SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2022
ORGANIZATION AND MANAGEMENT SERIES NO. 162

THE ANALYSIS OF POLISH PATENT APPLICATIONS IN THE SOLAR ENERGY TECHNOLOGY WITH THE USE OF TEXT MINING METHODOLOGY

Karolina BĘBEN[1], Marzena NOWAKOWSKA[2*]

[1] Kielce University of Technology, Faculty of Management and Computer Modelling; k.beben@tu.kielce.pl,
ORCID: 0000-0002-2305-6723

[2] Kielce University of Technology, Faculty of Management and Computer Modelling; m.pajecki@tu.kielce.pl,
ORCID: 0000-0002-6934-523X

* Correspondence author

Purpose: Knowledge management belongs to the most important elements of organisational management, including manufacturing enterprises. Patent information plays an increasingly important role in this area. Identification of the main directions of invention activity may inspire new product and process ideas, and can help to improve existing solutions. The above is particularly important in the energy sector, which is currently struggling with increasing problems. In this context, solar energy is the subject of interest to inventive communities. The paper discusses patent applications related to solar energy, taking up the task of discovering the main tendencies of technological solutions in this area.

Design/methodology/approach: In the work, a pilot study of the research aimed to indicate the directions of technological development in the field in Poland was undertaken. Shortened descriptions of selected patent documents from the Polish Patent Office (PPO) were the subject of the investigation. The descriptions were reduced to the form of a vector space model by using text mining tools. The exploration of such prepared data was done applying unsupervised text mining techniques. Hierarchical cluster analysis enabled the identification of groups of similar inventions. An algorithm to detect outliers within individual patent groups was also developed and applied.

Findings: Five patent clusters were identified covering the following thematic areas: PV panel designs, PV panel component designs, the improvement of solar-heat conversion device performance, and solar collector designs. Six patent applications stood out thematically in four of the five clusters.

Research limitations/implications: The research is limited to a selected number of patent documents form PPO. However, the presented method and research area are promising. It is planned to extend the analyses to a larger set of patent documents and solve the problem related to the language uniformity of patent applications along with merging data from various sources. In this aspect, a full patent description will be consider as well.

Originality/value: In relation to solar energy issues, main patent areas and patent outliers that may be indicators of special interests of inventors were identified. In relation to methodology issues, new solutions within consecutive research steps were proposed.

http://dx.doi.org/10.29119/1641-3466.2022.162.1

Keywords: solar energy patent analysis, patent documents processing, patent clustering, outliers detection, Polish Patent Office.

Category of the paper: Research paper.

1. Introduction

Knowledge management is one of the most important elements of organisational management, including manufacturing enterprises. Patent information plays an increasingly important role in this area. Patent documents reveal the essence of inventions in a clear, unambiguous and understandable way for a specialist. Data contained in patent databases can be employed for strategic planning purposes and the proper use of these data can contribute to the increase of a company's market value and to achieve competitive advantage of a company. Identification of the main directions of inventive activity in a given area is one of elements supporting the creation and disposal of substantive (intellectual) competences and practical skills, which are intangible assets of the enterprise. Such knowledge may inspire new product and process ideas, and can help to improve existing solutions.

The above is particularly important in the energy sector, which is currently struggling with increasing problems. Climate changes and the devastation of the natural environment caused by the production of dirty energy, on the one hand, and the consequences of geopolitical conditions, also in connection with the war in Ukraine, on the other hand, justify the necessity to support and develop technologies concerning renewable energy sources (RES). One of the most popular RES that are of interest to inventive communities is solar energy. The paper discusses patent applications related to solar energy, taking up the task of identifying the main directions of technological solutions in this area.

A patent is the right to the exclusive use of a technical solution (invention) for a specified period of time, for profit (Journal of Laws, 2000). It is valid only in those countries where patent offices have granted protection for inventions. Patent protection can be obtained in a national, regional or international mode. Such a protection is a signal of an effective inventive activity in a given field. Information on patent documents is made available through free databases, created mainly by national patent offices (for example, Polish Patent Office), international or regional intellectual property organizations (for example, WIPO or ESPACENET). There are also commercial databases containing additional information about inventions, such as legal data (like in PATSTAT). However, there is no repository that includes all patent documents published at any time.

The structure of a patent document depends on the country (or the organization) specificity, but it always contains certain standardized information, which, in addition to the basic data identifying the application, includes technical information, such as International Patent

Classification (IPC) symbols (WIPO, 2015). The IPC symbols are assigned by professional patent attorneys and allow inventions to be classified according to technology domains. Several symbols can be used for one patent document, but it is assumed that the symbol that represents the invention most adequately appears first in the symbol list (WIPO, 2022). In the study, the International Patent Classification was used in order to investigate patent applications within the solar energy technology. The IPC symbols considered in the paper are the result of literature research and analysis of WIPO (2021) indications. The symbols cover the following list: E04D 13/18, F03G 6/06, F24J 2/, F24S, G05F 1/67, H01L 25/00, H01L 31/00, H01L 31/04, H01L 31/05, H01L 31/18, H02J 7/35, H02N 6/00, H02S. The description of the codes is presented in Appendix A.

There were 24,685 patent applications for the solar energy technology identified in the PATSTAT database, which were submitted to patent offices of countries located in Europe, between 01-01-2001 and 28-02-2021 (the data were acquired on 2022-04-24; fall 2021 edition). Among the applications, 16,958 have at least one inventor from a European country – such documents are hereinafter referred to as "European applications". Similarly, if any inventor is a citizen of a particular country, their patent document is assigned to that country (for example, German applications, Polish applications). The ranking of countries according to the number of patent applications of inventors from a given country submitted in the field of solar energy is shown in Figure 1 (first 15 places).

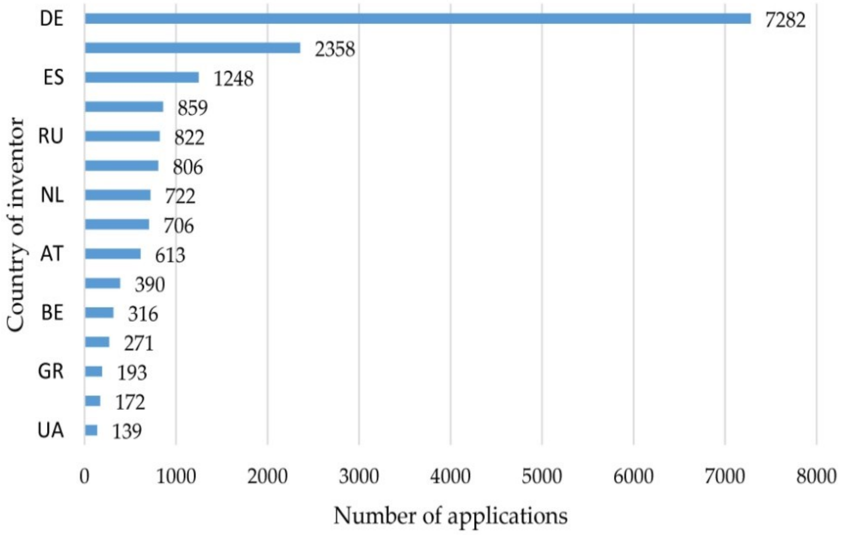

Figure 1. European patent applications in the solar energy technology submitted between 01-01-2001 and 28-02-2021 to national patent offices located in Europe, reported in PATSTAT.

Source: authors' own elaboration.

The presented summary should be considered as a picture of the inventive activity of Europeans, since some documents may be counted several times (for several countries), taking into account that several people from different European countries may be inventors assigned to the same application. However, the results presented in the figure confirm, to some extent, the outcome published in other papers, e.g. Binz et al. (2017), Breyer et al. (2013), Sampaio et al. (2018). Germany, France and Spain took the top three places, but the predominance of Germany is significant. The number of applications by German inventors is greater than the sum of applications by inventors from the next six countries in the ranking. The share of documents of Polish inventors in the total number of European inventors' applications amounted to 1.6% (390), yet Poland was ranked tenth out of total 46 countries.

The number of Polish applications against the number of European ones in given years is presented in Figure 2. The year 2020 is not discussed as the data from this year does not include all the applications. The highest activity of Europeans occurs between 2009 and 2011, but the share of Poles in the number of applications does not exceed 2% in this period. A decreasing tendency in the number of European applications has been observed since 2010, while Polish applications increased from 2% to almost 6%.

