с экономическими и энергетическими недостатками ветровой энергетики [8].

Признавая важность искусственного интеллекта (ИИ), данное исследование было проведено для семи различных энергетических систем и их различных применений, включая: (I) производство электроэнергии; (II) доставку электроэнергии; (III) распределительные сети; (IV) хранение энергии; (V) энергосбережение, новые энергетические материалы и устройства; (VI) энергоэффективность и нанотехнологии; и (VII) энергетическую политику и экономику [20].

Примечание: две статьи [7] и [8], в силу их широкой тематики, отнесены системой Carrot2 как к подкластеру «источники возобновляемой энергии», так и к подкластеру «энергетические системы». Алгоритмы кластеризации текстов, основанные на предположении, что каждый текст относится к нескольким темам, а каждая тема описывается набором ключевых терминов, приводит к вышенаблюдаемому результату.

## **Заключение**

### **Выводы по второй части публикации**

Проведенные исследования показали, что в публикациях по вопросам энергетической политики и безопасности издательств Elsevier и MDPI за 2021–2023 гг. доминирует следующая тематика:

*Возобновляемая энергетика*

(912 документов, 7 подкластеров)

Возобновляемые источники энергии (422)

Энергетический переход (247)

Энергетические технологии (314)

Энергетические ресурсы (255)

Производство энергии (280)

Устойчивое развитие (206)

Потребление возобновляемой энергии (208)

*Энергетическая безопасность*

(768 документов, 7 подкластеров)

Энергетическая политика (335)

Энергетические ресурсы (221)

Глобальная энергия (195)

Изменение энергии (234)

Источники энергии (232)

Энергетические технологии (219)

Увеличение энергии (194)

*Энергетическая система*

(807 документов, 7 подкластеров)

Энергетическая система (325)

Энергетический переход (221)

Энергетические технологии (304)

Спрос на энергию (234)

Энергетика стран (260)

Возобновляемые источники энергии (191)

Энергия будущего (202)

*Потребление энергии*

(511 документов, 7 подкластеров)

Энергоэффективность (152)

Потребление электроэнергии (161)

Потребление возобновляемой энергии (208)

Спрос на энергию (137)

Производственное потребление (171)

Устойчивое развитие (118)

Энергетика Китая (118)

*Устойчивая энергетика*

(623 документа, 7 подкластеров)

Энергетический переход (170)

Энергетические технологии (209)

Энергетические ресурсы (174)

Спрос на энергию (156)

Экономическое развитие (177)

Эффективная энергетика (177)

Национальная энергетика (159)

*Энергетический сектор*

(548 документов, 7 подкластеров)

Энергоэффективность (154)

Электроэнергетический сектор (214)

Переход к энергетике (140)

Спрос на энергию (154)

Энергетические технологии (185)

Устойчивое развитие (128)

Эффект энергетике (156).

Показана возможность использования именных групп извлеченных из текстов заголовков и аннотаций для кластеризации публикаций как основы выявления актуальных тем исследований. Именные группы, как термины, имеют даже большее пересечение с авторскими ключевыми словами, чем термины, полученные в первой части публикации. Более подробно об использовании именных групп для кластеризации документов можно ознакомиться в статье [21].

На примере трех подкластеров: «возобновляемые источники энергии», «энергетическая политика» и «энергетические системы», показана рациональность использования технологии экстрактивного реферирования для описания доминирующих тем в подкластере путем выявления публикаций, содержащих тексты (предложения), имеющие высокий ранг согласно алгоритму TextRank.

### **Общий вывод по двум частям публикаций**

Учитывая, что исследование выполнено в формате демонстрации возможности, а не всестороннего исследования тематики публикаций и сравнения метода анализа их текстов, можно дать следующие рекомендации:

При составлении вводной части исследования, чтобы снизить предвзятость в выборе актуальных задач в рамках рассматриваемой темы, целесообразно использовать современные методы, основанные на больших, предварительно обученных языковых моделях, таких как ChatGPT. Сгенерированные с использованием таких моделей тексты носят весьма общий характер, но хорошо отражают широко обсуждаемые вопросы по изучаемой тематике. Фактически большое внимание к той или иной задаче может служить мерой ее актуальности.

