UNIVERSIDADE NOVE DE JULHO - UNINOVE
INFORMATICS AND KNOWLEDGE MANAGEMENT GRADUATE PROGRAM - IKGP

ALAYDES MIKAELLE DE MORAIS

TECHNOLOGICAL TRAJECTORY ANALYSIS IN WIND AND SOLAR ENERGY FROM TEXT MINING TECHNIQUES IN PATENTS AND PAPERS

São Paulo
2021

UNIVERSIDADE NOVE DE JULHO - UNINOVE
INFORMATICS AND KNOWLEDGE MANAGEMENT GRADUATE PROGRAM - IKGP

ALAYDES MIKAELLE DE MORAIS

TECHNOLOGICAL TRAJECTORY ANALYSIS IN WIND AND SOLAR ENERGY FROM TEXT MINING TECHNIQUES IN PATENTS AND PAPERS

Master's degree dissertation presented to the Informatics and Knowledge Management Graduate Program at Universidade Nove de Julho - UNINOVE, in partial fulfillment of the requirements for the degree of Master in Informatics and Knowledge Management.

ADVISOR: Prof. Dr. Cleber Gustavo Dias

São Paulo
2021

Morais, Alaydes Mikaelle de.
Technological trajectory analysis in wind and solar energy from text mining techniques in patents and papers. / Alaydes Mikaelle de Morais. 2021.
106 f.
Dissertação (Mestrado) – Universidade Nove de Julho - UNINOVE, São Paulo, 2021.
Orientador (a): Prof. Dr. Cleber Gustavo Dias.
1. Solar energy. 2. Wind energy. 3. Text mining. 4. Patent analysis. 5. Renewable energies.
Dias, Cleber Gustavo. II. Título.
CDU
004

UNINOVE
Universidade Nove de Julho

PARECER – EXAME DE DEFESA

Parecer da Comissão Examinadora designada para o exame de defesa do Programa de Pós-Graduação em Informática e Gestão do Conhecimento, a qual se submeteu a aluna regularmente matriculada **Alaydes Mikaelle de Morais.**

Tendo examinado o trabalho apresentado para obtenção do título de "Mestre em Informática e Gestão do Conhecimento", com Dissertação intitulada "Technological Trajectory Analysis in Wind and Solar Energy from Text Mining Techniques in Patents and Papers", a Comissão Examinadora considerou o trabalho:

(X) Aprovado **( ) Aprovado condicionalmente**
( ) Reprovado com direito a novo exame **( ) Reprovado**

Parecer:
A aluna foi considerada aprovada em seu exame de defesa.

EXAMINADORES

Prof. Dr. Cleber Gustavo Dias (Orientador – UNINOVE)

Profa. Dra. Anapatrícia de Oliveira Morales Vilha - (Membro Externo - UFABC)

Prof. Dr. Renato José Sassi (Membro Interno – UNINOVE)

Prof. Dr. Wonder Alexandre Luz Alves (Membro Interno – UNINOVE)

São Paulo, 19 de fevereiro de 2021.

ACKNOWLEDGMENTS

Thanking everyone involved in this dissertation would be impossible, but I want all of you to know how grateful I am.

I would like to first say thank you God for allowing me to achieve this dream, then offer this work to my husband, who for his unconditional support never let me give up, my mother who always supported study as a form of growth, and my father and brother who were present throughout this journey.

I would like to thank my advisor who showed me the way to get here, who became a friend for life.

I am thankful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) and Nove de Julho University (UNINOVE) for their support.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Financing Code 001 and by the Universidade Nove de Julho (UNINOVE).

AGRADECIMENTOS

Agradecer a todos os envolvidos nessa dissertação seria impossível, mas eu gostaria que todos soubessem o quanto grata eu estou.

Primeiramente gostaria de agradecer a Deus por me permitir realizar esse sonho, e oferecer esse trabalho ao meu esposo, que me deu suporte incondicional e nunca me permitiu desistir, minha mãe que sempre disse que estudar era uma forma de crescimento, e ao meu pai e irmão que estavam presentes durante toda essa jornada.

Ao meu orientador que me mostrou o caminho para chegar até aqui, que virou um amigo que levarei para a vida, só tenho a dizer muito obrigada.

Estou agradecida a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) e a Universidade Nove de Julho (UNINOVE) pelo suporte.

Esse estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001 e a Universidade Nove de Julho (UNINOVE).

Abstract

Renewable energies have become relevant assets for many countries during the past years. More particularly, some research institutions and companies around the world have invested efforts and significant resources to foster scientific discoveries in various areas of renewable developments, two of the most important being wind and solar energy. Therefore, to determine if there is a connection between the studied energies or databases, the objective of this work is to present a technological trajectory analysis in wind and solar energy fields, using text-mining techniques to correlate the findings presented in granted patents and patent application documents from 2010 until 2019 and papers published by the IEEE Xplore Digital Library, considering the same period. The IEEE library has more than 5 million items and it considers itself the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity, and the World Intellectual Property Organization database has around 84 million patent documents. The present research has processed 68,064 patents documents for wind energy and 59,224 patents documents for solar energy. Moreover, more than 13,158 papers have been processed for wind and solar energy during the aforementioned period. Both qualitative and quantitative methods were used to explore the findings and a literature review was done including the main text-mining fields, areas and tasks addressed by the authors in the state of the art. The Keyword search task was chosen and it was used in the abstracts and other textual information for both repositories. The results have shown not only scientific discoveries and technologies developed by researchers, inventors, Universities and patent applicants but also potential trends of the distinct stakeholders on each type of renewable energy. Thus indicating a relationship between the two energies and the database in the indicated period.

Keywords: Solar energy, Wind energy, Text mining, Patent analysis, Renewable energies

Resumo

As energias renováveis se tornaram ativos relevantes para muitos países durante os últimos anos. Mais particularmente, algumas instituições de pesquisa e empresas ao redor do mundo investiram esforços e recursos significativos para fomentar descobertas científicas em várias áreas de desenvolvimento renovável, sendo duas das mais importantes a energia eólica e solar. Portanto, para avaliar se existe uma relação entre as energias ou base de dados estudadas, o objetivo deste trabalho foi apresentar uma análise da trajetória tecnológica nos campos da energia eólica e solar, utilizando técnicas de mineração de texto para correlacionar os achados presentes em documentos de patentes e pedidos de patentes concedidos de 2010 até 2019 e trabalhos publicados pelo IEEE Xplore Digital Library, considerando o mesmo período. O IEEE Xplore Digital Library tem mais de 5 milhões de itens e se considera a maior organização técnica profissional do mundo dedicada ao avanço da tecnologia em benefício da humanidade, e o banco de dados da Organização Mundial da Propriedade Intelectual tem cerca de 84 milhões de documentos de patentes. A pesquisa atual processou 68.064 documentos de patentes para energia eólica e 59.224 documentos de patentes para energia solar. Além disso, mais de 13.158 artigos foram processados para a energia eólica e solar durante o período acima mencionado. Foram utilizados métodos qualitativos e quantitativos e foi feita uma revisão da literatura incluindo os principais campos, áreas e tarefas da mineração de texto endereçadas por autor dentro do estado da arte. A abordagem com a tarefa baseada em pesquisa em palavras-chave foi escolhida, e foi utilizada nos resumos e outras informações textuais para ambos os repositórios. Os resultados mostraram não apenas descobertas científicas e tecnologias desenvolvidas por pesquisadores, inventores, universidades e requerentes de patentes, mas também tendências potenciais dos distintos interessados em cada tipo de energia renovável estudada. Indicando assim uma relação entre as duas energias e as base de dados no período indicado.

Palavras-chave: Energia solar, Energia eólica, Mineração de texto, Análise de patentes, Energias Renováveis

Contents

List of Figures

List of Abbreviations

CNREC	*China National Renewable Energy Centre*
CSP	*Concentrated solar power*
DB	*Database*
EIA	*U.S. Energy Information Administration*
EPO	*European Patent Office*
IEEE	*Institute of Electrical and Electronics Engineers*
IRENA	*International Renewable Energy Agency*
NREL	*National Renewable Energy Laboratory*
NPE	*National Phase Entries*
PCT	*Patent Cooperation Treaty*
PV	*Photovoltaics*
U.S.	*United States of America*
E.U.	*European Union*
WIPO	*World Intellectual Property Organization*
KDT	*Knowledge-Discovery in Text*

Chapter 1

Introduction

Concerning the discussions on the energy aspect, deepened by the international scenario of scarcity of oil and the changes in the climate, caused by through the burning of fossil fuels, research and technical, economic and impact studies arise socioeconomic and environmental enterprises of renewable energy focused on the development of alternatives in energy production, such as from the power of the winds, the called wind power and through the capture of light of the sun, solar power (PÍTSICA, 2015). The next sections describe a contextualization of the present scenario of renewable energies, an overview of the text mining approach and the relevance of the patent system and journals to achieve the main goal of this research.

1.1 A contextualization of the current scenario of renewable energies

Renewable Energies are experiencing an impacted growth. As resources that are naturally replenishing but flow limited, and virtually inexhaustible in duration but limited in the amount of energy that is available per unit of time (BANOS et al., 2011), renewable energy sources minimize environmental impacts, and once they are considered as clean sources of energy it is possible to consider that they produce minimum secondary wastes and are sustainable based on current and future economic and social needs (OLAJIRE, 2020).

Bringing some global scenarios, as two of the most influential nations in the world, in the United States, in mid-2019, renewable energy sources provide only about eleven percent of the electricity in the country's grid (EIA, 2019). However, high penetration of renewable energy is becoming a trend and researchers have shown that supplying all the energy needs of the United States from it could be achievable in the future (NEILL et al., 2018). The National Renewable Energy Laboratory - NREL suggested that in 2050 renewable energy potentially will support about 80 percent of the total electricity consumption in the U.S. (HAND et al., 2012). China, following the same path, has a goal to generate 85 percent of electricity from renewable sources in 2050 (PENG et al., 2020).

Across the world, it is possible to find government agencies and institutes working to find, study and improve most of renewable energies. The U.S. Energy Information Administration - EIA, International Renewable Energy Agency - IRENA, National Energy Administration - NEA in the U.S. and China National Renewable Energy Centre - CNREC are some examples of such governmental bodies. They encourage, promote and enrich researchers who are looking for renewable energy data (NEWBERY et al., 2018).

As described by KABIR et al.(2018), China and the United States are the main coun-

tries which non-hydro renewable energy capacities, including wind and solar energies, are protagonists. Figure 1.1, for example, shows the top 6 countries in renewable power capacities in 2016 and its clear the performance of the China and the United States in the application of these type of energies.

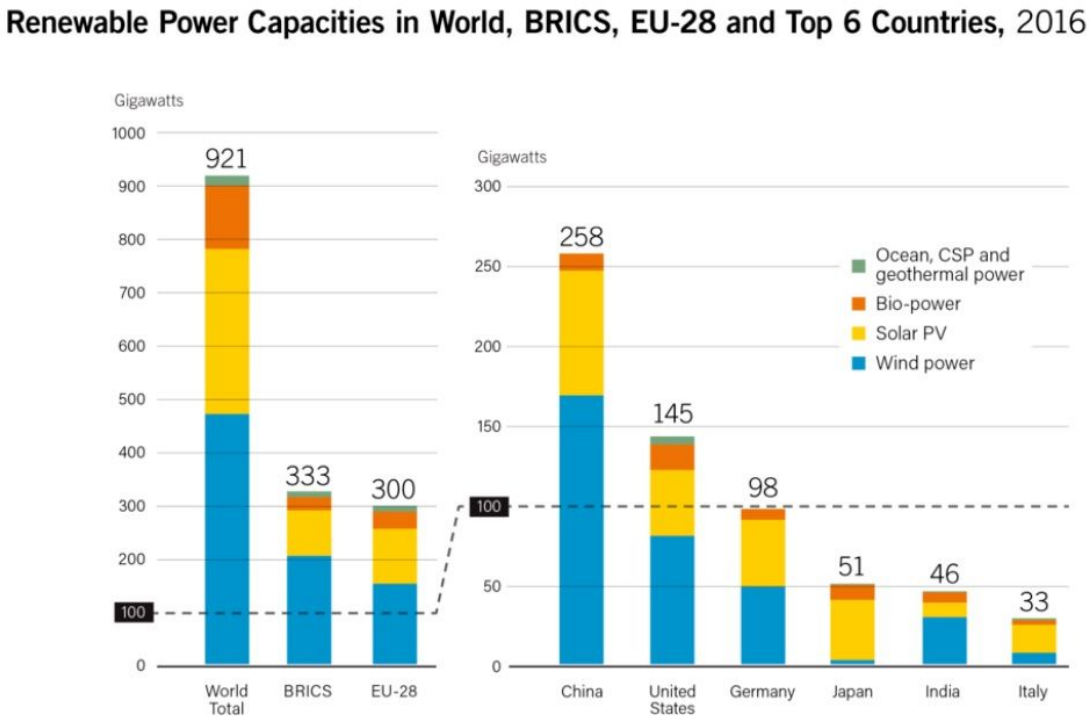

Figure 1.1: *Top 6 Countries in Renewable Power Capacities (2016) (Source: Adapted from REN21,2020).*

Moreover, the data published by (REN21)(2020), for example, highlighted the relevance of those countries considering the investment in renewable energies in 2019 (Figure 1.2). However, it should be noted that other countries are also interested in the use of solar PV and wind power capacities, as well as other types of renewable technologies.

Annual Investment / Net Capacity Additions / Production in 2019
Technologies ordered based on total capacity additions in 2019.

	1	2	3	4	5
Investment in renewable power and fuels capacity (not including hydropower over 50 MW)	**China**	United States	Japan	India	Chinese Taipei
Solar PV capacity	**China**	United States	India	Japan	Vietnam
Wind power capacity	**China**	United States	United Kingdom	India	Spain
Hydropower capacity	**Brazil**	China	Lao PDR	Bhutan	Tajikistan
Geothermal power capacity	**Turkey**	Indonesia	Kenya	Costa Rica	Japan
Concentrating solar thermal power (CSP) capacity	**Israel**	China	South Africa	Kuwait	France
Solar water heating capacity	**China**	Turkey	India	Brazil	United States
Ethanol production	**United States**	Brazil	China	India	Canada
Biodiesel production	**Indonesia**	United States	Brazil	Germany	France

Figure 1.2: *Top Five Countries Distributed by Technology in 2019 (Source: Adapted from REN21,2020).*

More particularly, wind and solar power are the leaders when the topic is a clean source of energy, companies and countries are running to understand, study and get more information and knowledge to produce more with less and to avoid non-renewable, more traditional energy sources such as Oil (INFIELD; FRERIS, 2020). As shown in Figure 1.3, wind power and solar PV capacity are the main renewable fields wherein the investments were made during the past two years, not including hydropower.

		2018	2019
INVESTMENT			
New investment (annual) in renewable power and fuels[1]	billion USD	296.0	**301.7**
POWER			
Renewable power capacity (including hydropower)	GW	2,387	**2,588**
Renewable power capacity (not including hydropower)	GW	1,252	**1,437**
Hydropower capacity[2]	GW	1,135	**1,150**
Wind power capacity	GW	591	**651**
Solar PV capacity[3]	GW	512	**627**
Bio-power capacity	GW	131	**139**
Geothermal power capacity	GW	13.2	**13.9**
Concentrating solar thermal power (CSP) capacity	GW	5.6	**6.2**
Ocean power capacity	GW	0.5	**0.5**

Figure 1.3: *Renewable Energy Indicators 2019 (Source: Adapted from REN21,2020).*

Therefore, in view of the recent interest of the wind and solar energies applications around the world, the present study has been carried out and motivated to understand the main correlations between wind and solar energies, in order to establish a technological trajectory in both fields, by retrieving scientific publications and patents documents as better disclosed in the next chapters. For this purpose, text mining techniques were used to obtain relevant information from public patents and journal databases.

1.2 An overview of the text mining approach

Text Mining can be described as a discovery of what is new, when previously unknown information is automatically extracted from different written resources. The key is the link between what is extracted and the set of information to form new facts or new hypotheses to be further explored. Text mining is different from performing a search on some search site. In search, the user is typically looking for something that is already known and was written by another person, having as a problem only to remove all the material that is currently not relevant to their needs in order to find the relevant information. In text mining, the goal is to discover unknown, new information, something that no one knows yet and therefore could not yet have in writing or be a source of search on sites(GUPTA; BANSAL; SHARMA, 2019).

GUPTA; BANSAL; SHARMA(2019) also explain that being a field of data mining, text mining tries to find interesting patterns from large databases, such as those of patents

and journals, and one point to consider is that text mining, also known as Intelligent Text Analysis, Text Data Mining or Knowledge-Discovery in Text (KDT), which generally refers to the process of extracting interesting and non-trivial information and knowledge from unstructured text, which will be better detailed in the methods and materials section of this work.

As mentioned above, the databases chosen to conduct KDT were those of patents and journals, since they are public and free databases and enable the replication of this and other works, if the academic community so chooses.

1.3 The fundamentals of the patent system and its importance for the retrieval of technological information

A patent is an exclusive right granted that can be applied, in general, when someone finds a new way of doing something, or offers a new technical solution to a problem. If the inventor wants to get that right, technical information about the invention must be disclosed to the public in a patent application that will be held in a patent database (VOET, 2020). That information helps others to know what is new, avoid duplication of work; also, it protects the original author of the invention if someone tries to use its protected rights.

Patents can be created for a number of reasons, for example to support the researchers and their research in case it evolves into a profitable product. Others are created with that goal in mind: having a commercial application of a technology or technique. In addition, patents help universities, industries and even governments who are interested in protecting their new technologies, processes and methods (STIM, 2020). RAFFOUL; BRION(2011) also explained reasons to patent:

> When used properly, patents can provide a useful means to obtain technology, funding, and other assets that a company may be in need of. Historically, they have been used as corporate weapons and, in quite a few cases, have been quite successful in denying markets and opportunities to competitors. As well, patents have been used to add value to a company whether it be through the sale or the licensing of a company's patent portfolio. Obtaining patent protection need not be overly expensive – an effective patent filing strategy that leverages the existing international patent treaties can defer patenting costs. (RAFFOUL; BRION, 2011).

COHEN; NELSON; WALSH(2000) brings a different look at the reasons for patenting, a more economic and even political one, thus clarifying some other common uses of patents and their relevance:

One broader use of patents observed particularly in chemical (apart from-drugs) and other discrete product industries is their combination to build patent fences around some patented core invention. Such fence building involves the patenting, though not licensing (nor necessarily even commercializing), of variants and other inventions that might substitute for the core innovation in order to preempt rivals from introducing competing innovations.56 Firms do not, however, build such patent fences because individual patents effectively prevent imitation or substitution, but because they do not. A second common use of patents which also goes beyond the licensing or commercialization of the patented invention is observed in complex product industries such as electronics. This is patenting to become or remain a major competitor (i.e., "player") in an industry, often via the amassing of large portfolios. The fact that the same patents are often used for both blocking and negotiations in such industries suggests that firms patent not only to protect their own technology, but to hold their rivals hostage by controlling technology that they need. The threatened sanction is often less the expectation of exclusion pending the resolution of a lawsuit, but the more certain and immediate economic harm due to the legal action alone. The ransom demanded by the firm is either formal access to rival technology realized through liberal cross-licensing, or at least the ability to do work similar to that of its rivals without being sued. In this fashion, patents confer the reciprocal access to one another's technologies which enables firms to steadily improve and expand their product lines and processes–something which firms must do to be major competitors in complex product industries subject to rapid technological change. By conferring nonexclusive access to a market in such settings, patents are less an instrument for appropriating rents directly from the firm's own patented inventions (via their commercialization or licensing), and more an instrument for appropriating a share of the oligopolistic rents accruing to the new technologies of all incumbents. How big a share, however, will often be affected by the size and quality of a firm's patent portfolio which affects the terms of trade between rival technologies and its own.

The threat, often implicit, of infringement suits and countersuits underpins almost all the uses of patents, whether to force participation in cross-licensing negotiations in complex product industries, to build patent fences in discrete product industries, or to protect the ability to license or commercialize a new technology in the drug and other industries. Since patents themselves are the best countermeasure to the threat of litigation, it is not surprising, therefore, that we observe the prevention of suits to be one of the most important uses of patents across all industries, notwithstanding the nature of the technology.

> What is interesting is that patents appear to be used as the basis for either threatening or defending against litigation independent of whether they are considered to be effective in their more conventional applications. This suggests that patents can be used either defensively or offensively even if they are weak or untested. They need only confer the right to sue and thus impose the costs of litigation and possible injunctive relief on rivals. But presumably the effectiveness of patents in such litigation-intensive strategies depends upon the firm's ability to support the requisite legal talent, as suggested by our findings that smaller firms are disproportionately dissuaded from applying for patents due to the costs of their defense and disproportionately report patents to be ineffective. The broad issue posed, however, by the pervasive defensive use of patents is whether the social value of patenting is substantially reduced "because it requires all to assume the overhead of defensive patenting" (von Hippel [1988, p. 53]). (COHEN; NELSON; WALSH, 2000)

Existing works demonstrate certain influences of patents on universities and the private sector, and HAASE; ARAÚJO; DIAS(2005) presents some policy suggestions are inferred through studies on improving the information infrastructure on patents and their transposition to the market, a greater incorporation of researchers in commercial exploitation strategies, the orientation of research activities to market demand, a cooperation of companies in areas of intensive research, and even the granting of exclusive licenses to companies, to lead to an optimization of the commercialization of patents through flexible systems of stimulus and contact networks.

Seeking to reveal the importance of the search for information in this type of database, the universal industrial property system, with emphasis on patent documents, presents, analyzes the history, objectives and characteristics of the statutes of trademarks and patents, as well as the national laws pertinent to the subject, explains the organization of this system at the level of international organizations, discusses the importance of the patent document as a source of technological information, analyzing the possibilities and restrictions of its use, the facilitating mechanisms for retrieving patent information and access to specialized databases(FRANÇA, 1997).

An international or universal system of intellectual property places the rules in the larger context of international law and public policy objectives, leading governments, industry and interest groups to achieve a jurisprudence of international dispute resolution mechanisms. This internally, with the creation of national and regional courts that take the discussion on patents to a much more friendly layer, seeking international legislation and agreement to resolve possible divergences, bringing with it an understanding of disputes between countries, investment, and technology transfer(ABBOTT; COTTIER; GURRY, 2019).

The World Intellectual Property Organization - WIPO has the biggest International

Patent System, which is ruled by the Patent Cooperation Treaty - PCT, then, WIPO holds patent applications from most countries and has around 84 million patent documents, including 3.8 million published international patent applications. It is from there that this research got part of renewable energies information.

1.4 The Technological Trajectory Analysis

A chronological analysis of the technological trajectory was performed. GROSSOEHME; LIPSTEIN(2016) brings the importance of this type of analysis as a way to seek an understanding and clarification of situations using time as a metric:

> Time-ordered displays have been previously described as a method to help preserve "chronological flow" and permit understanding of what led to what. Trajectory analysis expands upon this base through the use of sequential matrices(GROSSOEHME; LIPSTEIN, 2016).

LUDVIGSEN et al.(2010) also wrote about the trajectory analysis with a chronological aspect, how it is possible to analyze groups, people or professions looking at a period of time, aiming at the deconstruction of a learning trajectory, he also reports that according to the time unit chosen there will be limitations of what is or is not possible to be explored.

> The literature on educational learning trajectories draws mainly from "hard core" information processing models of human cognition or constructivist accounts with a sensitivity for contextual contingencies. The lifetime perspective leans on narratology and research on life histories. In line with different "turns" within the social sciences this tradition has moved in the direction of contextualism during the last decades. The same could apply to the concept "community learning trajectories" that includes both theories of social participation that take individual minds interacting with others as a point of departure – and those that focus on social context as an infrastructure for human learning. The latter may coincide with approaches that address knowledge issues, and that are indicated in the right column(LUDVIGSEN et al., 2010)).

DARWIN; SCHUMPETER(2015) complements that from historical observations, technological revolutions emerge in the industrialized countries and spread belatedly to the more underdeveloped countries. In the evolutionary interpretation, each successive technological revolution leads to a massive substitution of the set of existing technologies by superior technologies, as well as to the revitalization of industrial sectors. This reinforces the importance of a chronological analysis in the renewable energy field.

1.5 Research problem

This research intends to contribute and offer for the community in general a better understanding of correlations of what is being studied between universities, industry and government regarding the research and applications of wind and solar energies. Trying to answer: What was the trajectory of wind and solar energy in the patent and paper databases during the period studied?

Therefore, the present work intends to answer some questions related to the investments and use of wind and solar power around the world, considering different kind of players in the past decade.

In this sense, this study is not only focused on understanding the state of the art for such technologies, but also highlight potential trends of the distinct stakeholders on each type of renewable energy, establishing, therefore, a technological trajectory analysis.

Among the questions answered by this work, it can be highlighted: i-) the relation between papers and patents based in both renewable energies searched, such as wind and solar energy to show the trajectory of these renewable energies in the chosen period; ii-) how those researchers and patent applicants are working with those different types of energy over a decade, based in those results; and iii-) also, this research wants to bring into evidence a correlation based in a quantitative study grounded in a text mining investigation of keywords from journals and patents cross years, energy and source of study.

1.6 Justification

There is a growing body of literature that recognizes the importance of renewable energies around the globe. However, much uncertainty still exists about the relation between those studies in different areas of interest, such as scientific manuscripts and patent documents. Moreover, how governments, industry and universities are working at the same subjects in and highlight points where those could work together (TWIDELL; WEIR, 2015).

In the literature review some studies were found that aimed to find correlations between wind and solar energy in databases of patents and journals, which motivated even more interest in this study.

The use of text mining as an attempt to find patterns was a way to automate this type of collection of information, giving the academic, scientific and industrial community a faster and more reproducible way to find patterns and trends.

The importance of finding a relationship between what is studied by researchers who focus on patenting and academics in the area of wind and solar energy is to give the community seeking information on the subject researched a source of reference of trajectory

of these two energies. Overcoming the state of the art and focusing on research until more practices.

The world woke up about the importance of Renewable Energies, mainly solar and wind power. Reaffirming that, OMER(2008) related that in the 1997 Montreal Protocol governments agreed to phase out chemicals used as refrigerants that have the potential to destroy stratospheric ozone. As such, countries are making rules and pacts to maintain the global warming stable, industries are looking to be more sustainable and research are fetching for new ways to do more wasting less resources(HALL, 2000).

The goal to transition to a 100 percent renewable energy future demands the participation of everyone. HEINBERG; FRIDLEY(2016) suggest that achieving of that as rapidly, equitably, and sustainably as possible must begin with new ideas and asking the right questions. For that, all kinds of research should work together, spreading information to improve and help each other to achieve that goal.

Besides looking to help those researchers, the aim is to search for data and information to base new technologies and try to lead them, in the future, to new paths of renewable energy sources. Also, when it is possible to analyse how others are working on the same subject it could help to decide new ways to the same goals, or even new places to go.

Reliable databases, as Patents and Journals, could assist new researchers to improve their studies showing what it is new and less unveiled in the research field. Also, if Solar or Wind Power are in the center of knowledge today or even why it is happening can be questions that would be worth answering.

Knowing that, even researchers who did not decide what kind of renewable energies to study yet could ask more questions and choose, in a more certain way, if the studies are worth trying or should not be pursued. As well as help to understand if some subject could be more relevant to delve into its field that it currently is.

A few other future avenues of research are also possible, some of which will be dealt with during the course of this dissertation.

1.7 Delimitation of the theme and organization of this study

Based on the aforementioned, this research presents an analysis between the years of 2010 to 2019, of renewable energies studies focusing on wind and solar power in journals and patent documents, looking primarily to the amount of each in both data sources in each year.

The period was chosen trying to bring a trajectory until the year closest to the development of the research, trying to bring actuality and a history without forgetting the viability to process and download the files, and the two renewable energies chosen were the ones that showed the highest number of research and patents in a previous investigation.

The WIPO database was chosen because it has the largest number of applications and

granted patents and has a global aspect. The IEEE database was chosen for the content of the research with a focus on renewable energy and the relevance of this database to the academic community.

The importance of this study lies in bringing to the academic and scientific community the state of the art of research focusing on wind and solar power in the contexts and periods already indicated, through the use of text mining.

The academic and scientific community and the industry in general benefit from this work because it will be able to observe the trends and trajectories of the renewable energies studied and will help them in taking decisions on future subjects to be studied and/or researched.

This research was divided into sub-sections and chapters to better understand the reader, still in the Introduction (Chapter 1) the subdivisions present the research problem, the justification and the objectives that led to the production of this scientific text.

The next chapters bring how the academic community is working on subjects similar to the study in the Review of Literature (Chapter 2), the most important concepts for understanding the work, the theoretical framework (Chapter 3), the method used and described (Chapter 4), an analysis and discourse of the results (Chapter 5), and the conclusion(Chapter 5) and bibliography used.

1.8 Objectives

The objective of this work is to present an analysis of the technological trajectory in the wind and solar energy fields, using text mining techniques to correlate the results presented in granted patents and patent application documents based on the relationship between papers and patents to determine if there is a connection between the studied energies or databases.