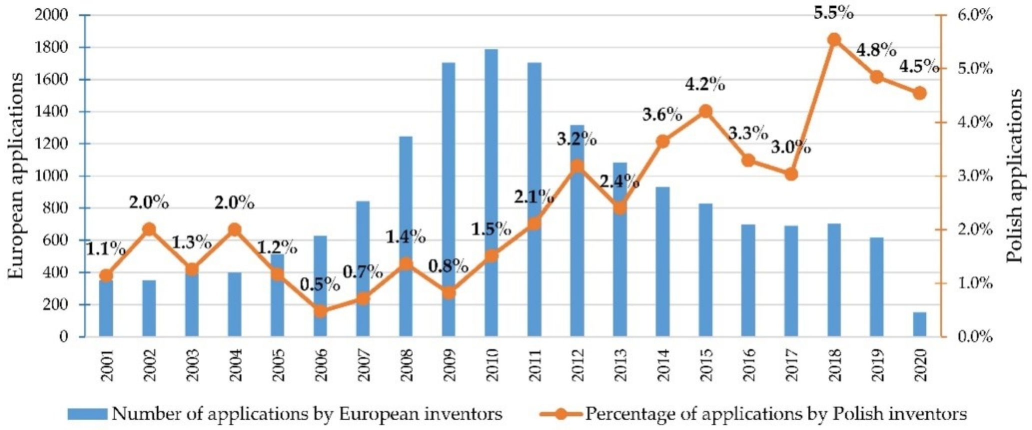

Figure 2. European patent applications in the solar energy technology submitted between 01-01-2001 and 28-02-2021 to national patent offices located in Europe, registered in PATSTAT.

Source: authors' own elaboration.

Data in the PATSTAT database do not contain the information necessary for the analyses in the study; in Polish inventors' applications submitted to PATSTAT, short descriptions of inventions are missing in more than one quarter (107 items) of the documents. Therefore the national database of the Polish Patent Office (PPO) was chosen as the data source for the pilot research; on the basis of the analysis of selected patent applications related to solar energy, the characterisation of Polish inventors activity in the area was conducted.

The article consists of six parts. After this introductory section, a literature review discussing patents related to solar energetics is presented. The third section describes the research process, in particular the method of isolating thematic areas of the solar energy technology explored by Polish inventors in Polish conditions, and the method of identifying outliers. Section four shows the obtained results, which are: the patent clusters and outliers detected within them. The fifth section profiles and discusses the clusters and comments on the outlier cases. Finally, a summary and conclusions are presented.

2. Literature review

The literature on patents and solar energy is extensive and, especially in recent years, has grown exponentially. Since the information in all patent databases has a similar structure, research analyses based on it cover two approaches. One can be defined as quantitative, usually supported by frequency analysis as regards selected patent attributes. It is usually devoted to the identification of quantitative trends in the invention landscape. The other approach can be described as qualitative, typically with the use of text mining methods to classify and identify patterns in patents. The two approaches are often combined in the research on patented applications. Green energy, and solar energy in particular, became a common topic in the field of patent research more than a decade ago. A brief literature review on patent analysis for solar energy is presented below. Due to the subject matter of the analyses conducted in the study and the used methodology, the review focuses on the works in which text mining methods in examining patent application descriptions were used.

The work by Liu et al. (2011) examined the photovoltaic technology development from the perspective of patent growth trajectories. The patent data were taken from the database operated by the United States Patent and Trademark Office (USPTO). The research focused mainly on materials used in photovoltaics. Keyword co-occurrence analysis was applied to classify patents into five groups. Three of them, labelled as: Emerging PV, Group III–V, and Silicon, were described as constantly and strongly growing ones. The remaining two, called CdTe and CIS/CIGS were defined as being at the mature stage. Yoon and Kim (2012) proposed a semantic approach method to detect outliers among granted patents, thus indicating potential technology opportunities. The approach was illustrated using organic photovoltaic cells (OPV) related patents retrieved from USPTO database. Nine inventions were indicated as having a strong possibility of being unusual. The following three ones were discussed: (1) a method of fabricating charge-transport structures, (2) an invention concerning "solar networks and power grids", (3) the use of a "poly cross linked phthalocyanine compound". It was outlined that a final review by an expert is necessary to diagnose the selected patents as delivering new technology opportunities. Chi and Ying (2012) discussed the technology evolution of building integrated photovoltaics (BIPV). In the analysis of patent descriptions, they transformed patent documents into structured data and used patent matrix map analysis to develop R&D strategy of related industries. The technology appeared to be growing with a long-term development potential. The patent text analysis enabled indicating the technology key points in the future as follows: durability, ease of maintenance, customization of PV modules and the BIPV security. Lizin et al. (2013) made a quantitative overview of a global patent activity on the organic photovoltaic (OPV) solar cell types, their substrates and encapsulation materials on the basis of patents retrieved from the FamPat database. The authors found OPV still residing in the fluid technology development phase, following an exponential growth path. Two main technological solutions were identified: (1) a group of semiconductors suitable for energy conversion or control (IPC: H01L-031), and (2) glass, then paper and textiles for photovoltaic cells substrates. Venugopalan and Rai (2015) proposed a hierarchical technique based on natural language processing for classifying patents and for identifying linkage between them. The approach was exemplified on USPTO database issued patents and patent application on PV balance of system technology categories. Topic modelling was used to identify inventions technology areas. Binz et al. (2017) addressed the question about spatial dynamics in new clean technology fields in relation to existing industry lifecycle models and globalization as a creator of new lifecycle patterns. They investigated solar energy technology patents from Thomson Innovation and Derwent World Patent Index global patent databases. The authors discovered that the largest number of patents was related to core technologies and associated with the production of solar cells and modules, while patents related to extraction technology (silicon, ingots and water sawing) made up a small proportion of the knowledge base. However, the authors acknowledged that recently emerging thin-film and organic PV technologies were not the focus of their research. De Paulo et al. (2018) analysed patent data on solar photovoltaics from the Derwent Innovation (DI) database. They filtered out co-ownership patents and use social network analysis to find PV technology development networks. The following patent areas were found as the most frequent ones: "Devices adapted for the conversion of radiation energy into electrical energy", "Assemblies of a plurality of solar cells", and "Silicon; single-crystal growth". Sampaio et al. (2018) used the same data source (DI database) in the discovery of the technological development of photovoltaic cells, applying a frequency analysis. The following areas were indicated as the ones on which predominant patents concentrate: semiconductors for the conversion of solar radiation into electric energy, generators for the direct conversion of light energy into electric energy, and adaptation of solar panels for roof structures. Polymer-based photovoltaic cell technologies, carbon nanostructures, III-V compounds, cadmium telluride and amorphous silicon cells had a predominance of deposited patents as regards the technological issue. Due to the rapid development in the field of photovoltaic technology, Li et al. (2019) selected perovskite solar cell technology as the case study in their work. They used Derwent Innovations database and Twitter data. The evolutionary path of the technology was presented in the form of a map developed on the results of patent topic clustering and relying on the experts' knowledge in the field. Trappey et al. (2019) analysed WoS scientific literature and DI database patent documents to identify key issues and the patent evolution of the solar power technology. The authors combined multiple, unsupervised machine learning algorithms in their research. The conclusion was that more novel technologies describing integration of renewable energy generation systems and simulation of grid-connected energy storage systems could be found in the literature, while technologies describing solar hydropower storage system with subsystems for indirect solar connection technology were presented in patents.

In the vast majority of cases, solar patent topic investigations focus on photovoltaic technology that has developed considerably fast in the last thirty-plus years. In this aspect, cluster analysis, mapping technology and topic modelling were engaged either in overall PV issue or in its selected subareas. The application of text mining technology to the content analysis of patents that deal not only with the photovoltaic aspect, but take into account the full patent landscape of solar energy technology, is presented less frequently in the scientific literature. To the best of the authors' knowledge, there are no works on this type investigation concerning the invention activity in a selected country, in particular with regard to the EU area. The study proposes a research process in which an enriched approach to text mining of patent data was developed. This is a pilot study of the research aimed to indicate the areas of technological development in the field of solar energy in Poland.

3. Materials and methods

The scheme of the research process carried out in the study concerning Polish inventions in the solar energy technology is shown in Figure 3. The proposed approach consists of several steps and includes: acquiring and preparing patent data for the analysis, preprocessing textual data, generating clustering models and selecting the best one, and finally, detecting outliers in the best clustering model. The results obtained in any step constitute data for the next step. The left part of the figure refers to individual steps of the research process, whereas the right part indicates the results of the steps.

