energies

MDPI

Article

Information Analysis on Foreign Institution for International R&D Collaboration Using Natural Language Processing

Jihoo Jung [1,†], Jehyun Lee [2,*,†], Sangjin Choi [1] and Woonho Baek [1]

[1] Global Strategy Team, Korea Institute of Energy Research, Daejeon 34129, Republic of Korea
[2] Computational Science & Engineering Laboratory, Korea Institute of Energy Research, Daejeon 34129, Republic of Korea
* Correspondence: jehyunlee@kier.re.kr
† These authors equally contributed to this study.

Abstract: The number of international collaborations in research and development (R&D) has been increasing in the energy sector to solve global environmental problems—such as climate change and the energy crisis—and to reduce the time, cost, and risk of failure. Successful international project planning requires the analysis of research fields and the technology expertise of cooperative partner institutions or countries, but this takes time and resources. In this study, we developed a method to analyze the information on research organizations and topics, taking advantage of data analysis as well as deep learning natural language processing (NLP) models. A method to evaluate the relative superiority of efficient international collaboration was suggested, assuming international collaboration of the National Renewable Energy Laboratory (NREL) and the Korea Institute of Energy Research (KIER). Additionally, a workflow of an automated executive summary and a translation of tens of web-posted articles is also suggested for a quick glance. The valuation of the suggested methodology is estimated as much as the annual salary of an experienced employee.

Keywords: open API; international cooperation; data analysis; R&D planning; text mining

Citation: Jung, J.; Lee, J.; Choi, S.; Baek, W. Information Analysis on Foreign Institution for International R&D Collaboration Using Natural Language Processing. *Energies* **2023**, *16*, 33. https://doi.org/10.3390/en16010033

Academic Editor: Jaroslaw Krzywanski

Received: 1 November 2022
Revised: 14 December 2022
Accepted: 15 December 2022
Published: 21 December 2022

1. Introduction

An international collaboration between institutions for research and development (R&D) is becoming essential. According to the National Scientific Board, the percentage of worldwide science and engineering articles produced with international collaboration rose from 17% to 23% between 2008 and 2018 [1]. Although most of the global co-authorship is placed by low-income countries (86% in 2014), the European Union, one of the most developed societies, also marks 46% in 2014, where the greatest leap took place in the Arabian States, from 44% (2005) to 77% (2014) [2]. Despite the active global collaboration, it is challenging for an institution to find a proper research partner due to the increasing complexity, convergence, and multi- and interdisciplinarity [3–5].

International collaboration is becoming increasingly important in the energy sector because many problems cannot be solved by a single country or company alone. For example, the global climate crisis, greenhouse gas reduction, energy transition, and the realization of a hydrogen economy are topics where international collaboration is needed [6–8]. There are several key objectives of international cooperation. The first is achieving common goals, such as overcoming the climate crisis or reducing greenhouse gas emissions. These include economic factors that could reduce the costs and risks of technology development and market entry, achieve economies of scale in production, and reduce the time required for new product development and commercialization. Because of these benefits, many countries are aggressively promoting international cooperation as a major national R&D strategy, and Korean national research institutes and companies are managing various international cooperation projects [9,10].

In order to plan a successful international cooperation project, a systematic analysis of the detailed research areas, technological expertise, and joint research demands of partner countries or institutions is required. When a mutual understanding between the international cooperation partners is achieved, joint projects that benefit each party can be implemented [11]. Various studies have been conducted on international project-planning strategies. Jung et al. compared the technology and productivity of research institutes and companies by using patent indicators to select promising international cooperation partners in the hydrogen energy field [12]. Lee and Chung compared the technology expertise of the two countries using expert survey data to reveal cooperative technologies in the Korean and Indian defense sectors [13]. Chang and Yun took advantage of bibliometric information to select targets for international cooperation in the field of energy fusion from a comprehensive point of view [14]. The current status of international collaboration countries was analyzed using details of the paper's co-authors, and a list of promising cooperative countries was derived from several publications and citations. National research institutes of Korea have continuously conducted international joint projects; however, the performance is often unsatisfactory. The main reason is suspected to be inadequate information on the collaborating partners because, still, most of the decision relies on the individual researchers' human network in the same sectors in which the global view is absent. In order to overcome the limit, it is necessary to establish R&D strategies based on the sufficient analysis of the records, including research topics, technology expertise, and capabilities—including the human resources and facilities. Providing an executive summary of recent research has been conducted; however, it is challenging for human resources to collect and analyze enormous volumes of data such as academic publications and reports from partner institutions [15,16]. Moreover, proficiency in technical terminology, as well as unfamiliar fields, is another barrier regarding the department taking charge of international collaboration, which is an administrative organization that has to follow the advances of almost all sectors of the institute [17,18]. Although some studies utilized literature analysis for international collaboration partner targeting, studies on individual works of literature were overlooked.

In this study, we suggest a decision support method based on a mass literature analysis for international collaboration with an example of the National Renewable Energy Research (NREL). Data analysis on more than 13,000 academic publications for the last 10 years and 50 articles posted on its website was conducted not only from the perspective view, such as institutional and international collaboration trends, but also in detail, such as research topic classifications for each literature compared with the ones from the Korea Institute of Energy Research (KIER). Additionally, an automatic report generation, including the translated summary of the articles, is provided.

Many deep learning-based literature summaries and translations have been published, but their seamless application on research comparison as well as analysis [19–22]. It should be noted that our approach can be applied to any institution producing academic papers and websites for public use. Machine translation is also able to be applied to most languages because the deep learning translation models these days support multi-language translation and are being developed for the analysis of non-native English-speaking foreign institutes [23].

2. Materials and Methods

2.1. Literature Data Acquisition and Analysis

NREL's recent research archives were collected from two sources. One is Scopus, an abstract and citation database of scientific literature [24]. Publications from the NREL and KIER were acquired using Scopus API with the query "AF-ID ("National Renewable Energy Laboratory" 60030451)" and "AF-ID ("National Renewable Energy Laboratory" 60087822)", and the time span of the publication was limited to the last decade (2013–2022). Finally, 9219 works of literature were obtained for NREL and 3667 for KIER. The other is the official website of NREL. While the literature retrieved from Scopus contains academic achievements, the visions, social activities, as well as attempts at industrial applications were found in web-posted articles. Fifty-four articles posted on August 2022 were categorized into nine groups by the main subject—the main page, hydrogen, photovoltaics, energy storage, decarbonization in transporting systems, building energy, materials, wind energy, and grid technology [25]. The list of the web articles to be reviewed is summarized in the form of a Microsoft Excel file.

The Python programming language, with a number of libraries, took advantage of the retrieval and analysis of the text data. Scopus data was retrieved using *pybliometrics* [26], preprocessed and analyzed with *pandas* [27], and visualized with *matplotlib* and *seaborn* [28,29]. Conventional data analysis, such as grouping and trend analysis, have been performed with the python packages.

Network analysis was applied to examine the collaboration between institutes and countries using the *VOSviewer*, not only as a visualizer but also for relational analysis, such as co-authorship and citation analysis [30]. The results were exported and revisited for quantitative review for further analysis.

2.2. Topic Modeling by Latent Dirichlet Allocation (LDA)

Research topics of each literature were determined by Latent Dirichlet Allocation (LDA), one of the most common tasks for extracting latent topics from text documents [31,32]. A graphical model of LDA is shown in Figure 1, where M denotes the number of documents, N_m is the number of words in a given document of which id is m, $\vec{\theta}_m$ is the expected topic proportion of document m, generated by a Dirichlet distribution parameter $\vec{\alpha}$. $\vec{\varphi}_k$ is the word distribution of topic k, which is generated by a Dirichlet parameter $\vec{\beta}$. $z_{m,n}$ and $w_{m,n}$ are the topic for nth word and the word in nth position in document m, respectively [33]. In this study, LDA was used as a form of implementation in the *gensim* package [34].

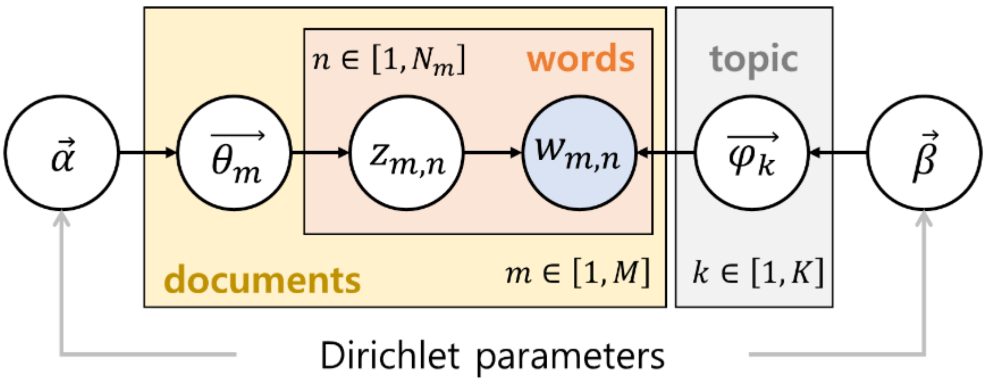

Figure 1. Graphical model of Latent Dirichlet Allocation (LDA).