1.8.1 Research Specific Objectives

The main objectives of the present research can be divided into the following three groups:

1.8.1.1 Group I: Objectives related to information retrieved from patent documents from 2010 to 2019

- Investigate the trajectory of patents on solar and wind energy using the keywords extracted with text mining techniques;
- Highlight and understand the main applicants for wind and solar energy;

- Clarify the countries/offices that have received more applications for the studied renewable energies;
- Investigate the subjects that were most patented using the IPC code for wind and solar energy.

1.8.1.2 Group II:Objectives related to information retrieved from papers from 2010 to 2019

- Investigate where within the IEEE Library the solar and wind energy Journals were being applied;
- Identify which institutes were conducting the most research and publishing on solar and wind energy;
- Understand the trajectory of the journals on wind and solar energy from the keywords extracted with text mining techniques.

1.8.1.3 Group III:Objectives related to information retrieved from both data repositories from 2010 to 2019

- Clarify which renewable energy, solar or wind, were most studied in the IEEE and WIPO databases;
- Investigate if there is any relationship on what is studied about wind energy in the WIPO and IEEE databases;
- Investigate if there is any relationship on what is studied about solar energy in the WIPO and IEEE databases;
- Clarify if wind and solar energy can be related in the WIPO database;
- Clarify if wind and solar energy can be related in the IEEE database;
- Identify if there is any pattern on what was researched about wind and solar energy in the IEEE and WIPO database during the study period.

Chapter 2

Literature Review

Patent databases are one of the many types of database available on the Internet for studies to be made upon, but it is not a very easy database to be analyzed, as ARAS et al.(2014) explained in his study on the difficulties and challenges of performing text mining in patents database, in the universe of researchers who file patents there is a legal layer to be met, and requirements, and a language of its own.

Seeking to find patterns and anticipate movements in the industry, KIM et al.(2015) conducted a work mining a large database and comparing how industries behaved between and within them, suggesting that transformations could be anticipated and a method to do so, his studies however did not find a convergence between the various industries researched or when dynamic patterns were applied, but brought findings on a transformation within the universe of each industry separately.

RODRIGUEZ-ESTEBAN; BUNDSCHUS(2016) worked researching about biomedical knowledge into patents databases and described that mining on patent bases was not very mature compared to the same type of work on academic articles, for example, and that it was necessary to complement it by making a not so automated review.

Also thinking about patent analysis, the system to classify Japanese patent documents, the F-term System, which performs the classification of the patent according to the technical characteristics of the inventions described in them, SONG; KIM; LEE(2017) brought another type of report, from which it was able to extract technical information available in patents, and also useful for the process of creating ideas for large and small innovations, bringing possible tendencies that other researchers could benefit from. The author prefers to use F-Term to understand the techniques used on the patent, as SONG; KIM; LEE(2017) describes and explain:

> The F-term is a patent classification system that classifies patent documents according to the technical attributes of the inventions described therein. Since the technical attributes are analyzed according to various perspectives, it is possible to investigate a much wider range of technologies than is possible with other classification systems (Schellner, 2002). Rapidly developing technologies, such as ICT and new material, require the investigation of huge quantities of documents addressing the existing technology. The F-term, which can be classified from a range of perspectives, can utilize the preferred classification.
>
> The structure of the F-term features nine digits. The leftmost two digits are the "theme group" (e.g., 3D: Automotive and transportation vehicles), while the five digits from the left are referred to as the "theme code" (e.g., 3D013: Vehicle waterproofing, decorations, and sanitation devices). The "view code"

> (e.g., AA: Arrangements)" consists of the subsequent two blocks of digits, and the last two digits, called the "number code" (e.g., AA01: Side door parts, AA02: Body sides). This paper uses the "theme group" for an industry, the "theme code" as a technology, the "view code" as a dimension of a technical attribute (TA i), and the "number code" as a shape of technical attributes (TA i(j)) for the dimension of TA i (see Fig. 1).
>
> Both the F-term and the IPC system are patent classifications for technical taxonomy. While IPC searches by using a single view, the F- term is capable of detailed investigation through multiple views (see Fig. 2). Since the F-term supports five times the number of classifications as IPC, the effectiveness of the technical classification is relatively high. The number of F-term codes is currently 350,000, while the number of views is 22,000 (Schellner, 2002). Compared to keywords extracted from patent documents, F-term codes are more structured and thus enable systematic analysis(SONG; KIM; LEE, 2017).

PARK; REE; KIM(2013), seeking promising patents, also sought to deepen his study with text mining techniques by combining findings on patents databases with the adoption of the theory of inventive problem solving (also known as TRIZ). Using floating wind turbines to validate his approach, a new method to identify promising patents for technology transfer was proposed to the academic community.

PARK; REE; KIM(2013) in his work brought two concepts that fit into his research using text mining, TRIZ and SAO-based, the first one its important to understand his choice on the SAO-based approach:

> TRIZ was originally developed by the Soviet inventor Genrich Al'tshuller and his associates by analyzing a vast number of patents across many different fields (Al'tshuller, 1984; Salamatov, Souch- kov, Strogaia, Yakovlev, 1999). TRIZ is a powerful tool to provide systematic and innovative ideas for problem-solving and technology analysis (Mann, 2001). The TRIZ evolution trends, is a TRIZ tool that reveals the patterns of evolution of business and technology systems, and is useful to evaluate the status of the system today and how it will evolve in the future. Classical TRIZ discovered eight patterns of evolution of technical systems: Completeness of parts of the system, Energy conductivity of a system, Harmonizing the rhythm of the system's parts, Increasing ideality, Uneven development of the system's parts, Transition to a super-system, Transition from macro- to micro-level, and Increasing the s-field development. Recently, Mann (2002) suggested updated TRIZ trends (Fig. 1). In particular, what makes TRIZ evolution trends a useful tool for technology evaluation and forecasting is that almost every TRIZ trend follows the basic principle of the TRIZ philosophy, Increasing Ideality, which means that technology systems

> evolve to increasing benefits while reducing harm, and that most technologies and systems evolved only in this direction (left-to-right direction in the TRIZ trends). For instance, the 'Dynamization' evolution trend entails evolution from an immobile system to a jointed system, to a flexible system, and ultimately to a field system(PARK; REE; KIM, 2013).

Thinking about the stages, an innovation stage, through a growth stage, to a maturity stage, the next concept presented has a correlation with those evolution trends and also explained why he did not use the keyword-base approach, the approach presented in this research:

> SAO-based text mining: To automatically analyze unstructured technological information in patents, it should be transformed into an abstracted form which includes the technological key concepts and structural relations among components. To this end, Natural Language Processing (NLP)-based text mining techniques are required. Text mining approaches can be broadly classified into two types: keyword-based, and SAO-based. Although much previous research has employed a keyword-based approach due to its simplicity and ease of use, it is insufficient to reflect the specific technological key concepts and structural relationships among components, because the keyword vector abstracted from the patent by keyword-based text mining, is composed only of frequency of keyword occurrence. In contrast, an SAO-based approach can reflect specific technological key concepts and relations (Yoon Kim, 2011b). S represents 'solution' and AO represents 'problem' in the technological sentence, and thus an SAO structure reflects specific key findings in the patent (Cascini Zini, 2008). In addition, S and O represent the components and A denotes the effect or relations between them; thus an SAO structure describes the structural relationships among components in a patent (Moehrle et al., 2005). In particular, an SAO-based approach is appropriate for this study because the rule base of RFJ of TRIZ evolution trends is expressed in AO structure, so functional semantic similarity between the rule base and AO structures from extracted SAOs can be measured directly.
>
> Recent studies which adopted text mining based patent analysis for technology management have started to exploit an SAO-based approach instead of a keyword-based approach. Examples include analyzing patent risk (Bergmann et al., 2008; Park, Yoon, & Kim, 2012), profiling inventors (Moehrle et al., 2005), monitoring technology (Gerken & Moehrle, 2012), constructing technology trees (Choi, Park, Kang, Lee, & Kim, 2012), analyzing technological trends (Choi, Yoon, Kim, Lee, & Kim, 2011; Yoon & Kim, 2011c) and detect-

ing signals of new technological opportunities (Yoon & Kim, 2012)(PARK; REE; KIM, 2013).

KIM; BAE(2017) also conducted a research that tried to find information by looking at primary technologies on patent databases, its new method was to cluster documents and validate with the patent indicators, showing an academic desire of extracting knowledge from those databases.

Therefore, all these works were somehow seeking to extract intelligence from a patent database, and that text mining or its pillars were used in their great majority, indicating that data extracted from patents are an important source of competitive intelligence, SHIH; LIU; HSU(2010) explains that the business world could use this type of information, and with the same criteria, the academic world can also benefit from it.

Another database that researchers tend to benefit from, besides being easy to access on the Internet, is that of Journals, which is currently more exploited than that of patents precisely because of its contribution to the academic world, as SUNIKKA; BRAGGE(2012) brought a study of the importance of mining information in this type of database, even to extract basic information such as definition and comparison of concepts.

Using text mining to search for works related to text mining in selected academic journals, JUNG; LEE(2020) was able to identify papers that had relevant academic contribution, this research brought the keywords most used by academics in a certain period of history and in a more automated way, that could be reproduced and influence other researchers.

Trying to map which locations in Asia have developed more in communication technologies, and understand which key words were more frequent in this type of study, ZHENG et al.(2016) revealed, using text mining in communication journals, that it was possible to find patterns and try to clarify future projects with their theoretical contributions looking into those databases using that method.

Along the same path, building an overview of information sciences journals and a retrospective analysis of what was done by the academic community, YU et al.(2017) built a map of trends and patterns using text mining in his research, suggesting that it is a valid and promising way to seek this kind of information.

AMADO et al.(2018) also conducted a research in academic literature using text mining and brought to the community a research on applications of Big Data in Marketing, these results were able to demonstrate that this subject is still in an embryonic phase, indicating the importance of deepening the study and use of Big Data in the area indicated.

In order to understand trends in mathematics education, JIN; KO(2019) used text mining in journals and correlated the key words with the subjects of this type of education to find patterns and try to predict the topics of future research and what would be

worthwhile to continue researching, building up which topics of research the academic community could stride for.

Performing an analysis comparing the two databases, Patents and Journals, MADANI; WEBER(2016) sought an evolution of what had already been done in the analysis of patents by text mining, bringing results focused on information retrieval and recognition and analysis of patterns.

BUKOWSKI et al.(2020) brought another type of analysis looking at the databases of patents and publications, sought to understand who performed them and with the type of information that was contained within them if it was possible to develop understandable profiles of experts in biomedical engineering for later evaluation on the training of these experts, their study indicated that it was possible to find and rank these characteristics with text mining.

The importance of a patent study is validated with relevant and scientific reports as presented by NEUHÄUSLER; ROTHENGATTER; FRIETSCH(2019). In this case, authors looked at the state of the art and brought to the academic community a different point of view about the use of that type of information and its applications in the scientific circle.

The work published by STERLACCHINI(2020) also connects patents to clarify trends and determinants of energy innovations. In that paper it is suggested that a discussion about a decrease in the number of energy patenting should be made, targeting a reduction of Oil price. As ALBINO et al.(2014) shows, a comprehensive overview of the evolution of a specific type of eco-innovations is proposed, namely low-carbon energy technologies using patent document analysis.

Other works addressed the use of patent citation networks to study the dynamics of some technical issue and identify its technological development, including, for example, mapping technological trajectories of crystalline silicon PV, data communication standards and energy storage devices, as MUBAROK; NAFIZAH; PERMANA(2019), VERSPAGEN(2007), FONTANA; NUVOLARI; VERSPAGEN(2009) and KUMAR et al.(2018) worked before. In other studies, a patent landscape analysis was provided to identify the technological development of PV cells, as showed by LIZIN et al.(2013) and SAMPAIO et al.(2018).

Papers showing renewable energy trends and how the world is working with that theme in the academic universe were found. SINGH(2013), for example, shows a review of the progress made in solar power generation research and development since its inception and a comprehensive review of the state of the art of wind energy conversion systems and technologies have been proposed by CHENG; ZHU(2014).

In ALEIXANDRE-TUDÓ et al.(2019) work, it has a recent and another important example, this study highlights which countries publish more about renewable energies, which kind of renewable energy is on focus at the moment and the academic cooperation

between the U.S., China and the E.U., while also pointing that there still is much to study about this.

Recently, it is important to mention that, the patent documents have also been used to evaluate other energy technological fields, as the solutions disclosed by BIENVENIDO-HUERTAS et al.(2018) and TSAI; LIU; HSUEH(2020) show.

Table 2.1 shows the the difference between the works studied and what is being proposed regarding the topics covered:

Table 2.1: *Summary of the literature review.*

Author	Year	Text Mining	Patents	Journals	Renewable Energies	Solar Power	Wind Power
Aggarwal, Charu C and Zhai, ChengXiang	2012	X					
Hidir Aras, René Hackl-Sommer, Michael Schwantner and Mustafa Sofean	2014	X	X				
Rodriguez-Esteban, Raul and Bundschus, Markus	2016	X	X				
Song, Kisik and Kim, Karp Soo and Lee, Sungjoo	2017	X	X				
Park, Hyunseok and Ree, Jason Jihoon and Kim, Kwangsoo	2013	X	X				X
Kim, Namil and Lee, Hyeokseong and Kim, Wonjoon and Lee, Hyunjong and Suh, Jong Hwan	2015	X					
Sunikka, Anne and Bragge, Johanna	2012	X		X			
Kim, Gabjo and Bae, Jinwoo	2017		X				
Shih, Meng-Jung and Liu, Duen-Ren and Hsu, Ming-Li	2010	X	X				
Jung, Hoon and Lee, Bong Gyou	2020	X		X			
Zheng, Pei and Liang, Xuan and Huang, Guanxiong and Liu, Xun	2016	X		X			
Yu, Dejian and Xu, Zeshui and Pedrycz, Witold and Wang, Wanru	2017	X		X			
Amado, Alexandra and Cortez, Paulo and Rita, Paulo and Moro, Sérgio	2017	X		X			
Jin, Mireu and Ko, Ho Kyoung	2019	X		X			
Madani, Farshad and Weber, Charles	2016	X	X	X			
Bukowski, Mark and Geisler, Sandra and Schmitz-Rode, Thomas and Farkas, Robert	2020	X	X	X			
Neuhäusler, Peter and Rothengatter, Oliver and Frietsch, Rainer	2019		X				
Sterlacchini, Alessandro	2020		X		X		
Albino, Vito and Ardito, Lorenzo and Dangelico, Rosa Maria and Petruzzelli, Antonio Messeni	2014		X		X		
Mubarok, Muhammad Husni and Nafizah, Ully and Permana, Muhammad Yorga	2019		X		X		
Verspagen, Bart	2007		X		X		
Fontana, Roberto and Nuvolari, Alessandro and Verspagen, Bart	2009		X		X		
Kumar, Vimal and Lai, Kuei-Kuei and Chang, Yu-Hsin and Lin, Chien-Yu	2018		X		X		
Lizin, Sebastien and Leroy, Julie and Delvenne, Catherine and Dijk, Marc and De Schepper, Ellen and Van Passel, Steven	2013		X		X	X	
Sampaio, Priscila Gonçalves Vasconcelos and González, Mario Orestes Aguirre and de Vasconcelos, Rafael Monteiro and dos Santos, Marllen Aylla Teixeira and de Toledo, José Carlos and Pereira, Jonathan Paulo Pinheiro	2018		X		X	X	
Singh, Girish Kumar	2013			X	X	X	
Cheng, Ming and Zhu, Ying	2014			X	X		X
Aleixandre-Tudó, José Luis and Castelló-Cogollos, Lourdes and Aleixandre, José Luis and Aleixandre-Benavent, Rafael	2019			X	X		
Bienvenido-Huertas, David and Quiñones, Juan Antonio Fernández and Moyano, Juan and Rodríguez-Jiménez, Carlos E	2018		X		X		
Tsai, Chieh-Wa and Liu, Tung-Kuan and Hsueh, Po-Wen	2020		X		X	X	
This research	2021	X	X	X	X	X	X

Therefore, the aforementioned works usually addressed patent documents or scientific manuscripts to extract valuable information and better understand the technological evolution of some renewable energies. In a distinct way, the present research has performed text-mining procedures in abstracts for both patents and papers, in order to evaluate some technological trends in wind and solar energy during the last decade.

Chapter 3

Theoretical Framework

3.1 The International Patent System

The work published by GRILICHES(1998) defines a Patent as a document, that grants the right to exclude anyone else from the production or use of a specific new device, apparatus, or process for a stated number of years. It is issued by an authorized governmental agency only after an examination focused on novelty and inventive step of the claimed object and its potential utility.

WIPO, previously defined in the introduction of this paper, holds most of patent applications in the globe in their International Patent System. It has over 150 Contracting States based in its PCT as noted by POTTELSBERGHE(2009). It also informs that the PCT helps inventors when they want a Patent protection in a large number of countries, simultaneously, by filing a single "international" patent application. It is an international treaty that has national and international phases to seek and skip the obligation of filing several separate national or regional patent applications (POTTELSBERGHE, 2009). In the PCT system, the inventor applies in a patent office just once, and the application is transmitted to the international authorities and published in the PatentScope, i.e., a patent search tool that allows user to search for patents in the WIPO database. Figure 3.1 demonstrates the process:

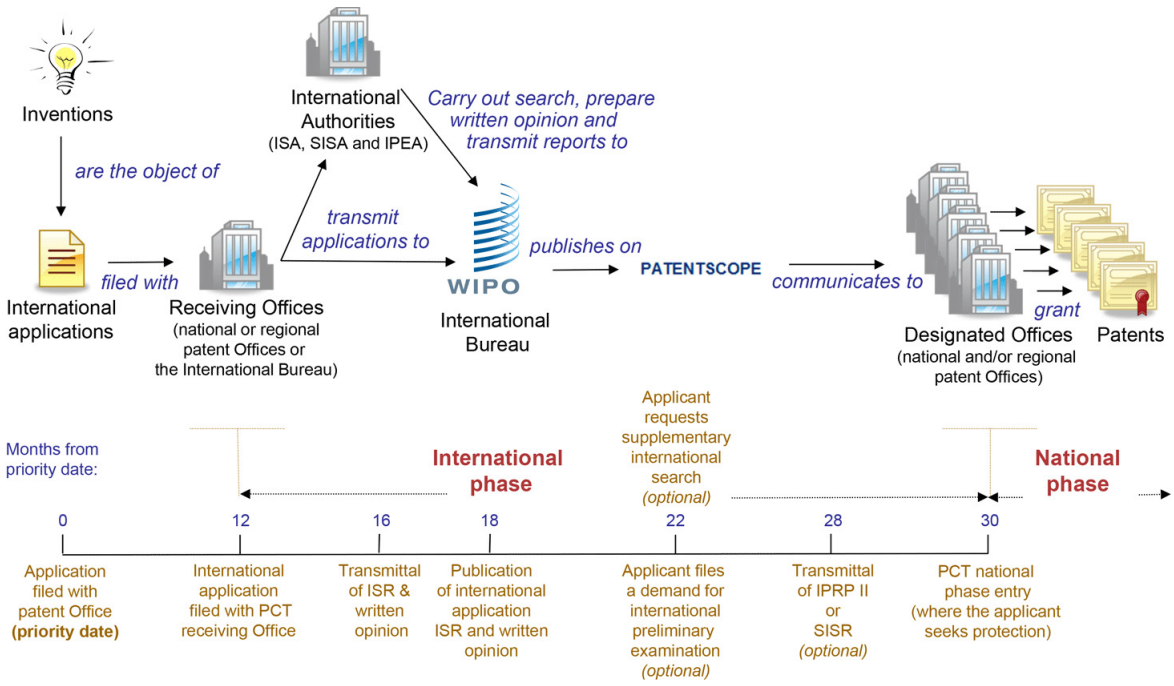

Figure 3.1: *Overview of the PCT System (WIPO, 2019).*

Looking for more options to understand the patents universe and its indicators, SHIH;

LIU; HSU(2010) shows the indicators that could be used to find information in that type of database:

> Patents are one of the major sources of technological and competitive information because such data are easy to access and the content is highly innovative. Since the value of patents is rarely observable, scholars and research organizations have defined a number of patent indicators to determine the value of patents (Brockhoff, 1991; CHI-Research; Reitzig, 2004; Tuomo, Hermans, and Kulvik, 2007)(SHIH; LIU; HSU, 2010).

Depending on the choice of the search, it is possible to use any of the codes or indicators presented, or even a combination of them. In the present work the IPC code has been used due to the ease in retrieving such information from the patent databases.

3.2 Text-mining fields and practice areas to extract textual information

As cited by WEISS; INDURKHYA; ZHANG(2015), the text-mining procedures typically involve the feature extraction of a text in a spreadsheet format. In general, the text source is an unstructured data and some techniques should be done to obtain the features commonly generated from text. VIJAYARANI; ILAMATHI; NITHYA(2015) explained the concepts and issues that could be solved using text mining:

> Text mining is the process of seeking or extracting the useful information from the textual data. It is an exciting research area as it tries to discover knowledge from unstructured texts. It is also known as Text Data Mining (TDM) and knowledge Discovery in Textual Databases (KDT). KDT plays an increasingly significant role in emerging applications, such as Text Understanding. Text mining process is same as data mining, except, the data mining tools are designed to handle structured data whereas text mining can able to handle unstructured or semi-structured data sets such as emails, HTML files and full text documents etc. Text Mining is used for finding the new, previously unidentified information from different written resources.
>
> Structured data is data that resides in a fixed field within a record or file. This data is contained in relational database and spreadsheets. The unstructured data usually refers to information that does not reside in a traditional row-column database and it is the opposite of structured data. SemiStructured data is the data that is neither raw data, nor typed data in a conventional database system(VIJAYARANI; ILAMATHI; NITHYA, 2015).

Basically, the text-mining techniques can be divided into some steps, such as described as follows: i-) Collecting Documents: in this first step the data should be collected from the

relevant documents according to the problem description; ii-) Document Standardization: in this second step the documents collected in the previous step could be in a variety of different formats, depending on how the documents were generated, thus, it is necessary to convert them to a standard format; iii-) Tokenization: in this third step the handling text is applied to break the stream of characters into words or, tokens. This is an important step to extract higher-level information from the document (WEISS; INDURKHYA; ZHANG, 2015); iv-) Stemming or Lemmatization: in this step different tokens are combined to a standard form in order to eliminate some forms of the same word before further processing, i.e., to normalize both words to the single form, including grammatical variants such as singular/plural and present/past cases.

However, text mining can be defined more precisely. As described by MINER(2012), there are six fields of text mining (oval representation) divided into Data Mining, Databases, Statistics, Library and Information Sciences, Computational Linguistics and AI and Machine Learning. In these fields there are seven different text mining practice areas, i.e., seven very different things that a professional could have in mind when talking about text mining, as shown in Figure 3.2. The practice areas are divided into i) Document Classification; ii) Document Clustering; iii) Information Retrieval; iv) Web Mining; v) Concept Extraction; vi) Natural Language Processing and vii) Information Extraction.

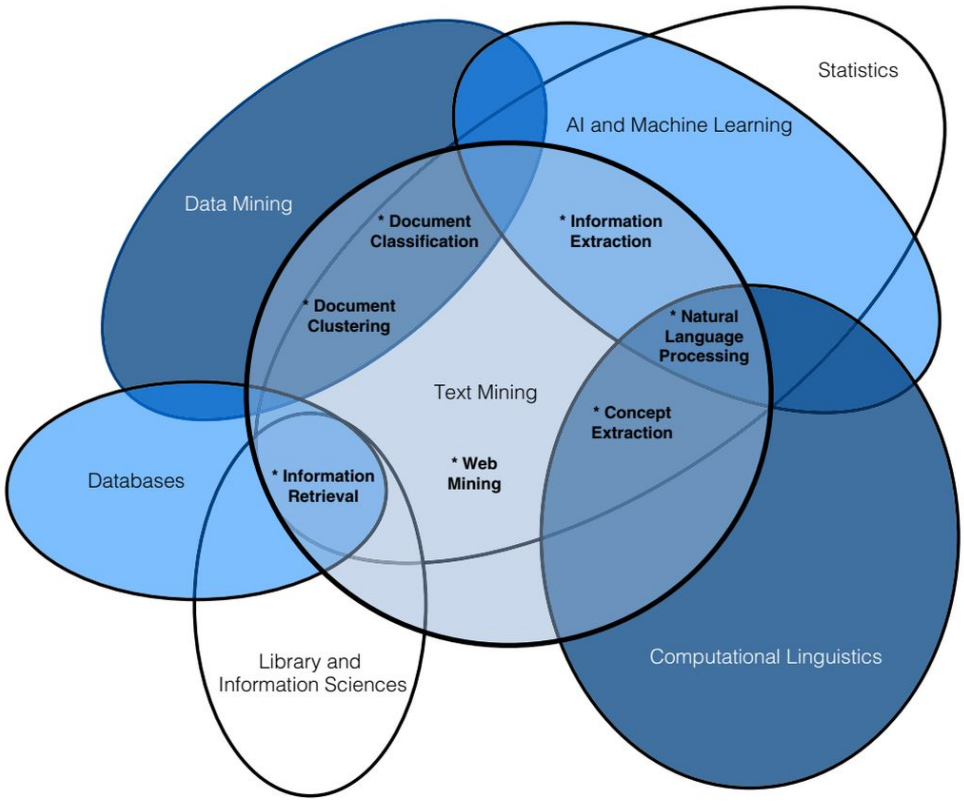

Figure 3.2: *A Venn diagram of the intersection of text mining fields and its related practice areas (MINER, 2012).*

It should be noted that the seven practice areas, as cited by MINER(2012), overlap considerably in specific text mining tasks (labels within ovals), as shown in Figure 3.3.

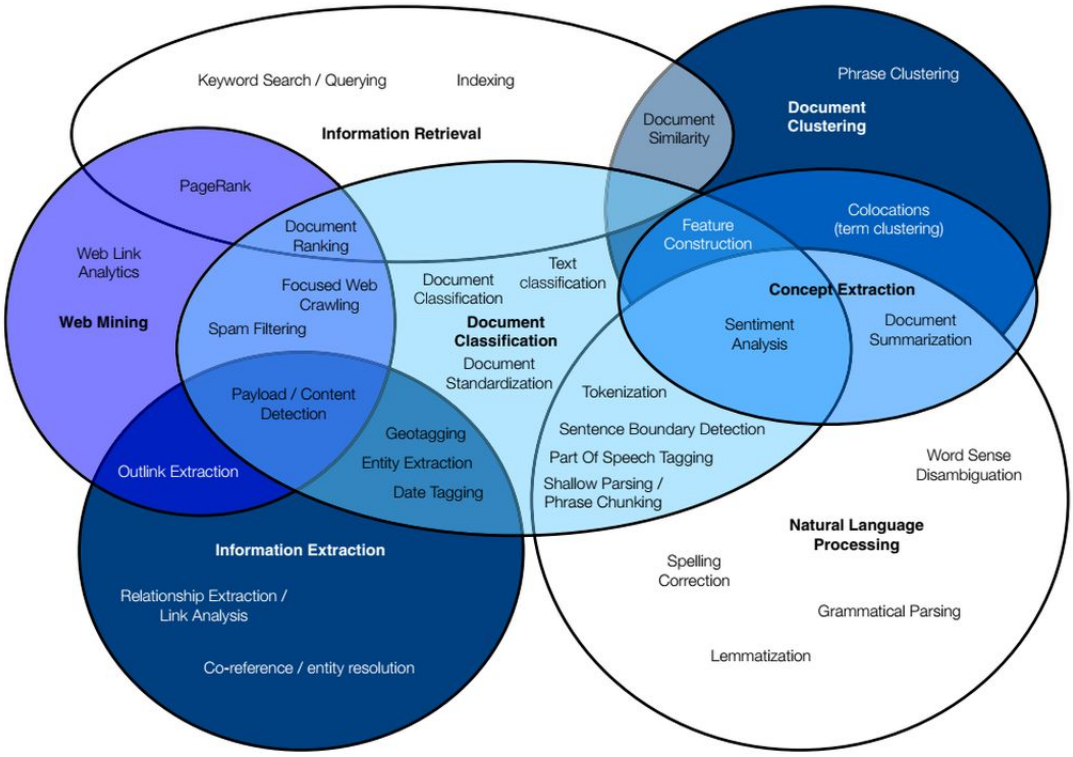

Figure 3.3: *The practice areas and intersections between some specific text mining tasks (MINER, 2012).*

The collective set of features, or words, is typically called a dictionary (WEISS; INDURKHYA; ZHANG, 2015). In some cases reduction techniques are also used to reduce the size of the dictionary, such as the compilation of stopwords to remove and the use of frequency information on the words that can be quite useful in reducing dictionary size. In this sense, the most frequent words are often stopwords and can be deleted (WEISS; INDURKHYA; ZHANG, 2015). Obviously, the remaining most frequent used words are often the important words that should remain in a local dictionary (DIAS, 2018).