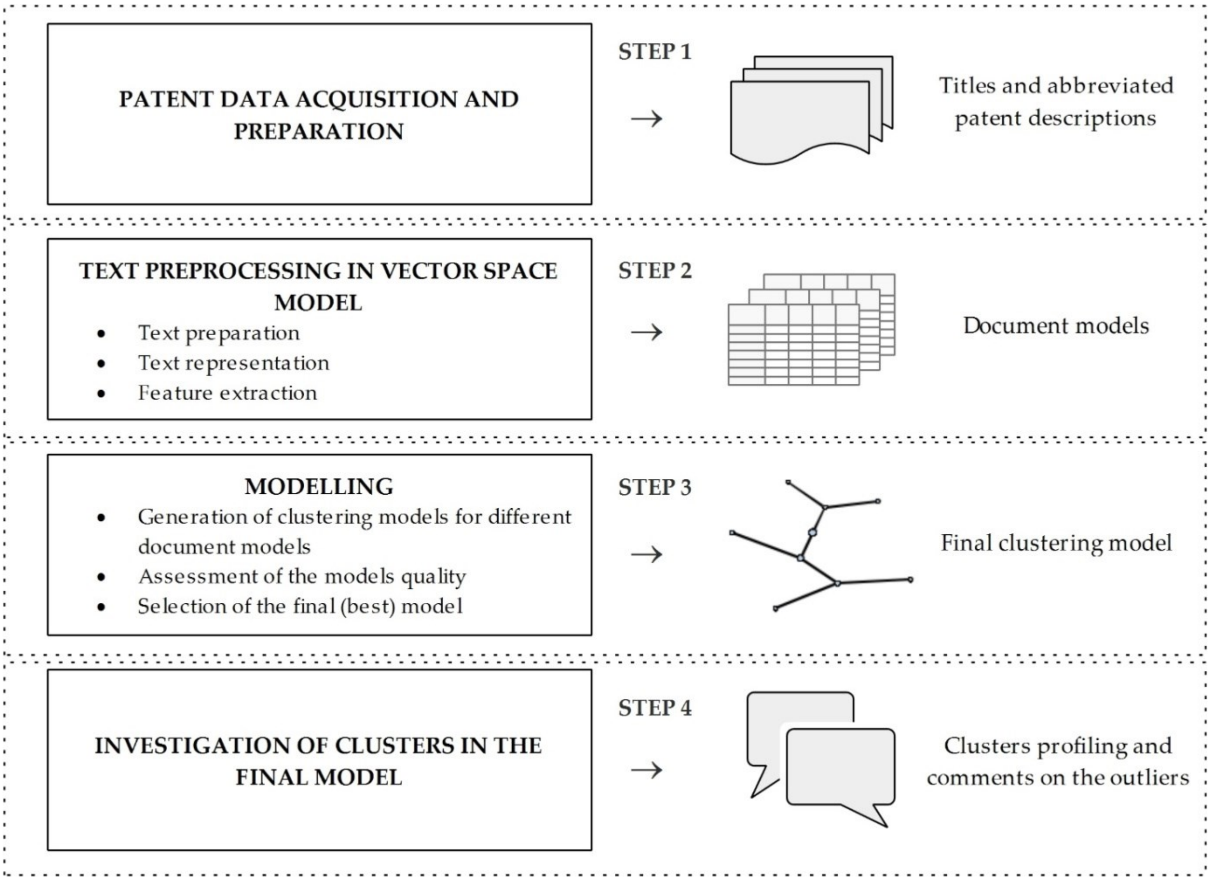
PATENT DATA ACQUISITION AND PREPARATION
STEP 1
Titles and abbreviated patent descriptions
TEXT PREPROCESSING IN VECTOR SPACE MODEL
Text preparation
Text representation
Feature extraction
STEP 2
Document models
MODELLING
Generation of clustering models for different document models
Assessment of the models quality
Selection of the final (best) model
STEP 3
Final clustering model
INVESTIGATION OF CLUSTERS IN THE FINAL MODEL
STEP 4
Clusters profiling and comments on the outliers

Figure 3. The research process for the identification of Polish inventiveness areas in the solar energy technology.

Source: authors' own elaboration.

The following subsections describe the individual steps. The used methodologies and solutions are explained and relevant references are given.

3.1. Patent data acquisition and preparation

The Polish patent document structure consists of the following elements:

- a title page containing bibliographical data that identifies the document and a short description of the invention,
- other pages containing a full invention description, patent claims and drawings, if any.

In order to acquire the data, the following bibliographic attributes were used in the definition of search criteria: the inventor's country, the patent application date, International Patent Classification symbols (WIPO, 2021), and keywords. Roots of Polish words related to solar energy have been introduced in the list of keywords, taking into account the fact that the used IPC symbols refer to an area broader than just the technology field under investigation. Table 1 presents the conditions whose logical conjunction was used to search for documents in the PPO repository. The search and retrieval of the resources was performed on 2022-02-17.

Table 1.
Criteria for searching patent documents in the solar energy technology

Attribute	Value
Inventor's country	PL
Application date	≥ 2001-01-01
IPC symbols	E04D 13/18, F03G 6/06, F24J 2/, F24S, G05F 1/67, H01L 25/00, H01L 31/00, H01L 31/04, H01L 31/05, H01L 31/18, H02J 7/35, H02N 6/00, H02S
Words for title and abbreviated patent description search	fotow, solar, słon, słoń, świat, świetl[1]

[1] full English equivalents: photovoltaic, solar, sunny, sun, light, luminous/illuminated.

Source: authors' own elaboration.

A dataset containing 7 fields and 368 records was obtained, where each record represents a separate patent application. After transforming selected fields, the set of the following variables subjected to further analyses, was obtained:

- a title and abbreviated description of patent application,
- the decision on granting an exclusive right,
- the patent application date,
- the list of IPC symbols assigned to the document.

The primary research process involved the text variables: the title and the abbreviated description. The remaining variables were used in the final step of the analysis of results.

3.2. Text preprocessing in vector space model

In the study, the vector model was chosen for processing text data, in which text documents became the source of "a bag of words" extracted from these documents. Defining "a bag of words" includes lexical processing and the reduction of the word list extracted from patent documents. Due to the fact that the Polish language is inflectional, lemmatization was carried out in terms of lexical processing. Next, nouns and verbs were selected from the obtained parts of speech, a stop-list, and a list of synonyms were used. As a result, a set of words, called terms, was created, on the basis of which the representation of documents in the form of a term – document matrix (TDM, TD matrix) was defined (Kadhim et al., 2014).

TDM is an array in which row headings identify individual patent documents, that is observations, while column headings identify terms, or variables in the multidimensional space, that characterize the observations. The elements of the matrix provide information about the relationship between the term and the document. To specify these elements, weighting functions that transform information about the frequency of terms occurrence are used (Lan et al., 2009). The choice of the weighting function determines the final form of the TD matrix representation. Therefore, the creation of the representation may result in several different representations of the same document collection, depending on a weighting function.

Regardless of the representation, the TD matrix is sparse and has a large dimension, thus implying the need to reduce this dimensionality. The dimension reduction of the TD matrix involves identifying a minimum number of features that best describe the dataset variability. The latent semantic indexing method, which conducts a decomposition of the TD matrix according to its singular values, was applied (Albright, 2004). The matrix decomposition is used to introduce so-called pseudo-terms, which characterize documents and whose number is smaller than the number of original terms. The key task is to determine the degree of reduction in the number of terms. This can be established from a scree plot of the singular values of the TD matrix, ordered in a descending manner. The point on the plot indicating a slowdown in the dynamics of the singular values change defines the degree of reduction. It was proposed to identify this point considering the change in the angle between the horizontal axis and the straight lines connecting the first point of the scree plot to its subsequent points.

The preprocessing stage of textual data is predominantly a decision-making process and may, to some extent, require human involvement. This is particularly visible when thesauri (stop-list, synonyms) are built or when the degree of TDM dimension reduction is determined. The end result is a document model in the form of a reduced TD matrix, where column headings are pseudo-terms and row headings represent documents.

In the study, three different reduced TD matrices were examined – the ones obtained from three different TDM representations in which the matrix elements were defined by the following weighting functions (Bęben, 2020): binary (BIN), log term frequency – inverse document frequency product (LOG-IDF), and term frequency (TF).