2.3. Keywords Extraction, Executive Summary, and Translation

A couple of transformer deep learning models were applied for keyword extraction from the abstract. TextRank, suggested by Mihalcea et al., is a graph-based word ranking algorithm motivated by PageRank on websites by Brin and Page, in which a weight of a website is determined by relative importance that "more important" website has more links from others [35,36]. Likewise, in TextRank, a word is determined by weights of words calculated in the same manner,

$$WS(V_i) = (1-d)d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j), \quad (1)$$

where $WS(V_i)$ is a TextRank for word V_i, and w_{ij} is a weight. d, a damping factor representing the probability of jumping from a given word to another one in the graph, was set to 0.85. TextRank is implemented and released as a form of *pyTextRank*, an extension of Python's *spaCy* pipeline for natural language processing (NLP) [37,38]. Due to the discrepancies in keywords extraction by methodologies, we have applied *KeyBert* as a complementary [39]. The main algorithm of *KeyBERT* is based on *BERT*, an abbreviation of Bidirectional Encoder Representations from Transformers, regarding context, which is often ignored by statistics-based language models, and therefore more meaningful keywords would be acquired [40]. Lemmatization of the keywords was performed using *spaCy*, and unification of the synonyms as standard terms was implemented, taking advantage of Wikipedia's redirection process [38,41–43].

In this study, abstract summary and sequential translation are one of the most crucial parts. TLDR, a transformer-based executive summary model, was chosen as a summarization model due to its specification on the SciTLDR dataset, with 5400 summaries on over 3200 scientific papers [22]. Translation of the summary was dependent on *Google Translate*, which served as an Application Programming Interface (API) service through *RapidAPI* [44].

Finally, the analysis results were summarized in the form of a Microsoft Word document, in which word clouds—comprising the keywords extracted from the website articles—were included [45,46].

The analysis flowchart used in this study is shown in Figure 2, performed on Jupyter Lab, an integrated development environment for Python [47]. Contrary to academic literature data, web-posted articles are studied in detail. The paragraphs are summarized and translated to improve delivery, and the words are counted, transformed, and redirected to find reliable messages, and finally visualized as a word cloud to note the beginning of a chapter in a generated report. Most of the processes are automatic, except for data analysis and error handling processes, including redirection.

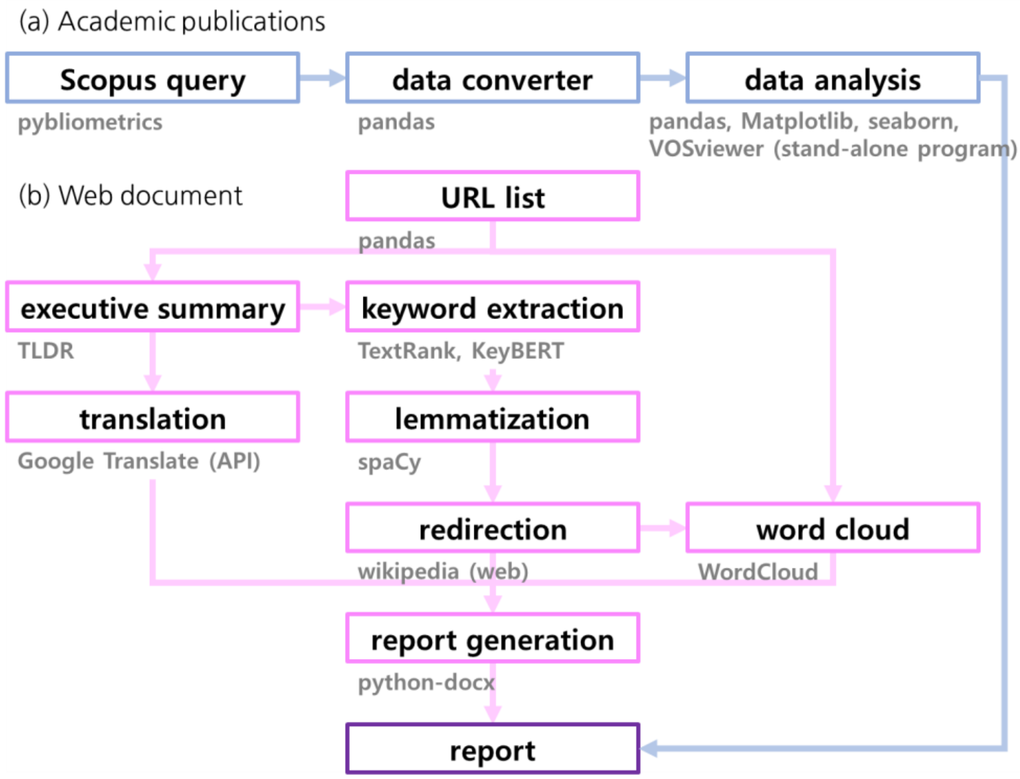

Figure 2. Schematic of the analysis of a foreign institution, (**a**) academic publications, and (**b**) web documents. Names of the Python packages and data analysis programs are annotated below each stage; the types of usage are parenthesized if not a Python package.

3. Results

3.1. Overview: International Collaboration Status of an Institution

Academic publications of the NREL are classified by contributing affiliations and countries, and domestic (inside the United States) publications are subdivided into (1) stand-alone research performed by the NREL and the United States Department of Energy (DOE) and (2) collaborations between the laboratories governed or established by the DOE [48], (3) including other institutes outside of the DOE (for example, the Illinois Institute of Technology or University of Georgia) [49,50]. The number and proportion of publications during a given period are shown in Figure 3. The number of publications gradually increases, with the proportion in each category being practically constant. It should be noted that the articles for the year 2022 are still in progress. The only remarkable change as a function of time was the proportion of publications among DOE laboratories. The average for the first three years was 2.0% (region ①); however, for the last three years, this value doubled to 4.4% (region ②). Approximately 29% of the studies were conducted in collaboration with foreign institutions.

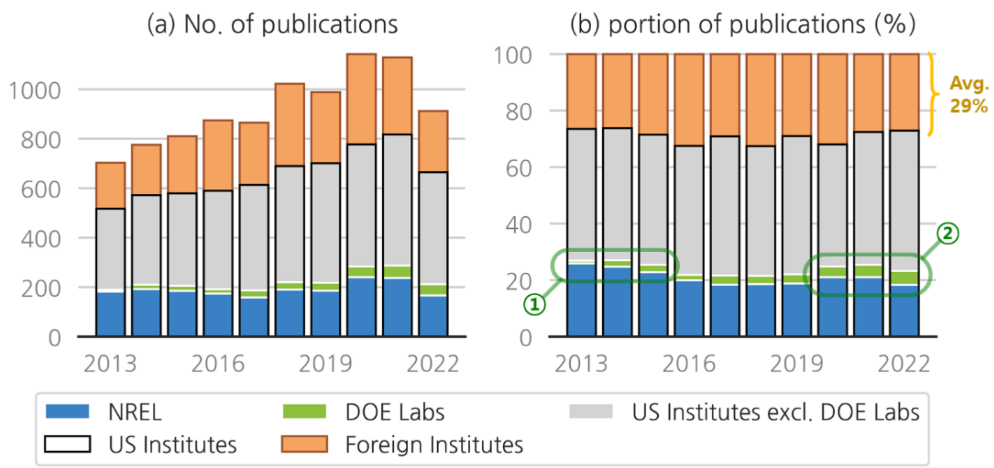

Figure 3. Types of NREL collaborative research (2013~2022). The same data is displayed from two points of view: (**a**) number and (**b**) proportion of publications. "US Institutes" includes publications of the NREL alone or the NREL and DOE together (co-authorship with other laboratories of the DOE), and with other institutions excluding the DOE laboratories but within the United States.

Figure 4 shows the NREL's international collaboration network. Eighty countries, including the United States, are clustered into twenty-two groups by frequency of co-authorship, denoted by color. The size of the bubbles and the width of the connecting lines represent the number of publications by country and co-authorship, respectively. Apart from the United States, China, Germany, and the United Kingdom are the countries with the strongest total collaborative links, a summation of other countries being weighted by co-authored publications.

The number of collaborative publications in the top 12 countries is shown in Figure 5. Germany, Denmark, and Australia follow the NREL's total publications trend shown in Figure 3a, where the United Kingdom is rapidly rising as a close partner. This is more clearly illustrated by a direct comparison of the number of co-authored publications between 2013 and 2022 (Figure 6). In most countries, the number of collaborative publications with other countries also increases as the total number of publications increases. There are a few countries with decreasing collaborations, namely South Korea.