Bringing insights about three text mining tools for patents designed for information ARAS et al.(2014) shows concepts and utilities for Numeric Property Extraction, Keyword Extraction, Text Segmentation. Starting with Numeric Property Extraction:

> NUMERIC PROPERTY EXTRACTION: In many technical fields, key information is provided in the form of figures and units of measurement. However, when these data appear in full text, they are almost certainly lost for search and retrieval purposes. The reason for this is that full text is indexed in a way that makes it searchable with strings. In that manner, only the string representation of a numeric property would be searchable, which is, of course, wholly unsatisfactory(ARAS et al., 2014).

Using that tool, the steps are the Identification of numbers, Identification of units of measurements, Unit normalisation, Identification of Intervals, Enumerations and Ratios and Noise Reduction.

Another tool presented by the same author was Text Segmentation:

> TEXT SEGMENTATION: Patent documents are lengthy, abundant, and full of details, such that it may hinder the topic analysis for humans and for machines as well. One of the text mining techniques which can ease these intricacies is text segmentation [3]. The automatic structuring of patent texts into predefined sections will serve as a pre-processing step to patent retrieval and information extraction, as well as enable the interested people to understand easily the structure of a patent that leads to fast, efficient, and easy access to specific information which they are looking for. Furthermore, noun phrases of important sections in the patent texts could be used as main features for patent classification and clustering to achieve a good performance.
>
> The textual part of a patent contains title, abstract, claims, and the detailed description (DetD) of the invention. The latter includes the summary, embodiment, and the description of figures and drawings of the invention. As of the amount of information in DetD, there is a need for auto- mated tools, which can determine the document-level structure of the DetD, identify the different sections and map them automatically to known section types. There has been previous work which showed that the semantic of the patent document structure is valuable in patent retrieval [6], but it only focused on structured patent text which is labelled by specific tags in the original text. The work in [1] presented a rule-based information extraction system to automatically annotate patents with relevant metadata including section titles. In this section, we describe our text segmentation method which is used to recognise the structure of the DetD.
>
> There are many challenges that arise in patent text segmentation, for example measuring the similarity between the sentences is difficult to use because there are a lot of identical terms in the sentences. Another challenge is that the patent contains a lot of new technical terminologies which are hard to collect when using a term matching technique. To meet these challenges, we currently develop a patent text segmentation tool which automatically segments the patent text into semantic sections by discovering the headers inside the texts, identifying the text content which is related to each header, and determining the meaning of the header(ARAS et al., 2014).

The steps to Text Segmentation are Dataset and Preprocessing and Header Detection and Meaning.

The last tool presented for ARAS et al.(2014) is the Keyword Extraction, similar to the one applied in this research:

> KEYWORD EXTRACTION: Keywords extracted from a document are of great benefit for search and content analysis. In the patent domain important keywords can be utilised for searching as well as getting an overview of the topics and the focus of a single patent document or an answer set. In both cases they can avoid unnecessary time-consuming and costly analysis e.g. in prior art or freedom to operate scenarios. Existing methods for keyword extraction – be it automatic or supervised – use either statistical features for detecting keywords based on the distribution of words, sub-words and multi-words, or exploit linguistic information (e.g. part-of-speech) over a lexical, syntactic or discourse analysis. Furthermore, hybrid approaches exist, which try to combine the various types of algorithms and apply additional heuristic rules, e.g. based on position, length or layout(ARAS et al., 2014).

Chapter 4

Materials and Methods

The main objective of this research is to study a technological trajectory analysis in wind and solar energy from text mining techniques, amongst patents and papers. To achieve this objective a quantitative and qualitative research method will be used. Quantitative research methods are more adequate when a large data set is available, such as the one resulting from the initial data extraction of this research. Qualitative research methods are designed to help reveal the behavior and perception to a particular topic, in this case topics related to renewable energies.

For a trend analysis purposes, in wind and solar energy, this research has applied some text-mining techniques in abstracts of granted patent and patent application documents from 2010 until 2019 and abstracts of papers published by IEEE Xplore Digital Library, considering the same period. As mentioned before, the IEEE library has been chosen due to the large size of its database, including wind and solar researches and patents have been retrieved from the WIPO database, since its repository has more than 80 million patent documents. In addition, these data sources were selected because these are generally considered as reputable sources and provide easy to use tools to search and extract data.

Therefore, the present methodology, for patents and papers information extraction, has been carried out using the main text-mining methods (WEISS; INDURKHYA; ZHANG, 2015) and it consists of six stages as shown in Figure 4.1.

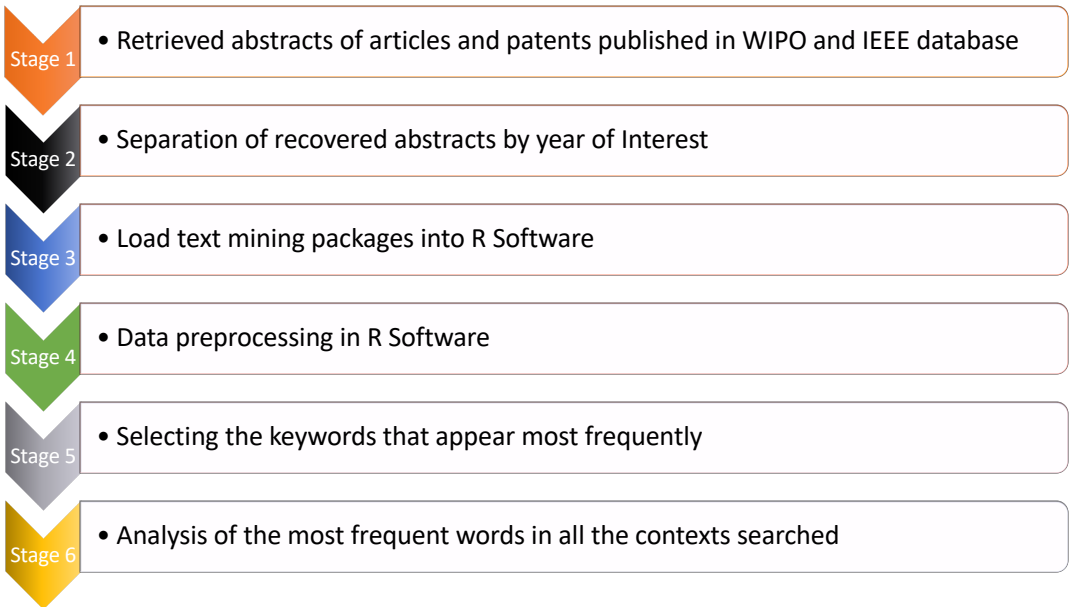

Figure 4.1: *Main stages used in the present methodology.*

Stage 1 and Stage 2 are related to the collecting documents and standardization

methods, since articles and patents abstracts have been retrieved and data have been distributed by year of interest.

Stage 1 is mostly a decision-making phase: Which types of renewable energy to search for, what databases to use, and what filters to apply in the search, such as a period and terms, and what to look inside the Journals and Patents for the information.

Of the renewable energy types, this paper chose the following, presented here along the terms filter used to search data over the databases:

- **Solar** - Terms: `'solar power'`
- **Wind** - Terms: `'wind power'`

For Patents databases this paper opted for the WIPO Database (also known as Patentscope) and as a Journal database, the IEEE Xplore Digital Library was selected. These data sources were selected because these are generally considered as reputable sources, have an extensive archive and provide easy to use tools to search and extract data.

Stage 2 separates the abstracts for year of interesting, so the period chosen was between 2010 and 2019, the last decade, and the data extracted from the results found in Patents and Journals are Abstracts.

Stage 2 also is composed by the extraction of data from the chosen databases, the filtration process and data pruning as to remove irrelevant entries and duplicates, and grouping the extracted data.

A pruning process was applied over the WIPO and IEEE datasets. The pruning process consists in removing duplicate data from the databases and organizing what will be fed into the next step.

To correlate WIPO and IEEE the Abstract is the piece of data chosen as it is common to both and contains roughly the same: a brief explanation of what the Patent or Article entails.

The files have been processed from text files (.csv format) in stage 3 and text-mining packages have been loaded into R software, as showed:

- `library("tm")`
- `library("SnowballC")`
- `library("wordcloud")`
- `library("RColorBrewer")`

It is important to notice that "abstract" below is the list that will be reduced to wordclouds and frequented terms, and that the aforementioned libraries are to use these functionalities as Figure 4.2 shows:

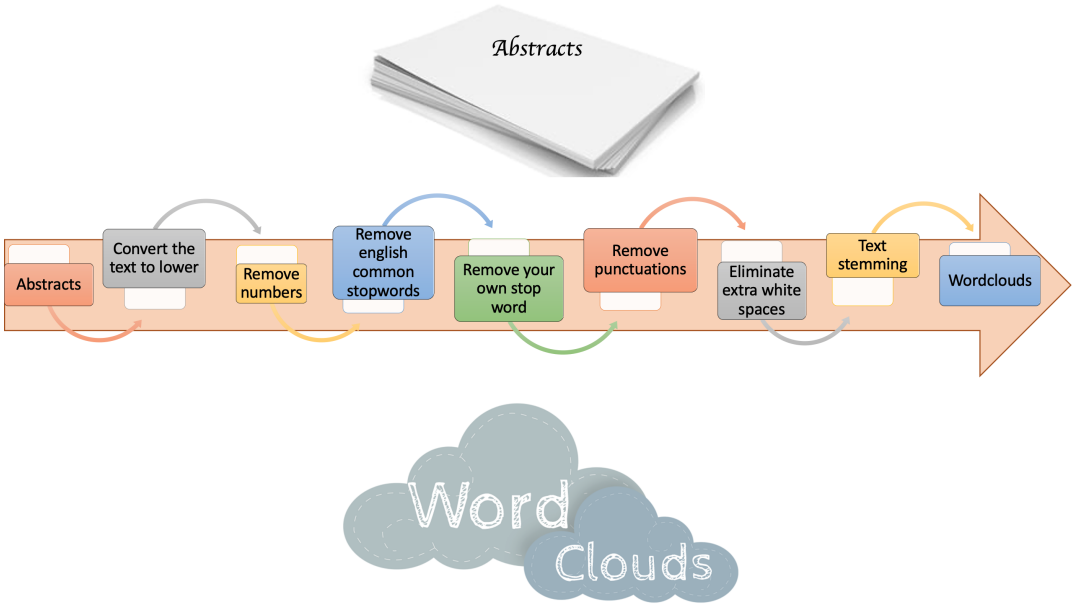

Figure 4.2: *Main steps used on the abstracts.*

Abstracts are input, the whole text is converted to lowercase, numbers and frequent English words are removed, a list of common words to the types of text presented and punctuation are also dropped, and only the reduced inflected words to their word base are used.

In stage 4, the tokenization and lemmatization are applied, after that it is possible to collect a dictionary in stage 5. That process was made for each renewable energy data extracted from IEEE and WIPO. After that, in stage 6 it is possible to compare the dictionary in each renewable energy field and database and correlate them.

The hardware utilized on the whole process was: Processor Name: Dual-Core Intel Core i5; Processor Speed: 2,3 GHz; Number of Processors: 1; Total Number of Cores: 2; L2 Cache (per Core): 256 KB; L3 Cache: 4 MB; Hyper-Threading Technology: Enabled; Memory: 8 GB; Boot ROM Version: 428.0.0.0.0; SMC Version (system): 2.43f10.

The programming language used was R, and RStudio as the open software chosen. Working with Structured data, as Countries/Officers, Authors, Proponents, and codes such as IPC, but also Unstructured data, the summaries of the texts of patents and journals.

This work has a Keyword Search Text Mining Task and it is based in a Databases Text Mining Field and an Information Retrieval Practice Area, as explained by MINER(2012). First the search was done on the abstract of journals and patents in two specific databases, WIPO and IEEE, and looking for trends on a period of time with defined subjects, Solar and Wind Power, seeking information retrieval and to guide other authors with this area trajectory using those extracted keywords.

Being the technical approach a keyword-base, and having BUKOWSKI et al.(2020) work as a start, the framework created is presented on the Figure 4.3:

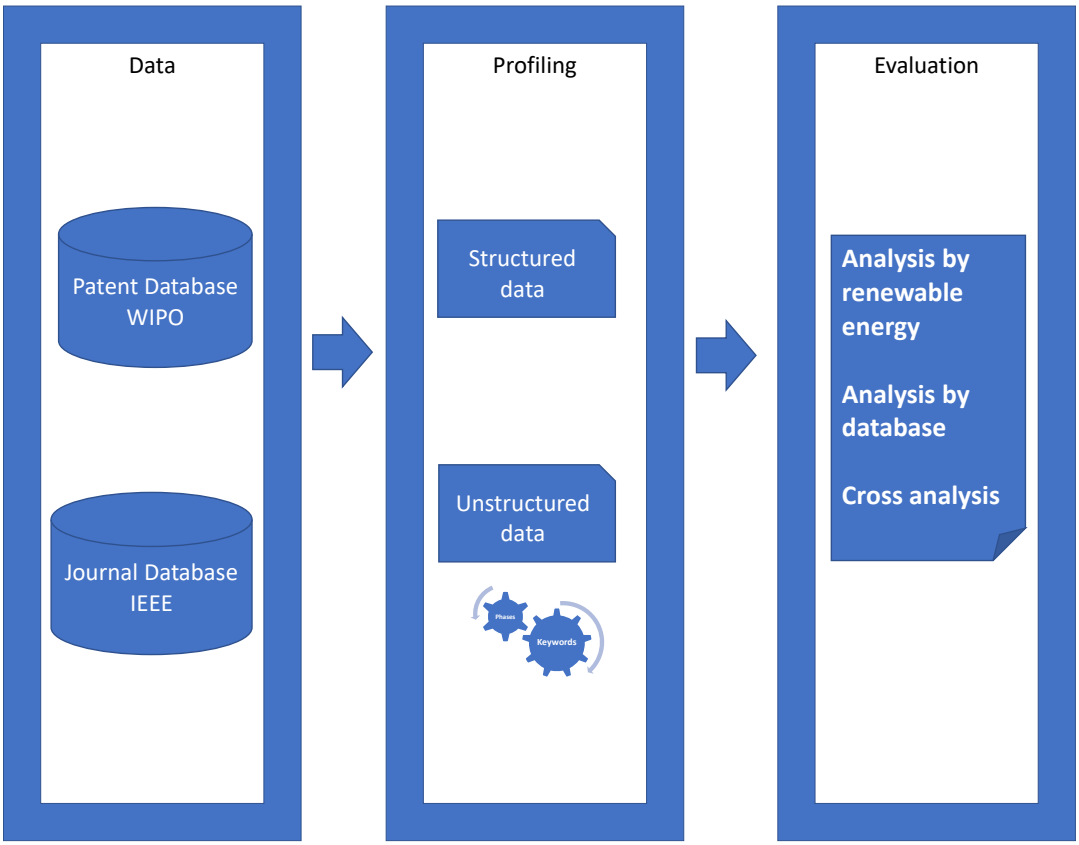

Figure 4.3: *Overview of main steps from profile development to evaluation on Keyword-based approach.*

4.1 The relevance of the IEEE journals for retrieving renewable energies information

One of the important journal databases in the renewable energy sources field is the Institute of Electrical and Electronics Engineers - IEEE, which is a nonprofit institution based in the U.S. and it is one of the biggest organizations dedicated to improve technological advance and to help humanity with its issues (FETTWEIS, 2014). IEEE is a reliable source for international researchers and because of that, it was chosen as the source for searching information related to renewable energies in this work.

As cited by IEEE(2020), the 2019 Journals Citation Reports®(JCR) study, released in June 2020, reveals that IEEE journals continue to maintain rankings at the top of their fields.

Scientific texts, including manuscripts, publications, patents and proposals, are a source of enormous wealth of knowledge, and only with a thorough work of hypothesis development, search for answers, analysis of results, which can take months and even years of dedicated work of a researcher, become relevant to the academic area. Being even one of the challenges in data mining research, to develop effective methods to transform the researched text so that it can be interpreted and perform the validation of the

searched hypothesis, converting data into relevant information that can become future knowledge(JIANG; SHANG, 2020).

4.2 Correlated works that applied text mining approaches for retrieving technological information

Text Mining has intensified over the years (AGGARWAL; ZHAI, 2012), which confirms and informs that this has happened in great part due to the popularization of the Internet and the ease of finding and downloading databases that enable it. Therefore, a great advantage of text mining is that it can be used by several areas and for multiple purposes such as Machine Learning, Data Mining and Information Retrieval, including aggregating or serving as a basis for other types of research.

As cited by MINER(2012), there are six related fields and seven text mining practice areas at the major intersections with its six related fields. Therefore, Table 4.1 shows a summary of the related works that applied some text mining fields and respective practice areas. In chapter 3, the text mining fields and areas were better explained.

Analyzing only the authors who used text mining in their work, it was possible to verify the KDT approaches and compare them with what was done in this work, according to Table 4.1:

Table 4.1: *Summary of the text mining approaches.*

Number	Author	Year	Text Mining Field	Practice Area	Text Mining Task
1	Aggarwal, Charu C and Zhai, ChengXiang	2012	Computational Linguistics	Concept Extraction	Sentiment Analysis
2	Hidir Aras, René Hackl-Sommer, Michael Schwantner and Mustafa Sofean	2014	Databases	Information Retrieval	Keyword Search
3	Rodriguez-Esteban, Raul and Bundschus, Markus	2016	Databases	Information Retrieval	Document Similarity
4	Song, Kisik and Kim, Karp Soo and Lee, Sungjoo	2017	Data Mining	Document Classification	Document Classification
5	Park, Hyunseok and Ree, Jason Jihoon and Kim, Kwangsoo	2013	Text Mining	Information Extraction	Payload/ Content detection
6	Kim, Namil and Lee, Hyeokseong and Kim, Wonjoon and Lee, Hyunjong and Suh, Jong Hwan	2015	Statistics	Information Retrieval	Indexing
7	Sunikka, Anne and Bragge, Johanna	2012	Text Mining	Information Extraction	Payload/ Content detection
8	Shih, Meng-Jung and Liu, Duen-Ren and Hsu, Ming-Li	2010	Statistics	Information Retrieval	Document Ranking
9	Jung, Hoon and Lee, Bong Gyou	2020	Databases	Information Retrieval	Keyword Search
10	Zheng, Pei and Liang, Xuan and Huang, Guanxiong and Liu, Xun	2016	Text Mining	Information Extraction	Relationship Extraction/ Link Analysis
11	Yu, Dejian and Xu, Zeshui and Pedrycz, Witold and Wang, Wanru	2017	Text Mining	Information Extraction	Co-Reference/ entity resolution
12	Amado, Alexandra and Cortez, Paulo and Rita, Paulo and Moro, Sérgio	2017	Databases	Information Retrieval	Indexing
13	Jin, Mireu and Ko, Ho Kyoung	2019	Statistics	Information Retrieval	Keyword Search
14	Madani, Farshad and Weber, Charles	2016	Databases	Information Retrieval	Keyword Search
15	Bukowski, Mark and Geisler, Sandra and Schmitz-Rode, Thomas and Farkas, Robert	2020	AI and Machine Learning	Information Extraction	Payload/ Content detection
16	**This research**	2021	Databases	Information Retrieval	Keyword Search

AGGARWAL; ZHAI(2012) brought concepts and research from the universe of text mining. In his text it is shown which calculations and formulas are most used and the reasons for their use for each case, while ARAS et al.(2014), also with a more conceptual approach but focusing on a Patent Database, demonstrates the challenges of using

some techniques, including Numeric Property Extraction, Keyword Extraction and Text Segmentation, with emphasis on the universe of patents.

RODRIGUEZ-ESTEBAN; BUNDSCHUS(2016), again on a conceptual approach, clarifies in his work the difference of performing a search on a database of patents and journals, then exemplifying what, in his understanding, is easier to be researched in his field of research (biochemistry). For example, recognition by entity name and names of chemical compounds are some of them and the method that, in his knowledge, best fits to search this data (Information Retrieval) using document similarities, such as Automatic Classification and Clustering, which are methods that are close to what was done in this research.

Using text mining techniques based on keywords SONG; KIM; LEE(2017) shows a similar approach to what was done in this research but he limited his search within the context of patents and their F-Term. An F-Term classifies a patent by the technologies and attributes that inventors indicate were used.

In 2017 PARK; REE; KIM(2013) showed the SAO-based approach, an information extraction area, and justified in his research that even the most used model, Keywords-base which is easier to implement and simple, was not enough to reflect the solution and problem found with his extractions. He also explains that with SAO it is possible to describe the structural relations between the components of patents studied which is justified and possible by the use of a TRIZ tool, identifying its content.

KIM et al.(2015) used a Cooperative Patent Classification (CPC) to classify data extracted from a patent database before clustering data in a two-dimensional space with k-means, dividing the extracted data with common aspects into numerical groups. Thus mapping the industries and indexing the results to show the dynamic patterns of industry convergence.

Looking at the work of SUNIKKA; BRAGGE(2012) it is possible to conclude that his text mining approach was used to find information in the contents of a text, even using a tool to extract the words and the knowledge to answer his questions of who, what and where other authors were working. It is worth noting that some authors would rather use tools that are already implemented to speed up the results.

Using a set of techniques to mine changes that occurred in patents, SHIH; LIU; HSU(2010) brought in his study a model called Patent Trend Change Mining (PTCM), which helped him to identify a competitive intelligence when applied to a specific industry of a country. It is a singular approach for the kind of research that he is conducting and it helped him to evaluate and rank patents by their degree of change since the output information would be used as a starting point for decision makers, generating competitive intelligence to help managers develop appropriate business strategies.

JUNG; LEE(2020) sought to look at studies that conducted research on text mining, revealing which aspects or subjects were most studied, when and by whom using tools

that showed the approximation between keywords mined meaning that this work can be classified as a keyword search, in the Information Retrieval area, and since it looked up the most frequent words in specific databases, Web of Science and Scopus, as in the field of Databases.

Searching in the specific period between 1955-2014, ZHENG et al.(2016) used text mining in his research to map a particular field of research, Asian communication, within a journal, looking for relationships or links on the keywords most used in titles and abstracts that identified patterns or trends in the way they communicated and in which region.

Aiming to show a retrospective analysis with text mining and bibliometry, YU et al.(2017) sought to conduct his studies in over seven thousand articles and using data visualization tools built an analysis of patterns which provided a map of the citations made in the body of academic works studied.

AMADO et al.(2018) with an approach of searching for trends in the Marketing universe, searching over big data, did his searches in more than one thousand and five articles between 2010 and 2015 also looking at the most frequent keywords. He portrayed an understanding of the most studied topics in this context, from the creation of his dictionaries of words to the indexing of the terms found, divided over topics in his field of research.

Seeking understanding of his field of study from the mining of text in journals related to education in Mathematics, JIN; KO(2019) also used some analysis tools to statistically verify the relationships and correlations between the keywords found, attempting to evaluate trends for future publications.

Studying the Patents area MADANI; WEBER(2016) used text mining techniques based on keywords in more than one hundred publications in a database of articles and then used bibliometric analysis to deal with the evolution of patents and to research recognition patterns.

BUKOWSKI et al.(2020) conducted his studies and research in databases of journals and patents focusing on the area of biomedical engineering, but emphasizing selected experts with particular focus on academic studies and patents of these individuals. He also used techniques related to machine learning, discovering that patents usually bring new information when compared with answers from other databases when looking for the same subject and that it was possible to trace a profile of the Researcher study by looking at one's publications and works.

This work is a hybrid of most of authors with a Databases Text Mining Field, Information Retrieval Practice Area and Keyword Search Text Mining Task, because the search was made on the abstract of journals and patents in specific databases, looking for trends on a period of time with defined subjects, Solar and Wind Power, and it is trying to guide other authors with this area trajectory using those extracted keywords. The next section shows the mains concepts applied during the research.

On profiling it is possible to divide the data that do not need to pass for the phases described above on Figure 4.1, and the data that will follow all steps. The keywords and the information already structured will be delivered to be analysed on the next Chapter.

Chapter 5

Analysis and Results Discussion

This chapter summarizes the results obtained from text mining of abstracts using the patents documents and papers. Therefore, in order to achieve the objectives of the research, this section is divided into three subsections, as follows: i-) Data retrieved and processed from WIPO patents; ii-) Data retrieved and processed from IEEE papers database; and iii-) Cross analysis from WIPO patents x IEEE scientific articles as shown in Figure 5.1.

Figure 5.1: *Subsections that summarize this Chapter.*

5.1 Results obtained from WIPO patents

This section shows the main results obtained from the analysis of patents documents during the past decade, in order to achieve the objectives of the Group I, as described in Chapter 1.

5.1.1 Trajectory analysis of patents on wind and solar energy using extracted keywords

The first analysis, extracted from WIPO patents database, enabled a visualization of frequent words that appear and disappear in the course of time, and others who grows or not during the decade, as Figure 5.2 and Figure 5.3 elucidate:

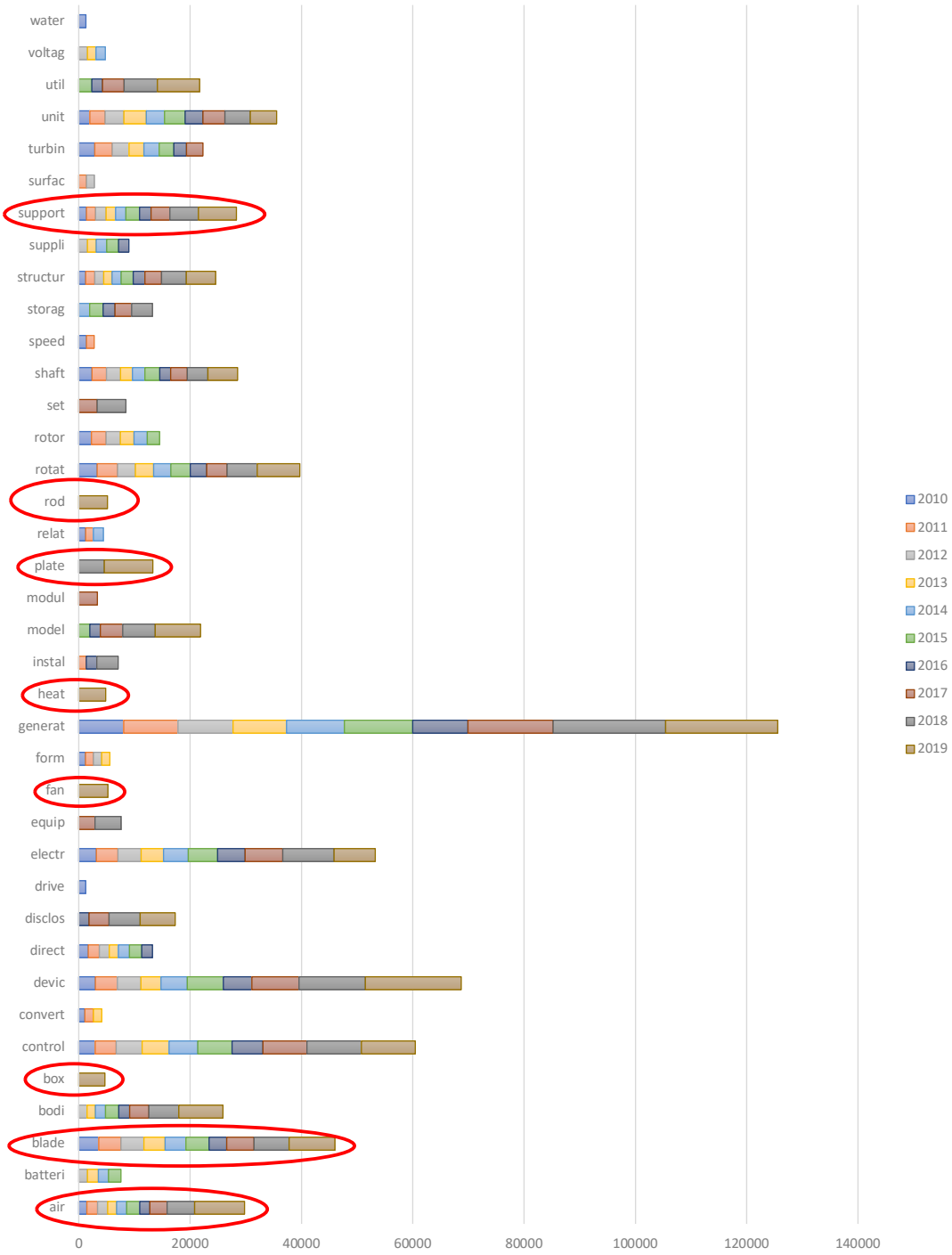

Figure 5.2: *All Words extracted from Wind Power Search on WIPO.*

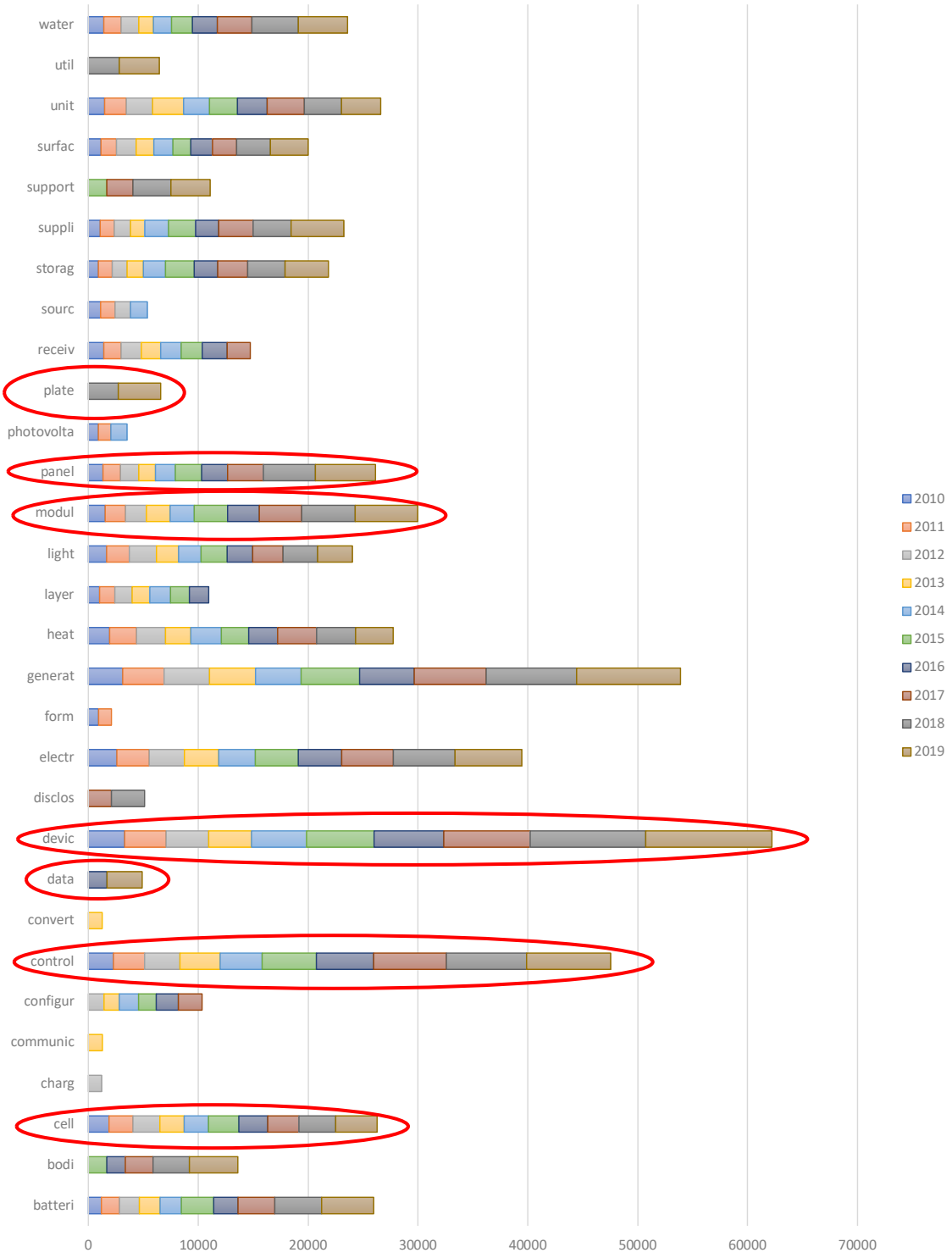

Figure 5.3: *All Words extracted from Solar Power Search on WIPO.*

As a first observation, regarding the figures 5.2 and 5.3, it is possible to observe some words (highlighted with a red ellipse) that came up in recent years or increased during the past two years. These words were further investigated in more detail in section 5.1.2, using n-gram structures. The second one, extracted from WIPO patents database, made

it possible to identify frequent words that are repeated year by year in the course of a decade, in both searches, i.e., solar and wind power, as shown in Table 5.1, Table 5.2 and Table 5.3. In Table 5.1, it is possible to verify that seven words keep appearing each year during the decade searched, considering the results retrieved from “Wind Power” terms.