3.3. Modelling

In the process of identifying thematic areas of the solar energy technology explored by Polish inventors, various data-mining tools were used:

- Hopkins test (Hopkins, Skellam, 1945; Banerjee, Dave, 2004) to check for a clustering tendency in a collection of patent documents,
- cluster analysis to extract thematic groups of inventions,
- a global measure for evaluating the quality of a clustering model to identify the best model relative to other ones created by the same clustering algorithm but using different representations of the TD matrix – the authors' proposal.

A hierarchical grouping algorithm – the Ward method (Vijaya et al., 2019), enabling the identification of clusters of similar documents, was used in the research. In the Ward's algorithm, the number of clusters in the final clustering model is deter-mined automatically with the use of the criterion referring to the pseudo F and pseudo t^2 statistics. One can find the method description in the SAS Reference Help (SAS Institute Inc., 2016).

Since clustering was performed for BIN, LOG-IDF, and TF representations of the TD matrix, three clustering models were obtained. To select the best one, the *GQA* global quality assessment measure of the clustering model was proposed.

The *GQA* measure is determined from the following sub-measures: mean silhouette width *MS*, percentage of positive silhouette width values *PPS*, modified Dunn index *MDI*, and clustering effectiveness measure *CEM*.

- Mean silhouette width *MS*:

$$MS = \frac{1}{n} \cdot \sum_{j=1}^{l} \sum_{i=1}^{n_j} S_j(x_i) \tag{1}$$

where:

$S_j(x_i)$ – silhouette width for the (x_i) point (i-th observation) in a j-th cluster (Rousseeuw, 1987):

$$S_j(x_i) = \frac{b_j(i) - a_j(i)}{\max\{b_j(i), a_j(i)\}}$$

$$a_j(i) = \frac{1}{n_j - 1} \cdot \sum_{\substack{m=1 \\ m \neq i}}^{n_j} d\left(x_{(j)i}, x_{(j)m}\right) \tag{2}$$

$$b_j(i) = \min_{\substack{k=1\cdots l \\ k \neq i}} \frac{1}{n_k} \cdot \sum_{m=1}^{n_k} d(x_{(j)i}, x_{(k)m})$$

$j = 1 \dots l$,

$i = 1 \dots n_j$,

l – number of clusters in the model,

n_j – number of observations in cluster j,

n – total number of observations,

$d(x_{(j)i}, x_{(k)m})$ – Euclidean distance between the i-th observation in the j-th cluster and the m-th observation in the k-th cluster.

In equation (2), *a*(*i*) is the average distance of the ($\mathrm{x_i}$) point from all the other points in the same cluster and *b*(*i*) is the mean distance of that point from all the points in the closest cluster to its cluster.

- Percentage of positive silhouette width values *PPS*:

$$PPS = \frac{1}{n} \cdot \sum_{j=1}^{l} \sum_{i=1}^{n_j} P_j(x_i) \tag{3}$$

where:

$$P_j(x_i) = \begin{cases} 1 \text{ for } S_j(x_i) > 0 \\ 0 \text{ for } S_j(x_i) \leq 0 \end{cases} \tag{4}$$

- Modified Dunn index *MDI* (Halkidi, 2001):

$$MDI = \frac{\underset{1\le j\le l-1}{min}\{\underset{j+1\le k\le l}{min}\{AD(j,k)\}\}}{\underset{1\le j\le l}{max}\{AWCD(j)\}} \tag{5}$$

where:

$$AD(j,k) = \frac{\sum_{i=1}^{n_j}\sum_{m=1}^{n_k} d(x_{(j)i}, x_{(k)m})}{n_j \cdot n_k} \tag{6}$$

$$AWCD(j) = \frac{1}{\frac{(n_j^2 - n_j)}{2}} \cdot \sum_{i=1}^{n_j-1}\sum_{m=1}^{n_j} d(x_{(j)i}, x_{(j)m}) \tag{7}$$

In equations (6) and (7), *AD(j, k)* is the average distance between clusters *j* and *k*, and *AWCD(j)* is the average within cluster distance.

- Clustering effectiveness measure *CEM*:

$$CEM = \frac{ADB}{ADW} \tag{8}$$

where:

$$ADB = \frac{\sum_{j=1}^{l-1}\sum_{k=j+1}^{l}(\sum_{i=1}^{n_j}\sum_{m=1}^{n_k} d(x_{(j)i}, x_{(k)m}))}{\prod_{j=1}^{l} n_j} \tag{9}$$

$$ADW = \frac{\sum_{j=1}^{l} AWCD(j) \cdot n_j}{n} \tag{10}$$

In equations (8) and (9), *ADB* is the average distance between clusters (which is the cluster separation measure) and *ADW* is the average distance within clusters (which is the intra-cluster compactness measure) respectively.

All of the aforementioned sub-measures have an equal interpretation of monotonicity: the larger their values are, the better the clustering quality is. The *GQA* global measure of the quality of the clustering model is given by the following equation:

$$GQA = \frac{1}{2} \cdot (MSI_N + MDI_N) \cdot (CEM_N + PPS_N) \tag{11}$$

where *N* is *min-max* normalization of a measure to the interval <1, 2>, which makes it easier to illustrate *GQA* graphically and to interpret it.

The best clustering model is the one for which the *GQA* measure has the highest value. A graphical illustration of *GQA* is a radar plot, in which the normalized sub-measures are variables and their values for a given model form the vertices of the quadrilateral. The area of the quadrilateral reflects the quality of a clustering model under consideration.

3.4. Investigation of clusters in the final model

The final step of the research refers to the definition of the main subject areas in the set of the Polish patent documents, and then to the identification an outlier documents in each of these areas. The main topic area was assumed to be defined by patent documents belonging to the same cluster. Characterization of these areas was carried out using descriptive terms determined by the binomial probability of a given term belonging to a cluster (SAS Institute Inc., 2012) and the IPC symbols assigned to the documents creating the cluster.

According to the information given in the SAS software documentation (SAS Institute Inc., 2012), the following algorithm is used to select descriptive terms for clusters. For specified *m* descriptive terms for each cluster, the top $2 \cdot m$ most frequently occurring terms in each cluster are used to compute the descriptive terms. For each of the $2 \cdot m$ terms, a binomial probability for each cluster is computed. The probability of assigning a term to cluster *j* is *prob* = F(*k*|*N*, *p*), where:

- F is the binomial cumulative distribution function,
- *k* is the number of times when the documents containing the term appear in cluster *j*,
- *N* is the number of documents in cluster *j*,
- *p* is equal to (*sum-k*)/(*total-N*); *sum* is the total number of times when the documents containing the term appear in all the clusters, and *total* is the total number of documents.

The *m* descriptive terms are those with the highest binomial probabilities.

In order to identify documents that do not match the nature of the cluster, an outlier patent detection procedure was proposed. An outlier patent is represented by points in multidimensional space (after dimension reduction) whose location is significantly away from the main clustering tendency. The main tendency is determined by the centre of gravity, whose coordinates are the average of the coordinates of the documents for which the expert positioned the IPC code relating to solar energy (see Table 1) as the first in the list of codes classifying the invention. The Mahalanobis distance was used in diagnosing outlier documents (De Maesschalck, 2000). The following algorithm was proposed for the identification of outlier patent documents.

The outlier patent document search algorithm

- Calculate the gravity centre of points in the reduced multidimensional space.
- Sort the set of document distances from the centre of gravity in a non-decreasing order.
- Calculate the mean value of the first 95% of the distances in this set as the cut-off point for outliers, the point is denoted as *TO* (threshold for outlier).

- Compare the difference between two consecutive distances (points on the distance axis) with the *TO* value.
- If the difference between a given distance and the distance preceding it is greater than *TO*, then this distance (point on the axis) defines the critical distance value *CDV*. The *CDV* and all distances (points on the axis) greater than that identify documents in the cluster which can be treated as potential outliers.