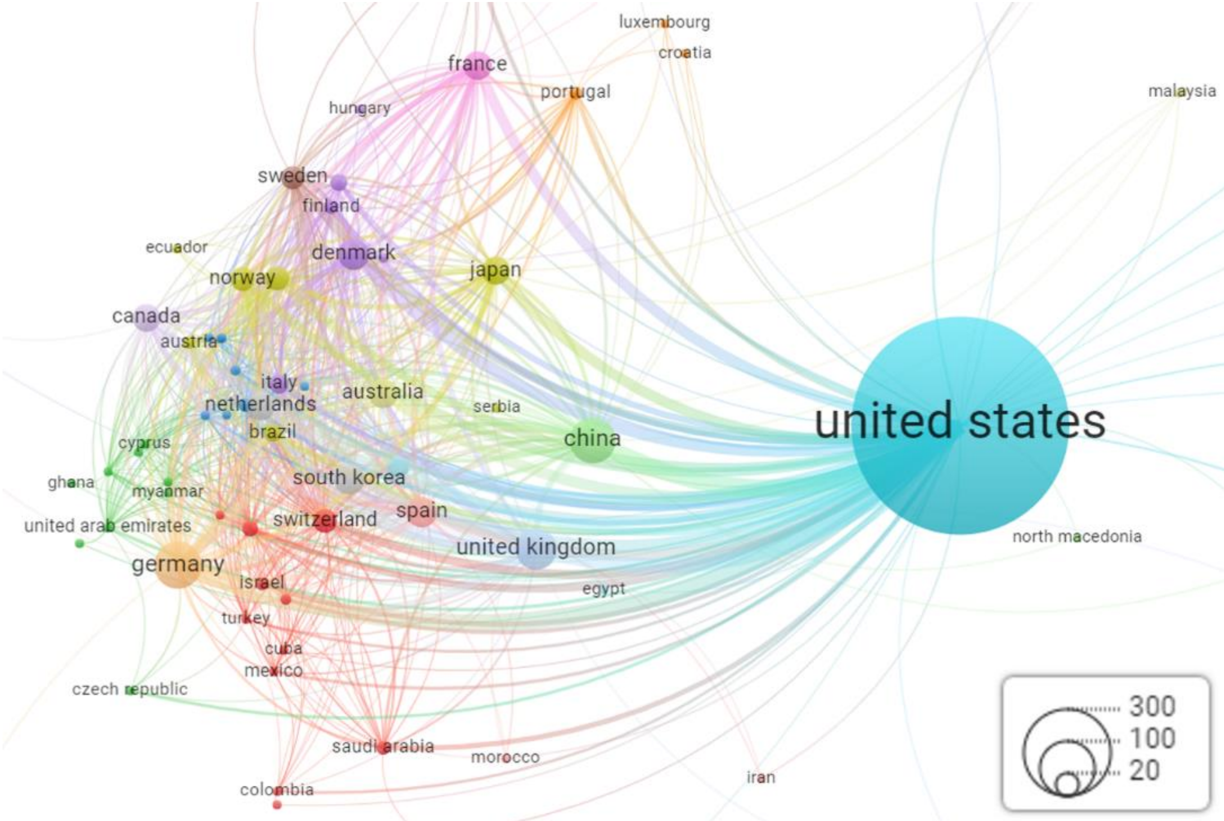

Figure 4. The NREL's collaboration network. The size of the nodes and the width of the edges represent the number of publications of a country and the co-authored ones, respectively.

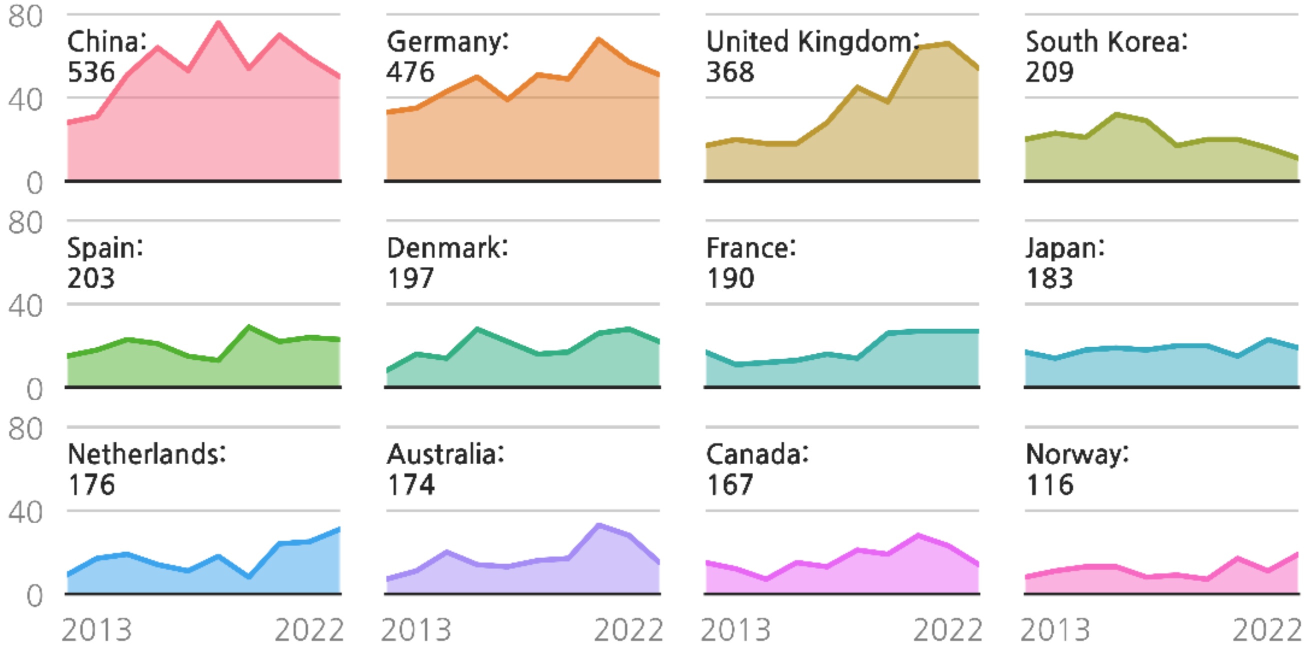

Figure 5. The number of co-authored publications in the top 12 countries. The number of publications with the NREL is annotated beneath the country name.

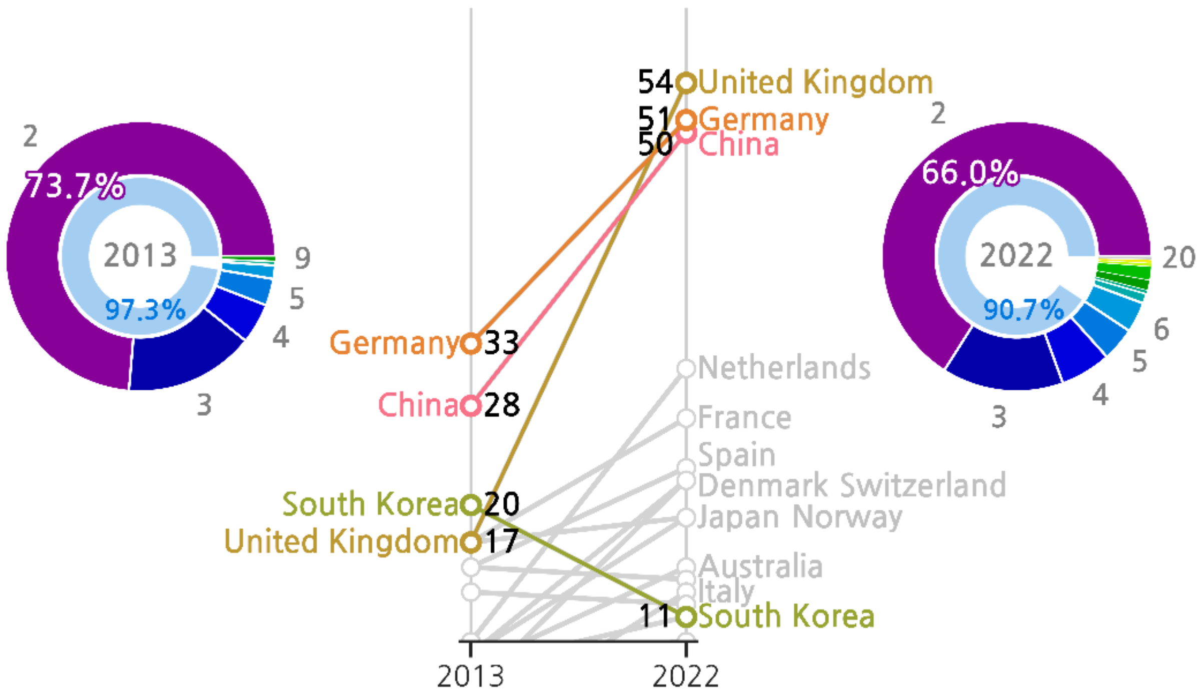

Figure 6. Change in the number of co-authored publications by country between 2013 and 2022. Not all country names are displayed to avoid overlapping, particularly in 2022. Inset donut charts represent the proportions of publications by international collaboration. The number of co-authored countries is, attached next to the corresponding wedges; some annotations are omitted for better visualization of the maximum number.