Если в основной части исследования ставится задача, например, составления аналитического отчета, то в данном случае целесообразно использовать методы анализа, максимально сохраняющие связь с исходными публикациями, потому что именно по ним и будет составляться отчет.

В этом случае целесообразнее использовать кластеризацию библиометрических записей, а не кластеризацию значимых терминов, например, авторских ключевых слов и т. п. То же касается преимущества использования автоматического экстрактивного реферирования (суммирования) по сравнению с абстрактным реферированием, так как полученные первым методом высокоранговые предложения сохраняют связь с исходными текстами, к которым всегда можно вернуться.

Общий вывод: современные методы работы с большими объемами текстов могут существенно ускорить составление тематических аналитических отчетов. Снижение предвзятости в выборе актуальных задач в рамках рассматриваемой темы предполагает привлечение экспертов и наличия формальных методов оценки, что весьма затратно. С другой стороны, экспертиза предполагает этап проверки рассматриваемых материалов на минимальное соответствие общим требованиям, в этом случае, применение даже таких, еще мало используемых методов, как ChatGPT, может разгрузить работу экспертов, освободив их от части рутинных действий.

*Статья написана в рамках выполнения государственного задания ИПНГ РАН (тема «Фундаментальный базис энергоэффективных, ресурсосберегающих и экологически безопасных, инновационных и цифровых технологий поиска, разведки и разработки нефтяных и газовых месторождений, исследование, добыча и освоение традиционных и нетрадиционных запасов и ресурсов нефти и газа; разработка рекомендаций по реализации продукции нефтегазового комплекса в условиях энергоперехода и политики ЕС по декарбонизации энергетики (фундаментальные, поисковые, прикладные, экономические и междисциплинарные исследования)», № 122022800270-0).*

### Литература

1. *Behpour S., Mohammadi M., Albert M.V. et al.* Automatic trend detection: Time-biased document clustering // *Knowledge-Based Systems*. 2021. Vol. 220. P. 106907. <https://doi.org/10.1016/j.knosys.2021.106907>
2. *Ibrahim Altmami N., El Bachir Menai M.* Automatic summarization of scientific articles: A survey // *Journal of King Saud University – Computer and Information Sciences*. 2022. Vol. 34, No. 4. P. 1011–1028. <https://doi.org/10.1016/j.jksuci.2020.04.020>
3. *Pawar S., Manjula Gururaj H., Chiplunar N.N.* Text summarization using document and sentence clustering // *Procedia Computer Science*. 2022. Vol. 215. P. 361–369. <https://doi.org/10.1016/j.procs.2022.12.038>
4. *Milek D., Nowak P., Latosińska J.* The development of renewable energy sources in the European Union in the light of the European Green Deal // *Energies*. 2022. Vol. 15, No. 15. P. 5576. <https://doi.org/10.3390/en15155576>
5. *Asim M., Qamar A., Kanwal A. et al.* Opportunities and challenges for renewable energy utilization in Pakistan // *Sustainability*. 2022. Vol. 14, No. 17. P. 10947. <https://doi.org/10.3390/su141710947>
6. *Shahnazi R., Alimohammadlou M.* Investigating risks in renewable energy in oil-producing countries through multi-criteria decision-making methods based on interval type-2 fuzzy sets: A case study of Iran // *Renewable Energy*. 2022. Vol. 191. P. 1009–1027. <https://doi.org/10.1016/j.renene.2022.04.051>
7. *Huang W.-C., Zhang Q., You F.* Impacts of battery energy storage technologies and renewable integration on the energy transition in the New York State // *Advances in Applied Energy*. 2023. Vol. 9. P. 100126. <https://doi.org/10.1016/j.adapen.2023.100126>
8. *Msigwa G., Ighalo J.O., Yap P.-S.* Considerations on environmental, economic, and energy impacts of wind energy generation: Projections towards sustainability initiatives // *Science of The Total Environment*. 2022. Vol. 849. P. 157755. <https://doi.org/10.1016/j.scitotenv.2022.157755>
9. *Abdul D., Wenqi J., Tanveer A.* Prioritization of renewable energy source for electricity generation through AHP-VIKOR integrated methodology // *Renewable Energy*. 2022. Vol. 184. P. 1018–1032. <https://doi.org/10.1016/j.renene.2021.10.082>
10. *Wang S., Sun L., Iqbal S.* Green financing role on renewable energy dependence and energy transition in E7 economies // *Renewable Energy*. 2022. Vol. 200. P. 1561–1572. <https://doi.org/10.1016/j.renene.2022.10.067>
11. *Phoumin H., Kimura F., Arima J.* ASEAN's energy transition towards cleaner energy system: Energy modelling scenarios and policy implications // *Sustainability*. 2021. Vol. 13, No. 5. P. 2819. <https://doi.org/10.3390/su13052819>
12. *Heffron R.J., Körner M.-F., Sumarno T. et al.* How different electricity pricing systems affect the energy trilemma: Assessing Indonesia's electricity market transition // *Energy Economics*. 2022. Vol. 107. P. 105663. <https://doi.org/10.1016/j.eneco.2021.105663>
13. *Gunnarsdottir I., Davidsdottir B., Worrell E., Sigurgeirsdottir S.* Sustainable energy development: History of the concept and emerging themes // *Renewable and Sustainable Energy Reviews*. 2021. Vol. 141. P. 110770. <https://doi.org/10.1016/j.rser.2021.110770>