4. Results

Preprocessing 368 pieces of text data on patents (titles and abbreviated patent descriptions) was performed in *SAS Enterprise Miner* using the tools of the *Text Miner* package. As a result, a set of 1746 terms was obtained, on the basis of which three forms of TD matrix were created: BIN, LOG-IDF and TF. The dynamics of changes in the ordered descendingly singular values (SV) of the TD matrix were examined so as to determine the degree of reduction in the number of terms. Figure 4 presents the scree plots of the singular values (large plot) and the change in the angle between the horizontal axis and the straight lines connecting the first point of the plot to its subsequent points in the scree plot (small one) for each representation of the TD matrix. The plots indicate the number of singular values equal to 5, 7 and 8, implying the reduction degree of TD matrices for the BIN, LOG-IDF and TF representations respectively. Their reduced TD matrices are denoted by the *BIN_5SV*, *LOG-IDF_7SV* and *TF_8SV* symbols further on.

For the analysed set of text patent data represented by full-size and reduced TD matrices, it was checked whether there is a tendency to form clusters. A Hopkins test was conducted in which the null hypothesis that the dataset follows a multivariate uniform distribution (i.e., there are no distinct clusters) is verified. It is assumed that if the test statistic is greater than (aprox.) 0.5 then the data tends to cluster (Hopkins, Skellam, 1954).

In the study, for the full-size (original) TD matrices, the test statistic values are greater than 0.83 and for the reduced TD matrices greater than 0.71, indicating a statistically significant tendency to cluster, which means that the data are significantly clusterable.

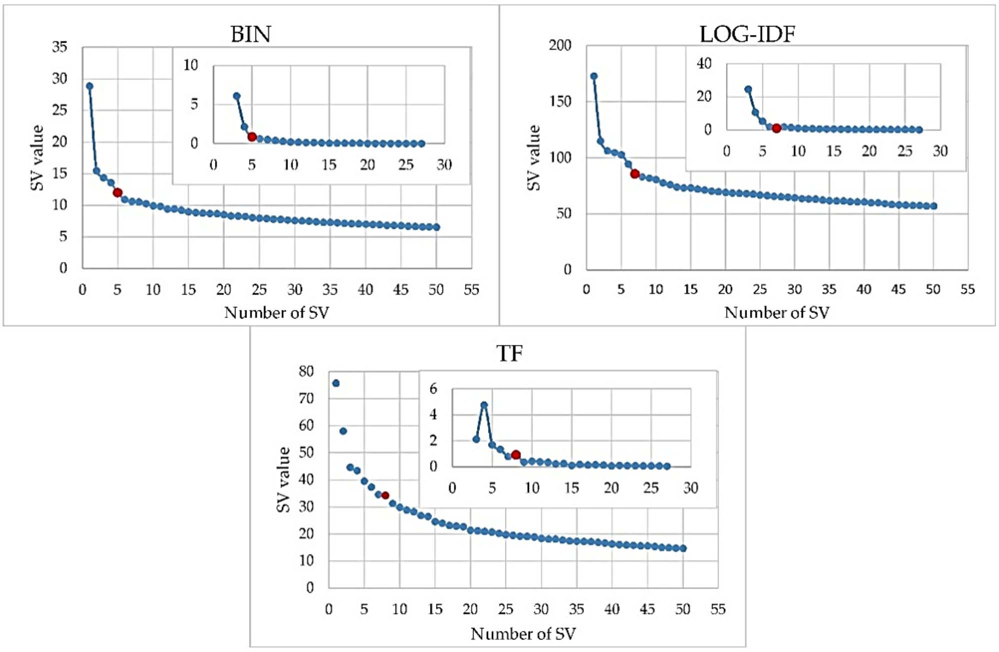

Figure 4. Dynamics of changes in singular values depending on the TD matrix representation.

Source: authors' own elaboration.

The *BIN_5SV*, *LOG-IDF_7SV*, and *TF_8SV* matrices were subjected to the cluster analysis resulting in three clustering models. Global measures of clustering quality (*GQA*) were calculated for them. Figure 5 shows a radar plot in which the variables are the normalized sub-measures of the model quality. The model created from the *BIN_5SV* TD matrix, which has the largest *GQA* value equal to 6.47, was selected for further analysis; the model is denoted as *BIN_5SV*.

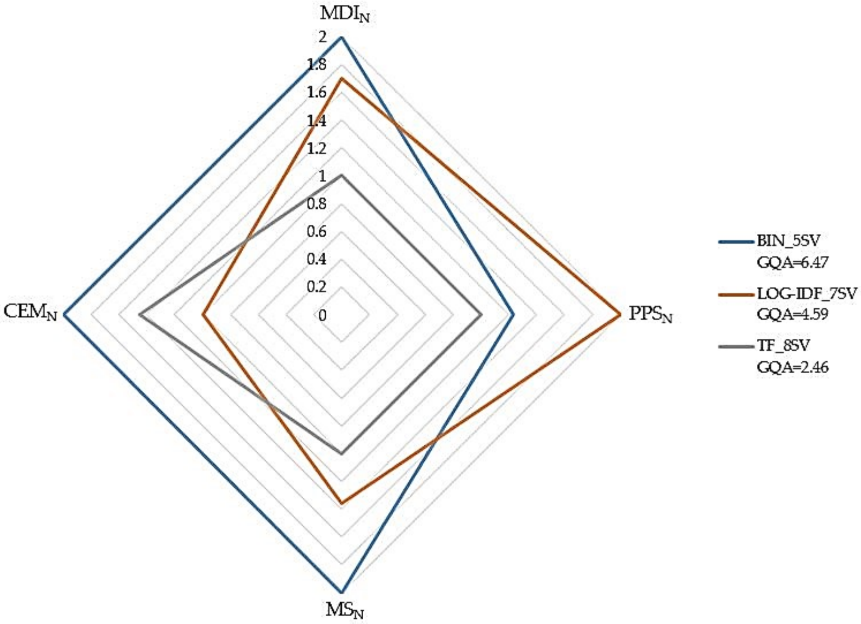

Figure 5. Graphical illustration of the *GQA* measure for different clustering models.

Source: authors' own elaboration.

The Ward's algorithm determined five clusters in the *BIN_5SV*-model, which are further identified by the symbols: *C*1-*C5*. They allowed the identification of leading thematic areas in the field under investigation. The areas are typically characterized by a set of most important terms extracted from documents assigned to individual clusters. In the study, the *m* parameter (determining the number of the terms) was set apriori as equal to 15. The terms are listed in Table 2, where certain statistics characterizing the clusters are also given. The "+" sign indicates that the term includes multiple grammatical forms of the word.

Table 2.
Description of the patent clusters BIN_5SV model

Cluster	Cluster size (# patents granted)	Extracted terms	Total number of IPC codes (per document)	Number of solar code / first-positioned IPC code
C1	94 (35)	+cell, +substrate, +time, +semiconductor, +dye, +generator, +voltage, +foil, +photoelectrode, +contact, +material, +layer, +current, +form, +cover	320 (3.4)	142 / 57 (44% / 18%)
C2	29 (13)	+photodiode, +bend, +shape, +centre, +hole, +meander, +element, +mechanism, +area, aluminium, +thickness, +frame, +lens, +layer, +part	102 (3.5)	46 /18 (45% / 18%)
C3	77 (27)	+battery, +valve, +mechanism, +installation, +hole, +water, +net, +clutch, +container, +construction., +heat, +sensor, +device, +engine, +panel	223 (2.9)	102 / 44 (48% / 20%)
C4	62 (25)	+evaporator, +exchanger, +valve, +pump, +layout, +circulation, +heat, +outfall, +controller, +water, +container, +installation, +collector, +meander, +application	167 (2.7)	79 / 41 (47% / 25%)
C5	106 (47)	+collector, +absorber, +pipe, +wall, +divider, +radiation, +medium, +canal, +space, +coverage, +chamber/compartment, +air, +plate, +application, +surface	253 (2.4)	124 / 89 (49% / 35%)

Source: authors' own elaboration.

In addition to the extracted terms characterizing the clusters, the IPC solar-related codes were taken into account to profile the main subject areas of Polish inventiveness. The codes are presented in Figure 6.