Looking to the first year, 2010, and the last one, 2019, all those words increase frequency, giving a strong suggestion that, in those subjects, inventors still going at the same path looking for "air", “blade”, “control”, “devic”, “electr”, “generat”, “rotat”, “shaft”, “structur”, “support” and "unit".

A careful reading about patent abstracts allows us to observe that some frequent words are applied for solving particular technical problems. For example, the word “blade” is related to inventors and applicants whose concern is about the design of wind turbine rotor blades and how to improve its aerodynamic performance (Table 5.1).

In most cases, the patents disclose some blade apparatus, devices and methods to construct and/or install its structure. Therefore the words “device” and “connected”, for example, are quite cited in the patent abstracts due to the connection of the components commonly found during the construction of a wind rotor. The word “control” is frequently associated to other words, such as “blade” and “shaft”, since the inventors have developed technologies to improve the pitch control of the blade pitch angle, for example.

Table 5.1: *Frequent words that repeat from 2010 to 2019. WIPO, Wind Power Search.*

	YEAR									
WORDS	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
air	1437	1910	1841	1597	1803	2355	1826	3102	4941	8947
blade	3568	4011	4100	3828	3732	4165	3156	4925	6303	8207
control	2926	3801	4639	4831	5174	6162	5535	7918	9779	9677
devic	2959	3950	4253	3602	4727	6473	5113	8474	11918	17229
electr	3155	3890	4167	3991	4472	5257	4948	6772	9201	7407
generat	8061	9744	9899	9598	10395	12281	9944	15292	20241	20115
rotat	3292	3672	3206	3271	3109	3467	2973	3664	5397	7660
shaft	2375	2584	2522	2138	2236	2675	1978	2965	3723	5326
structur	1213	1666	1573	1524	1615	2163	2115	2997	4427	5300
support	1354	1650	1896	1696	1868	2446	2075	3372	5114	6845
unit	1968	2760	3384	4015	3257	3712	3213	3941	4553	4752
	FREQUENCY									

Table 5.2: *Recent wind energy patents documents where the word "blade" was often cited.*

Patent Aplication Id	Title	Country	IPC	Application Date
WO2019209313	Tip extensions for wind turbine rotor blades and methods of installing same	WO	F03D 1/06; F03D 80/30	27/04/18
US275482224	Joint for connecting a wind turbine rotor blade to a rotor hub and associated methods	United States	F03D 1/06	13/12/17
WO2019210330	Flexible wind turbine blade with actively variable twist distribution	WO	F23D 1/06/; F03D 7/02	29/04/19
CN276202370	Fan blade assembly structure for new energy wind power generation	China	F03D 13/10	25/02/19
CN276141839	Glue blocking device, blade and blade forming method	China	F03D 1/06	20/06/19

When the study is about "Solar Power", it is possible to suggest the same affirmation, looking at Table 5.3, fifteen words showed up each year.

Table 5.3: *Frequent words that repeat from 2010 to 2019. WIPO, Solar Power Search.*

	YEAR									
WORDS	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
batteri	1184	1653	1802	1881	1946	2942	2207	3358	4259	4739
cell	1876	2180	2447	2219	2205	2769	2638	2819	3354	3774
control	2286	2822	3222	3671	3812	4940	5213	6614	7297	7658
devic	3308	3767	3881	3891	5003	6157	6344	7859	10518	11495
electr	2589	2944	3199	3130	3333	3904	3953	4697	5616	6084
generat	3131	3751	4138	4185	4159	5312	4977	6559	8229	9459
heat	1927	2465	2628	2305	2778	2486	2627	3553	3559	3420
light	1664	2062	2462	2021	2047	2372	2333	2764	3139	3161
modul	1531	1840	1901	2174	2187	3038	2888	3855	4875	5677
panel	1325	1599	1666	1507	1809	2403	2354	3255	4736	5475
storag	914	1257	1345	1486	2015	2624	2140	2721	3396	3949
suppli	1073	1286	1482	1289	2163	2464	2102	3127	3467	4791
surfac	1148	1405	1800	1601	1737	1636	1966	2204	3064	3437
unit	1473	1979	2391	2823	2363	2535	2708	3378	3383	3572
water	1411	1570	1611	1330	1641	1909	2263	3150	4224	4486
	FREQUENCY									

An important finding is that five words are present in both searches and in all ten years, and it strongly points to a correlation between inventors in solar and wind power fields of study over a decade. Getting close to the results to look at the applicants, for example, as shown in the next subsection, it is possible to evaluate the main stakeholders on each technical field.

5.1.2 Evaluating relationships between words using N-grams

As cited by SILGE; ROBINSON(2017), it is possible to use some individual words to extract important information from the documents, as shown in the previous section. However, the relationship between words, i.e., which words tend to follow others immediately or words that tend to co-occur within the same documents are cases particularly interesting in the text mining field.

Therefore, in the present research the N-gram approach has been applied to identify some important technological concepts belonging to some keywords in patent documents.

5.1.2.1 Analyzing some recent technological concepts in wind power using bigrams

In this section, the relationship between some words has been evaluated as highlighted in figure 5.2, i.e., considering some technological concepts extracted for wind power using bigrams. In this case, some bigrams have been observed, including words like "fan", "plate", "box", "rod", "blade", "support" and "air", for example.

In figure 5.4 it is possible to identify some important technological concepts when the most frequent bigrams are shown.

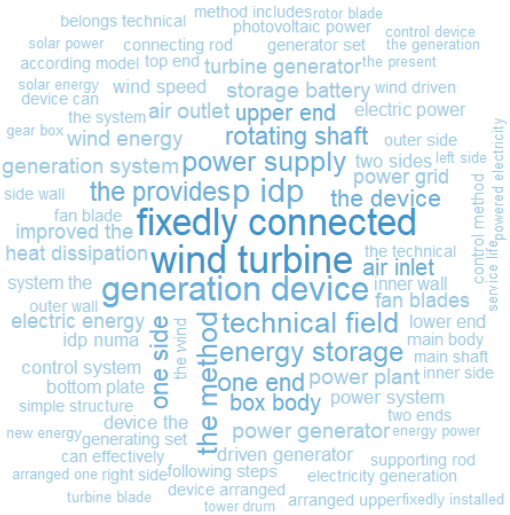

Figure 5.4: *Most frequent bigrams in Wind Power (2019).*

In addition, as cited by SILGE; ROBINSON(2017), the bigrams can be arranged in a network of words, being possible to visualize a combination of connected nodes, as shown in figure 5.5.

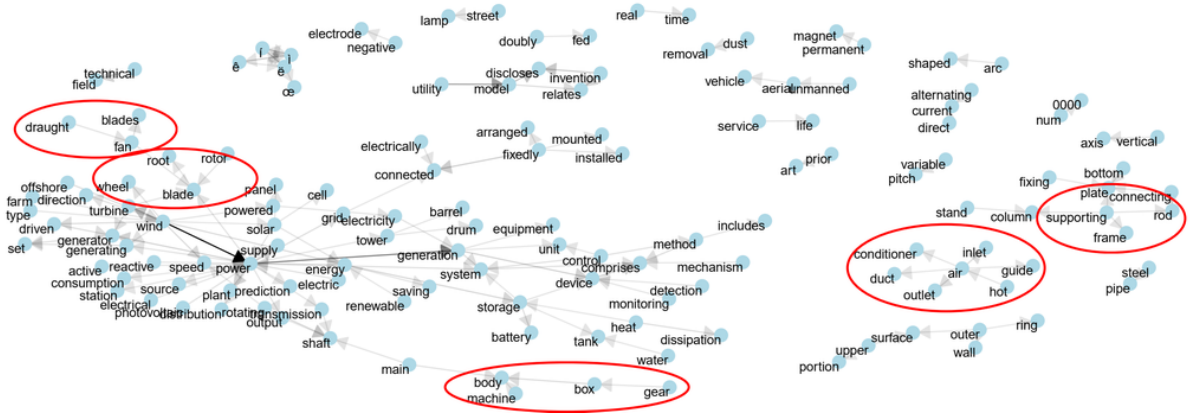

Figure 5.5: *Network of bigrams in Wind Power (2019).*

As highlighted in red ellipses, the network of bigrams allow us to investigate the relationship between some frequent words and some technological ideas or concepts around

the inventions in patent documents. For example, the word "fan" is associated with words "draught" and "blades", as well as the word blade is also related to words "root", "rotor" and "wheel", which suggests some specific structures for wind turbine. The word "box", as shown in the network, is related to words "body machine" and "gear". The word "rod" is associated to the word "supporting", which helps understanding some mechanical arrangements. For example, the following structures are found in the abstracts: i) "...a supporting rod is welded at the upper end of the bottom rod..." ; ii) "...the fixing rings are fixedly inserted into the vertical supporting rods in a penetrating mode..."; and iii) "...one ends of the bottoms of the vertical supporting rods are fixedly connected with the corresponding movable balls...".

For researchers and inventors, these bigrams and word graphs can be a guide to new technical solutions and/or improving the state of the art patents related to a particular wind power field.

5.1.2.2 Analyzing some technological concepts in solar power using bigrams

As presented in the previous section, some technological concepts have also been investigated in solar power using the abstracts of patent documents, considering inventions published in 2019. In this case, words like "bottom", "panel", "control" and "device", for example, are growing in the past two years. Figure 5.6 shows some frequent bigrams found for solar power in 2019, being possible to observe for the word "control", for example, the combination "control device", "control unit", "control system", among others.

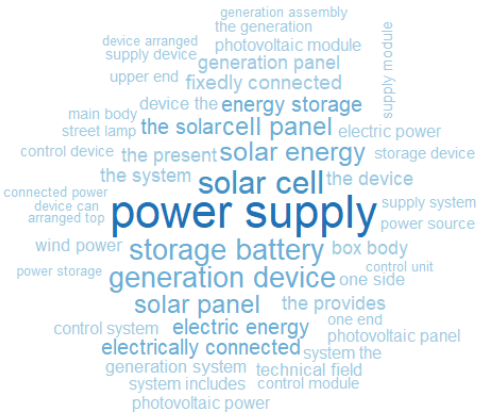

Figure 5.6: *Most frequent bigrams in Solar Power (2019).*

In figure 5.7, it is possible to have a broader assessment of the connections (nodes) between the most frequent bigrams. Therefore, these results have suggested an important concentration in some specific issues for solar power developments, as can be seen in the

results presented for wind power.

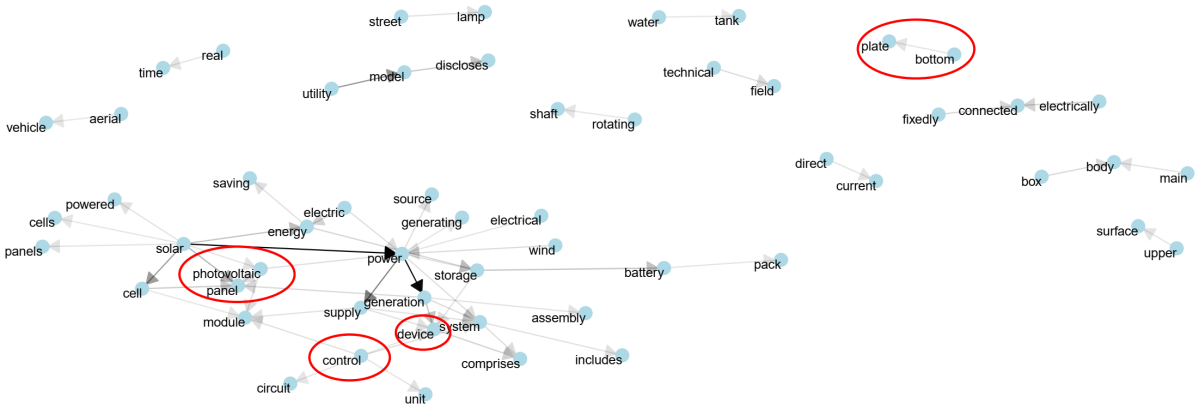

Figure 5.7: *Network of bigrams in Solar Power (2019).*

5.1.3 The main applicants for wind and solar energy

By observing the results shown in Table 5.4 and Table 5.5, for example, some interesting information about the top ten companies on each energy sector are disclosed. It is possible to observe that two applicants are present in both lists, General Electric Company and Siemens Aktiengesellschaft, which indicates that companies that are patenting are interested in both renewable energies in those years searched too.

Table 5.4: *Top ten applicants from 2010 to 2019. WIPO, Wind Power Search.*

Applicants	**Frequency**
STATE GRID CORPORATION OF CHINA	1,279
GENERAL ELECTRIC COMPANY	1,195
VESTAS WIND SYSTEMS A/S	714
WOBBEN PROPERTIES GMBH	678
SIEMENS AKTIENGESELLSCHAFT	589
MITSUBISHI HEAVY INDUSTRIES, LTD.	545
MITSUBISHI HEAVY IND LTD	427
GEN ELECTRIC	422
HITACHI LTD	334
HITACHI, LTD.	327

Table 5.5: *Top ten applicants from 2010 to 2019. WIPO, Solar Power Search.*

Applicant	**Frequency**
SUZHOU CSI SOLAR POWER TECHNOLOGY CO., LTD.	645
GENERAL ELECTRIC COMPANY	514
CANADIAN SOLAR INC.	353
INTERNATIONAL BUSINESS MACHINES CORPORATION	307
INTEL CORPORATION	292
GOOGLE INC.	246
SIEMENS AKTIENGESELLSCHAFT	239
CHANGSHU CSI ADVANCED SOLAR INC.	229
SONY CORPORATION	229
APPLE INC.	219

Illustrating year by year both the WIPO website and PatentScope gives time series of all applicants as Figure 5.8 and Figure 5.9 elucidate, there it is possible to affirm that new applicants are arising such as the State Grid Corporation of China, that reached the first position in 2014 and it is keeping it in the Wind Power scenario. Solar Power shows another relevant applicant, Suzhou Csi Solar Power Technology Co., Ltd., that got the first position only in 2017 and stood there until 2018.

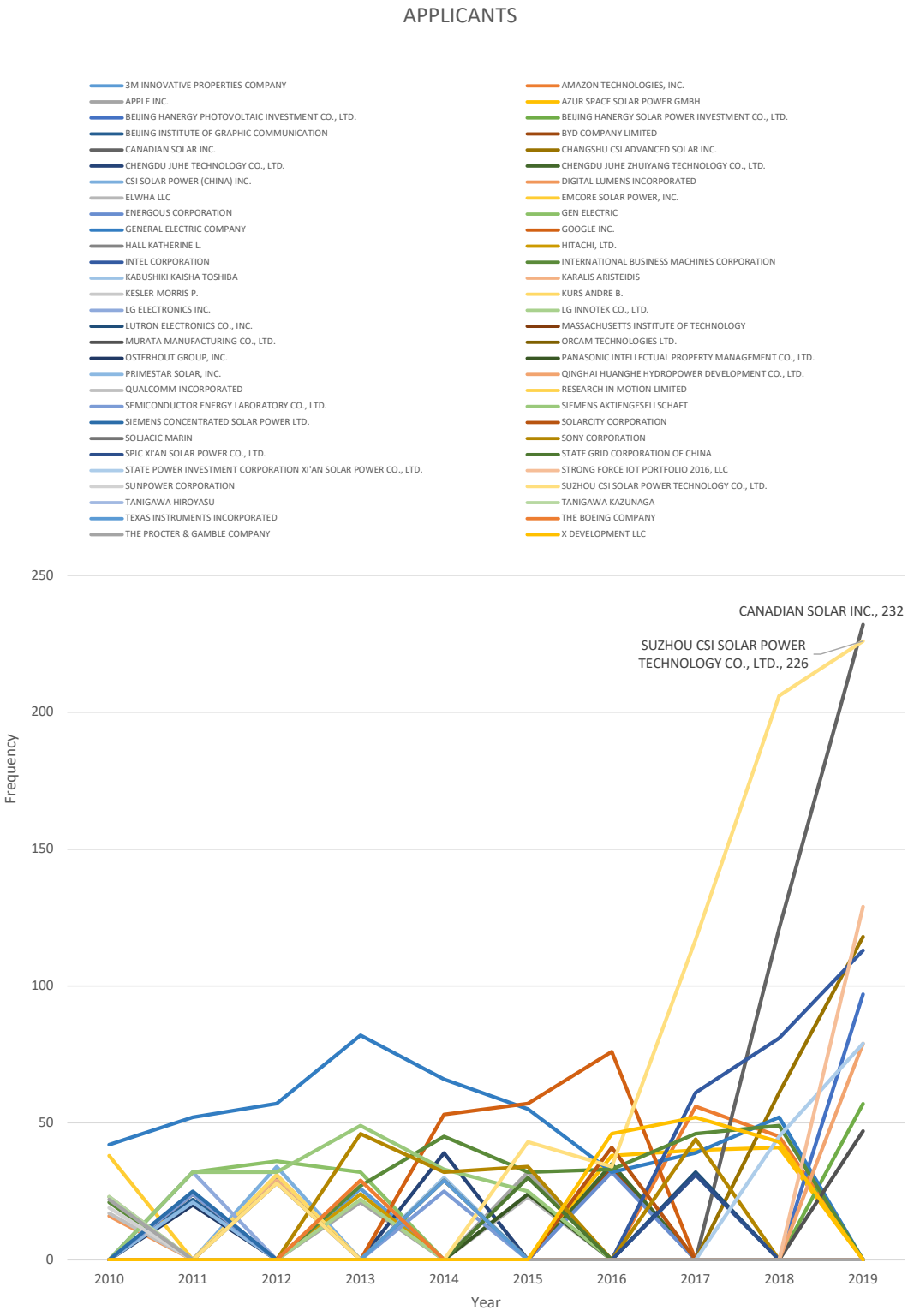

Figure 5.8: *Top applicants year by year. WIPO, Wind Power Search.*

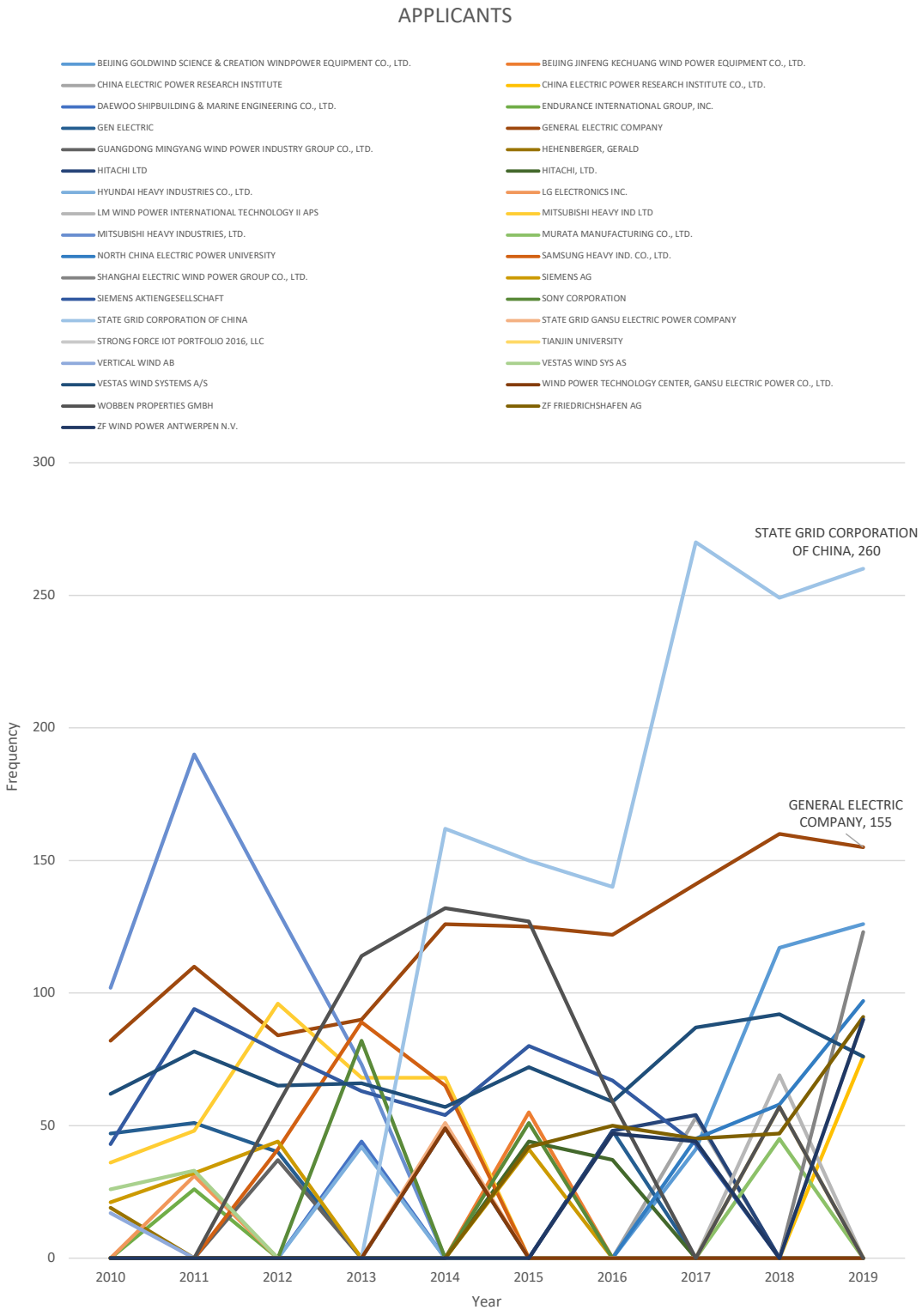

Figure 5.9: *Top applicants year by year. WIPO, Solar Power Search.*

5.1.4 Countries and offices that received more applications for the studied renewable energies

Table 5.6 presents the results related to Countries and/or Offices and revels that the top results are the same, China got the first position in Wind search and U.S. in Solar, but PCT and the European Patent Office have the same position. It indicates that those ten Countries/Offices are investing in Solar and Wind Power research in those ten years.

Table 5.6: *Top ten Countries/Offices from 2010 to 2019. WIPO, Solar and Wind Power Search.*

Wind		**Solar**	
Country/Office	Frequency	Country/Office	Frequency
China	28,862	United States of America	23,854
United States of America	15,545	China	12,876
PCT	5,969	PCT	8,412
European Patent Office	5,961	European Patent Office	4,479
Republic of Korea	3,968	Australia	2,731
Japan	1,815	Republic of Korea	2,056
Australia	1,698	Canada	1,324
Canada	1,358	Japan	1,012
India	591	United Kingdom	687
United Kingdom	435	India	671

5.1.5 Investigating the technological field of wind and solar developments using IPC code

The IPC code is another important information retrieved from patent documents, since this code is related to the different areas of technology. This data have been recovered

after processing the proposed text-mining techniques. Thus, in both scenarios, solar and wind, it was possible to note that it has some similarities, since six codes are presented in Column Wind and in Column Solar, as shown in Table 5.7, looking into a whole decade.

Table 5.7: *Top ten IPC Code from 2010 to 2019. WIPO, Solar and Wind Power Search.*

Wind		**Solar**	
IPC	Frequency	IPC	Frequency
F03D	23,769	H02J	9,094
H02J	10,141	H01L	8,507
H02K	2,952	H02S	6,755
H01M	2,776	F24J	3,514
G06Q	2,452	G06F	2,879
H02P	2,157	H01M	2,350
F03B	2,056	G06Q	2,187
G06F	2,026	H02M	1,846
H02M	1,728	H04L	1,741
H02S	1,619	H04W	1,733

Dividing the results by year, as seen in Figure 5.10 and Figure 5.11, the IPC codes line did not change a lot over the years in the Wind Scenario when compared to Solar. While in Wind Power search the IPC Code F03D was the first position in all years, Solar Power has a new one each 2-4 years, that indicates that in one scenario has more constancy than the other.