The country names In 2013 and overlapped ones in 2022 are omitted except for the topmost three countries in 2013 and 2022. As reported in other research on international collaboration, the number of countries participating in a publication is increasing [4], despite the constant proportion of international collaborations of the NREL (Figure 3b). In 2013, approximately 74% of the research was conducted by two countries, between the United States and another, but in 2022 this proportion dropped to 66%, whereas the number of publications by equal or smaller than five countries decreased from 97.3% to 90.7%. Additionally, it was recorded that 20 countries participated in a worldwide photovoltaic efficiency study in 2022, a phenomenon that can be inferred to be a form of strengthening international collaboration [51].

3.2. Research Topics: Collaboration Subject Candidates

Topic Modeling

In order to figure out research trends by topic as well as the strengths of each institution, LDA was applied to the two institutions' 12,548 abstracts of the last decade. By dropping the articles of which the abstract was absent, the number of articles published by the NREL and KIER slightly dropped to 8922 and 3626, respectively. The number of topics was determined as 9, which almost maximizes the coherence score (C_v) as 0.54, where the values above 0.5 are fairly good [52]. The maximum C_v, 0.55 was found at 29 topics (Figure 7), but we judged that the cost of analyzing 19 more categories exceeds the benefits.

The first five keywords contributing to the topic definition are summarized in Table 1 with the corresponding top three references with the largest sum of keyword weights on this topic. For example, the first category consisted of the keywords "cell", "solar", "device", "film", "efficiency", and others. Reviewing the articles belonging to the articles of the first topic group [53–55], the name of the topic was chosen as "solar cell", and all other topic names were determined in the same manner.

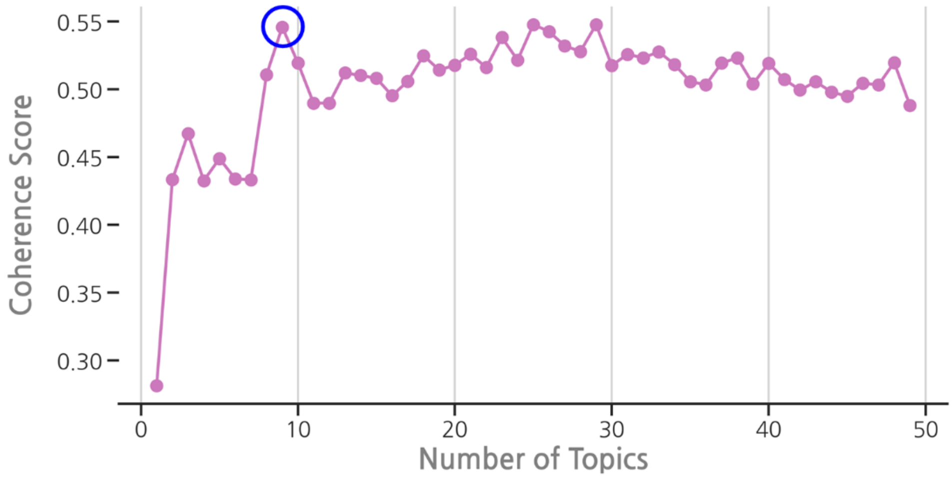

Figure 7. Coherence score (C_v), taken by varying the number of topics from 1 to 49.

Table 1. The main topic and top 5 keywords for each topic category.

Topic No.	Topic Name	Keywords (Top 5)	Reference (Top 3)
1	solar cell	cell, solar, device, film, efficiency	[53–55]
2	catalysts and electrodes	catalyst, surface, membrane, electrode, reaction	[56–58]
3	bioproduction	lignin, biomass, production, enzyme, cellulose	[59–61]
4	energy systems	energy, system, power, cost, model	[62–64]
5	materials science	material, module, structure, property, energy	[65–67]
6	biomass fuel	fuel, gas, process, catalyst, temperature	[68–70]
7	heat system	heat, thermal, system, power, temperature	[71–73]
8	wind turbine	model, wind, turbine, datum, method	[74–76]
9	photophysics	carrier, si, electron, surface, film	[77–79]

The topics were visualized on a principal component analysis (PCA) space Figure 8, which consisted of the first and second principal components (PC1 and PC2, respectively) with assistance from the python visualization package for LDA, *pyLDAvis* [80]. The size of the circles denotes the probability of documents. In the PCA space, it is expected that the topics located nearby are physically close. Nine categories might be regarded as four clusters as a function of the azimuthal locations on the PCA space. The right half (positive PC1) is almost of systems—heat system, energy system, and wind turbine. The topmost group is biomass—biomass fuel and bioproduction, whereas the bottommost one is photoelectric physics, materials, and application (solar cell). Another circle close to the electrochemical reaction (catalyst and electrode) is isolated.

The smallest category is photophysics (6.49%), whereas the largest is energy systems (19.66%) of entire works of literature. If we can find differences between the NREL and the KIER in each category, it would be beneficial for the international collaboration strategy decision. In addition, if we can trace the number of publications as a function of time, it would also be helpful to figure out if a category has been popular in the past but decaying or emerging recently. Since works of literature are decomposed as a sum of keywords through LDA, the number of papers should be expressed as a sum of weights. Analysis of these viewpoints is visualized in Figure 9. Horizontal and vertical axes denote time and category, respectively. The sum of weights is represented as a colormap—the lighter, the more. Due to the fact that the total number of publications of the NREL is about 2.5 times larger than the KIER, the left panel (NREL) is lighter than the right one in general.

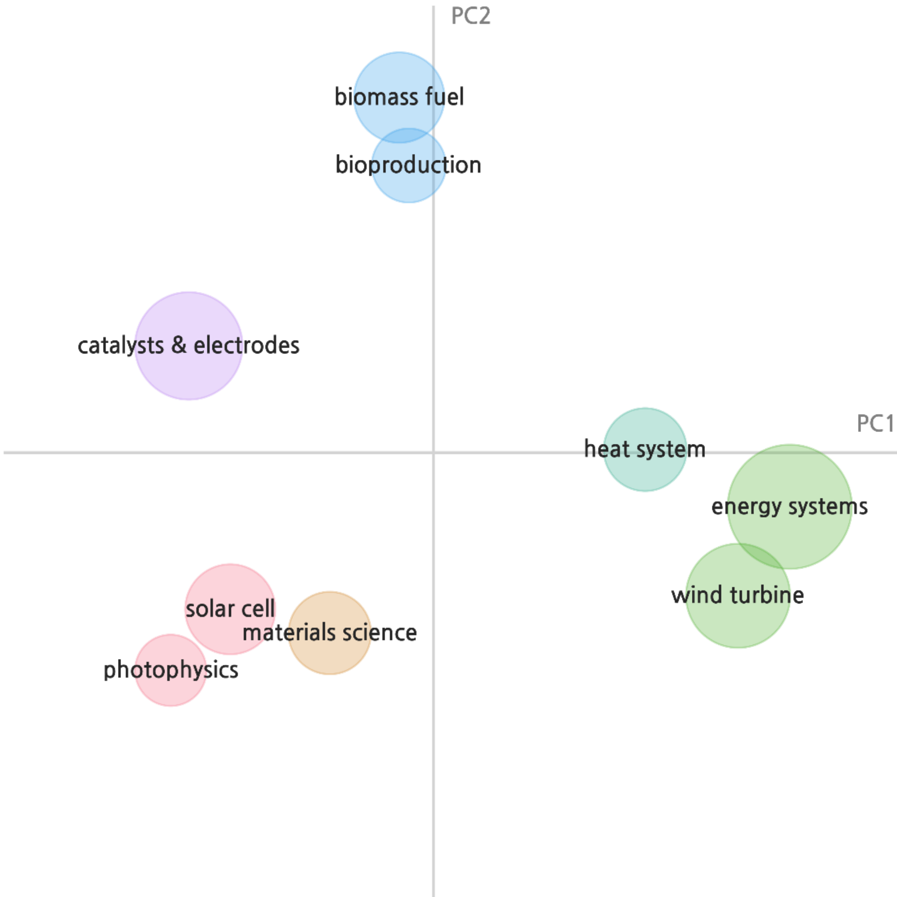

Figure 8. Intertopic distance map of topics on the PCA space.

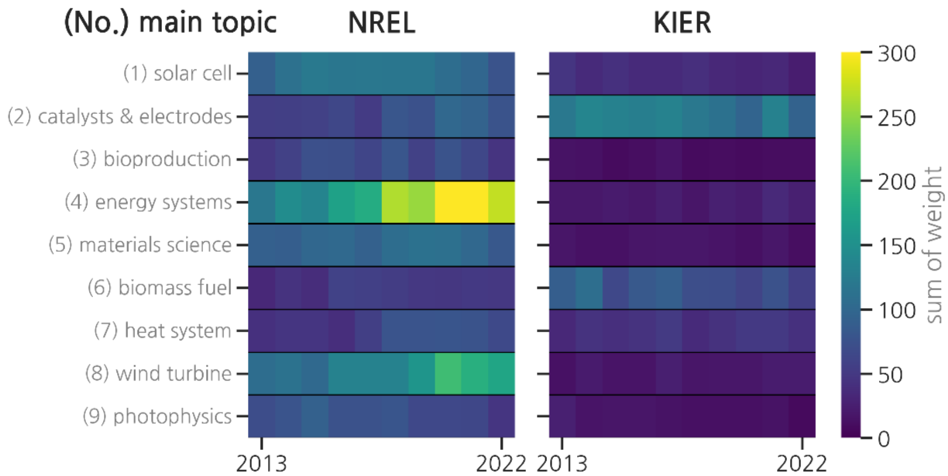

Figure 9. Weight summation as a function of time for nine topics as a function of time, with comparisons between the two institutions.