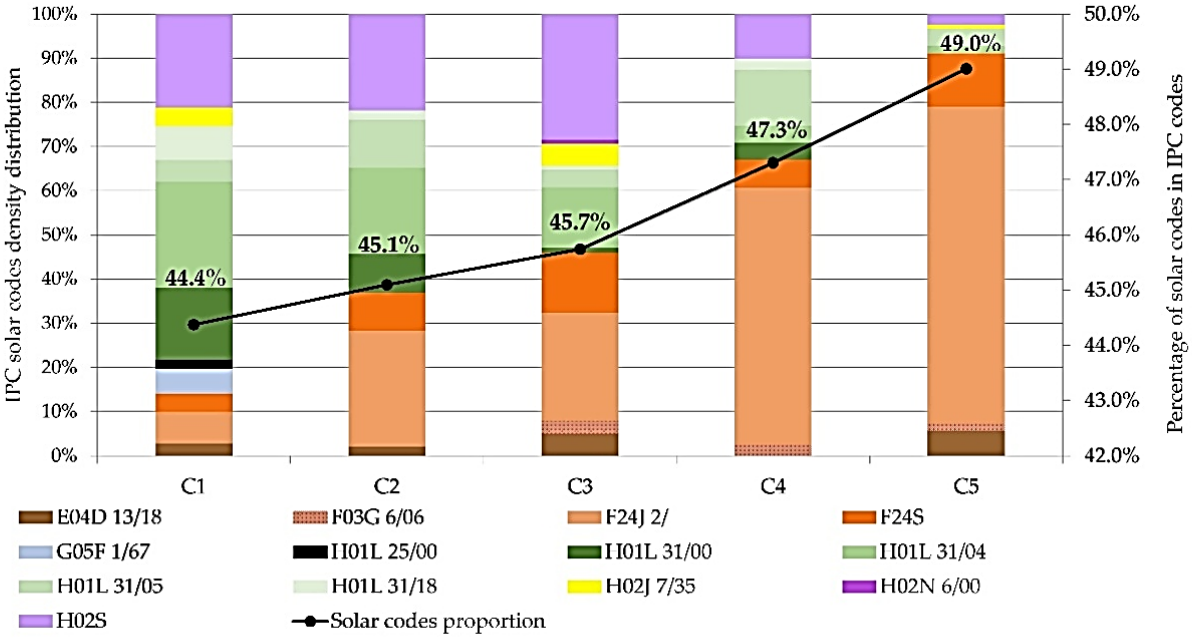

Figure 6. Information on IPC solar related codes in the *BIN_5SV* model clusters.

Source: authors' own elaboration.

Bar charts and the corresponding left-hand axis in the figure show the distribution of the IPC solar related codes in individual clusters of the *BIN_5SV* model. The green group refers to codes related to photovoltaics, while the orange group refers to codes related to solar thermal energy. The line plot and the corresponding right-hand axis illustrate the percentage of solar energy related codes in the total number of IPC codes as regards the documents of a given cluster. In each cluster there are IPC codes not assigned to the solar thematic area; patent applications cover a wider area than just solar energy – the percentage of non-solar codes ranges from 51% in cluster *C*5 to 56% in cluster *C*1.

Additional information on the frequency of IPC codes occurrence in documents of particular clusters is presented in Table 3; non-solar related IPC codes form the *Other codes* separate category. For each cluster, two values are given for each code: the first is the number of occurrences of the code in the cluster documents, and the second one (enclosed in parentheses) is the number of those occurrences in the first position of the IPC classification. Due to the very rare occurrence, the information about the H01L 25/00 and H02N 6/00 codes has been moved to the end of the table.

Table 3.

IPC codes frequency distribution by the clusters of the BIN_5SV model

Cluster	E04D 13/18	F03G 6/06	F24J 2/	F24S	G05F 1/67	H01L 31/00	H01L 31/04	H01L 31/05	H01L 31/18	H02J 7/35	H02S	Total solar codes	Other codes
C1	4 (1)	0 (0)	10 (5)	6 (2)	8 (5)	23 (15)	34 (11)	7 (2)	11 (2)	6 (1)	30 (13)	**142** (57)	178 (37)
C2	1 (1)	0 (0)	12 (5)	4 (3)	0 (0)	4 (3)	9 (1)	5 (1)	1 (0)	0 (0)	10 (4)	**46** (18)	56 (11)
C3	5 (0)	3 (1)	25 (12)	14 (8)	0 (0)	1 (1)	14 (4)	4 (0)	1 (1)	5 (1)	29 (16)	**102** (44)	121 (33)
C4	0 (0)	2 (0)	46 (31)	5 (4)	0 (0)	3 (1)	3 (0)	10 (1)	2 (0)	0 (0)	8 (4)	**79** (41)	88 (21)
C5	7 (3)	2 (0)	89 (71)	15 (12)	0 (0)	0 (0)	2 (1)	5 (0)	0 (0)	1 (0)	3 (2)	**124** (89)	129 (17)
Total	**17 (5)**	**7 (1)**	**182 (124)**	**44 (29)**	**8 (5)**	**31 (20)**	**62 (17)**	**31 (4)**	**15 (3)**	**12 (2)**	**80 (39)**	**493 (57)**	**572 (119)**
H01L 25/00 is not the first-positioned IPC code; it appeared 3 times, only in cluster *C1*. H02N 6/00 is not the first-positioned IPC code; it appeared once, only in cluster *C3*.													

Source: authors' own elaboration.

For each cluster, the proposed outlier search algorithm was applied. The results are summarized in Table 4 and illustrated in Figure 7 consisting of two parts. One part, in the form of a numerical axis, is an illustration of how the algorithm for detecting outlier points in a cluster works. Individual points on the numerical axis correspond to the distances of cluster documents from the centre of gravity. Red colour indicates the distances that identify outliers. The second part of the figure is an illustration on a plane of the clustering results showing the location of documents in each cluster. The figure is an approximate visualization as points (documents) in 5-dimensional space were dropped into 2-dimensional space. The centre of gravity is marked with a green circle, while outlier points are marked with red circles.

Table 4.

Outliers detection results

Cluster	TO	CDV	Identifiers of potential outlier documents
C1	4.34	9.19	361, 363, 354
C2	4.6	7.5	38
C3	4.6	11.73	360
C4	4.35	9.79	267
C5	4.51	-	-

Source: authors' own elaboration.