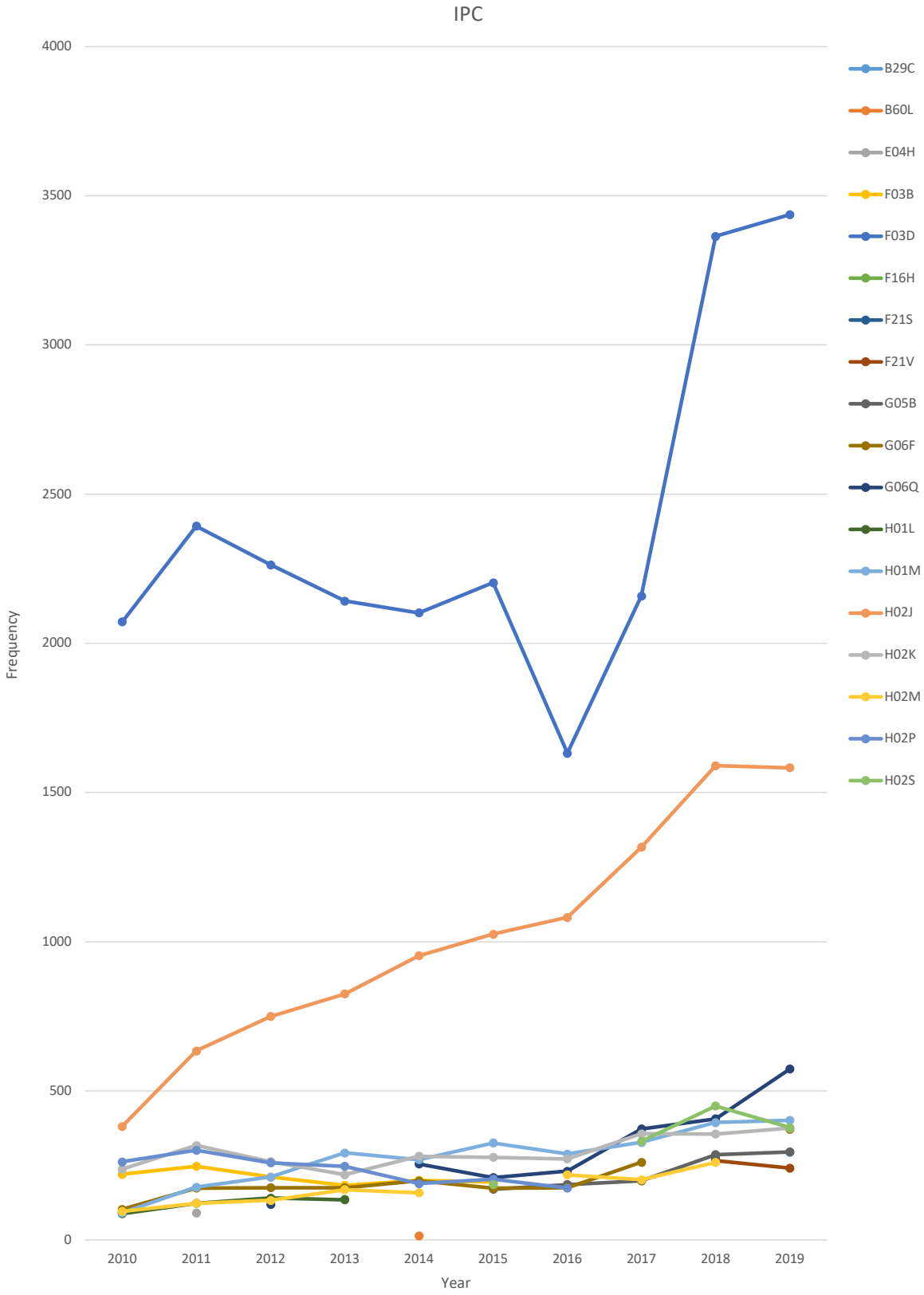

Figure 5.10: *Top IPC Codes year by year. WIPO, Wind Power Search.*

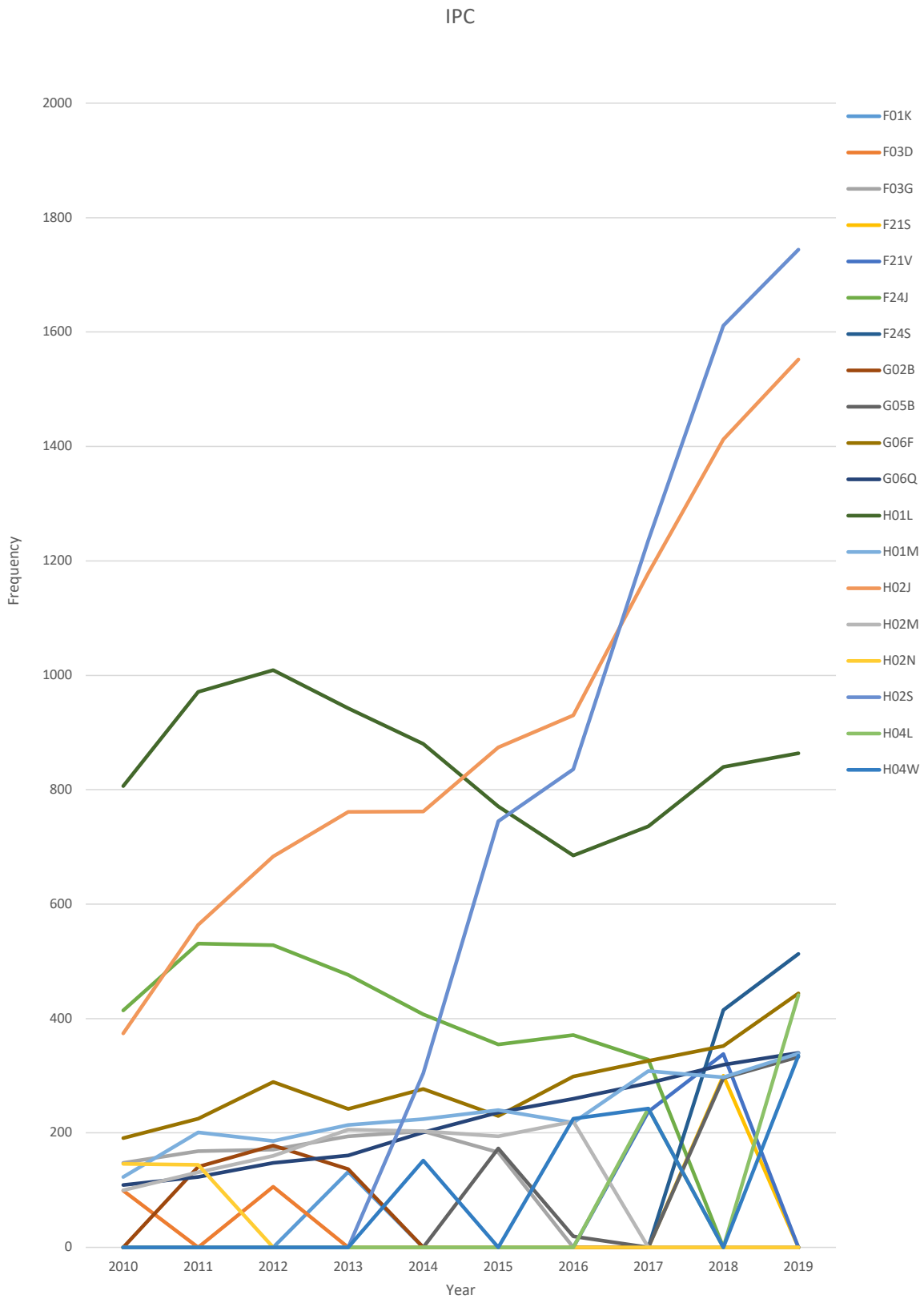

Figure 5.11: *Top IPC Codes year by year. WIPO, Solar Power Search.*

The IPC code F03D is related to "wind motors", which covers mechanisms for convert-

ing the energy of wind into useful mechanical power and the transmission of such power to its point of use. It is important to note that, in solar energy inventions, over the past five years, the applicants have been investing efforts and resources in technologies related to the IPC codes H02S and H02J. These codes are related respectively to generation of electric power by conversion of infrared radiation, visible light or ultraviolet light, e.g., using photovoltaic modules, and circuit arrangements or systems for supplying or distributing electric power.

5.2 Results obtained from IEEE papers database

By studying the IEEE papers abstracts and other journal information, it is possible to identify the publication title of all articles in both areas, solar and power energy, over a decade. The Publication title shows where the paper was published or, in other terms, which area of IEEE it was applied.

When researchers are looking to write and publish an article, they lookup electronic journals, magazines, or newspapers and for that end, a publication title. By looking at the Publication title applied by the researchers choosing to write about solar and wind power, it is possible to find matches in the results.

Understanding where other authors are focusing their effort when they are studying the same subject, solar and power, can help other researchers in their decision of which publisher they should apply or choose to reveal their findings.

5.2.1 The main IEEE Title Journals related to wind and solar energies subjects

Over a decade, filtering the top twenty publication titles to solar and doing the same to wind search, Table 5.8 and Table 5.9 bring the information that from the forty results, only six rows in each table are not in the other table. Leading to the understanding that researchers from both renewable energies are publishing at the same places.

Table 5.8: *Top twenty publication titles for papers published from wind energy issue.*

Wind	
Publication Title	Frequency
IEEE Transactions on Power Systems	859
IEEE Transactions on Sustainable Energy	701
IET Renewable Power Generation	688
IEEE Transactions on Industrial Electronics	585
IEEE Transactions on Industry Applications	567
IEEE Transactions on Applied Superconductivity	562
IEEE Transactions on Power Electronics	532
IEEE Transactions on Energy Conversion	530
IET Generation, Transmission & Distribution	420
IEEE Access	389
The Journal of Engineering	375
IEEE Transactions on Power Delivery	361
IEEE Transactions on Magnetics	320
IEEE Transactions on Smart Grid	295
IET Electric Power Applications	280
Monthly Notices of the Royal Astronomical Society	212
IET Power Electronics	177
Journal of Modern Power Systems and Clean Energy	150
IEEE Transactions on Dielectrics and Electrical Insulation	130
IEEE Journal of Emerging and Selected Topics in Power Electronics	129

Table 5.9: *Top twenty publication titles for papers published from solar energy issue.*

Solar	
Publication Title	Frequency
IEEE Journal of Photovoltaics	448
IET Renewable Power Generation	293
IEEE Transactions on Power Electronics	222
IEEE Transactions on Sustainable Energy	197
IEEE Access	184
IEEE Transactions on Industry Applications	169
IEEE Transactions on Industrial Electronics	166
IEEE Transactions on Smart Grid	151
IEEE Transactions on Power Systems	122
Monthly Notices of the Royal Astronomical Society	113
IET Generation, Transmission & Distribution	100
IET Power Electronics	78
IEEE Transactions on Energy Conversion	73
IEEE Latin America Transactions	70
IEEE Transactions on Industrial Informatics	70
The Journal of Engineering	68
Proceedings of the IEEE	48
IEEE Sensors Journal	45
IEEE Transactions on Plasma Science	43
IEEE Journal of Emerging and Selected Topics in Power Electronics	42

5.2.2 The main institutions that are researching and publishing on solar and wind energy

Moreover, applying the text-mining techniques in the IEEE papers database it was possible to better understand the main research stakeholders. The results shown in Tables 5.11 and 5.10 disclose the top ten Universities and Research Institutions or Affiliations, which are working in the both issues, i.e., solar and wind energy. For example, the Aalborg University in Denmark has an important contribution in wind and solar energy researches. As cited in (GRAHAM, 2018), Aalborg University has been recognized as one of the 10 institutions most consistently cited as "current leader" in engineering education, including others such as Olin College of Engineering, Massachusetts Institute of Technology (MIT), Stanford University, among others.

Table 5.10: *Top ten Affiliations from 2010 to 2019. IEEE, Wind Search.*

Wind	
Affiliation	Frequency
Aalborg University, Aalborg, Denmark	107
Southeast University, Nanjing, China	73
Zhejiang University, Hangzhou, China	64
North China Electric Power University, People's Republic of China	62
Tsinghua University, Beijing, China	59
Huazhong University of Science and Technology, Wuhan, China	57
China Electric Power Research Institute, People's Republic of China	51
Tsinghua University, People's Republic of China	44
National Renewable Energy Laboratory, Golden, CO, USA	40
Nanyang Technological University, Singapore	38

Table 5.11: *Top ten Affiliations from 2010 to 2019. IEEE, Solar Power Search.*

Solar	
Affiliation	Frequency
Indian Institute of Technology Delhi, New Delhi, India	107
Aalborg University, Aalborg, Denmark	38
National Renewable Energy Laboratory, Golden, CO, USA	38
National University of Singapore, Singapore	21
University of Waterloo, Waterloo, ON, Canada	18
Nanyang Technological University, Singapore	30
National Institute of Technology, India	15
Sandia National Laboratories, Albuquerque, NM, USA	15
National University of Singapore, Singapore	14
Zhejiang University, Hangzhou, China	12

The affiliation field, meaning the Educational and/or Research Institutions that are researching and applying to Journals in both renewable energies, has four similarities in the top ten listed, consistent with the discovery that wind and power are walking together in the academic field.

5.2.3 Evaluating extracted keywords from IEEE abstract documents in solar and wind power

The graphs represented on Figure 5.12 and Figure 5.13 show the evolution of the most frequent words extracted year by year during the researched decade and with them it is possible to understand that some words were focused in some specific years and then no longer while others remained frequent throughout the researched period:

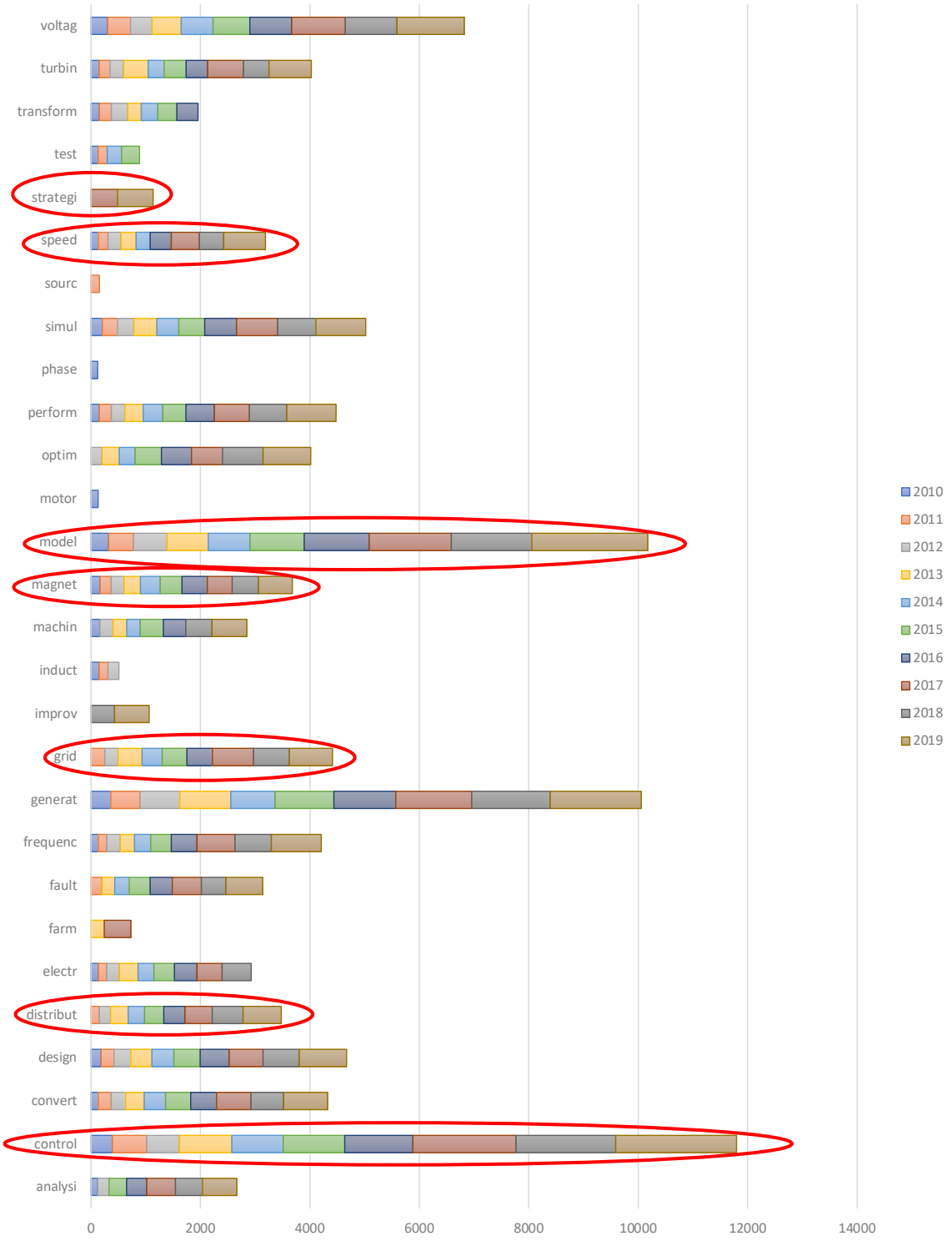

Figure 5.12: *All Words extracted from Wind Power Search on IEEE.*

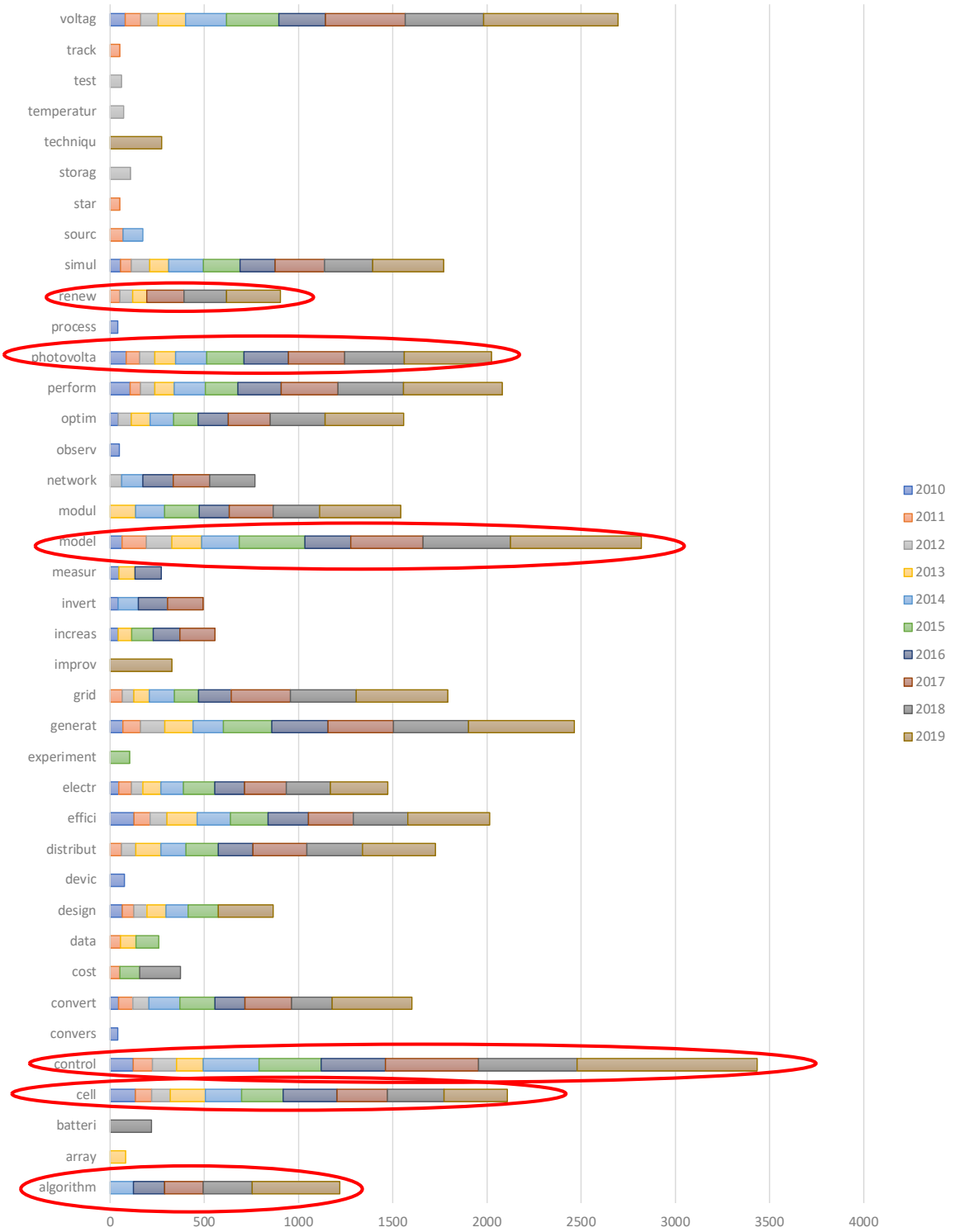

Figure 5.13: *All Words extracted from Solar Power Search on IEEE.*

In figures 5.12 and 5.13 some words have been highlighted (red ellipses), in order to show mainly their importance in the past two years. By analyzing the word network in both energy fields, it is possible to extract some technical and/or scientific concepts related to the results. The n-grams approach has also been used to better understand the

technical concepts in the abstracts. In figure 5.14, for example, it is clear that the word "strategy" is related to, as shown in figure 5.12, a control strategy in the wind energy developments when we observe its respective network of bigrams. In addition, it is possible to extract other word semantic considerations, when distinct bigrams are considered such as the word "grid", related to "power" and "connected", and "magnet" related to the word "permanent" and synchronous generator.

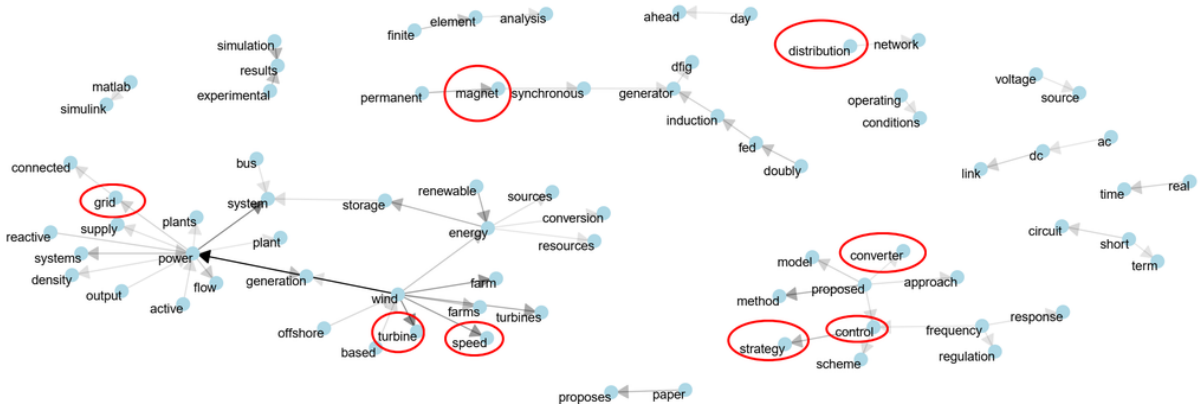

Figure 5.14: *Network of bigrams in IEEE abstracts for Wind Power (2019).*

In figure 5.15, some technical concepts have been investigated using the bigrams network. In this case, some bigrams like "solar cell", "solar photovoltaic", "control scheme", among others, allow us to understand the state of the art technology developments in the solar power field.

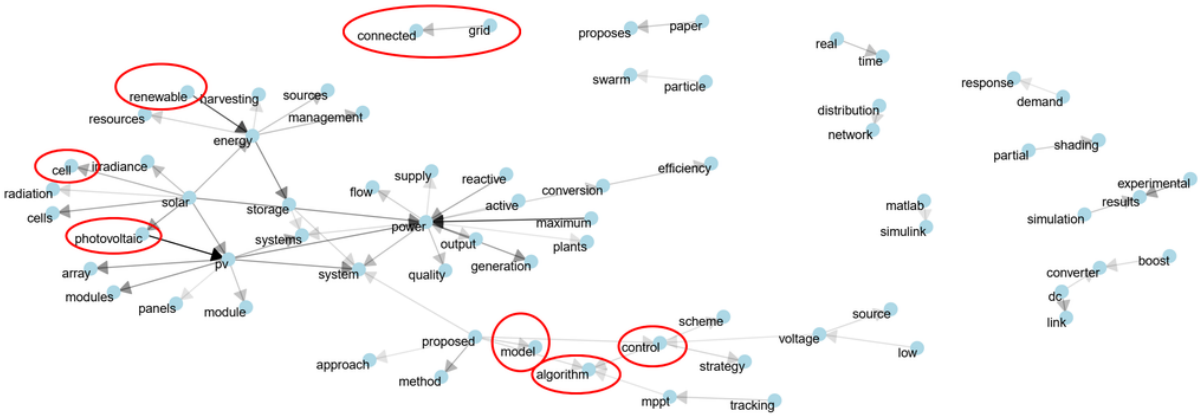

Figure 5.15: *Network of bigrams in IEEE abstracts for Solar Power (2019).*

The IEEE frequency words that appear on all searches, as showed in Table 5.12 and Table 5.13, elucidate that in both cases, Wind and Power, researches have grown, each year, in specifics areas.

Table 5.12: *Frequent words that repeat from 2010 to 2019. IEEE, Wind Power Search.*

	YEAR									
WORDS	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
control	391	629	595	957	941	1121	1247	1884	1822	2206
convert	133	235	262	341	392	459	472	628	596	805
design	185	235	305	388	400	480	534	617	660	866
frequenc	132	156	246	259	299	376	468	695	665	910
generat	360	535	727	931	812	1071	1131	1390	1432	1662
magnet	165	203	236	300	359	399	461	456	480	620
model	322	455	613	752	763	991	1188	1499	1471	2118
perform	150	221	248	334	357	424	521	638	684	900
simul	205	278	300	421	401	472	583	752	695	912
turbin	148	197	252	448	291	401	393	654	467	775
voltag	302	421	392	532	582	671	769	978	944	1228
	FREQUENCY									

Table 5.13: *Frequent words that repeat from 2010 to 2019. IEEE, Solar Power Search.*

	YEAR									
WORDS	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
cell	134	86	99	186	192	221	286	267	301	336
control	122	102	128	142	296	330	341	494	523	956
effici	126	86	89	160	177	200	214	238	289	435
electr	45	68	60	96	120	166	158	222	234	304
generat	67	94	129	149	162	257	298	348	398	562
model	63	129	135	158	200	348	244	384	464	694
perform	106	55	76	103	165	173	230	301	348	524
photovolta	85	72	80	110	165	198	236	299	316	463
simul	55	57	98	101	184	194	186	262	256	377
voltag	80	82	92	146	217	278	248	423	416	713
	FREQUENCY									

Observing the frequency of those words at the first and last year, it is possible to state that all words, in both tables, grew.

Six words appeared in both results, that information strongly indicates that during the period searched, solar and wind power researchers were working in paths with similarities. More particularly, it is possible to note that the words "generat" and "control", for example, appears as a frequent word in both energy sectors, as the next subsection will detail.

5.3 Cross analysis from WIPO patents x IEEE scientific articles

This subsection discloses some correlations between data extracted from WIPO patents and IEEE papers. As shown by Figure 5.16, in the past decade there was an important growth of technologies protected by patents and those published by scientific manuscripts. The solar power researches found in papers had a relative growth greater than wind power discoveries, but in absolute numbers wind power solutions is greater when compared to solar power technologies in papers and patents, as shown in Table 5.14.

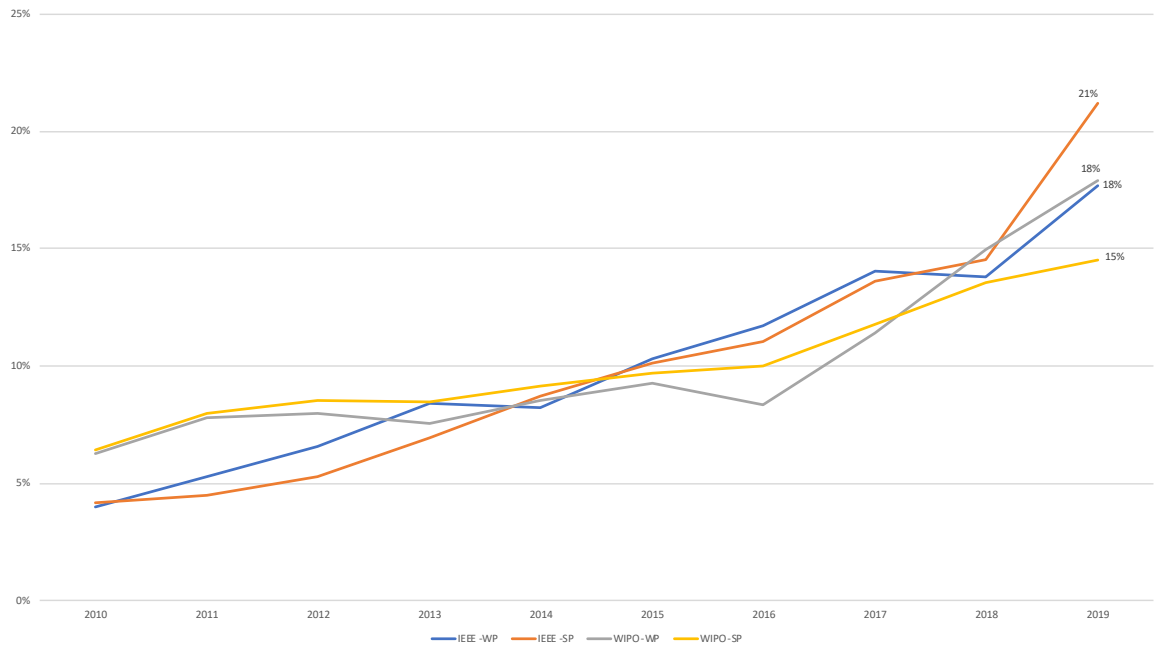

Figure 5.16: *Percent of data extracted from databases.*

Table 5.14: *Amount of data extracted from databases.*

	YEAR									
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
IEEE - Wind Power	379	503	627	799	782	980	1112	1334	1310	1682
IEEE - Solar Power	152	163	192	253	319	369	402	496	530	774
WIPO - Wind Power	4262	5301	5432	5157	5797	6294	5686	7759	10186	12190
WIPO - Solar Power	3801	4719	5056	5001	5405	5745	5919	6961	8018	8599
	FREQUENCY									

Comparing what was extracted from Wind Power and Solar Power, separately, looking

at WIPO and IEEE, in those ten years, the results are Table 5.15 and 5.16 :

Table 5.15: *Unique words that always repeat in WIPO and IEEE from 2010 to 2019. Wind and Solar Power Search.*

IEEE and WIPO - Wind Power	IEEE and WIPO - Solar Power
control	cell
generat	control
	electr
	generat

Table 5.16: *Unique words that repeat at least once in WIPO and IEEE from 2010 to 2019. Wind and Solar Power Search.*

IEEE and WIPO - Wind Power	IEEE and WIPO - Solar Power
control	batteri
convert	cell
electr	control
generat	convert
model	data
speed	devic
turbin	electr
voltag	generat

In the wind field, eight words appears in both cases, WIPO and IEEE at least once, and two repeat themselves all years. This can indicates a potential correlation between the two types of scientists, inventors and academic ones, and the work that they are producing at the same time.

