Unveiling technology clusters and prominent investors of home automation networking through patent analysis

Konstantinos Charmanas[a], Konstantinos Georgiou[a], Nikolaos Mittas[b] and Lefteris Angelis[a]

[a] *School of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece*
[b] *Department of Chemistry, International Hellenic University, 65404 Kavala, Greece*

Abstract

Home automation systems and networks aim to boost the quality of life and support the automation of industrial operations. Their wide applicability has attracted the interest of researchers and companies. In the last decades, home automation systems constitute an emerging domain, and the corresponding developed technologies are available to both organizations and infrastructures as well as common users. The importance of home automation systems is apparent as their technologies are utilized in various important domains such as smart homes, Internet of Things (*IOT*), vehicles and healthcare. Companies that develop products of this nature aim to patent their most valuable technologies in order to protect their investments against rival competitors. The granted patents are recorded and stored by different patent offices that provide comprehensive information and constitute the knowledge base in much research. In this study, we analyze patent records from the United States Patent and Trademark Office (USPTO) that are related to home automation networks in order to reveal the most relevant technologies and domains of application as well as the respective prominent stakeholders. Our findings provide insights on the current state of home automation technologies and support researchers and organization in tasks that are related to competitor analysis, strategic planning and technology forecasting.

Keywords 1

Patent analysis, Home automation networks, Cluster analysis

1. Introduction

Ubiquitous computing is a concept that encompasses multiple different areas of the computer science domain, including distributed computing, sensors and IoT technologies as well as artificial intelligence. A primary aspect of ubiquitous computing is its objective on affecting the daily aspects of human life, providing time-saving and innovative alternatives to otherwise tiring tasks. An example of ubiquitous computing that highlights its presence in the modern society are home automation systems, expressed in devices that range from digital assistants to smart lightbulbs and devices that can be found in an average household.

The importance and benefits of home automation systems are more than evident in everyday activities, as they provide capabilities that facilitate and improve home security, management, flexibility and remote control. The main components that encompass a complete home automation system are the user interface, the participating electronical devices and a mode of transmission [1]. The additional components of common home automation systems include voice recognition, mobile communication through messaging and *IOT*, as well as popular Information and Communication Technologies (ICTs) like Bluetooth and Wi-Fi [2]. Such a plethora of applied concepts indicates that the different home automation technologies should be

Published in the Workshop Proceedings of the EDBT/ICDT 2023 Joint Conference (March 28-March 31, 2023, Ioannina, Greece).
EMAIL: kcharman@csd.auth.gr (A. 1); konsgeor@csd.auth.gr (A. 2); nmittas@chem.ihu.gr (A. 3); lef@csd.auth.gr (A. 4)
ORCID: 0000-0001-5743-609X (A. 1); 0000-0002-9851-1196 (A. 2); 0000-0003-3061-7864 (A. 3); 0000-0002-6677-4039 (A.4)

evaluated based on indicators that measure both their capabilities and potential costs, which include data rates, transmission range, energy costs, security, complexity and more [3,4]. Hence, creating and establishing home automation systems that can effectively exploit these indicators through efficient technologies and inventions consists of a challenging task.

Given that inventions with innovative concepts of home automation technologies provide organizations and companies with ample opportunities of exploiting them in the market, the early patenting of inventions has become a staple for investors of not only home automation, but also in any other technological domain. Thus, patent data can be a potent indicator of technological growth, development and business activity [6].

In this study, we explore the main applications and technologies that are related to home automation systems through a patent analysis methodology that employs dimensionality reduction and cluster analysis techniques. To do so, we collect and analyze information of technological patents, derived from the USPTO, that are related to home automation systems. Complementary, we also reveal the main stakeholders that invest in home automation technologies and contribute to the landscape profiling of the investigated technological area and subareas.

The results of this study provide insights for both practitioners and industries in identifying the main domains, applications, methodologies and competitors of home automation technologies. In the context of disseminating knowledge, this study serves as a steppingstone in merging the disciplines of ubiquitous computing and patent analysis, offering valuable information that can foster and promote additional research on the topic. In addition, our methodologies constitute a robust example of applied Machine Learning on ubiquitous computing data, indicating breakthrough findings related to patents of the field via a streamlined data analytics pipeline.