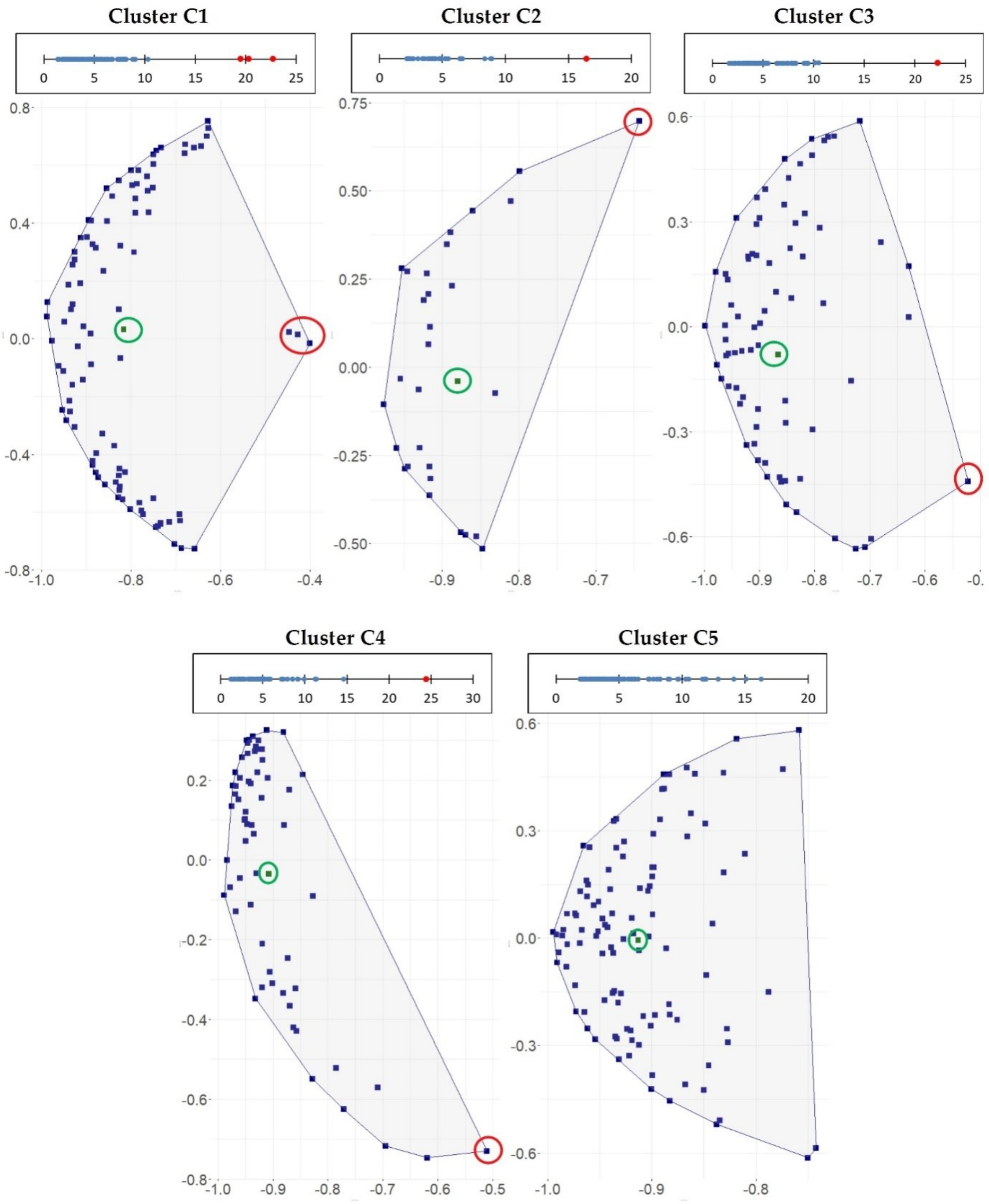

Figure 7. Distances of the documents from the cluster gravity point and the illustration of the documents location in each cluster in the BIN_5SV model – up and down graph respectively.

Source: authors' own elaboration.

5. Discussion

By observing the cluster characteristics shown in Table 2, Figure 6 and Table 3, the following profiles can be derived from the Polish patents related to solar energy in the period 01/2001-03/2021.

Cluster *C*1 is the most heterogeneous one as regards both all and the first-positioned IPC solar-related codes. The number of the codes assigned to the patent documents belonging to this cluster is greater than in the other clusters (see Table 3). Three out of them: H01L 31/00, H01L 31/04, and H02S, account for over 61% of solar classification (68% in the case of first-positioned codes). Considering the description of these IPC codes and the extracted terms (such as *cell*, *generator*, *voltage*, *current*, *form*, *cover*) it can be assumed that cluster *C*1 is devoted to technical solutions for the design of photovoltaic cells and the use of photovoltaic cells in dedicated technical solutions. The cluster generally refers to the conversion of the solar radiation energy, mainly captured by PV devices, into electrical energy or for the control of electrical energy by such radiation.

In cluster C1, three potential outliers with a high mutual similarity were identified. They concern technical solutions for 3D spatial photovoltaic panels. They describe projects the purpose of which is the improvement of photoelectric efficiency. 3D spatial panels make a better use of the same base area than fixed flat solar panels, capturing more amount of light by the light-absorbing material. There are no other 3D photovoltaic panel solutions in the cluster.

In cluster *C*2, the subject of patent applications according to the IPC classification, regardless of the *Other codes* category, is divided into two types of inventions – those concerning solar heat utilization (F24J 2/, F24S) and the ones regarding solar electricity utilization (all the codes starting with H01L). The cluster refers to the design solutions for solar devices or their components (*photodiode*, *bend*, *shape*, *centre*, *hole*, *meander*, *aluminium*, *lens*).

There is one potential outlier identified in cluster *C*2 and it concerns the solution that can have an application in photovoltaic panels. That is a silicon diode in which the active layer has a thickness greater than the visible radiation penetration layer, and the substrate layer has the same type of conductivity as the active layer. There is another application as regards a photodiode in the cluster, but its description concerns the reception of infrared radiation.

In cluster *C*3, more than 42% of the solar-related codes (45% first-positioned codes) refer to photovoltaic devices (H01L 31/04, H02S), nearly 40% (45% respectively) to solar heat devices (F24J 2/, F24S). The cluster concerns inventions relating to improvements in the operation of solar solutions and the accumulation and collection of energy generated from solar radiation (*battery*, *valve*, *mechanism*, *installation*, *sensor*, *clutch*, *engine*, *container*, *heat*). In terms of the type of a solar device, heterogeneity is a feature of the cluster.

There is one potential outlier identified in cluster *C3* and describing the hybrid design of a 3D spatial photovoltaic panel with cooling and heating for enhancing its performance. The panel is cooled by receiving thermal energy and using it to heat domestic water. In contrast, heating is used to remove snow and ice from the panel. In cluster *C3* there are no other applications with such a solution.

Cluster *C*4 refers to solar thermal radiation utilization systems for dedicated solutions. It contains proposals to improve the operation of solar heat receivers (*evaporator*, *exchanger*, *pump*, *controller*, *water*, *meander*, *collector*). Among all solar-related codes, 65% are direct references to solar collectors (F24J 2/, F24S).

The only potential outlier in cluster *C*4 describes composite zinc oxide nanowires for the production of electrodes in dye-sensitized photovoltaic cells and the production method. There are no other applications that contain such a solution.

Cluster *C*5 is the clearest one as regards both all and the first-positioned IPC solar-related codes. Its documents contain design solutions devoted to devices or their components connected with the conversion of solar radiation into heat, or for heat accumulation (*collector*, *absorber*, *pipe*, *divider*, *radiation*). Nearly 84% of all the solar-related codes are direct references to solar collectors (F24J 2/, F24S).

No potential outliers were identified in cluster 5. The distances of the documents from the cluster centre of gravity are evenly distributed, which is the consequence of the homogeneity of the cluster.

After reviewing the applications, a comment can be added that clusters *C*2, *C*3 and *C*4 contain also inventions relating to hybrid solar energy solutions.

6. Conclusions

In the period 01/2001-03/2021, 368 applications in the field of solar energy by Polish inventors were submitted to the Polish Patent Office. The number of the applications increased gradually over the years; at the end of the period, there were more than three times as many applications annually as at the beginning. The research on patent activity and the search for correlations within patent classes and groups allowed identifying the directions of technological solutions in the field of solar energy in Poland.

Most of the patent applications relate to solutions for converting solar energy into electricity, slightly fewer relate to converting solar energy into thermal energy, and there are also hybrid solutions. The clustering analysis made it possible to distinguish among them thematic areas including: PV panel designs, PV panel component designs, the improvement of solar-heat conversion device performance, and solar collector designs. In the collection of the analysed patent applications, there were sporadic references to increasingly common technologies focusing on solar cell materials that would have both high efficiency and low cost.

Seeking for outliers in the analysed sets of documents may be beneficial for research and development departments in various organizations or for individual inventors. The proposed outlier patent document search algorithm allowed the identification of six patent applications that stood out thematically in four out of the total number of five clusters. None have been rejected in the patent granting process; 4 applications are pending (2022/05/22) and 2 have received patent protection. It seems that the direction of development in the solar energy technology may be 3D spatial panels and material technologies for photovoltaics. All outliers concern photovoltaic solutions and none of them refer to the technology of converting solar radiation into thermal energy. The proposed algorithm for identifying outliers seems to work well, but it was applied to a limited study topic and should be verified for other cases. It is also advisable to seek expert opinion in this respect.

In the research process, new solutions regarding the research methodology were proposed in the consecutive steps (see Figure 3). They constitute the authors' methodology contribution as follows:

- the development of the method for selecting the degree of the Term Document Matrix dimensionality reduction (step 2),
- the proposition of the *GQA* global measure of clustering quality assessment (step 3),
- the elaboration of the outlier patent detection algorithm (step 4).

There is the following substantive contribution (as regards the research subject):

- the characterization of the main tendencies of inventiveness in the Polish solar energy technology,
- the detection of outliers in the Polish patent applications, indicating potential directions of technology development in the field of solar energy in Poland.

The presented research is a pilot study and it discusses the issue of solar energy patents for one country, that is Poland. However, the results may be interesting as a country case study and can be referred to in the context of analogous studies for other countries. The developed research methodology is also very promising.

It is planned to extend the analyses to a larger set of patent documents and at the same time solve the problem related to the language of the patent applications (the necessity of having all the analyzed documents in one language) along with merging data from various sources. The data extension concerns a wider range of the inventors' nationality and a wider range of patent databases (such as USPTO and EPO). Considering the fact that patent titles and patent abstracts can be insufficient for the analysis, full patent descriptions are planned to be considered as well. In this aspect, patent claims should also be taken into account.

A properly edited dictionary plays a key role in cluster identification. Therefore, it will also be the subject of the further work, in which parts of speech other than nouns and verbs should be taken into account.