14. *Huang B., Zhang L., Ma L. et al.* Multi-criteria decision analysis of China's energy security from 2008 to 2017 based on Fuzzy BWM-DEA-AR model and Malmquist Productivity Index // *Energy*. 2021. Vol. 228. P. 120481. <https://doi.org/10.1016/j.energy.2021.120481>
15. *Belaid F., Al-Sarhi A., Al-Mestneer R.* Balancing climate mitigation and energy security goals amid converging global energy crises: The role of green investments // *Renewable Energy*. 2023. Vol. 205. P. 534–542. <https://doi.org/10.1016/j.renene.2023.01.083>
16. *Zhang J., Wang Z.* Does inter-provincial energy exportation affect local energy affordability? Evidence from China // *Energy Reports*. 2022. Vol. 8. P. 13848–13858. <https://doi.org/10.1016/j.egyr.2022.10.299>
17. *Pintér G., Zsiborács H., Baranyai N.H.* Aspects of determining the energy storage system size linked to household-sized power plants in Hungary in accordance with the regulatory needs of the electric energy system // *Sustainability*. 2022. Vol. 14, No. 5. P. 2622. <https://doi.org/10.3390/su14052622>
18. *Pandey A.K., Kalidasan B., Reji Kumar R. et al.* Solar energy utilization techniques, policies, potentials, progresses, challenges and recommendations in ASEAN countries // *Sustainability*. 2022. Vol. 14, No. 18. P. 11193. <https://doi.org/10.3390/su141811193>
19. *Johansen K.* Blowing in the wind: A brief history of wind energy and wind power technologies in Denmark // *Energy Policy*. 2021. Vol. 152. P. 112139. <https://doi.org/10.1016/j.enpol.2021.112139>
20. *Ahmad T., Zhu H., Zhang D. et al.* Energetics Systems and artificial intelligence: Applications of industry 4.0 // *Energy Reports*. 2022. Vol. 8. P. 334–361. <https://doi.org/10.1016/j.egyr.2021.11.256>
21. *Zheng H.-T., Kang B.-Y., Kim H.-G.* Exploiting noun phrases and semantic relationships for text document clustering // *Information Sciences*. 2009. Vol. 179, No. 13. P. 2249–2262. <https://doi.org/10.1016/j.ins.2009.02.019>

## ПРИЛОЖЕНИЕ

## Приложение 1

Перевод текстов табл. 2. Примеры абзацев текстов и строк, составленных из терминов именных групп.