Confirming that, solar power also brings eight words in WIPO that are the same as showed in the IEEE extraction at least once, and four that are present each year, looking at the same period. Revealing that researchers are studying those two renewable energies with many similarities.

On Table 5.17 and Table 5.18 it is possible to observe that researchers that applied to IEEE papers to Wind and Solar Power has a potential correlation, and also it is possible to applied the same indicative to inventors from WIPO:

Table 5.17: *Unique words that always repeat in Wind and Solar Power from 2010 to 2019. WIPO and IEEE Search.*

IEEE - Wind and Solar Power	WIPO - Wind and Solar Power
control	control
generat	devic
model	electr
perform	generat
simul	unit
voltag	

Table 5.18: *Unique words that repeat at least once in Wind and Solar Power from 2010 to 2019. WIPO and IEEE Search.*

IEEE - Wind and Solar Power	WIPO - Wind and Solar Power
control	batteri
convert	bodi
design	control
distribut	convert
electr	devic
generat	disclos
grid	electr
improv	form
model	generat
optim	heat
perform	modul
simul	plate
sourc	storag
test	suppli
voltag	support
	surfac
	unit
	util
	water

IEEE cross analysis on Wind and Solar power brought fifteen words and WIPO's nineteen that appears at least once on the period, and six and five, respectively, all years. Revealing that researchers are studying those two renewable energies, but now on the same basis, with strong similarities too.

Looking now year by year, all words that appears on the frequency word clouds made

from the extracted results, now increase the limit from twenty to two hundred words, it is possible to find other similarities. Figures 5.18 and 5.17 show new words:

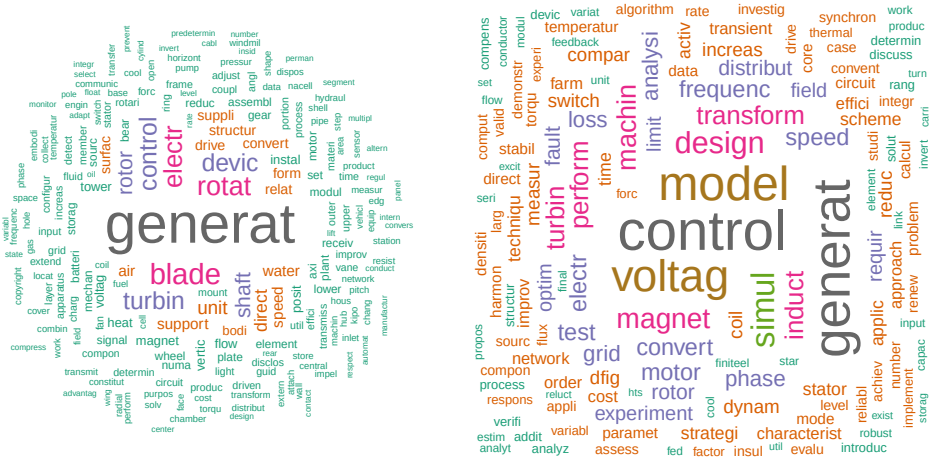

Figure 5.17: *Word clouds from WIPO and IEEE in 2010, Wind Power Search.*

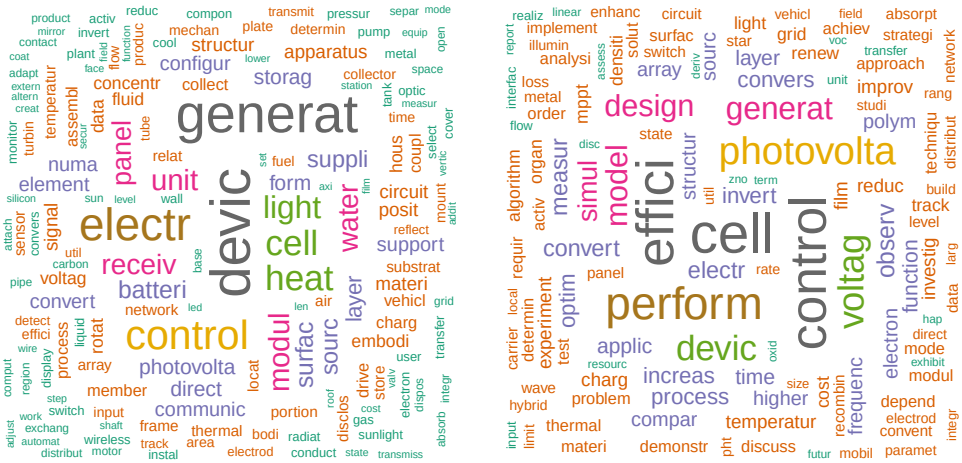

Figure 5.18: *Word clouds from WIPO and IEEE in 2010, Solar Power Search.*

Looking at the color palettes, it is possible to say that the search for Wind Power in WIPO brought more diverse words than IEEE's by checking the most frequent words.

In the Solar Power search, in both databases, the frequency of words is concentrated in a smaller amount of words, revealing a pattern between the two Solar Power searches in 2010. And extending the search on those word cloud, that have up to two hundred frequent words, the year 2010 brought some new relevant words, forty-three words appear on all four word clouds, as Table 5.19 shows:

Table 5.19: *2010 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search.*

WORDS							
circuit	data	distribut	grid	level	process	switch	work
combin	design	effici	improv	materi	reduc	temperatur	
configur	detect	electr	increas	measur	signal	time	
control	determin	field	input	modul	sourc	unit	
convert	devic	flow	integr	network	structur	util	
cost	direct	generat	invert	perform	suppli	voltag	

In 2011, also focusing on the color palette, Wind Power and Solar Power are more similar in terms of diversity and frequency of words found in the two databases. More words divided the ranking of the most frequent looking at the previous year. Figure 5.20 and Figure 5.19 show forty-nine words that are present in both databases and searches, and Table 5.20 discriminates them:

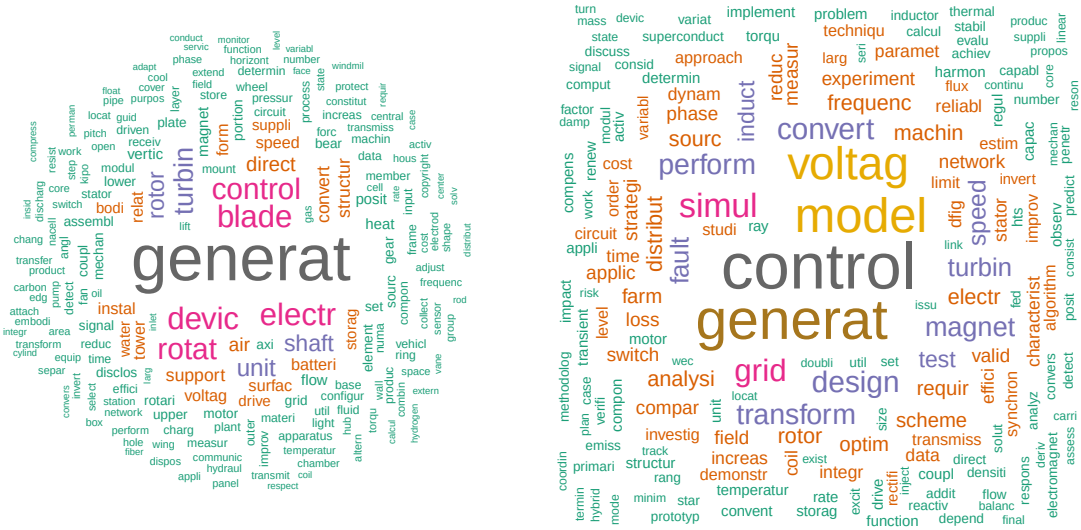

Figure 5.19: *Word clouds from WIPO and IEEE in 2011, Wind Power Search.*

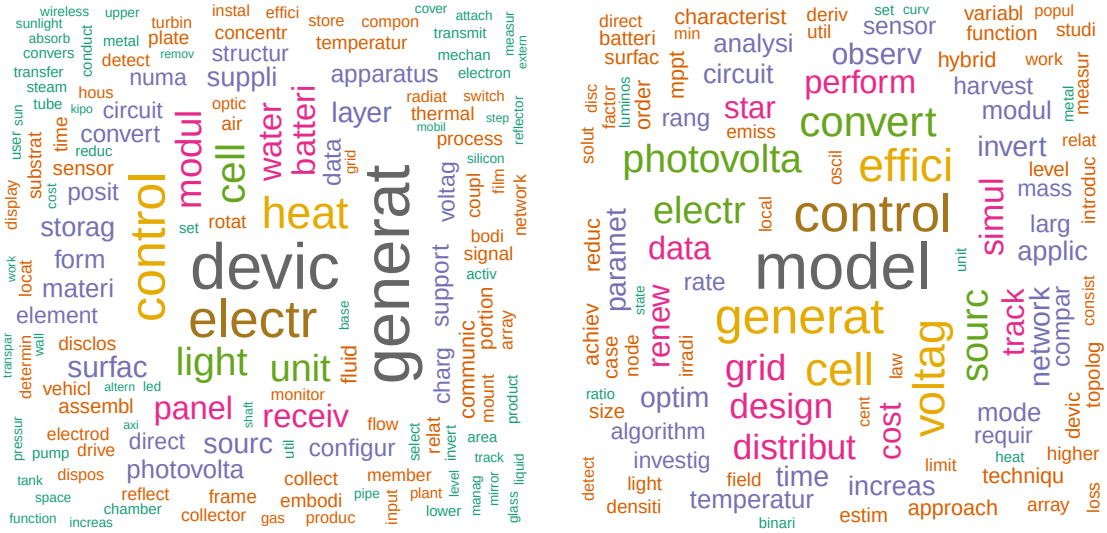

Figure 5.20: *Word clouds from WIPO and IEEE in 2011, Solar Power Search.*

Table 5.20: *2011 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

WORDS						
activ	convert	distribut	grid	locat	reduc	temperatur
circuit	cost	effici	improv	measur	set	time
combin	data	electr	increas	modul	sourc	transmiss
compon	detect	field	input	network	storag	unit
conduct	determin	flow	integr	perform	structur	util
control	devic	function	invert	phase	suppli	voltag
convers	direct	generat	level	process	switch	work

The year 2012 displays fewer words that are highlighted in relation to amount observed, and more words diluting the ranking of more frequent ones, which may indicate that more subjects are being brought to this research field. The other three searches bring words that are more frequent when compared to others, as Figure 5.22, Figure 5.21 and Table 5.21 demonstrate, and only forty-five words appear on the four searches:

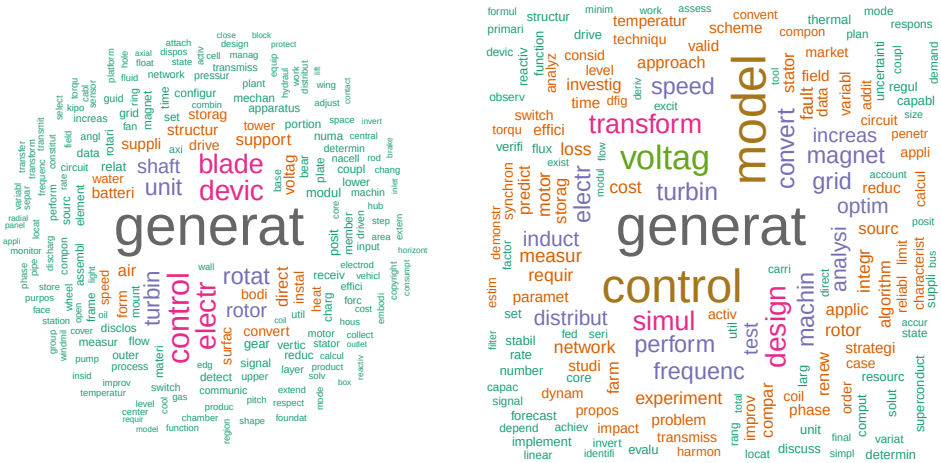

Figure 5.21: *Word clouds from WIPO and IEEE in 2012, Wind Power Search.*

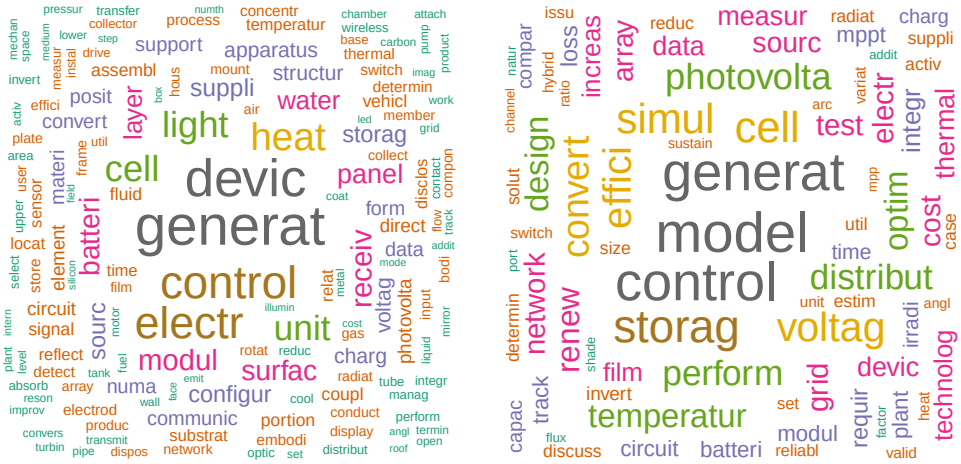

Figure 5.22: *Word clouds from WIPO and IEEE in 2012, Solar Power Search.*

Table 5.21: *2012 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

WORDS							
activ	convert	distribut	grid	measur	produc	suppli	util
batteri	cost	effici	improv	mode	reduc	switch	voltag
circuit	data	electr	increas	modul	requir	temperatur	work
combin	design	field	invert	network	set	time	
compon	determin	function	level	perform	sourc	transmiss	
control	devic	generat	locat	phase	storag	unit	
convers	direct	generat	level	process	switch	work	

Growing again, the number of words that appears on the searches increase to forty-nine in 2013, but the behavior about the frequency of words and their colors is very similar to the previous year, with Wind Power's research in WIPO with few words much more frequent and its great majority diluted in smaller frequencies. Figure 5.24, Figure 5.23 and Table 5.22 reveal them:

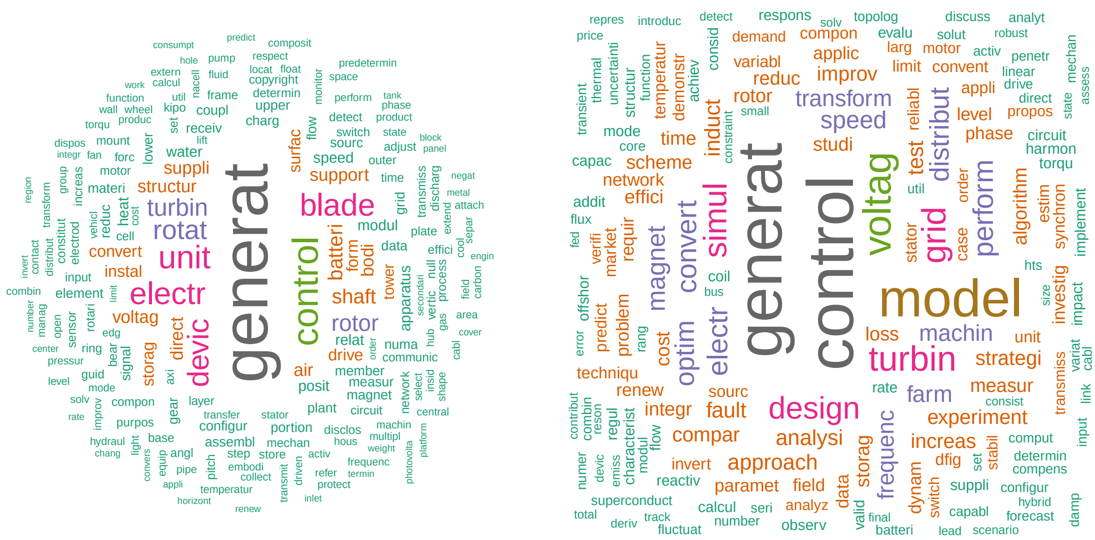

Figure 5.23: *Word clouds from WIPO and IEEE in 2013, Wind Power Search.*

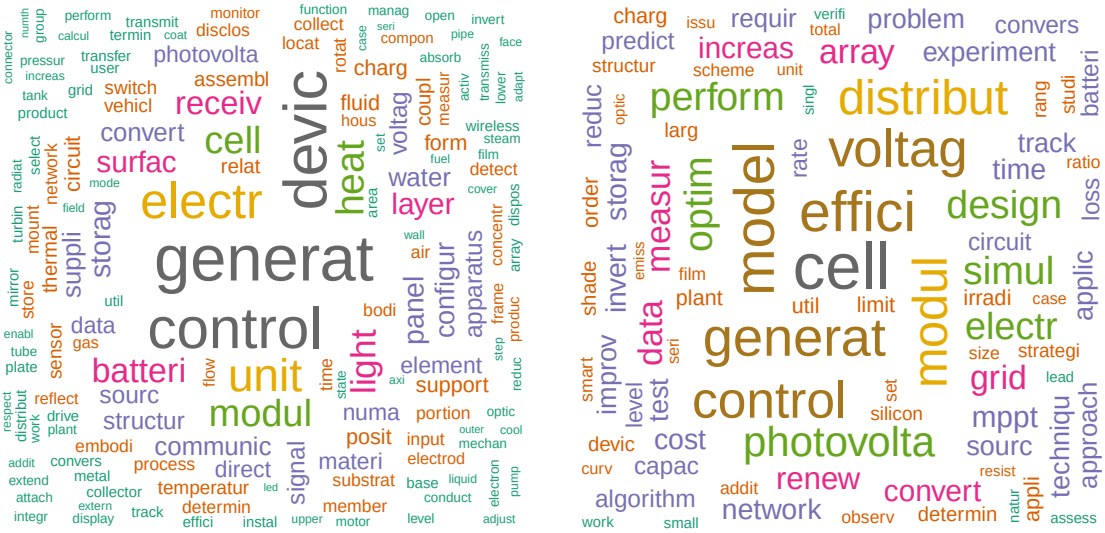

Figure 5.24: *Word clouds from WIPO and IEEE in 2013, Solar Power Search.*

Table 5.22: *2013 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

WORDS						
batteri	control	effici	improv	modul	set	temperatur
calcul	convert	electr	increas	network	sourc	time
charg	data	field	input	perform	state	transmiss
circuit	detect	flow	integr	phase	storag	unit
combin	determin	function	invert	plant	structur	util
compon	devic	generat	level	produc	suppli	voltag
configur	distribut	grid	measur	reduc	switch	work

In 2014, the number of words that is identical on all searches jumped to fifty-four, all those words can be seen on Figure 5.26, Figure 5.25 and Table 5.23, but Solar Power in WIPO searches changes a little the behavior looking at previous years and brings less words that were repeated more, diluting more the frequencies, being able to demonstrate the search for new interests of researchers in this area.

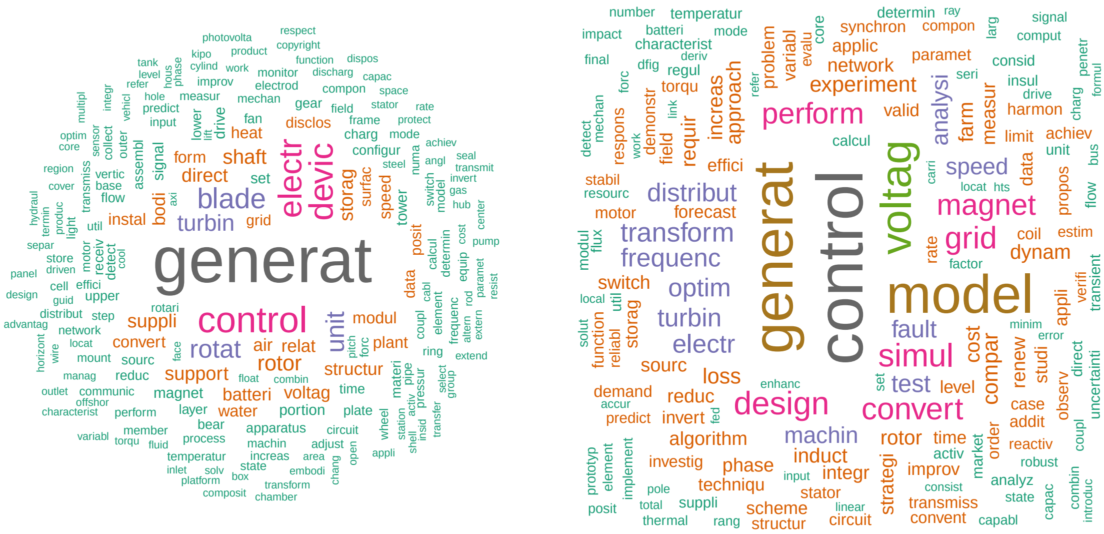

Figure 5.25: *Word clouds from WIPO and IEEE in 2014, Wind Power Search.*

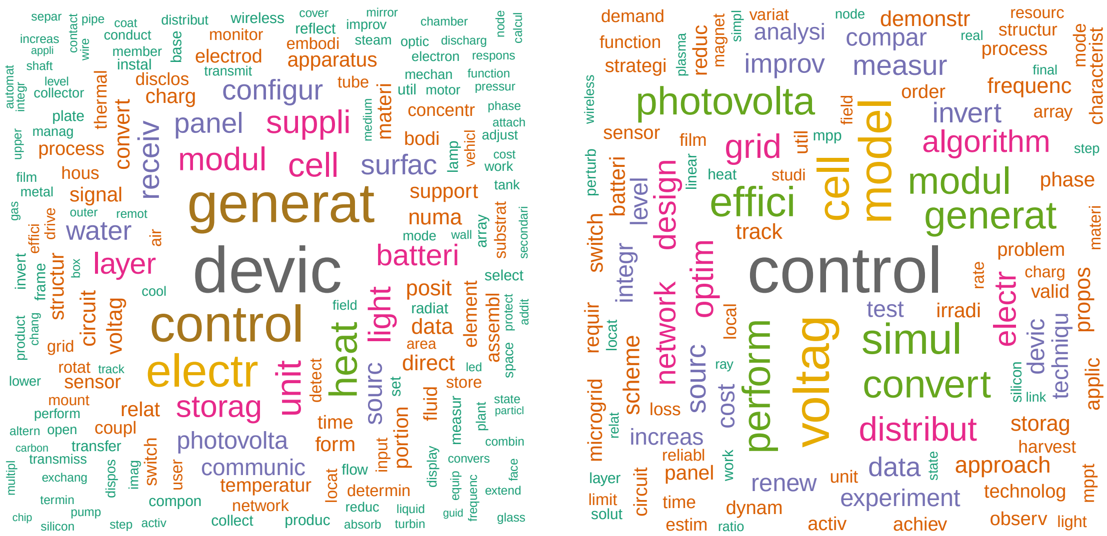

Figure 5.26: *Word clouds from WIPO and IEEE in 2014, Solar Power Search.*

Table 5.23: *2014 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

WORDS							
activ	compon	direct	generat	level	network	state	time
appli	control	distribut	grid	locat	perform	storag	unit
batteri	convert	effici	improv	manag	phase	structur	util
calcul	cost	electr	increas	measur	process	suppli	voltag
charg	data	field	input	mechan	produc	support	work
circuit	determin	frequenc	integr	mode	reduc	switch	
combin	devic	function	invert	modul	sourc	temperatur	

Table 5.24 together with Figure 5.28 and Figure 5.27 show sixty-two words that are a result of all four extracted word clouds in 2015, and Solar Power in WIPO searches continues with the same characteristic of the previous year, confirming a possible search for new interests of these researchers.