Interestingly in some topics—catalyst and electrode, and biomass fuel—the KIER has a larger sum of weights than the NREL. In addition, the difference between the two institutions in the heat system is not as much as the global publication amount. It can be hypothesized that the KIER or the Korean government has focused on a few topics to be more competitive. On the other hand, the sum of weights in the NREL is quite diverse. Energy systems and wind turbines occupy the largest portion, particularly the growth of energy systems in the recent few years is outstanding.

The difference between the two institutions provides a clear guideline for international collaboration.

3.3. Web Articles: Information a Step Forward towards Public and Industry

3.3.1. Summarization and Translation

Contrary to the academic literature analyzed from a global view, web articles are processed individually for efficient filtering. The one who is in charge of international collaboration is supposed to read all web-posted articles because the articles are likely to be friendlier than academic papers for non-researchers; however, the problem is that the total amount could be too much for manual processing. Consequently, our strategy for web articles was summary and translation, as shown on the left branch of Figure 2b. An example of the process is represented in Figure 10. The article was about a heavy-duty hydrogen truck that consisted of 876 words [81]. *TLDR* summarizes it into 69 words in which the main message and important details, i.e., keywords and abbreviations, are kept and finally translated to Korean. The translation not only substituted English sentences with Korean but also maintained the most important words using parenthesis, for example, " 국립 재생 에너지 연구소 (National Renewable Energy Laboratory)" and " 연료 전지 전기 자동차 (FCEV)".

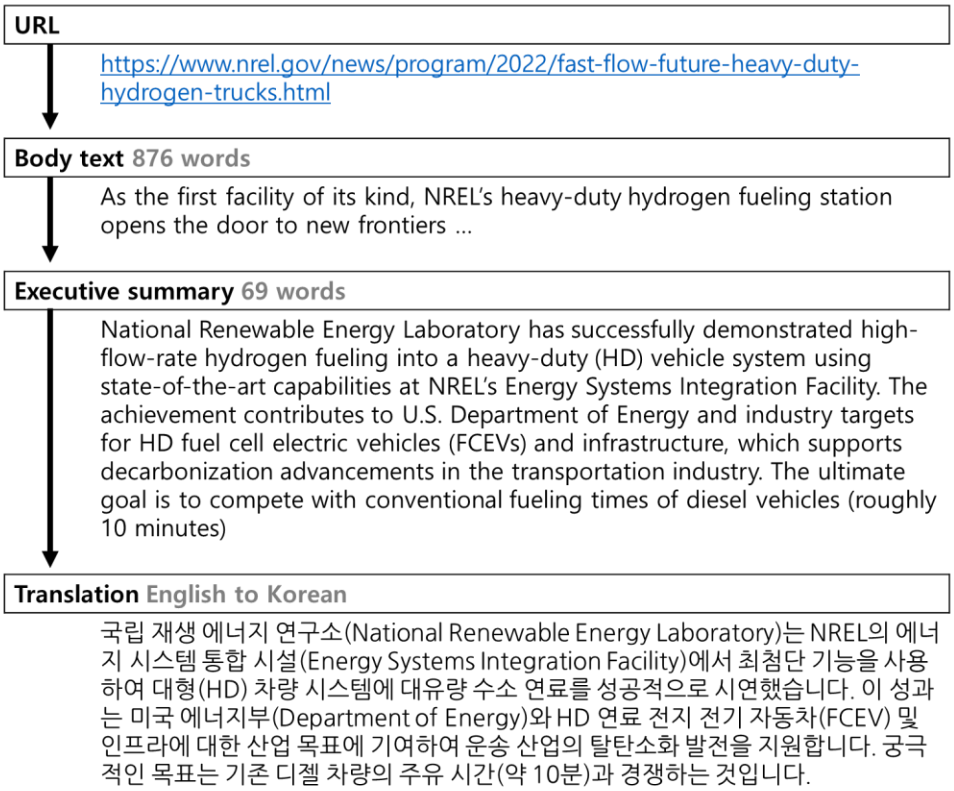

Figure 10. An example of the text translation process followed by an executive summary. The paragraph of the last step is the translation from English to Korean version.

3.3.2. Keyword Extraction

The other way, other than summarization and translation in Figure 2b, is for keyword extraction. The process with the same example is shown in Figure 11. The two keyword extraction algorithms, *TextRank* and *KeyBERT*, extract different keywords from an independent aspect. Both results are passed through lemmatization and redirection processes using *spaCy* and *Wikipedia* and finally merged. It is worthwhile to take a look before and after *Wikipedia* redirection. This process is for the treatment of synonyms utilizing the dictionary of Wikipedia; therefore, the keyword "clean hydrogen" is replaced with "green hydrogen". The inputs and outputs of redirection are added to the dictionary to be saved in local storage to reduce redundant requests from the Wikipedia website [42]. Finally, the keywords are merged and placed beside the summary and translation, as well as converted to a word cloud for report decoration and insights delivery [46]. The abundant keywords extracted by *pyTextRank* can be cleaned up using a similarity measurement such as cosine similarity [82]; however, they can be used as raw materials for knowledge databases in the future. Nevertheless, with the possible usefulness, it is probable for readers to feel redundancy. In further studies on and using the described methodology, the number of similar words is going to be controlled by readers' feedback.

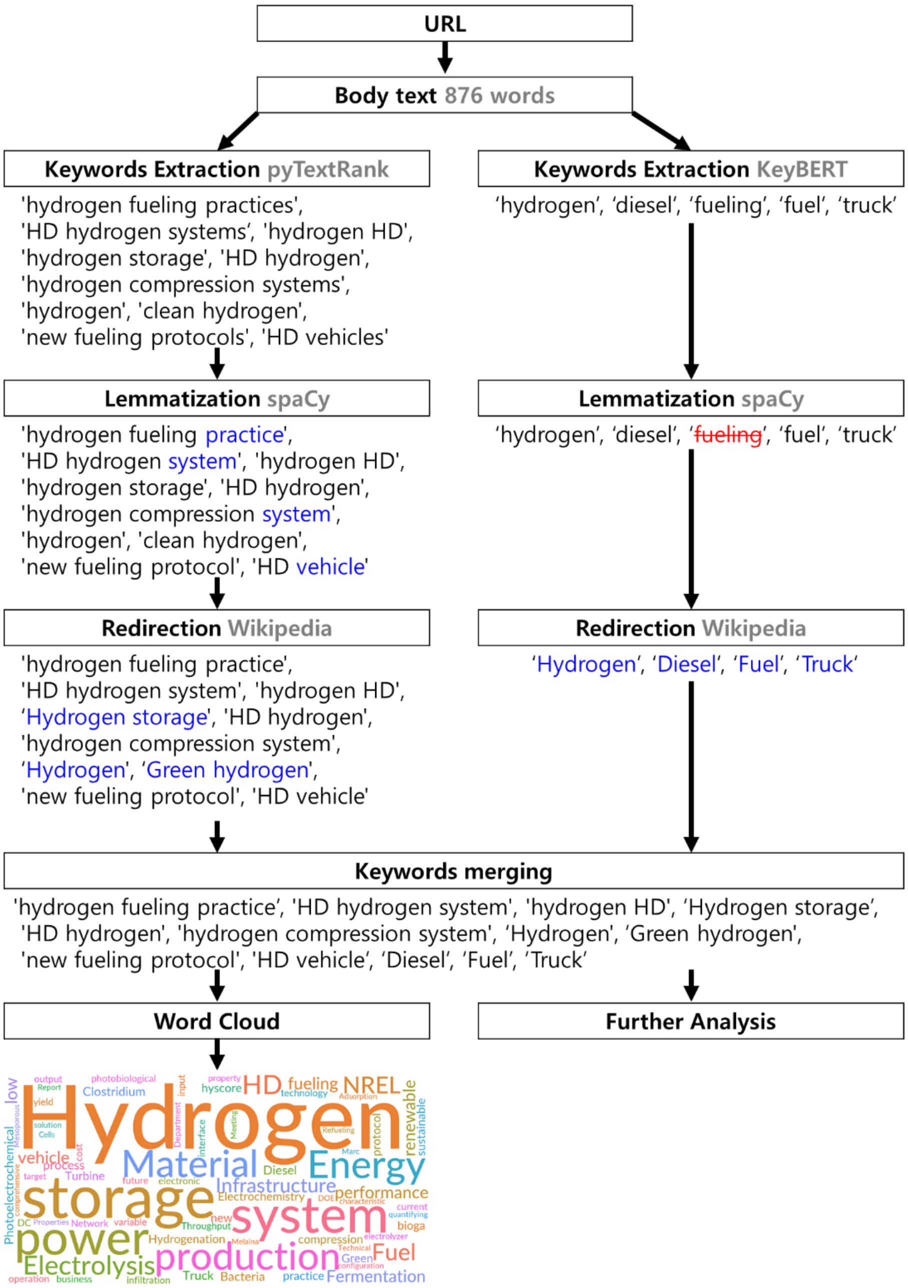

Figure 11. A flowchart of the keyword extraction process. The different approaches of pyTextRank and KeyPERT resulted different keywords, those are finally merged to form a word cloud and further analysis.