Заголовок статьи	Строка, содержащая термины из 5000 именных групп
1	2
Эффективность программы «Мое электричество» в поддержку развития фотоэлектрических установок в Польше	электрическая программа эффективность развитие PV установка Польша исследования показывают легитимность механизмы возобновляемой энергии проблемы распределение средств статья ситуация средства фотоэлектрические микроустановки Польша электрическая программа статья результаты анализ неравенство использование средств электрическая программа неравенство интенсивность поддержка условия максимизация производство электроэнергии электрическая программа PV микроустановки средняя мощность PV мощность финансовые ресурсы электрическая программа средства распределение излишки средний доход малые домохозяйства потенциальный проект критерии выбора анализ критерии доходы домохозяйств
Может ли Верхний полуостров Мичигана добиться справедливости при переходе на 100% возобновляемую электроэнергию? Исследование общественного мнения в области социально-технических изменений	верховная юстиция возобновляемая электроэнергия опрос общественное восприятие изменение стоимость энергии западная верхняя северная часть штата США ситуация ВВП жители счета за электричество интерес возобновляемая электроэнергия регион вопросы факторы ВВП жители переход поддержка переход ввп исследование факторы жители переход ВВП исследование общественное восприятие опрос жителей результаты регрессии показывают инициатива жителей вероятность потребления поддержка развитие ветроэнергетики переход поддержка высокая тенденция ВВП проект предпочтения результаты исследования дорожная карта будущее планирование с участием сообщества регион
Влияние государственной политики на пробелы в энергоэффективности: данные по Украине	влияние государственная политика энергетический разрыв доказательство документ проверка влияние инвестиции энергетическая эффективность разрыв выводы библиометрический анализ рост исследования интересы ядро энергетическая эффективность разрыв центральное качество институты увеличение зеленых инвестиций энергетический сектор отношения инвестиции синергетический эффект энергетическая эффективность разрыв национальная экономика анализ временные ряды мировой банк данных Евростат период методы единичный радикальный признак данных тест вес-моделирование коинтеграционный анализ выводы энергетическая эффективность разрыв год увеличение зеленых инвестиций энергетика год стабильность общественное восприятие увеличение общественного точки восприятия точки стабильности скорость восстановления энергетический сектор политика энергетическая эффективность разрыв государственный уровень общественное восприятие стабильность
Решение проблемы эффекта городских тепловых островов: межстрановая оценка роли зеленой инфраструктуры	городские острова тепла эффект межстрановая оценка зеленая инфраструктура городские острова тепла эффект городские районы повышение температуры локальные тепловые волны интенсивность снижение уровней качество жизни документ отзывы зеленая инфраструктура интенсивность влияние человеческий комфорт всесторонний обзор литературы документ отчеты анализ тематические исследования набор городов 13 стран географические регионы

## Продолжение таблицы

1	2
	климатические зоны результаты общая зеленая инфраструктура городские районы условия воздействие дополнение исследование воздействие городские проблемы город последствия документ комплекс тепловые волны изменение климата городские зеленые ресурсы преимущества влияние документ необходимость градостроители эффекты проекты новых зданий
Декомпозиционный анализ эволюции местной энергетической системы как инструмент оценки влияния местных действий: Методология и пример Мальме, Швеция	декомпозиционный анализ эволюция местной энергетической системы инструмент эффект местные действия методология документ использование декомпозиционный анализ эффект местные действия связанные с энергетикой смягчение изменения климата политика оценка планирование местный уровень оценка влияние местные действия вызов перспектива факторы изменения местные выбросы парниковых газов связанные с энергетикой местные органы власти методология декомпозиция модель изменения местная энергетическая система причины/последствия методология эффект местные действия тематическое исследование
Моделирование поведения участников торгов на немецких фотоэлектрических аукционах	моделирование поведения на торгах фотоэлектрические аукционы статья возобновляемая энергия поддержка распределение типы аукционов основы теории аукционов правила фотоэлектрических аукционов работа стратегии торгов стоимость электроэнергии методология набор правил цена неравновесная стратегия торгов функция монотонные функции монотонные функции основы теории аукционов моделирование результаты PV поддержка распределение аукционная система стоимость ценообразование правила участников оценка аукцион неравновесная стоимость перспектива применение ценообразование правила политика решения

Автор еще раз подчеркивает, что в правой колонке представлен не полный текст, а перевод слов, по которым осуществлялось установление близости текстов в процессе их кластеризации с использованием демоверсии программы Cartot2.

**Приложение 2**

*В этом приложении даны исходные тексты на английском языке тех предложений, которые рассматривались как высокорейтинговые, по которым находились публикации аннотаций, включающие данные предложения. Найденные публикации отражали тематику рассматриваемого кластера второго уровня.*

The generation of primary energy from renewable sources, the share of energy from renewable sources in the final energy consumption, the structure of energy generation from renewable sources, and the share of energy from renewable sources in the energy used by the transport sector are discussed [4].

The purpose of this study is energy transition by designing a strategy for the adoption of renewable energy policies in the entire energy system by using all renewable energy resources to forecast future energy needs and carbon emission mitigation potential [5].