2. Related Work

Patent analysis is a process that is related to analyzing information of patent data and extracting knowledge for various purposes including trending analysis, technology forecasting, strategic positioning and competitor analysis [5]. The main properties of patent data provide information regarding the patent assignees (companies or organizations) and inventors, textual descriptions of the patent claims as well as citation relations with other patents, literature and applications. Also, each patent is assigned to one or more classes that characterize its technological objectives, using some popular classification schemes such as the Cooperative Patent Classification (*CPC*)[2] and International Patent Classification (*IPC*)[3] schemes.

Patent analysis is a well-known technique that has application in different technological domains that are related to augmented reality [6,7], *IOT* [8,9], healthcare [10,11] and vehicles [12,13] among many others. In general, the various employed methodologies and data mining techniques applied in patent analysis are usually based on citation networks [14,15,16], topic modelling [6,7,17,18], co-word analysis [19,20], deep learning [9,21] and clustering algorithms [9,22,23].

3. Methodology

The first goal of this study is to discover the general technological areas where home automation networks are applied and the methodologies that are leveraged to construct this type of networks. The second goal is to map the main assignees of technological patents, that are related to home automation networks, and identify the main stakeholders of each technological area, as measured by the overall patents that they possess. These two goals are summarized into the following Research Questions (*RQs*):

RQ_1: What are the main technological areas that are related to home automation networking technologies?

RQ_2: Who are the main stakeholders of the different technologies that are related to home automation networking?

The main framework of this study is presented in **Figure 1**.

[2] https://www.uspto.gov/web/patents/classification/cpc/html/cpc.html

[3] https://www.wipo.int/classifications/ipc/en/

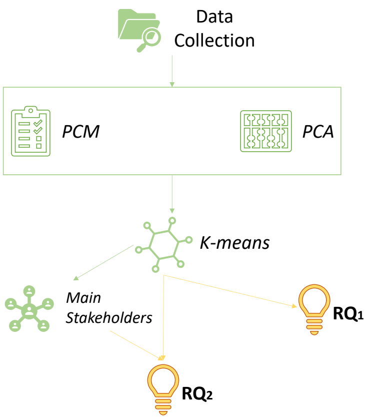

Figure 1: Main framework of the study

To satisfy these goals, we initially collect patent data from the USPTO which is proven to be reliable source, compared to other patent offices, that is often utilized for patent analysis tasks [24]. In order to retrieve only relevant patents, we manually searched the CPC scheme and identified nine prefixes of CPC identifiers that matched our criteria. As a result, 27 CPC subgroups are selected for this study and overall, a sample of 5740 patents were collected (*Data Collection*). The first patent of our dataset was granted in 1979 while the latest was granted in the June of 2022. The annual growth of home automation patents can be observed in **Figure 2**. As we can see, there is an upwards trajectory in later years, with the decrease after 2021 being attributed to the fact that many patent applications have not been granted and published yet.

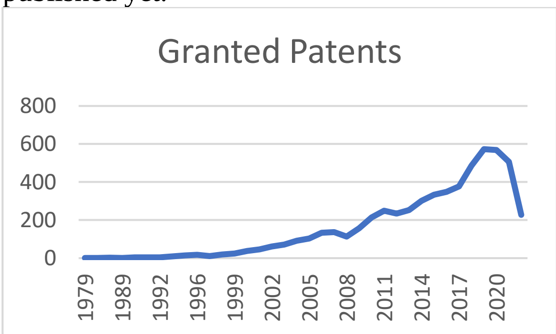

Figure 2: Annual Growth of Patents

It is hence apparent that home automation technologies have been a subject of interest since the dawn of computing, as they grant access to quality-of-life improvements. The short descriptions that characterize the aforementioned prefixes are presented in Table 1. We should mention that the general class of all the selected CPC subgroups is characterized as "*Data switching networks -characterized by path configuration, e.g. LAN [Local Area Networks] or WAN [Wide Area Networks] - Home automation networks*".