3.3.3. Report Generation

Features of the articles, title, URL, publication date, image, etc., crawled from web pages are sequentially composed with our NLP results—executive summary, translation, keywords, and word cloud—as a form of the Microsoft Word document utilizing the *Python-docx* package [45]. The composition of a page is carefully designed for readability. The hyperlink is assigned to the URL to allow interactivity, and the word cloud is placed between sessions to notice the beginning of a new topic. The word clouds are decorations rather than functional objects as shown in Figure 12. In order to maintain the reader's concentration, aesthetic stimulation is important for refreshment [83]. Therefore, we inserted a few word clouds as well as representative images from each of the web articles.

2. *"Fast Flow Future for Heavy-Duty Hydrogen Trucks"*

- **URL :** https://www.nrel.gov/news/program/2022/fast-flow-future-heavy-duty-hydrogen-trucks.html
- **keywords :** Hydrogen, Diesel, Fuel, Truck
- **keyphrases:** hydrogen fueling practice, HD hydrogen system, hydrogen HD, Hydrogen storage, HD hydrogen, hydrogen compression system, Hydrogen, Green hydrogen, new fueling protocol, HD vehicle
- **요약 (국문) :** 국립 재생 에너지 연구소(National Renewable Energy Laboratory)는 NREL 의 에너지 시스템 통합 시설(Energy Systems Integration Facility)에서 최첨단 기능을 사용하여 대형(HD) 차량 시스템에 대유량 수소 연료를 성공적으로 시연했습니다. 이 성과는 미국 에너지부(Department of Energy)와 HD 연료 전지 전기 자동차(FCEV) 및 인프라에 대한 산업 목표에 기여하여 운송 산업의 탈탄소화 발전을 지원합니다. 궁극적인 목표는 기존 디젤 차량의 주유 시간(약 10 분)과 경쟁하는 것입니다.
- 요약 (영문) : National Renewable Energy Laboratory has successfully demonstrated high-flow-rate hydrogen fueling into a heavy-duty (HD) vehicle system using state-of-the-art capabilities at NREL's Energy Systems Integration Facility. The achievement contributes to U.S. Department of Energy and industry targets for HD fuel cell electric vehicles (FCEVs) and infrastructure, which supports decarbonization advancements in the transportation industry. The ultimate goal is to compete with conventional fueling times of diesel vehicles (roughly 10 minutes)

Figure 12. Screenshot of the report generated using the processes discussed in this paper. Hyperlinks are applied to the URLs of the articles for readers to access the original article. The Korean parts marked with bullets stands for summary, for example, "요약 (국문)" stands for "summary (in Korean)" where "" for "summary (in English)".

Finally, literature data analysis on 12,886 academic publications and 54 web-posted articles were summarized as a 57-page report written in the native language of the readers, containing an overview of the target institution's international collaborations over a 10-year period and recent activities as well as a comparison with the author's institute.

4. Discussion

Throughout the process proposed in this study, it is evident that the efficiency of information acquisition is dramatically enhanced. Readers can capture a sector trend by academic literature analysis of thousands of publications and then determine the current activities by reading the abridged versions of web-based articles. In spite of crude word-forward translation due to the imperfect machine translation lacking knowledge of a specific research field, paragraphs written in the reader's native language improve understanding efficiency enormously.

Furthermore, taking advantage of machine translation opens another possibility—that is, the retrieval and generation of reports in any language. These days, many deep learning-based models support multilingual processes [23,40,84,85]. By simply changing the input and output language options, it is possible to obtain information from and to any language our model supports (e.g., Russian, Persian, Chinese, etc.).

Applying the research topic analysis process proposed in this study in planning a joint international project can effectively find joint research areas that are mutually understandable and beneficial to each other. In this study, assuming international collaboration between the NREL in the United States and the KIER, the field of superior comparative research for efficient international collaboration was evaluated. The total number of publications of the NREL is about 2.5 times larger than that of the KIER, so it is generally supposed that the NREL is more capable, as revealed in the fields of energy systems and wind turbines. However, the KIER has some superior topics, such as catalysts and electrodes, biomass fuel, and heat systems, so mutual benefit could be achieved by including the topics in the proposal. It is obvious that a sufficient understanding of each institution's relative advantages would reduce the risk of failure.

5. Conclusions

As a result of the analysis of academic literature and web-based articles, detailed research areas of two institutions—for instance, KIER and NREL—were identified. To understand the research areas that can be expected to be synergistic when conducting international joint research, the papers published by two institutions over the past 10 years were analyzed. Topic modeling inferred areas where the two institutions have focused their research capabilities over the past decade. As a result, catalysts and electrodes, biomass fuel, energy systems, and wind turbines are chosen as promising candidates for international collaboration. However, since this expectation is only from the quantitative study, an in-depth discussion with the specialists in those fields would be essential for successful international collaboration planning. Furthermore, utilizing recently developed deep learning NLP techniques, a workflow for summarizing a large number of web articles into a set of short reports was achieved in only a few minutes.

It should be stated that although we have presented the methodology, there is room for improvement in performance. This is because the purpose of this study is to secure feasibility, so the comparative analysis of various topic extraction algorithms or document summary algorithms was omitted.

Nevertheless, the valuation of this work can be calculated by converting the salary of employees dedicated to survey and report generation of foreign institutions for international collaboration. Assuming that it takes 10 min and 1 h to read an abstract and a web article, respectively, 12,548 papers and 54 web articles correspond to 2145 h (268 days assuming 8 h/day) of work, which exceeds the number of working days in a year. In conclusion, our work is worthwhile to be an annual salary of an experienced worker. The model developed in this study should be used in the international R&D planning process to analyze the technology expertise and research domain of partner institutions; the performance of international cooperation could be greatly improved.

6. Patents

The data processing and report generation process described in this article is a patent pending in Korea (application number: 10-2022-0080454).

Author Contributions: Conceptualization, J.L., J.J. and S.C.; methodology, J.L.; software, J.L.; validation, J.J., S.C. and W.B.; data curation, J.J.; writing—original draft preparation, J.L. and J.J.; writing—review and editing, J.L. and J.J.; visualization, J.L.; supervision, J.J. and S.C.; project administration, J.L. and W.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was conducted under the framework of Research and Development Program of the Korea Institute of Energy Research (KIER) (C2-2444 and C2-2447).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Publications Output: U.S. Trends and International Comparisons. 2019. Available online: https://ncses.nsf.gov/pubs/nsb20206/international-collaboration (accessed on 29 November 2022).
2. UNESCO. *UNESCO Science Report: Towards 2030*; UNESCO: Paris, France, 2015.
3. Kim, Y.-I.; Lee, B.-M. A Study on Improving The Outputs of International Cooperation in Science and Technology: The Case of International S&T Cooperation Programs of the Ministry of Education, Science and Technology (MEST). *J. Korea Technol. Innov. Soc.* **2009**, *12*, 545–563.
4. Herrmannova, D.; Knoth, P. *An Analysis of the Microsoft Academic Graph*; D-Lib Magazine: Reston, VA, USA, 2016.
5. Shin, H.-D.; Chung, T.; Ryu, C.; Lee, J. What Drives International Science and Technology Cooperation. *J. Korea Technol. Innov. Soc.* **2010**, *13*, 638–655.
6. Leal Filho, W.; Wall, T.; Barbir, J.; Alverio, G.N.; Dinis, M.A.P.; Ramirez, J. Relevance of International Partnerships in the Implementation of the UN Sustainable Development Goals. *Nat. Commun.* **2022**, *13*, 613. [CrossRef] [PubMed]
7. Wang, S.; Wang, X.; Chen, S. Global value chains and carbon emission reduction in developing countries: Does industrial upgrading matter? *Environ. Impact Assess. Rev.* **2022**, *97*, 106895. [CrossRef]
8. International 'Collaboration Gap' Threatens to Undermine Climate Progress and Delay Net Zero. 2022. Available online: https://www.irena.org/News/pressreleases/2022/Sep/International-collaboration-gap-threatens-to-undermine-climate-progress-and-delay-net-zero (accessed on 29 November 2022).
9. IEA. *International Collaboration in Energy Technology*; IEA: Paris, France, 2000.
10. Energiepartnershaft Deutchland-Korea. 2022. Available online: https://www.energypartnership-korea.org/home/ (accessed on 29 November 2022).
11. Jung, J.; Choi, S.; Baek, W. Development of Co-beneficial Global Research Process for International Cooperation Performance. In Proceedings of the Korea Energy Society Conference, Daegu, Republic of Korea, 28–30 April 2021.
12. Jung, J.; Choi, S.; Lee, A.; Seo, J.; Park, M. Analysis on Foreign Institution for R&D Planning of International Cooperation in Hydrogen Technology. In Proceedings of the Korea Energy Society Conference, Busan, Republic of Korea, 31 October–1 November 2019.
13. Lee, H.; Chung, S. Study on the Strategies for Promoting Defense International Technology Cooperation: Based on the Korean and Indian Defense Joint Technology Cooperation. *J. Korea Technol. Innov. Soc.* **2014**, *17*, 584–603.
14. Chang, H.; Yun, H. Selection of International Cooperation Partners Based on the Bibliometrics Analysis: A Case Study of Fusion Energy Field. *J. Korea Technol. Innov. Soc.* **2018**, *21*, 213–225.
15. Fire, M.; Guestrin, C. Over-optimization of academic publishing metrics: Observing Goodhart's Law in action. *GigaScience* **2019**, *8*, giz053. [CrossRef]
16. Lieberman, R.; Nelson, W.G.; Sakr, W.A.; Meyskens, F.L., Jr.; Klein, E.A., Jr.; Wilding, G.; Partin, A.W.; Lee, J.J.; Lippman, S.M. Executive summary of the National Cancer Institute workshop: Highlights and recommendations. *Urology* **2001**, *57*, 4–27. [CrossRef]
17. Hull, R.; Vaid, J. Bilingual language lateralization: A meta-analytic tale of two hemispheres. *Neuropsychologia* **2007**, *45*, 1987–2008. [CrossRef]
18. Tikofsky, R.S.; Eric, H. *Lenneberg, Biological Foundations of Language*; John Wiley and Sons: New York, NY, USA, 1967; Volume 13, pp. 493–495.
19. Paulus, R.; Xiong, C.; Socher, R. A Deep Reinforced Model for Abstractive Summarization. In Proceedings of the 2018 International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
20. Zhang, X.; Wei, F.; Zhou, M. HIBERT: Document Level Pre-training of Hierarchical Bidirectional Transformers for Document Summarization. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July—2 August 2019.