The results showed that most important macro-level risks facing Iran's renewable energy generation were "inflation and price fluctuations", "corruption", "complex licensing procedures", "research and development and the capacity of domestic technology", "sanctions", "exchange rate fluctuations", and "property rights and contractual risks" [6].

Although existing studies analyze regional dispatch of renewable energy sources and capacity planning, they do not fully explore the impacts of the energy storage system technology's technical and economic characteristics on renewable energy integration and energy transition, and the importance of energy storage systems to the energy transition is currently ignored [7].

In addition, we discussed that energy storage systems, setting up microgrids, combination of solar, wind and energy storage, and renewable energies policies are some of the ways to combat wind energy's economic and energy impacts [8].

Therefore, six main criteria, social, economic, quality of energy, political, technical, environmental, and sixteen sub-criteria have been employed, and an assessment model is developed for prioritizing the most pertinent renewable energy sources for electricity generation in developing countries in which four major resources, solar energy, wind energy, hydropower, and biomass energy, are considered and their electricity generation potential [9].

More specifically, for renewable energy transition, renewable energy demand and supply ratio, energy consumption to GDP ratio, energy production elasticity, energy consumption elasticity, energy conversion efficiency and R & D investment through green financing are very essential [10].

The Association of Southeast Asian Nations (ASEAN) faces tremendous challenges regarding the future energy landscape and how the energy transition will embrace a new architecture—including sound policies and technologies to ensure energy access together with affordability, energy security, and energy sustainability [11].

The future energy landscape of ASEAN will rely on today's actions, policies, and investments to change the fossil fuel-based energy system towards a cleaner energy system, but any decisions and energy policy measures to be rolled out during the energy transition need to be weighed against potentially higher energy costs, affordability issues, and energy security risks [11].

Indonesia's current energy policy, which relies on cheap fossil fuels and focuses on two out of the three horns of the energy trilemma, namely, energy security and energy equity, may impede its efforts to higher shares of renewable energy sources [12].

A thematic analysis of these most cited publications led to the identification of four interrelated themes of sustainable energy development; access to affordable modern energy services, sustainable energy supply, sustainable energy consumption, and energy security [13].

Technological progress and energy innovations can help to improve China's energy security by diversifying the energy supply sources/pathways and improving the energy efficiency, which is of great significance for the development of clean and sustainable energy [14].

The analysis highlights that while renewables could provide a promising solution to enhance energy security, the current energy crisis has revealed that the current scale of renewable energy investments is not ready to address ongoing energy needs, forcing countries to re-unlock further hydrocarbon fuel sources such as coal, putting global energy transition on hold [15].

Meanwhile, inter-provincial energy exportation poses an inverted U-shaped impact on local energy affordability, and because energy exportation structure positively affects local energy affordability, an increase in the share of electricity in total energy exportation by 1% can lead to a 0.05% improvement in local energy affordability [16].

The innovative novelty of this study is that it examines the quantity and power of Hungarian HMKEs in the districts of the various electric companies over time with a view of exploring a possible way of their efficient integration into the electric energy system by determining the nominal energy storage power and energy capacity of the proposed energy storage systems [17].

The ASEAN countries have taken visionary steps towards increasing the renewable energy mix with the conventional grid without hampering the ongoing development; this study presents the solar energy utilization policies, potential, progresses, and challenges adopted in ASEAN countries; furthermore, in these nations there is a huge potential of solar energy being located near the equator, therefore, they should focus on both solar to electrical and solar to thermal energy applications; however, in order to meet the peak demand and ensure the reliability of renewable energy like solar power, the development of advanced energy storage systems could be the key areas, and concrete efforts are required [18].

Therefore, this article is a spotlight on government policies and goals focusing on energy potential, major progress in terms of energy storage and challenges in implementation of renewable energy systems in ASEAN countries; furthermore the recommended highlights on policies to accelerate the exploitation of renewable energy usage among the people are also discussed in detail, besides, the insights on reduction of carbon footprints over the next decade through incorporation of advanced energy storage systems [18].

From the first primitive wind turbines in the late 1800s, to the world wars, through the energy-crisis in the 70s, and into the decades of growing environmental awareness and concern, this historical account describes how policy priorities in Denmark gradually translated from the focus on energy diversification, energy efficiency and energy independence to the focus on sustainability and renewable energy resources, with wind energy and wind power technologies at the top of the priority list [19].