Table 1
Main CPC prefixes of the study

CPC prefix	Short descriptions
H04L12/2803	Home automation networks
H04L12/2805	Home Audio Video Interoperability [HAVI] networks
H04L12/2807	Exchanging configuration information on appliance services in a home automation network
H04L12/2809	Exchanging configuration information on appliance services in a home automation network - indicating that an appliance service is present in a home automation network
H04L12/281	Exchanging configuration information on appliance services in a home automation network - indicating a format for calling an appliance service function in a home automation network
H04L12/282	Controlling appliance services of a home automation network by calling their functionalities - based on user interaction within the home
H04L12/283	Processing of data at an internetworking point of a home automation network
H04L2012/284	Characterized by the type of medium used
H04L2012/285	Characterized by the type of home appliance used- Generic home appliances, e.g. refrigerators

In order to discover the general areas of the investigated domain, we employ a methodology similar to [23] by clustering the collected patents based on their CPC subgroup ID assignments. Our choice in replicating this methodology stems from the fact that the results of [23] are promising and well structured, rendering this study as a leading baseline in patent analysis.

In this approach the first step is to construct the Patent CPC Matrix (*PCM*) as follows (1).

$$PCM_{ij} = \begin{cases} 1, \text{when } the\ i-th\ patent\ is\ assigned\ to\ the\ j-th\ CPC\ subgroup \\ 0, \quad otherwise \end{cases} \tag{1}$$

where i=1,2,…,N and j=1,2,…,M. N is equal to the number of the collected patents, 5740 in our case, and M is equal to the number of the distinct CPC subgroup IDs assigned to the collected patents (5523 overall).

In the next step, since the number of CPC subgroups is quite extensive, we apply the Principal Components Analysis – *PCA* [25] dimensionality reduction technique that helps in boosting the performance of the clustering analysis for high dimensional data [26].

Moreover, we employ the *K-means* clustering algorithm [27] to discover the main technological areas, expressed as groups of CPC subgroups, that are related to home automation networking. The selection of *K-means* was based on its wide application in similar patent analysis studies and on its interpretability. In this approach each patent is assigned to a single cluster based on its properties derived from the *PCA* technique. In order to pick the optimal number of clusters for our analysis we make use of two evaluation metrics: (i) the *silhouette coefficient* [28] and (ii) The Normalized Pointwise Mutual Information – *NPMI* [29]. The silhouette coefficient of a single data point is presented in (2)

$$\mathrm{s(i)} = \frac{b(i)-a(i)}{\max\left(b(i),a(i)\right)} \tag{2}$$

where $a(i)$ is the mean distance between the *i-th* data point and the points that belong to the same cluster while $b(i)$ denotes the minimum mean distance between the *i-th* data point and the points of the remaining clusters.

While the silhouette coefficient is a standard approach to evaluate clustering models, the *NPMI* will help us to interpret the extracted clusters in terms of general technological areas using the five most probable CPC subgroup IDs of each cluster. The *NPMI* between two features is presented in (3).

$$NPMI(\mathrm{i,j}) = \frac{\log_2 \frac{p(i,j)}{p(i)p(j)}}{-\log_2 p(i,j)} \tag{3}$$

where $p(i)$ denotes the probability of a patent being assigned to the *i-th* CPC subgroup identifier and $p(i,j)$ denotes the probability of a patent being assigned to both *i-th* and *j-th* CPC subgroup identifier.

Finally, we make use of the information related to the assignees of each patent, included in the patent data, to reveal the main stakeholders of each technological area. In this phase, we rank the linkage strength of an assignee and a cluster according to the number of patents and belonging to this cluster and under the ownership of this assignee (*Main Stakeholders*).

4. Results

4.1. RQ$_1$: What are the main technological areas that are related to home automation networking technologies?

By leveraging the patent properties extracted from the *PCA*, we conduct several experiments to choose the number of clusters that optimize the *NPMI* and the *silhouette coefficient*. In this study, we set the range clusters from 2 to 20. The evaluation of the extracted models is presented in **Figure 3**.