21. Aghajanyan, A.; Shrivastava, A.; Gupta, A.; Goyal, N.; Zettlemoyer, L.; Gupta, S. Better Fine-Tuning by Reducing Representational Collapse. *arXiv* **2020**, arXiv:2008.03156.
22. Cachola, I.; Cohan, A.; Weld, D.S. TLDR: Extreme Summarization of Scientific Documents. *arXiv* **2020**, arXiv:2004.15011.
23. Conneau, A.; Lample, G. Cross-Lingual Language Model Pretraining. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.
24. Guz, A.N.; Rushchitsky, J.J. Scopus: A system for the evaluation of scientific journals. *Int. Appl. Mech.* **2009**, *45*, 351. [CrossRef]
25. National Renewable Energy Laboratory. Available online: https://www.nrel.gov/ (accessed on 10 September 2022).
26. Rose, M.E.; Kitchin, J.R. Pybliometrics: Scriptable bibliometrics using a Python interface to Scopus. *SoftwareX* **2019**, *10*, 100263. [CrossRef]
27. McKinney, W. *Data Structures for Statistical Computing in Python*; SciPy: Austin, TX, USA, 2010.
28. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
29. Waskom, M. Seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [CrossRef]
30. Van Eck, N.J.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523–538. [CrossRef]
31. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
32. Blei, D.M.; Jordan, M.I. Modeling Annotated Data. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '03, Toronto, Canada, 28 July–1 August 2003. [CrossRef]
33. Lee, J.; Kang, J.H.; Jun, S.; Lim, H.; Jang, D.; Park, S. Ensemble Modeling for Sustainable Technology Transfer. *Sustainability* **2018**, *10*, 2278. [CrossRef]
34. Rehurek, R.; Sojka, P. *Gensim–Python Framework for Vector Space Modelling*; NLP Centre, Faculty of Informatics, Masaryk University: Brno, Czech Republic, 2011; p. 3.
35. Mihalcea, R.; Tarau, P. TextRank: Bringing Order into Text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, 25–26 July 2004; Association for Computational Linguistics: Stroudsburg, PA, USA, 2004; pp. 404–411.
36. Page, L.; Brin, S.; Motwari, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Technical Report 1999–66; Computer Science Department, Stanford University: Stanford, CA, USA, 1999.
37. Nathan, P. *PyTextRank, A Python Implementation of TextRank for Phrase Extraction and Summarization of Text Documents*; Derwen: Sebastol, CA, USA, 2016.
38. Honnibal, M.; Montani, I. SpaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing. 2017. Available online: https://github.com/explosion/spaCy/issues/5863 (accessed on 11 September 2022).
39. Grootendorst, M. KeyBERT: Minimal Keyword Extraction with BERT. 2020. Available online: https://github.com/MaartenGr/KeyBERT (accessed on 11 September 2022).
40. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. *BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding*; Association for Computational Linguistics: Minneapolis, MN, USA, 2018.
41. Lee, J.; You, S.; Kim, C.K.; Oh, M.; Kim, B.; Kang, Y.H.; Kim, H.G. Trend Summary by SCOPUS Literature Analysis on Machine Learning Assisted BIPV Research. *J. Korean Sol. Energy Soc.* **2022**, *42*, 13–32. [CrossRef]
42. Wikipedia. Available online: https://www.wikipedia.org/ (accessed on 12 September 2022).
43. Lee, J.; You, S.; Kim, C.K.; Kim, H.G. Accelerated Literature Analysis using Open API. In Proceedings of the Applied Artificial Intelligence Conference, Daejeon, Republic of Korea, 1–2 July 2022.
44. RapidAPI. Available online: https://rapidapi.com/ (accessed on 11 September 2022).
45. Python-Docx. Available online: https://python-docx.readthedocs.io/ (accessed on 11 September 2022).
46. Oesper, L.; Oesper, L.; Merico, D.; Isserlin, R.; Bader, G.D. WordCloud: A Cytoscape plugin to create a visual semantic summary of networks. *Source Code Biol. Med.* **2011**, *6*, 7. [CrossRef]
47. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter Notebooks–A publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*; Loizides, F., Schmidt, B., Eds.; IOP Press: Amsterdam, The Netherlands, 2016; pp. 87–90.
48. National Laboratories, Department of Energy. Available online: https://www.energy.gov/national-laboratories (accessed on 12 September 2022).
49. Huang, J.; Zhou, X.; Cui, B. Online distribution system state estimation via stochastic gradient algorithm. *Electr. Power Syst. Res.* **2022**, *213*, 108305. [CrossRef]
50. Hengge, N.N.; Black, I.M.; Atmodjo, M.A.; Mohanty, S.S.; Ryno, D.; Himmel, M.E.; Azadi, P.; Bomble, Y.J.; Mohnen, D. Composition and yield of non-cellulosic and cellulosic sugars in soluble and particulate fractions during consolidated bioprocessing of poplar biomass by Clostridium thermocellum. *Biotechnol. Biofuels Bioprod.* **2022**, *15*, 23.
51. Kinsey, G.S.; Riedel-Lyngskær, N.C.; Miguel, A.-A.; Boyd, M.; Braga, M.; Shou, C.; Cordero, R.R.; Duck, B.C.; Fell, C.J.; Feron, S.; et al. Impact of measured spectrum variation on solar photovoltaic efficiencies worldwide. *Renew. Energy* **2022**, *196*, 995–1016. [CrossRef]