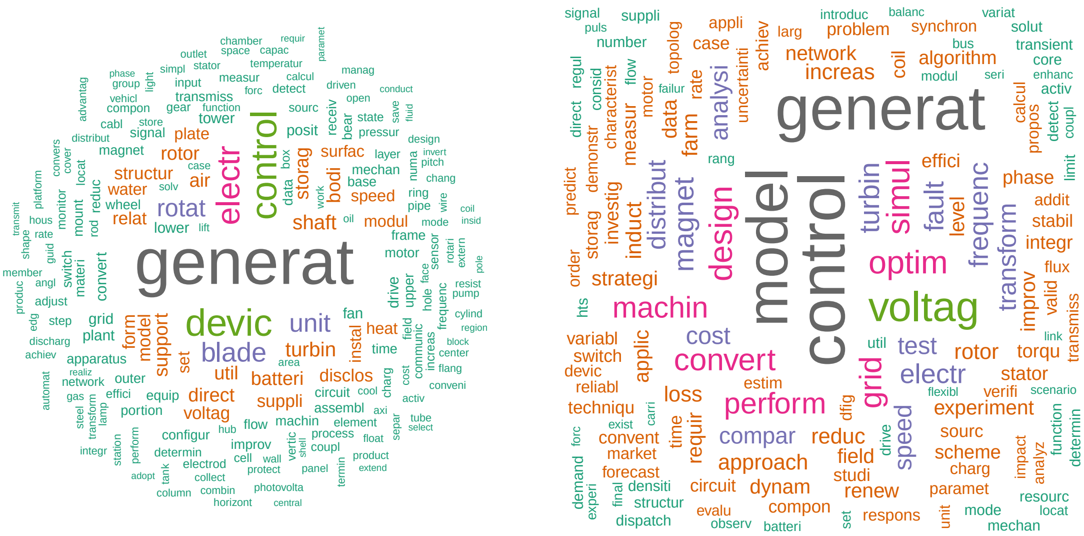

Figure 5.27: *Word clouds from WIPO and IEEE in 2015, Wind Power Search.*

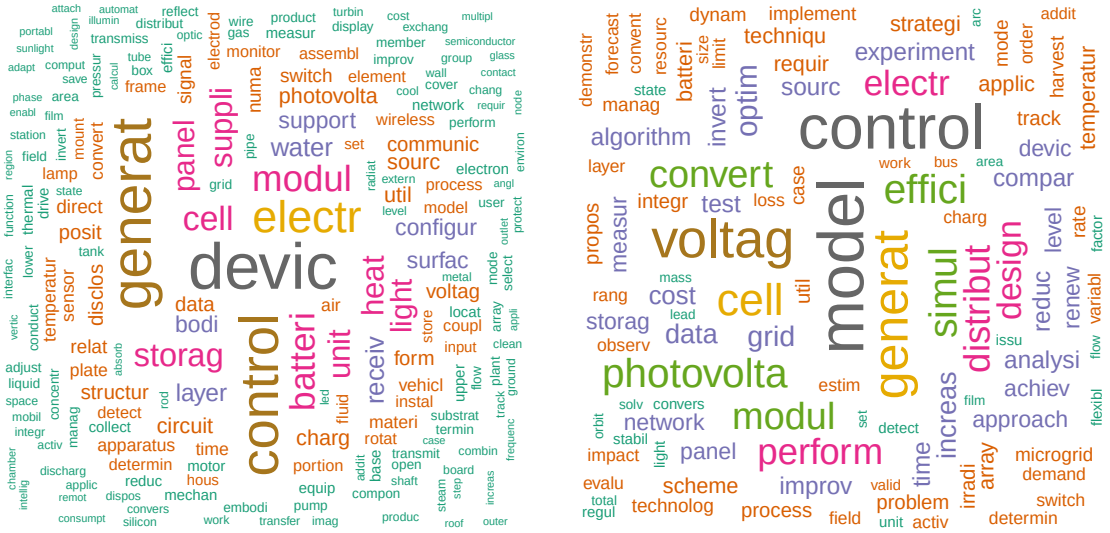

Figure 5.28: *Word clouds from WIPO and IEEE in 2015, Solar Power Search.*

Table 5.24: *2015 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

WORDS							
achiev	compon	determin	frequenc	invert	phase	signal	time
activ	configur	devic	function	manag	plant	sourc	transmiss
batteri	control	direct	generat	measur	process	state	unit
calcul	convert	distribut	grid	mode	produc	storag	util
case	cost	effici	improv	model	product	structur	voltag
charg	data	electr	increas	modul	reduc	suppli	work
circuit	design	field	input	network	requir	switch	
combin	detect	flow	integr	perform	set	temperatur	

Ending 2016 the amount of words that repeats on all searches were fifty-three, Solar Power and Wind Power searches at WIPO has brought considerable differences looking at IEEE, researchers show that they are diversifying into one and focusing their research on specific topics in another, as Figure 5.30, Figure 5.29 and Table 5.25 reveal:

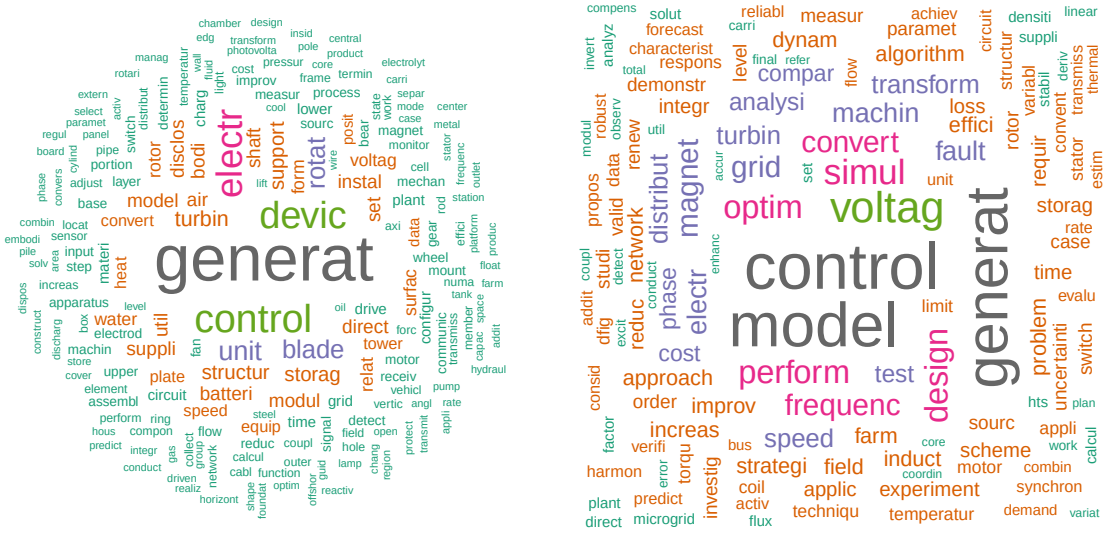

Figure 5.29: *Word clouds from WIPO and IEEE in 2016, Wind Power Search.*

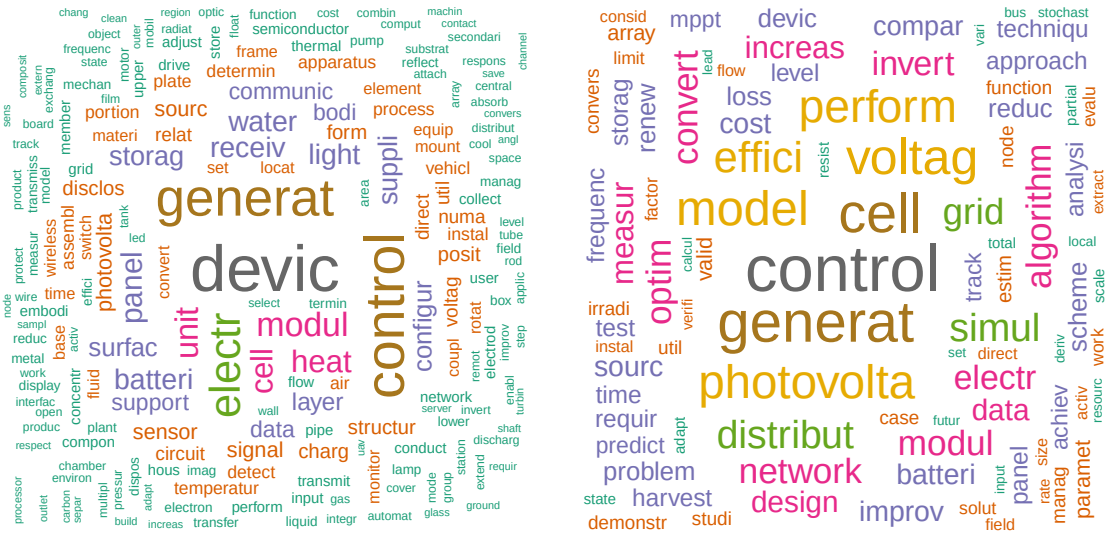

Figure 5.30: *Word clouds from WIPO and IEEE in 2016, Solar Power Search.*

Table 5.25: *2016 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

WORDS							
activ	convert	distribut	generat	locat	perform	storag	unit
batteri	cost	effici	grid	manag	plant	structur	util
circuit	data	electr	improv	measur	process	suppli	voltag
combin	detect	field	increas	mode	reduc	switch	work
conduct	determin	flow	input	model	set	temperatur	
configur	devic	frequenc	integr	modul	sourc	time	
control	direct	function	level	network	state	transmiss	

Sixty-five words repeat themselves in 2017, as it can be observed on Figure 5.32, Figure 5.31 and Table 5.26, Wind Power at IEEE reveals that more words were the target of research than in previous years, and the behavior of Solar, which already had the same behavior at both databases, is sustained.

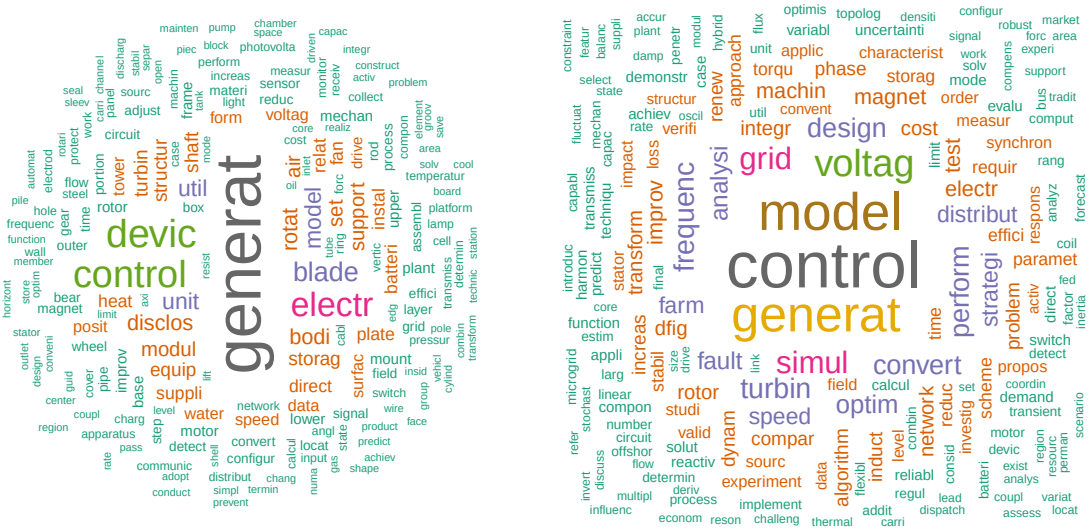

Figure 5.31: *Word clouds from WIPO and IEEE in 2017, Wind Power Search.*

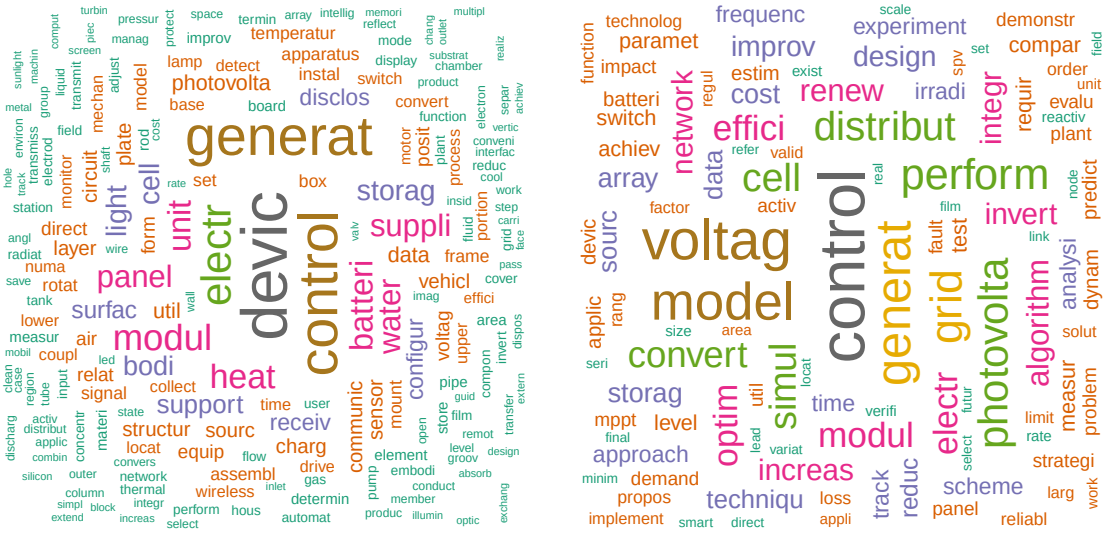

Figure 5.32: *Word clouds from WIPO and IEEE in 2017, Solar Power Search.*

Table 5.26: *2017 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

WORDS							
achiev	compon	determin	function	locat	plant	state	time
activ	control	devic	generat	measur	process	storag	transmiss
area	convert	direct	grid	mode	rate	structur	unit
batteri	cost	distribut	improv	model	reduc	suppli	util
case	data	effici	increas	modul	set	support	voltag
circuit	design	electr	integr	network	signal	switch	work
combin	detect	field	level	perform	sourc	temperatur	

Wind and Solar Power in IEEE and Wind in WIPO following the trend of the previous year, strongly indicating that in the year 2018 a greater amount of subjects were inserted as focus of studies by these researchers, Figure 5.34, Figure 5.33 and Table 5.27 show fifty-one words that repeated on all searches:

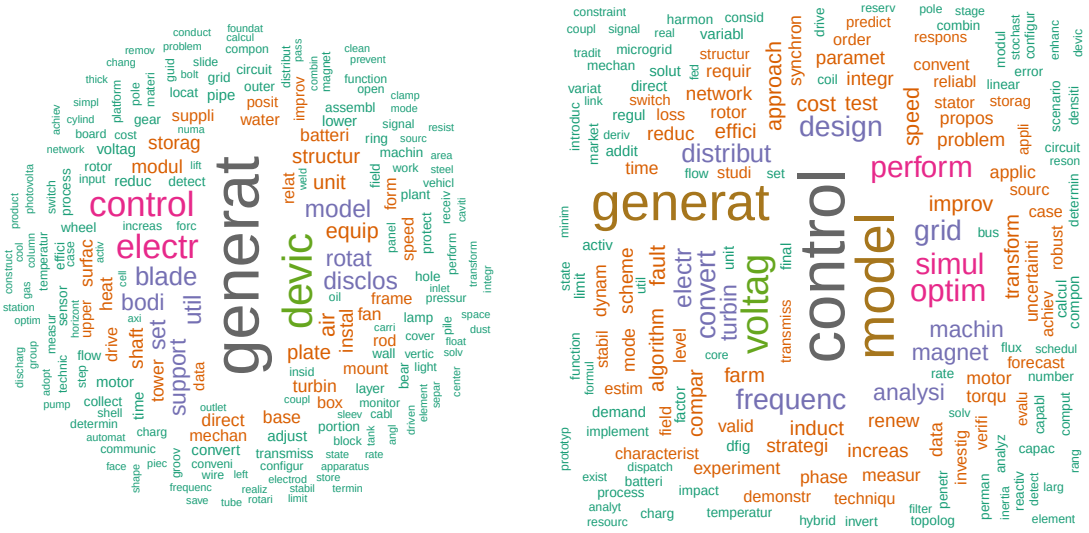

Figure 5.33: *Word clouds from WIPO and IEEE in 2018, Wind Power Search.*

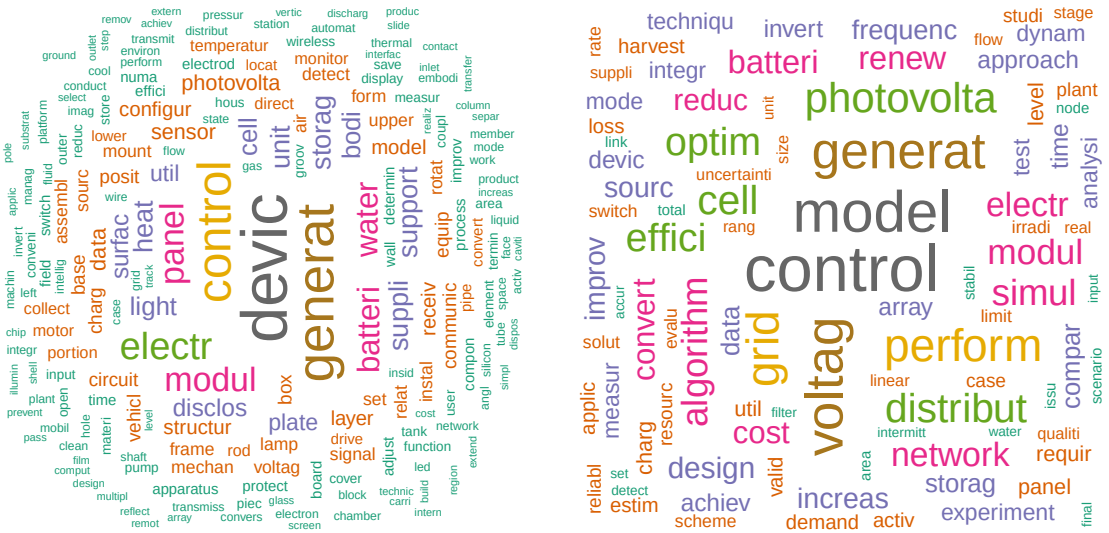

Figure 5.34: *Word clouds from WIPO and IEEE in 2018, Solar Power Search.*

Table 5.27: *2018 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

WORDS							
achiev	configur	devic	generat	mode	process	structur	util
activ	control	direct	grid	model	reduc	suppli	voltag
batteri	convert	distribut	improv	modul	set	switch	
case	cost	effici	increas	network	signal	temperatur	
charg	data	electr	input	perform	sourc	time	
circuit	detect	flow	integr	photovolta	state	transmiss	
compon	determin	function	measur	plant	storag	unit	

During the previous year more words are diluting the frequency in all the databases, raising the possibility of a change in the focus of the researches carried out, including Solar Power at IEEE, and the search demonstrated on Figure 5.36, Figure 5.35 and Table 5.28 showing fifty-nine identical words:

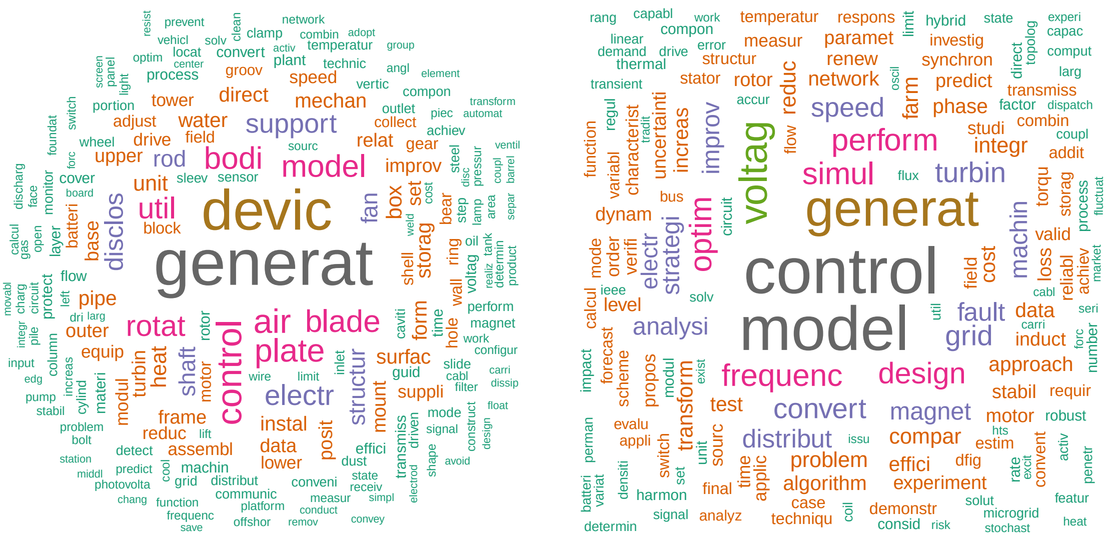

Figure 5.35: *Word clouds from WIPO and IEEE in 2019, Wind Power Search.*

Table 5.28: *2019 extracted words from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

WORDS							
achiev	configur	devic	grid	mode	reduc	structur	util
activ	control	direct	heat	model	sensor	suppli	voltag
batteri	convert	distribut	improv	modul	set	support	work
charg	cost	effici	increas	network	signal	switch	
circuit	data	electr	input	perform	solv	temperatur	
combin	design	flow	integr	plant	sourc	time	
compon	detect	function	limit	problem	state	transmiss	
conduct	determin	generat	measur	process	storag	unit	

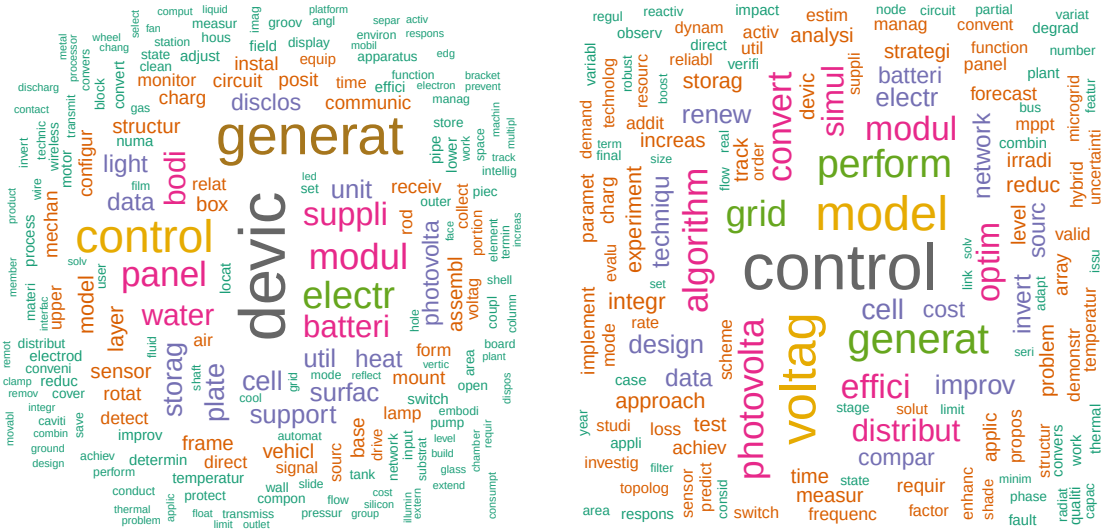

Figure 5.36: *Word clouds from WIPO and IEEE in 2019, Solar Power Search.*

Looking in all those years, at least twenty-one percent of the two hundred most frequent words repeat themselves as shows Table 5.29:

Table 5.29: *Amount of words that repeats from the four Word clouds (IEEE and WIPO on Wind and Solar Power Search).*

YEAR	QUANTITY	%
2010	43	21,5
2011	49	24,5
2012	45	22,5
2013	49	24,5
2014	54	27
2015	62	31
2016	53	26,5
2017	55	27,5
2018	51	25,5
2019	59	29,5

This indicates a potential correlation between inventors and academic scientists, and the work that they are producing at the same time for both studied renewable energies.

Bringing this lack of correlation into view can aid other researchers guiding their studies or new researches and eventually papers and patents showing how much and which research areas industry and academy differ the most.

The extracted knowledge demonstrates the trajectory accomplished in the two studied fields, solar and wind, with some words disappearing, appearing, intensifying or losing strength over the years. Some technologies were used in the studied renewable energies, and the proximity between the two types of researchers, the more academic and those directed to the industry that patents the discoveries is strong.

The contributions of this work are divided into sections for a better understanding of whether or not the proposed objectives have been achieved:

5.4 Group I: Objectives related to information retrieved from patent documents from 2010 to 2019

- It was possible to make an analysis of the trajectory of patents related to solar and wind energy by looking at the keywords found, some words/themes were stronger in the studied period and others were disappearing or appearing in some years. In wind power, for example, the patent documents have disclosed words related to some mechanical structures, such as blade, shaft and rod, among others. In the solar power field, the technological developments are strongly connected to solar cells, or other words related to power storage and battery packs;

- The main applicants of patents related to wind and solar energy were identified, as

well as positioned during the indicated period. The State Grid Corporation of China, that reached the first position in 2014 and have been keeping it in the Wind Power scenario. As for Solar Power, it showed another relevant applicant, Suzhou Csi Solar Power Technology Co., Ltd., that got the first position in 2017 and stood there until 2018. It is possible to observe that two applicants are present in both lists, General Electric Company and Siemens Aktiengesellschaft, which indicates that companies that are patenting are interested in both renewable energies in those years searched;

- The countries/offices that received the most patent applications for the renewable energy studied were listed and placed during the studied period. The ten countries/offices in each list are the same, indicating that those Countries/Offices are investing in Solar and Wind Power research in those ten years. In Wind search China got the first position and U.S. in Solar

- The subjects that were most patented using the IPC code for wind and solar energy were displayed and commented on. In the Wind Power search the IPC Code F03D was in the first position in all years. The F03D code is related to "wind motors". Solar Power, however, has a new one every 2-4 years. Over the past five years, the applicants have been investing efforts and resources in the IPC codes H02S and H02J. These codes are related respectively to generation of electric power by conversion of infrared radiation, visible light or ultraviolet light, e.g., using photovoltaic modules, and circuit arrangements or systems for supplying or distributing electric power.

5.5 Group II: Objectives related to information retrieved from papers from 2010 to 2019

- The publication titles helped to identify where within the IEEE Library the solar and wind energy Journals were being applied to. By observing the IEEE papers in Wind Energy, the words are more concentrated in control strategy, converter and simulation, among others, while inventions, as cited before, are concentrated toward mechanical improvements. In solar power IEEE journals, the developments have suggested more scientific efforts in control algorithms, solar cells and PV panels;

- The main institutes conducting the most research and publishing on solar and wind energy were identified, as well as positioned during the indicated period. The results have shown that Aalborg University, from Denmark, and Chinese Institutions have published the most papers about wind power during the past ten years. In addition, the Indian Institute of Technology Delhi is the most important research institution related to solar power field;

- It was possible to make an analysis of the trajectory of journals within the IEEE Library related to solar and wind energy by looking at the keywords found and its bigrams, showing that some words/themes were stronger in the studied period and others were disappearing or appearing in some years. Some words were highlighted, as the word "strategy", that came up strong on 2019 and it is related to a control strategy in the wind energy developments. In the solar power field, some technical concepts have been investigated and in this case some bigrams like "solar cell", "solar photovoltaic" and "control scheme", among others, allow us to understand the state of the art technology developments.

5.6 Group III: Objectives related to information retrieved from both data repositories from 2010 to 2019

- It was possible to identify the most researched/studied energy in the two databases, IEEE and WIPO. Wind energy had more articles published and more patents proposed;

- Some words were repeated when searching for wind energy in both databases during the whole period researched, as well as for solar energy in the same context, indicating a relationship of what was being studied in the renewable energy sought. Words such as "control" and "generat" could be found in both cross analysis;

- A connection about what was studied about solar and wind energy in the WIPO database was identified using the keywords extracted, as well as the IEEE database looking at the two renewable energies, indicating a potential correlation between the two types of scientists, inventors and academic ones, and the work that they are producing at the same time. In the wind field, eight words appears in both search scenarios, WIPO and IEEE, at least once and two repeat themselves all years. Solar power also brings eight words in WIPO that are the same as showed in the IEEE extraction at least once, and four that are present each year, looking at the same period;

- Some patterns were identified looking at wind and solar energy in the IEEE and WIPO database during the studied period, a year by year analysis of these patterns was also possible. Analysing the color palettes in the extracted wordclouds it possible to see that more words are diluting the frequency in all the databases over the years, raising the possibility of a change in the focus of the researches carried out in all four searches. And looking in all those years, at least twenty-one percent of the two hundred most frequent words repeat themselves, indicating a potential correlation between inventors and academic scientists, and the work that they are producing at the same time for Solar and Wind energy.

The evaluation of the extracted knowledge is that other researchers will be able to use the information obtained to improve their research with this data.

Chapter 6

Conclusions

This research proposed to show a technological trajectory analysis between Patents and Journals in the Renewable Energy field, Wind and Solar Power, during the past decade, 2010-2019. To that end, it utilised data extracted from the IEEE and WIPO databases.

Graphs, Tables and Charts were created as a result of the extracted data processing by utilising text mining and the relationship between Patents and Journals have been evidenced in all cases analyzed. Thus this research has identified the proposed technological trajectory analysis based in its find and aims to provide the first comprehensive investigation of technological trends in solar and wind energy considering an isolated assessment of these two source of information, as well as a correlation between the data retrieved from the two repositories.

It is worth remembering that the analysis of Renewable Energies undertaken here has extended the knowledge of how Patents and Journals databases relate for academic purposes and a text mining approach in Renewable Energies studies, as solar and wind energy, presented in Patents and Journals used a method and process that may be applied to other researches elsewhere in the world.

Prior to this study it was difficult to make assertions about how Inventors and Academic Researchers were working regarding specific renewable energies sources between the years of 2010 and 2019.

This study lays the groundwork for future research into other fields other than Wind and Solar Power or Renewable Energy.

The most important limitation lies in the fact that it is not easy to define if an English Top Word, one of most frequent words in that idiom, is relevant or not for each search.

It is possible to reveal that there are similarities in the extracted results, both by analyzing wind and solar energy separately and by cross-referencing them in the two databases, and it is also possible to indicate the extracted results as a source for further research on the subject. With this, it is reasonable to affirm that the specific objectives presented were met. And for having achieved the all proposed objectives it is thus possible to conclude that the study has fulfilled its purpose.

A natural progression of this work is to analyze the proximity of those extracted words in regard to context, identifying more information and not just words.

Chapter 7

Publications associated with this master's thesis

Dias, C.G.; Morais de, A.M.; Análise da Trajetoria Tecnológica da Energia Eólica a Partir de Técnicas de Mineração de Texto em Patentes. Revista Sodebras (Qualis B3), v.15, n.180, 2020.

Dias, C.G.; Morais de, A.M.; A Text Mining Approach in Patents and Papers: Analyzing Wind and Solar Energy Trends. International Journal of Computer Applications (IJCA) (Qualis A4), 175(37), pp. 1-12, 2020.

Bibliography

ABBOTT, F. M.; COTTIER, T.; GURRY, F. *International intellectual property in an integrated world economy*. [S.l.]: Aspen Publishers, 2019. Cited pag. 19.

AGGARWAL, C. C.; ZHAI, C. *Mining text data*. [S.l.]: Springer Science & Business Media, 2012. Cited pag. 43.

ALBINO, V. et al. Understanding the development trends of low-carbon energy technologies: A patent analysis. *Applied Energy*, Elsevier, vol. 135, p. 1–19, 2014. Cited pag. 29.

ALEIXANDRE-TUDÓ, J. L. et al. Renewable energies: Worldwide trends in research, funding and international collaboration. *Renewable Energy*, Elsevier, vol. 139, p. 268–278, 2019. Cited pag. 29.

AMADO, A. et al. Research trends on big data in marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*, Elsevier, vol. 24, n. 1, p. 1–7, 2018. Cited pag. 28, 45.

ARAS, H. et al. Applications and challenges of text mining with patents. In: *IPaMin@ KONVENS*. [S.l.: s.n.], 2014. Cited pag. 25, 36, 37, 38, 43.

BANOS, R. et al. Optimization methods applied to renewable and sustainable energy: A review. *Renewable and sustainable energy reviews*, Elsevier, vol. 15, n. 4, p. 1753–1766, 2011. Cited pag. 14.

BIENVENIDO-HUERTAS, D. et al. Patents analysis of thermal bridges in slab fronts and their effect on energy demand. *Energies*, Multidisciplinary Digital Publishing Institute, vol. 11, n. 9, p. 2222, 2018. Cited pag. 30.

BUKOWSKI, M. et al. Feasibility of activity-based expert profiling using text mining of scientific publications and patents. *Scientometrics*, Springer, p. 1–42, 2020. Cited pag. 29, 41, 45.

CHENG, M.; ZHU, Y. The state of the art of wind energy conversion systems and technologies: A review. *Energy Conversion and Management*, Elsevier, vol. 88, p. 332–347, 2014. Cited pag. 29.

COHEN, W. M.; NELSON, R. R.; WALSH, J. P. *Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)*. [S.l.], 2000. Cited pag. 17, 19.