References

1. Act of 30 June 2000. Industrial Property Law, Journal of Law 2000, No. 49 item 508 with later amendments (in Polish) (2000).
2. Albright, R. (2004). *Taming Text with the SVD*. Cary, NC: SAS Institute Inc.
3. Banerjee, A., Dave, R.N. (2004). Validating Clusters Using the Hopkins Statistic. *Proceedings of the 2004 IEEE International Conference on Fuzzy Systems*, *Vol. 1.* Budapest, Hungary: IEEE, pp. 149-153.
4. Bęben, K. (2020). Znaczenie Wyboru Reprezentacji Dokumentów Tekstowych i Miar Podobieństwa w Rankingu Wniosków Patentowych. *Inżynieria zarządzania. Cyfryzacja produkcji. Aktualności badawcze, 2, Vol. 2.* Warszawa: PWE, pp. 1-11.
5. Binz, C., Tang, T., Huenteler, J. (2017). Spatial Lifecycles of Cleantech Industries – The Global Development History of Solar Photovoltaics. *Energy Policy*, *Vol. 101*, pp. 386-402, doi:10.1016/J.ENPOL.2016.10.034.
6. Breyer, C., Birkner, C., Meiss, J., Goldschmidt, J.C., Riede, M. (2013). A top-down analysis: Determining photovoltaics R&D investments from patent analysis and R&D headcount. *Energy Policy*, *Vol. 62*, pp. 1570-1580, doi: 10.1016/j.enpol.2013.07.003.
7. Chiu, Y.-J., Ying, T.-M. (2012). A Novel Method for Technology Forecasting and Developing R&D Strategy of Building Integrated Photovoltaic Technology Industry. *Mathematical Problems in Engineering*, *Vol. 2012*, p. 24, doi:10.1155/2012/273530.
8. De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L. (2000). The Mahalanobis Distance. *Chemometrics and Intelligent Laboratory Systems*, *Vol. 50*, *Iss. 1*, pp. 1-18, doi:10.1016/S0169-7439(99)00047-7.
9. De Paulo, A.F., Ribeiro, E.M.S., Porto, G.S. (2018). Mapping Countries Cooperation Networks in Photovoltaic Technology Development Based on Patent Analysis. *Scientometrics*, *117*, pp. 667-686, doi:10.1007/s11192-018-2892-6.
10. European Patent Office. PATSTAT. *Worldwide Patent Statistical Database*. Retrived from https://www.epo.org/searching-for-patents/business/patstat.html, 15 May 2022.
11. Halkidi, M., Batistakis, Y., Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems 17*, pp. 107-145, doi: 10.1023/A:1012801612483.
12. Hopkins, B., Skellam, J.G. (1954). A New Method for Determining the Type of Distribution of Plant Individuals. *Annals of Botany*, *Vol. 18*, *No. 70*, pp. 213-227, doi:10.1093/oxfordjournals.aob.a083391.
13. Kadhim, A.I., Cheah, Y.-N., Ahamed, N.H. (2014). Text Document Preprocessing and Dimension Reduction Techniques for Text Document Clustering. *Proceedings of the 2014 4-th International Conference on Artificial Intelligence with Applications in Engineering and Technology* (pp. 69-73). Kota Kinabalu, Malaysia: IEEE.

14. Lan, M., Tan, C.L., Su, J.; Lu, Y. (2009). Supervised and Traditional Term Weighting Methods for Automatic Text Categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *Vol. 31*, *No. 4*, pp. 721-735, doi:10.1109/TPAMI.2008.110.
15. Li, X., Xie, Q., Jiang, J., Zhou, Y., Huang, L. (2019). Identifying and Monitoring the Development Trends of Emerging Technologies Using Patent Analysis and Twitter Data Mining: The Case of Perovskite Solar Cell Technology. *Technological Forecasting and Social Change*, *Vol. 146*, pp. 687–705, doi:10.1016/j.techfore.2018.06.004.
16. Liu, J.S., Kuan, C., Cha, S.-C., Chuang, W.-L., Gau, G.J., Jeng, J. (2011). Photovoltaic Technology Development: A Perspective from Patent Growth Analysis. *Solar Energy Materials and Solar Cells, Vol. 95*, *Iss. 11*, pp. 3130-3136, doi:10.1016/J.SOLMAT. 2011.07.002.
17. Lizin, S., Leroy, J., Delvenne, C., Dijk, M., Schepper, E., Passel, S. (2013). A Patent Landscape Analysis for Organic Photovoltaic Solar Cells: Identifying the Technology's Development Phase. *Renewable Energy*, *Vol. 57*, pp. 5-11, doi:10.1016/ j.renene.2013.01.027.
18. Rousseeuw, P.J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, *Vol. 20*, pp. 53-65, doi:10.1016/0377-0427(87)90125-7.
19. Sampaio, P.G.V., González, M.O.A., de Vasconcelos, R.M., dos Santos, M.A.T., de Toledo, J.C., Pereira, J.P.P. (2018). Photovoltaic Technologies: Mapping from Patent Analysis. *Renewable and Sustainable Energy Reviews*, *Vol. 93*, pp. 215-224, doi:10.1016/ j.rser.2018.05.033.
20. SAS Institute Inc. (2012). *SAS Text Miner 12.1 Reference Help*. Cary, NC.
21. SAS Institute Inc. (2016). *SAS/STAT®14.2 User's Guide. The CLUSTER Procedure*. Cary, NC.
22. Trappey, A.J.C., Chen, P.P.J., Trappey, C.V., Ma, L. (2019). A Machine Learning Approach for Solar Power Technology Review and Patent Evolution Analysis. *Applied Sciences*, *9(7), 1478*, p. 25, doi:10.3390/app9071478.
23. Venugopalan, S., Rai, V. (2015). Topic Based Classification and Pattern Identification. *Technological Forecasting and Social Change*, *Vol. 94*, pp. 236-250. doi:10.1016/ j.techfore.2014.10.006.
24. Vijaya, Sharma, S., Batra, N. (2019). Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering. *Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon).* Faridabad, India: IEEE, pp. 568-573.
25. WIPO (2015). *WIPO Guide to Using PATENT INFORMATION; Patent brochures*. Geneva: World Intellectual Property Organization.
26. WIPO (2021). World Intellectual Property Indicators 2021. Geneva: World Intellectual Property Organization.

27. WIPO (2022). *Guide to the International Patent Classification*. Geneva: World Intellectual Property Organization.

28. Yoon, J., Kim, K. (2012). Detecting Signals of New Technological Opportunities Using Semantic Patent Analysis and Outlier Detection. *Scientometrics, 90*, pp. 445-461, doi:10.1007/s11192-011-0543-2.

Appendix A

Table 1a.
The description of selected IPC codes

IPC Code	Description[1]
E04D 13/18	Roof covering aspects of energy collecting devices, e.g. including solar panels
F03G 6/06	Devices for producing mechanical power from solar energy with solar energy concentrating means
F24J 2/	Use of solar heat, e.g. solar heat collectors
F24S	Solar heat collectors; solar heat systems
G05F 1/67	Automatic systems in which deviations of an electric quantity from one or more predetermined values are detected at the output of the system and fed back to a device within the system to restore the detected quantity to its predetermined value or values, regulating electric power to the maximum power available from a generator, e.g. from solar cell
H01L 25/00	Assemblies consisting of a plurality of an individual semiconductor or other solid state devices
H01L 31/00	Semiconductor devices sensitive to infra-red radiation, light, electromagnetic radiation of shorter wavelength, or corpuscular radiation and specially adapted either for the conversion of the energy of such radiation into electrical energy or for the control of electrical energy by such radiation; Processes or apparatus specially adapted for the manufacture or treatment thereof or of parts thereof; Details thereof
H01L 31/04	* Adapted as photovoltaic [PV] conversion devices
H01L 31/05	*** Electrical interconnection means between PV cells inside the PV module, e.g. series connection of PV cells
H01L 31/18	* Processes or apparatus specially adapted for the manufacture or treatment of these devices or of parts thereof
H02J 7/35	Parallel operation in networks using both storage and other dc sources, (e.g. providing buffering) with light sensitive cells
H02N 6/00	Generators in which light radiation is directly converted into electrical energy
H02S	Generation of electric power by conversion of infra-red radiation, visible light or ultraviolet light, e.g. using photovoltaic [pv] modules

[1] The code description was taken from https://onscope.com/ipowner/en/ipc.html (accessed on 01 March 2022).

Source: authors' own elaboration.