Although existing studies analyze regional dispatch of renewable energy sources and capacity planning, they do not fully explore the impacts of the energy storage system technology's technical and economic characteristics on renewable energy integration and energy transition, and the importance of energy storage systems to the energy transition is currently ignored [7].

In addition, we discussed that energy storage systems, setting up microgrids, combination of solar, wind and energy storage, and renewable energies policies are some of the ways to combat wind energy's economic and energy impacts [8].

Recognizing the importance of AI, this study was conducted on seven different energetics systems and their variety of applications, including: i) electricity production; ii) power delivery; iii) electric distribution networks; iv) energy storage; v) energy saving, new energy materials, and devices; vi) energy efficiency and nanotechnology; and vii) energy policy, and economics [20].

## Identification of actual energy policy and security research topics in Elsevier and MDPI publications for 2021–2023.

### Part 2. Clustering of publications. Automatic extractive referencing

**B.N. Chigarev**

Oil and Gas Research Institute, Russian Academy of Sciences, Moscow, Russia  
E-mail: [bchigarev@ipng.ru](mailto:bchigarev@ipng.ru)

**Abstract.** This article focuses on identifying relevant research topics in Elsevier and MDPI publications 2021–2023 on energy policy and security based on publication clustering.

It is shown that name groups extracted from titles and abstracts can be used to cluster publications to identify relevant research topics. The extraction of noun phrases was carried out using the Sifaka software package. A table of the 5,000 most frequent name groups was constructed, the terms of which were retained in the title and abstract texts, and the remaining terms were deleted. The prepared texts were used to cluster the bibliometric records using a demo version of Carrot2 and the Lingo3G algorithm. The annotation texts for each cluster were subjected to automatic extractive referencing using the sumy program with the Text-rank algorithm and the stopwords-en-ISO.txt list. The possibility of using extractive referencing technology to describe the dominant topic of individual publication clusters is demonstrated.

A list of dominant themes in energy policy and security publications in Elsevier and MDPI publications for 2021–2023 is given.

**Keywords:** energy policy, research topics, bibliometric analysis, Sifaka, Carrot2, publication clustering, extractive referencing.

**Citation:** *Chigarev B.N.* Identification of actual energy policy and security research topics in Elsevier and MDPI publications for 2021–2023. Part 2. Clustering of publications. Automatic extractive referencing // Actual Problems of Oil and Gas. 2023. Iss. 2(41). P. 84–107. <https://doi.org/10.29222/ipng.2078-5712.2023-41.art6> (In Russ.).

#### References

1. *Behpour S., Mohammadi M., Albert M.V.* et al. Automatic trend detection: Time-biased document clustering // Knowledge-Based Systems. 2021. Vol. 220. P. 106907. <https://doi.org/10.1016/j.knosys.2021.106907>
2. *Ibrahim Altmami N., El Bachir Menai M.* Automatic summarization of scientific articles: A survey // Journal of King Saud University – Computer and Information Sciences. 2022. Vol. 34, No. 4. P. 1011–1028. <https://doi.org/10.1016/j.jksuci.2020.04.020>
3. *Pawar S., Manjula Gururaj H., Chiplunar N.N.* Text summarization using document and sentence clustering // Procedia Computer Science. 2022. Vol. 215. P. 361–369. <https://doi.org/10.1016/j.procs.2022.12.038>
4. *Milek D., Nowak P., Latosińska J.* The development of renewable energy sources in the European Union in the light of the European Green Deal // Energies. 2022. Vol. 15, No. 15. P. 5576. <https://doi.org/10.3390/en15155576>