DARWIN, C.; SCHUMPETER, J. A. Padrões históricos da mudança tecnológica e ondas longas do desenvolvimento capitalista. *Ricardo Dathein (org.)*, p. 223, 2015. Cited pag. 20.

DIAS, C. G. uzzy systems applications: a literature review revealed by patents and papers using text mining. *Interdisciplinary Scientific Journal*, LinkSciencePlace, vol. 5, n. 3, p. 125–147, 2018. Cited pag. 36.

EIA. *Renewable energy explained*. 2019. Available from Internet: <https://www.eia.gov/energyexplained/renewable-sources/>. Cited 27 aug. 2020. Cited pag. 14.

FETTWEIS, G. P. The tactile internet: Applications and challenges. *IEEE Vehicular Technology Magazine*, IEEE, vol. 9, n. 1, p. 64–70, 2014. Cited pag. 42.

FONTANA, R.; NUVOLARI, A.; VERSPAGEN, B. Mapping technological trajectories as patent citation networks. an application to data communication standards. *Economics of Innovation and New Technology*, Taylor & Francis, vol. 18, n. 4, p. 311–336, 2009. Cited pag. 29.

FRANÇA, R. O. Patente como fonte de informação tecnológica. *Perspectivas em ciência da informação*, vol. 2, n. 2, 1997. Cited pag. 19.

GRAHAM, R. The global state of the art in engineering education. *Massachusetts Institute of Technology (MIT) Report, Massachusetts, USA*, p. 1–162, 2018. Cited pag. 65.

GRILICHES, Z. Patent statistics as economic indicators: a survey. In: *R&D and productivity: the econometric evidence*. [S.l.]: University of Chicago Press, 1998. p. 287–343. Cited pag. 33.

GROSSOEHME, D.; LIPSTEIN, E. Analyzing longitudinal qualitative data: the application of trajectory and recurrent cross-sectional approaches. *BMC research notes*, Springer, vol. 9, n. 1, p. 1–5, 2016. Cited pag. 20.

GUPTA, V.; BANSAL, N.; SHARMA, A. *Text summarization for big data: A comprehensive survey*. 2019. 503–516 p. Cited pag. 16.

HAASE, H.; ARAÚJO, E. C. de; DIAS, J. Inovações vistas pelas patentes: exigências frente às novas funções das universidade. *Revista Brasileira de Inovação*, vol. 4, n. 2, p. 329–362, 2005. Cited pag. 19.

HALL, J. Environmental supply chain dynamics. *Journal of cleaner production*, Elsevier, vol. 8, n. 6, p. 455–471, 2000. Cited pag. 22.

HAND, M. M. et al. *Renewable electricity futures study. volume 1. Exploration of high-penetration renewable electricity futures*. [S.l.], 2012. Cited pag. 14.

HEINBERG, R.; FRIDLEY, D. *Our Renewable Future: Laying the Path for One Hundred PercClean Energy*. [S.l.]: Island Press, 2016. Cited pag. 22.

IEEE. *IEEE Journals Continue to Excel in Citation Rankings*. [S.l.], 2020. Available from Internet: <https://www.ieee.org/publications/subscriptions/journal-citations.html>. Cited 03 dec. 2020. Cited pag. 42.

INFIELD, D.; FRERIS, L. *Renewable energy in power systems*. [S.l.]: John Wiley & Sons, 2020. 18-19 p. Cited pag. 16.

JIANG, M.; SHANG, J. Scientific text mining and knowledge graphs. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. [S.l.: s.n.], 2020. p. 3537–3538. Cited pag. 43.

JIN, M.; KO, H. K. Analysis of trends in mathematics education research using text mining. *Communications of Mathematical Education*, Korean Society of Mathematical Education, vol. 33, n. 3, p. 275–294, 2019. Cited pag. 28, 45.

JUNG, H.; LEE, B. G. Research trends in text mining: Semantic network and main path analysis of selected journals. *Expert Systems with Applications*, Elsevier, p. 113851, 2020. Cited pag. 28, 44.

KABIR, E. et al. Solar energy: Potential and future prospects. *Renewable and Sustainable Energy Reviews*, vol. 82, p. 894 – 900, 2018. Cited pag. 14.

KIM, G.; BAE, J. A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change*, Elsevier, vol. 117, p. 228–237, 2017. Cited pag. 28.

KIM, N. et al. Dynamic patterns of industry convergence: Evidence from a large amount of unstructured data. *Research Policy*, Elsevier, vol. 44, n. 9, p. 1734–1748, 2015. Cited pag. 25, 44.

KUMAR, V. et al. Mapping technological trajectories for energy storage device through patent citation network. In: IEEE. *2018 9th International Conference on Awareness Science and Technology (iCAST)*. [S.l.], 2018. p. 56–61. Cited pag. 29.

LIZIN, S. et al. A patent landscape analysis for organic photovoltaic solar cells: Identifying the technology's development phase. *Renewable energy*, Elsevier, vol. 57, p. 5–11, 2013. Cited pag. 29.

LUDVIGSEN, S. R. et al. *Learning across sites: New tools, infrastructures and practices*. [S.l.]: Routledge, 2010. Cited pag. 20.

MADANI, F.; WEBER, C. The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis. *World Patent Information*, Elsevier, vol. 46, p. 32–48, 2016. Cited pag. 29, 45.

MINER, G. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. [S.l.]: Elsevier, 2012. 1053 p. Cited pag. 11, 35, 36, 41, 43.

MUBAROK, M. H.; NAFIZAH, U.; PERMANA, M. Y. Mapping technological trajectories of crystalline silicon (c-si) pv using patent analysis. *International Journal of Renewable Energy Research (IJRER)*, vol. 9, n. 4, p. 1660–1671, 2019. Cited pag. 29.

NEILL, S. P. et al. Tidal range energy resource and optimization–past perspectives and future challenges. *Renewable energy*, Elsevier, vol. 127, p. 763–778, 2018. Cited pag. 14.

NEUHÄUSLER, P.; ROTHENGATTER, O.; FRIETSCH, R. *Patent Applications-Structures, Trends and Recent Developments 2018*. [S.l.], 2019. Cited pag. 29.

NEWBERY, D. et al. Market design for a high-renewables european electricity system. *Renewable and Sustainable Energy Reviews*, Elsevier, vol. 91, p. 695–707, 2018. Cited pag. 14.

OLAJIRE, A. A. The brewing industry and environmental challenges. *Journal of Cleaner Production*, Elsevier, vol. 256, p. 102817, 2020. Cited pag. 14.

OMER, A. M. Energy, environment and sustainable development. *Renewable and sustainable energy reviews*, Elsevier, vol. 12, n. 9, p. 2265–2300, 2008. Cited pag. 22.

PARK, H.; REE, J. J.; KIM, K. Identification of promising patents for technology transfers using triz evolution trends. *Expert Systems with Applications*, Elsevier, vol. 40, n. 2, p. 736–743, 2013. Cited pag. 26, 27, 28, 44.

PENG, F. et al. Bilateral coordinated dispatch of multiple stakeholders in deep peak regulation. *IEEE Access*, IEEE, vol. 8, p. 33151–33162, 2020. Cited pag. 14.

PÍTSICA, M. *ENERGIAS RENOVÁVEIS: o papel da Irena*. [S.l.]: Appris Editora e Livraria Eireli-ME, 2015. Cited pag. 14.

POTTELSBERGHE, B. V. *Lost property: the European patent system and why it doesn't work. Bruegel Blueprint Series No. 9, 29 June 2009*. [S.l.: s.n.], 2009. vol. 9. 49–66 p. Cited pag. 33.

RAFFOUL, N.; BRION, A. Reasons for patent protection and cost-effective patent filing options for smes. *Technology Innovation Management Review*, vol. 1, n. 3, 2011. Cited pag. 17.

(REN21), R. E. P. N. for the 21st C. *Renewables 2020 Global Status Report*. [S.l.], 2020. Available from Internet: <https://www.ren21.net/gsr-2020/>. Cited 03 dec. 2020. Cited pag. 15.

RODRIGUEZ-ESTEBAN, R.; BUNDSCHUS, M. Text mining patents for biomedical knowledge. *Drug discovery today*, Elsevier, vol. 21, n. 6, p. 997–1002, 2016. Cited pag. 25, 44.

SAMPAIO, P. G. V. et al. Photovoltaic technologies: Mapping from patent analysis. *Renewable and Sustainable Energy Reviews*, Elsevier, vol. 93, p. 215–224, 2018. Cited pag. 29.

SHIH, M.-J.; LIU, D.-R.; HSU, M.-L. Discovering competitive intelligence by mining changes in patent trends. *Expert Systems with Applications*, Elsevier, vol. 37, n. 4, p. 2882–2890, 2010. Cited pag. 28, 34, 44.

SILGE, J.; ROBINSON, D. *Text mining with R*. [S.l.]: O'Reilly, 2017. Cited pag. 52, 53.

SINGH, G. K. Solar power generation by pv (photovoltaic) technology: A review. *Energy*, Elsevier, vol. 53, p. 1–13, 2013. Cited pag. 29.

SONG, K.; KIM, K. S.; LEE, S. Discovering new technology opportunities based on patents: Text-mining and f-term analysis. *Technovation*, Elsevier, vol. 60, p. 1–14, 2017. Cited pag. 25, 26, 44.

STERLACCHINI, A. Trends and determinants of energy innovations: patents, environmental policies and oil prices. *Journal of Economic Policy Reform*, Taylor & Francis, vol. 23, n. 1, p. 49–66, 2020. Cited pag. 29.

STIM, R. *Patent, copyright & trademark: an intellectual property desk reference*. [S.l.]: Nolo, 2020. vol. 16. 7-163 p. Cited pag. 17.

SUNIKKA, A.; BRAGGE, J. Applying text-mining to personalization and customization research literature–who, what and where? *Expert Systems with Applications*, Elsevier, vol. 39, n. 11, p. 10049–10058, 2012. Cited pag. 28, 44.

TSAI, C.-W.; LIU, T.-K.; HSUEH, P.-W. Patent analysis of high efficiency tunneling oxide passivated contact solar cells. *Energies*, Multidisciplinary Digital Publishing Institute, vol. 13, n. 12, p. 3060, 2020. Cited pag. 30.

TWIDELL, J.; WEIR, T. *Renewable energy resources*. [S.l.]: Routledge, 2015. Cited pag. 21.

VERSPAGEN, B. Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in complex systems*, World Scientific, vol. 10, n. 01, p. 93–115, 2007. Cited pag. 29.

VIJAYARANI, S.; ILAMATHI, M. J.; NITHYA, M. Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, vol. 5, n. 1, p. 7–16, 2015. Cited pag. 34.

VOET, M. A. *The generic challenge: understanding patents, FDA and pharmaceutical life-cycle management*. [S.l.]: BrownWalker Press, 2020. 15-119 p. Cited pag. 17.

WEISS, S. M.; INDURKHYA, N.; ZHANG, T. *Fundamentals of predictive text mining*. [S.l.]: Springer, 2015. 13–21 p. Cited pag. 34, 35, 36, 39.

WIPO. *Patents*. 2019. Available from Internet: <https://www.wipo.int/patents/>. Cited 08 sep. 2020. Cited pag. 11, 33.

YU, D. et al. Information sciences 1968–2016: a retrospective analysis with text mining and bibliometric. *Information Sciences*, Elsevier, vol. 418, p. 619–634, 2017. Cited pag. 28, 45.

ZHENG, P. et al. Mapping the field of communication technology research in asia: content analysis and text mining of ssci journal articles 1995–2014. *Asian Journal of Communication*, Taylor & Francis, vol. 26, n. 6, p. 511–531, 2016. Cited pag. 28, 45.

Appendices

Chapter 8

Appendix

These are the queries that were made in order to gather data from the WIPO database:

- `ALL:("solar power") AND DP:([01.01.2010 TO 31.12.2010])`
- `ALL:("solar power") AND DP:([01.01.2011 TO 31.12.2011])`
- `ALL:("solar power") AND DP:([01.01.2012 TO 31.12.2012])`
- `ALL:("solar power") AND DP:([01.01.2013 TO 31.12.2013])`
- `ALL:("solar power") AND DP:([01.01.2014 TO 31.12.2014])`
- `ALL:("solar power") AND DP:([01.01.2015 TO 31.12.2015])`
- `ALL:("solar power") AND DP:([01.01.2016 TO 31.12.2016])`
- `ALL:("solar power") AND DP:([01.01.2017 TO 31.12.2017])`
- `ALL:("solar power") AND DP:([01.01.2018 TO 31.12.2018])`
- `ALL:("solar power") AND DP:([01.01.2019 TO 31.12.2019])`
- `ALL:("wind power") AND DP:([01.01.2010 TO 31.12.2010])`
- `ALL:("wind power") AND DP:([01.01.2011 TO 31.12.2011])`
- `ALL:("wind power") AND DP:([01.01.2012 TO 31.12.2012])`
- `ALL:("wind power") AND DP:([01.01.2013 TO 31.12.2013])`
- `ALL:("wind power") AND DP:([01.01.2014 TO 31.12.2014])`
- `ALL:("wind power") AND DP:([01.01.2015 TO 31.12.2015])`
- `ALL:("wind power") AND DP:([01.01.2016 TO 31.12.2016])`
- `ALL:("wind power") AND DP:([01.01.2017 TO 31.12.2017])`
- `ALL:("wind power") AND DP:([01.01.2018 TO 31.12.2018])`
- `ALL:("wind power") AND DP:([01.01.2019 TO 31.12.2019])`

These are the queries that were built to gather data from IEEE. As this database does not provide the option to filter the Period or the type with queries, the type as "Journals" and the year were applied manually for each search:

- `("All Metadata":solar power)`
- `("All Metadata":wind power)`

The code used to reduced the "Abstracts" to wordclouds and frequenced terms, and that the aforementioned libraries are to use these functionalities:

- **Convert the text to lower case** - Code:

  ```
  abstract <- tm_map(abstract, content_transformer(toLower))
  ```

- **Remove numbers** - Code:

  ```
  abstract <- tm_map(abstract, content_transformer(removeNumbers))
  ```

- **Remove english common stopwords** - Code:

  ```
  abstract <- tm_map(abstract, removeWords, stopwords("english"))
  ```

- **Remove your own stop word** - Code:

  ```
  abstract <- tm_map(abstract, removeWords, listOfCommonWords)
  ```

- **Remove punctuations** - Code:

  ```
  abstract <- tm_map(abstract, removePunctuation)
  ```

- **Eliminate extra white spaces** - Code:

  ```
  abstract <- tm_map(abstract, stripWhitespace)
  ```

- **Text stemming** - Code:

  ```
  abstract <- tm_map(abstract, stemDocument)
  ```

The list of words used to "Remove your own stop words":

"capable", "configure", "main", "type", "oper", "condit", "includ", "point", "idp", "plural", "arrang", "develop", "connect", "compris", "provid", "maximum", "low", "high", "method", "study", "output", "solar", "wind", "energy", "power", "current", "load", "based", "paper", "proposed", "a", "about", "above", "after", "again", "against", "ain", "all", "am", "an", "and", "any", "are", "aren", "aren't", "as", "at", "be", "because", "been", "before", "being", "below", "between", "both", "but", "by", "can", "couldn", "couldn't", "d", "did", "didn", "didn't", "do", "does", "doesn", "doesn't", "doing", "don", "don't", "down", "during", "each", "few", "for", "from", "further", "had", "hadn", "hadn't", "has", "hasn", "hasn't", "have", "haven", "haven't", "having", "he", "her", "here", "hers", "herself", "him", "himself", "his", "how", "i", "if", "in", "into", "is", "isn", "isn't", "it", "it's", "its", "itself", "just", "ll", "m", "ma", "me", "mightn", "mightn't", "more", "most", "mustn", "mustn't", "my", "myself", "needn", "needn't", "no", "nor", "not", "now", "o", "of", "off", "on", "once", "only", "or", "other", "our", "ours", "ourselves",

"out", "over", "own", "re", "s", "same", "shan", "shan't", "she", "she's", "should", "should've", "shouldn", "shouldn't", "so", "some", "such", "t", "than", "that", "that'll", "the", "their", "theirs", "them", "themselves", "then", "there", "these", "they", "this", "those", "through", "to", "too", "under", "until", "up", "ve", "very", "was", "wasn", "wasn't", "we", "were", "weren", "weren't", "what", "when", "where", "which", "while", "who", "whom", "why", "will", "with", "won", "won't", "wouldn", "wouldn't", "y", "you", "you'd", "you'll", "you're", "you've", "your", "yours", "yourself", "yourselves", "could", "he'd", "he'll", "he's", "here's", "how's", "i'd", "i'll", "i'm", "i've", "let's", "ought", "she'd", "she'll", "that's", "there's", "they'd", "they'll", "they're", "they've", "we'd", "we'll", "we're", "we've", "what's", "when's", "where's", "who's", "why's", "would", "able", "abst", "accordance", "according", "accordingly", "across", "act", "actually", "added", "adj", "affected", "affecting", "affects", "afterwards", "ah", "almost", "alone", "along", "already", "also", "although", "always", "among", "amongst", "announce", "another", "anybody", "anyhow", "anymore", "anyone", "anything", "anyway", "anyways", "anywhere", "apparently", "approximately", "arent", "arise", "around", "aside", "ask", "asking", "auth", "available", "away", "awfully", "b", "back", "became", "become", "becomes", "becoming", "beforehand", "begin", "beginning", "beginnings", "begins", "behind", "believe", "beside", "besides", "beyond", "biol", "brief", "briefly", "c", "ca", "came", "cannot", "can't", "cause", "causes", "certain", "certainly", "co", "com", "come", "comes", "contain", "containing", "contains", "couldnt", "date", "different", "done", "downwards", "due", "e", "ed", "edu", "effect", "eg", "eight", "eighty", "either", "else", "elsewhere", "end", "ending", "enough", "especially", "et", "etc", "even", "ever", "every", "everybody", "everyone", "everything", "everywhere", "ex", "except", "f", "far", "ff", "fifth", "first", "five", "fix", "followed", "following", "follows", "former", "formerly", "forth", "found", "four", "furthermore", "g", "gave", "get", "gets", "getting", "give", "given", "gives", "giving", "go", "goes", "gone", "got", "gotten", "h", "happens", "hardly", "hed", "hence", "hereafter", "hereby", "herein", "heres", "hereupon", "hes", "hi", "hid", "hither", "home", "howbeit", "however", "hundred", "id", "ie", "im", "immediate", "immediately", "importance", "important", "inc", "indeed", "index", "information", "instead", "invention", "inward", "itd", "it'll", "j", "k", "keep", "keeps", "kept", "kg", "km", "know", "known", "knows", "l", "largely" , "last", "lately", "later", "latter", "latterly", "least", "less", "lest", "let", "lets", "like", "liked", "likely", "line", "little", "ll", "look", "looking", "looks", "ltd", "made", "mainly", "make", "makes", "many", "may", "maybe", "mean", "means", "meantime", "meanwhile", "merely", "mg", "might", "million", "miss", "ml", "moreover", "mostly", "mr", "mrs", "much", "mug", "must", "n", "na", "name", "namely", "nay", "nd", "near", "nearly", "necessarily", "necessary", "need", "needs", "neither", "never", "nevertheless", "new", "next", "nine", "ninety", "nobody", "non", "none", "nonetheless", "noone", "normally", "nos", "noted", "nothing", "nowhere", "obtain", "obtained", "obviously", "often", "oh", "ok", "okay", "old", "omitted", "one", "ones", "onto", "ord",

"others", "otherwise", "outside", "overall", "owing", "p", "page", "pages", "part", "particular", "particularly", "past", "per", "perhaps", "placed", "please", "plus", "poorly", "possible", "possibly", "potentially", "pp", "predominantly", "present", "previously", "primarily", "probably", "promptly", "proud", "provides", "put", "q", "que", "quickly", "quite", "qv", "r", "ran", "rather", "rd", "readily", "really", "recent", "recently", "ref", "refs", "regarding", "regardless", "regards", "related", "relatively", "research", "respectively", "resulted", "resulting", "results", "right", "run", "said", "saw", "say", "saying", "says", "sec", "section", "see", "seeing", "seem", "seemed", "seeming", "seems", "seen", "self", "selves", "sent", "seven", "several", "shall", "shed", "shes", "show", "showed", "shown", "showns", "shows", "significant", "significantly", "similar", "similarly", "since", "six", "slightly", "somebody", "somehow", "someone", "somethan", "something", "sometime", "sometimes", "somewhat", "somewhere", "soon", "sorry", "specifically", "specified", "specify", "specifying", "still", "stop", "strongly", "sub", "substantially", "successfully", "sufficiently", "suggest", "sup", "sure", "take", "taken", "taking", "tell", "tends", "th", "thank", "thanks", "thanx", "thats", "that've", "thence", "thereafter", "thereby", "thered", "therefore", "therein", "there'll", "thereof", "therere", "theres", "thereto", "thereupon", "there've", "theyd", "theyre", "think", "thou", "though", "thoughh", "thousand", "throug", "throughout", "thru", "thus", "til", "tip", "together", "took", "toward", "towards", "tried", "tries", "truly", "try", "trying", "ts", "twice", "two", "u", "un", "unfortunately", "unless", "unlike", "unlikely", "unto", "upon", "ups", "us", "use", "used", "useful", "usefully", "usefulness", "uses", "using", "usually", "v", "value", "various", "'ve", "via", "viz", "vol", "vols", "vs", "w", "want", "wants", "wasnt", "way", "wed", "welcome", "went", "werent", "whatever", "what'll", "whats", "whence", "whenever", "whereafter", "whereas", "whereby", "wherein", "wheres", "whereupon", "wherever", "whether", "whim", "whither", "whod", "whoever", "whole", "who'll", "whomever", "whos", "whose", "widely", "willing", "wish", "within", "without", "wont", "words", "world", "wouldnt", "www", "x", "yes", "yet", "youd", "youre", "z", "zero", "a's", "ain't", "allow", "allows", "apart", "appear", "appreciate", "appropriate", "associated", "best", "better", "c'mon", "c's", "cant", "changes", "clearly", "concerning", "consequently", "consider", "considering", "corresponding", "course", "currently", "definitely", "described", "despite", "entirely", "exactly", "example", "going", "greetings", "hello", "help", "hopefully", "ignored", "inasmuch", "indicate", "indicated", "indicates", "inner", "insofar", "it'd", "keep", "keeps", "novel", "presumably", "reasonably", "second", "secondly", "sensible", "serious", "seriously", "sure", "t's", "third", "thorough", "thoroughly", "three", "well", "wonder", "a", "about", "above", "above", "across", "after", "afterwards", "again", "against", "all", "almost", "alone", "along", "already", "also", "although", "always", "am", "among", "amongst", "amoungst", "amount", "an", "and", "another", "any", "anyhow", "anyone", "anything", "anyway", "anywhere", "are", "around", "as", "at", "back", "be", "became", "because", "become", "becomes", "becoming", "been", "before",

"beforehand", "behind", "being", "below", "beside", "besides", "between", "beyond", "bill", "both", "bottom", "but", "by", "call", "can", "cannot", "cant", "co", "con", "could", "couldnt", "cry", "de", "describe", "detail", "do", "done", "down", "due", "during", "each", "eg", "eight", "either", "eleven", "else", "elsewhere", "empty", "enough", "etc", "even", "ever", "every", "everyone", "everything", "everywhere", "except", "few", "fifteen", "fify", "fill", "find", "fire", "first", "five", "for", "former", "formerly", "forty", "found", "four", "from", "front", "full", "further", "get", "give", "go", "had", "has", "hasnt", "have", "he", "hence", "her", "here", "hereafter", "hereby", "herein", "hereupon", "hers", "herself", "him", "himself", "his", "how", "however", "hundred", "ie", "if", "in", "inc", "indeed", "interest", "into", "is", "it", "its", "itself", "keep", "last", "latter", "latterly", "least", "less", "ltd", "made", "many", "may", "me", "meanwhile", "might", "mill", "mine", "more", "moreover", "most", "mostly", "move", "much", "must", "my", "myself", "name", "namely", "neither", "never", "nevertheless", "next", "nine", "no", "nobody", "none", "noone", "nor", "not", "nothing", "now", "nowhere", "of", "off", "often", "on", "once", "one", "only", "onto", "or", "other", "others", "otherwise", "our", "ours", "ourselves", "out", "over", "own", "part", "per", "perhaps", "please", "put", "rather", "re", "same", "see", "seem", "seemed", "seeming", "seems", "serious", "several", "she", "should", "show", "side", "since", "sincere", "six", "sixty", "so", "some", "somehow", "someone", "something", "sometime", "sometimes", "somewhere", "still", "such", "system", "take", "ten", "than", "that", "the", "their", "them", "themselves", "then", "thence", "there", "thereafter", "thereby", "therefore", "therein", "thereupon", "these", "they", "thickv", "thin", "third", "this", "those", "though", "three", "through", "throughout", "thru", "thus", "to", "together", "too", "top", "toward", "towards", "twelve", "twenty", "two", "un", "under", "until", "up", "upon", "us", "very", "via", "was", "we", "well", "were", "what", "whatever", "when", "whence", "whenever", "where", "whereafter", "whereas", "whereby", "wherein", "whereupon", "wherever", "whether", "which", "while", "whither", "who", "whoever", "whole", "whom", "whose", "why", "will", "with", "within", "without", "would", "yet", "you", "your", "yours", "yourself", "yourselves", "the", "a", "b", "c", "d", "e", "f", "g", "h", "i", "j", "k", "l", "m", "n", "o", "p", "q", "r", "s", "t", "u", "v", "w", "x", "y", "z", "A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L", "M", "N", "O", "P", "Q", "R", "S", "T", "U", "V", "W", "X", "Y", "Z", "co", "op", "research-articl", "pagecount", "cit", "ibid", "les", "le", "au", "que", "est", "pas", "vol", "el", "los", "pp", "u201d", "well-b", "http", "volumtype", "par", "0o", "0s", "3a", "3b", "3d", "6b", "6o", "a1", "a2", "a3", "a4", "ab", "ac", "ad", "ae", "af", "ag", "aj", "al", "an", "ao", "ap", "ar", "av", "aw", "ax", "ay", "az", "b1", "b2", "b3", "ba", "bc", "bd", "be", "bi", "bj", "bk", "bl", "bn", "bp", "br", "bs", "bt", "bu", "bx", "c1", "c2", "c3", "cc", "cd", "ce", "cf", "cg", "ch", "ci", "cj", "cl", "cm", "cn", "cp", "cq", "cr", "cs", "ct", "cu", "cv", "cx", "cy", "cz", "d2", "da", "dc", "dd", "de", "df", "di", "dj", "dk", "dl", "do", "dp", "dr", "ds", "dt", "du", "dx", "dy", "e2", "e3", "ea", "ec",

"ed", "ee", "ef", "ei", "ej", "el", "em", "en", "eo", "ep", "eq", "er", "es", "et", "eu", "ev", "ex", "ey", "f2", "fa", "fc", "ff", "fi", "fj", "fl", "fn", "fo", "fr", "fs", "ft", "fu", "fy", "ga", "ge", "gi", "gj", "gl", "go", "gr", "gs", "gy", "h2", "h3", "hh", "hi", "hj", "ho", "hr", "hs", "hu", "hy", "i", "i2", "i3", "i4", "i6", "i7", "i8", "ia", "ib", "ic", "ie", "ig", "ih", "ii", "ij", "il", "in", "io", "ip", "iq", "ir", "iv", "ix", "iy", "iz", "jj", "jr", "js", "jt", "ju", "ke", "kg", "kj", "km", "ko", "l2", "la", "lb", "lc", "lf", "lj", "ln", "lo", "lr", "ls", "lt", "m2", "ml", "mn", "mo", "ms", "mt", "mu", "n2", "nc", "nd", "ne", "ng", "ni", "nj", "nl", "nn", "nr", "ns", "nt", "ny", "oa", "ob", "oc", "od", "of", "og", "oi", "oj", "ol", "om", "on", "oo", "oq", "or", "os", "ot", "ou", "ow", "ox", "oz", "p1", "p2", "p3", "pc", "pd", "pe", "pf", "ph", "pi", "pj", "pk", "pl", "pm", "pn", "po", "pq", "pr", "ps", "pt", "pu", "py", "qj", "qu", "r2", "ra", "rc", "rd", "rf", "rh", "ri", "rj", "rl", "rm", "rn", "ro", "rq", "rr", "rs", "rt", "ru", "rv", "ry", "s2", "sa", "sc", "sd", "se", "sf", "si", "sj", "sl", "sm", "sn", "sp", "sq", "sr", "ss", "st", "sy", "sz", "t1", "t2", "t3", "tb", "tc", "td", "te", "tf", "th", "ti", "tj", "tl", "tm", "tn", "tp", "tq", "tr", "ts", "tt", "tv", "tx", "ue", "ui", "uj", "uk", "um", "un", "uo", "ur", "ut", "va", "wa", "vd", "wi", "vj", "vo", "wo", "vq", "vt", "vu", "x1", "x2", "x3", "xf", "xi", "xj", "xk", "xl", "xn", "xo", "xs", "xt", "xv", "xx", "y2", "yj", "yl", "yr", "ys", "yt", "zi", "zz"