52. McLevey, J. *Doing Computational Social Science: A Practical Introduction*; SAGE Publications Limited: Thousand Oaks, CA, USA, 2022.
53. Abusnina, M.; Matin, M.; Moutinho, H.; Al-Jassim, M. Impact of the Stack Order in Cu-Zn-Sn Metal Precursors on the Properties of Cu_2ZnSnS_4 Thin Films. In Proceedings of the 2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC), New Orleans, LA, USA, 14–19 June 2015.
54. Jain, N.; Oshima, R.; France, R.; Geisz, J.; Norman, A.; Dippo, P.; Levi, D.; Young, M.; Olavarria, W.; Steiner, M.A. Development of Lattice-Matched 1.7 eV GalnAsP Solar Cells Grown on GaAs by MOVPE. In Proceedings of the 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC), Portland, OR, USA, 5–10 June 2016.
55. Veinberg-Vidal, E.; Vauche, L.; Weick, C.; Da Fonseca, J.; Jany, C.; Morales, C.; Lecouvey, C.; Desrues, T.; Voarino, P.; Fournel, F.; et al. Wafer-Bonded AlGaAs///Si Dual-Junction Solar Cells. In Proceedings of the 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC), Washington, DC, USA, 25–30 June 2017.
56. Godínez-Salomón, F.; Albiter, L.; Alia, S.M.; Pivovar, B.S.; Camacho-Forero, L.E.; Balbuena, P.B.; Mendoza-Cruz, R.; Arellano-Jimenez, M.J.; Rhodes, C.P. Self-Supported Hydrous Iridium–Nickel Oxide Two-Dimensional Nanoframes for High Activity Oxygen Evolution Electrocatalysts. *ACS Catal.* **2018**, *8*, 10498–10520. [CrossRef]
57. Song, H.-S.; Kim, H.-S.; Hassan, M.H.; Joh, D.-W.; Song, R.-H.; Lim, T.-H.; Hong, J.-E.; Park, S.-J.; Lee, S.-B. A dynamic infiltration technique to synthesize nanolayered cathodes for high performance and robust solid oxide fuel cells. *J. Energy Chem.* **2022**, *70*, 201–210.
58. Ingole, P.G.; Baig, M.I.; Choi, W.; An, X.; Choi, W.K.; Lee, H.K. Role of functional nanoparticles to enhance the polymeric membrane performance for mixture gas separation. *J. Ind. Eng. Chem.* **2017**, *48*, 5–15. [CrossRef]
59. Li, M.; Pu, Y.; Yoo, C.G.; Gjersing, E.; Decker, S.R.; Doeppke, C.; Shollenberger, T.; Tschaplinski, T.J.; Engle, N.L.; Sykes, R.W.; et al. Study of traits and recalcitrance reduction of field-grown COMT down-regulated switchgrass. *Biotechnol. Biofuels* **2017**, *10*, 12. [CrossRef]
60. Dumitrache, A.; Akinosho, H.; Rodriguez, M.; Meng, X.; Yoo, C.G.; Natzke, J.; Engle, N.L.; Sykes, R.W.; Tschaplinski, T.J.; Muchero, W.; et al. Consolidated bioprocessing of Populus using Clostridium (Ruminiclostridium) thermocellum: A case study on the impact of lignin composition and structure. *Biotechnol. Biofuels* **2016**, *9*, 31. [CrossRef]
61. Paye, J.M.D.; Guseva, A.; Hammer, S.K.; Gjersing, E.; Davis, M.F.; Davison, B.H.; Olstad, J.; Donohoe, B.S.; Nguyen, T.Y.; Wyman, C.E.; et al. Biological lignocellulose solubilization: Comparative evaluation of biocatalysts and enhancement via cotreatment. *Biotechnol. Biofuels* **2016**, *9*, 8. [CrossRef]
62. Simeone, C.E.; Thornton, M. Cordon screen: A cordon-based congestion pricing policy evaluation method for U.S. cities. *J. Air Waste Manag. Assoc.* **2022**, 1–15. [CrossRef]
63. Weers, J.; Driscoll, F.; Copping, A.; Ruehl, K.; Lilje, A. Portal and Repository for Information on Marine Renewable Energy PRIMRE. In Proceedings of the Offshore Technology Conference, Austin, TX, USA, 6–9 May 2019.
64. Beiter, P.; Cole, W.J.; Steinberg, D.C. Modeling the value of integrated U.S. and Canadian power sector expansion. *Electr. J.* **2017**, *30*, 47–59. [CrossRef]
65. Dimitrievska, M.; Stavila, V.; Soloninin, A.V.; Skoryunov, R.V.; Babanova, O.A.; Wu, H.; Zhou, W.; Tang, W.S.; Faraone, A.; Tarver, J.D.; et al. Nature of Decahydro-closo-decaborate Anion Reorientations in an Ordered Alkali-Metal Salt: Rb2B10H10. *J. Phys. Chem. C* **2018**, *122*, 15198–15207. [CrossRef]
66. Baranowski, L.L.; Zawadzki, P.; Lany, S.; Toberer, E.S.; Zakutayev, A. A review of defects and disorder in multinary tetrahedrally bonded semiconductors. *Semicond. Sci. Technol.* **2016**, *31*, 123004. [CrossRef]
67. Zakutayev, A.; Zhang, X.; Nagaraja, A.; Yu, L.; Lany, S.; Mason, T.O.; Ginley, D.S.; Zunger, A. Theoretical Prediction and Experimental Realization of New Stable Inorganic Materials Using the Inverse Design Approach. *J. Am. Chem. Soc.* **2013**, *135*, 10048–10054. [CrossRef]
68. Magrini, K.; Mukarakate, C.; Budhi, S.; Shetty, M.; Nimlos, M.R.; Román-Leshkov, Y. Feedstock and catalyst impact on bio-oil production and FCC Co-processing to fuels. *Biomass Bioenergy* **2022**, *163*, 106502. [CrossRef]
69. Starace, A.K.; Black, B.A.; Lee, D.D.; Palmiotti, E.C.; Orton, K.A.; Michener, W.E.; ten Dam, J.; Watson, M.J.; Beckham, G.T.; Magrini, K.A.; et al. Characterization and Catalytic Upgrading of Aqueous Stream Carbon from Catalytic Fast Pyrolysis of Biomass. *ACS Sustain. Chem. Eng.* **2017**, *5*, 11761–11769. [CrossRef]
70. Zacher, A.H.; Elliott, D.C.; Olarte, M.V.; Santosa, D.M.; Preto, F.; Iisa, K. Pyrolysis of Woody Residue Feedstocks: Upgrading of Bio-oils from Mountain-Pine-Beetle-Killed Trees and Hog Fuel. *Energy Fuels* **2014**, *28*, 7510–7516. [CrossRef]
71. Muhammad, H.A.; Cho, J.; Cho, J.; Choi, B.; Roh, C.; Ishfaq, H.A.; Lee, G.; Shin, H.; Baik, Y.J.; Lee, B. Performance improvement of supercritical carbon dioxide power cycle at elevated heat sink temperatures. *Energy* **2022**, *239*, 122216. [CrossRef]
72. Oshman, C.; Hardin, C.; Rea, J.; Olsen, M.L.; Siegel, N.; Glatzmaier, G.; Parilla, P.; Ginley, D.; Toberer, E.S. Design of a thermosyphon-based thermal valve for controlled high-temperature heat extraction. *Appl. Therm. Eng.* **2017**, *126*, 1141–1147. [CrossRef]
73. Ma, Z.; Martinek, J. Fluidized-Bed Heat Transfer Modeling for the Development of Particle/Supercritical-CO_2 Heat Exchanger. In Proceedings of the ASME 2017 11th International Conference on Energy Sustainability Collocated with the ASME 2017 Power Conference Joint with ICOPE-17, the ASME 2017 15th International Conference on Fuel Cell Science, Engineering and Technology, and the ASME 2017 Nuclear Forum, Charlotte, NC, USA, 26–30 June 2017.

74. Coulling, A.J.; Goupee, A.J.; Robertson, A.N.; Jonkman, J.M. Importance of Second-Order Difference-Frequency Wave-Diffraction Forces in the Validation of a FAST Semi-Submersible Floating Wind Turbine Model. In Proceedings of the ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering, Hamburg, Germany, 5–10 June 2013.
75. Vorpahl, F.; Strobel, M.; Jonkman, J.M.; Larsen, T.J.; Passon, P.; Nichols, J. Verification of aero-elastic offshore wind turbine design codes under IEA Wind Task XXIII. *Wind Energy* **2014**, *17*, 519–547. [CrossRef]
76. Robey, R.; Lundquist, J.K. Behavior and mechanisms of Doppler wind lidar error in varying stability regimes. *Atmos. Meas. Tech.* **2022**, *15*, 4585–4622. [CrossRef]
77. Limpens, R.; Neale, N.R. Free electron-driven photophysics in n-type doped silicon nanocrystals. *Nanoscale* **2018**, *10*, 12068–12077. [CrossRef]
78. Jaehong, P.; Obadiah, G.R.; Garry, R. Photo-Induced Carrier Generation and Recombination Dynamics Probed by Combining Time-Resolved Microwave Conductivity and Transient Absorption Spectroscopy. In Proceedings of the Physical Chemistry of Interfaces and Nanomaterials XIV, San Diego, CA, USA, 9–12 August 2015.
79. Johnston, S.; Zaunbrecher, K.; Ahrenkiel, R.; Kuciauskas, D.; Albin, D.; Metzger, W. Simultaneous Measurement of Minority-Carrier Lifetime in Single-Crystal CdTe Using Three Transient Decay Techniques. *IEEE J. Photovolt.* **2014**, *4*, 1295–1300. [CrossRef]
80. Sievert, C.; Shirley, K. *LDAvis: A Method for Visualizing and Interpreting Topics*; Association for Computational Linguistics: Baltimore, MD, USA, 2014.
81. Fast Flow Future for Heavy-Duty Hydrogen Trucks. 2022. Available online: https://www.nrel.gov/news/program/2022/fast-flow-future-heavy-duty-hydrogen-trucks.html (accessed on 4 December 2022).
82. Charikar, M. Similarity estimation techniques from rounding algorithms. In Proceedings of the STOC '02: Thiry-Fourth Annual ACM Symposium on Theory of Computing, Quebec, QC, Canada, 19–21 May 2002.
83. Breckler, S.J.; Allen, R.B.; Konecni, V.J. Mood-Optimizing Strategies in Aesthetic-Choice Behavior. *Music Percept. Interdiscip. J.* **1985**, *2*, 459–470. [CrossRef]
84. Google Translate. Available online: https://translate.google.com/ (accessed on 11 September 2022).
85. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*; Larochelle, H., Ed.; Curran Associates Inc.: Red Hook, NY, USA, 2020; pp. 1877–1901.