5. *Asim M., Qamar A., Kanwal A. et al.* Opportunities and challenges for renewable energy utilization in Pakistan // *Sustainability*. 2022. Vol. 14, No. 17. P. 10947. <https://doi.org/10.3390/su141710947>
6. *Shahnazi R., Alimohammadlou M.* Investigating risks in renewable energy in oil-producing countries through multi-criteria decision-making methods based on interval type-2 fuzzy sets: A case study of Iran // *Renewable Energy*. 2022. Vol. 191. P. 1009–1027. <https://doi.org/10.1016/j.renene.2022.04.051>
7. *Huang W.-C., Zhang Q., You F.* Impacts of battery energy storage technologies and renewable integration on the energy transition in the New York State // *Advances in Applied Energy*. 2023. Vol. 9. P. 100126. <https://doi.org/10.1016/j.adapen.2023.100126>
8. *Msigwa G., Ighalo J.O., Yap P.-S.* Considerations on environmental, economic, and energy impacts of wind energy generation: Projections towards sustainability initiatives // *Science of The Total Environment*. 2022. Vol. 849. P. 157755. <https://doi.org/10.1016/j.scitotenv.2022.157755>
9. *Abdul D., Wenqi J., Tanveer A.* Prioritization of renewable energy source for electricity generation through AHP-VIKOR integrated methodology // *Renewable Energy*. 2022. Vol. 184. P. 1018–1032. <https://doi.org/10.1016/j.renene.2021.10.082>
10. *Wang S., Sun L., Iqbal S.* Green financing role on renewable energy dependence and energy transition in E7 economies // *Renewable Energy*. 2022. Vol. 200. P. 1561–1572. <https://doi.org/10.1016/j.renene.2022.10.067>
11. *Phoumin H., Kimura F., Arima J.* ASEAN's energy transition towards cleaner energy system: Energy modelling scenarios and policy implications // *Sustainability*. 2021. Vol. 13, No. 5. P. 2819. <https://doi.org/10.3390/su13052819>
12. *Heffron R.J., Körner M.-F., Sumarno T. et al.* How different electricity pricing systems affect the energy trilemma: Assessing Indonesia's electricity market transition // *Energy Economics*. 2022. Vol. 107. P. 105663. <https://doi.org/10.1016/j.eneco.2021.105663>
13. *Gunnarsdottir I., Davidsdottir B., Worrell E., Sigurgeirsdottir S.* Sustainable energy development: History of the concept and emerging themes // *Renewable and Sustainable Energy Reviews*. 2021. Vol. 141. P. 110770. <https://doi.org/10.1016/j.rser.2021.110770>
14. *Huang B., Zhang L., Ma L. et al.* Multi-criteria decision analysis of China's energy security from 2008 to 2017 based on Fuzzy BWM-DEA-AR model and Malmquist Productivity Index // *Energy*. 2021. Vol. 228. P. 120481. <https://doi.org/10.1016/j.energy.2021.120481>
15. *Belaïd F., Al-Sarihi A., Al-Mestneer R.* Balancing climate mitigation and energy security goals amid converging global energy crises: The role of green investments // *Renewable Energy*. 2023. Vol. 205. P. 534–542. <https://doi.org/10.1016/j.renene.2023.01.083>
16. *Zhang J., Wang Z.* Does inter-provincial energy exportation affect local energy affordability? Evidence from China // *Energy Reports*. 2022. Vol. 8. P. 13848–13858. <https://doi.org/10.1016/j.egy.2022.10.299>
17. *Pintér G., Zsiborács H., Baranyai N.H.* Aspects of determining the energy storage system size linked to household-sized power plants in Hungary in accordance with the regulatory needs of the electric energy system // *Sustainability*. 2022. Vol. 14, No. 5. P. 2622. <https://doi.org/10.3390/su14052622>

18. *Pandey A.K., Kalidasan B., Reji Kumar R.* et al. Solar energy utilization techniques, policies, potentials, progresses, challenges and recommendations in ASEAN countries // *Sustainability*. 2022. Vol. 14, No. 18. P. 11193. <https://doi.org/10.3390/su141811193>

19. *Johansen K.* Blowing in the wind: A brief history of wind energy and wind power technologies in Denmark // *Energy Policy*. 2021. Vol. 152. P. 112139. <https://doi.org/10.1016/j.enpol.2021.112139>

20. *Ahmad T., Zhu H., Zhang D.* et al. Energetics Systems and artificial intelligence: Applications of industry 4.0 // *Energy Reports*. 2022. Vol. 8. P. 334–361. <https://doi.org/10.1016/j.egyr.2021.11.256>

21. *Zheng H.-T., Kang B.-Y., Kim H.-G.* Exploiting noun phrases and semantic relationships for text document clustering // *Information Sciences*. 2009. Vol. 179, No. 13. P. 2249–2262. <https://doi.org/10.1016/j.ins.2009.02.019>