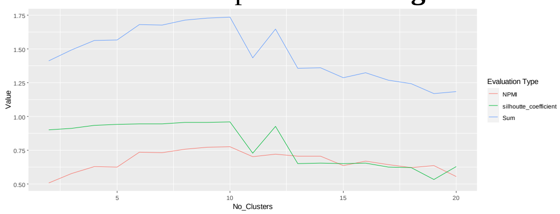

Figure 3: Evaluation of K-means models

An inspection of the Figure indicates that the optimal number of clusters for our experiments is equal to 10, as the extracted model for this number of clusters provides the highest values for both the *NPMI* and *silhouette coefficient* evaluation metrics. Next, we examine the ten most probable CPC subgroups included in each cluster in order to assign a representative description for the extracted technology clusters. In Figure 4 we display the descriptions of the ten clusters that comprise different aspects of home networking technologies

We have to mention that a large proportion of the collected patents belong to the tenth cluster that represents the general area of home automation networks as multiple patents are assigned only to the CPC subgroups that are related to the prefixes of Table 1. By inspecting the most probable CPC subgroups, we can conclude that the results do not indicate significant overlaps between clusters apart from some cases where the subgroups of the last cluster occur in other clusters as well. As the last cluster is related to the general investigated technological area, the identified overlaps should be characterized as insignificant and expected.

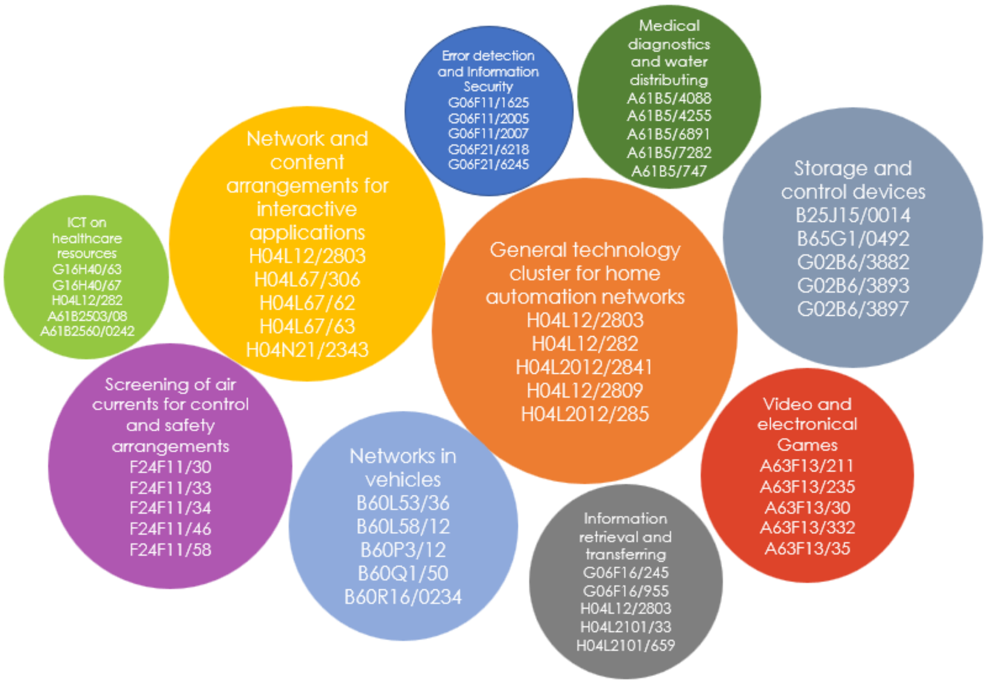

Figure 4: Patent Clusters and top CPCs

This observation shows that the applied approach discovered distinct and interpretable technological areas. These areas are related to both interesting and established techniques applied on home automation networks e.g., *Error detection and Information Security*, *Information retrieval and transferring* and important domains that leverage technologies of home automation networks e.g., *healthcare*, *vehicles*, *Storage and control devices*, *air screening*, *games*.

4.2. RQ$_2$: Who are the main stakeholders of the different technologies that are related to home automation networking?

In this subsection we present the main assignees contained in the retrieved patents. The involvement of the investigated assignees reveals their overall interest in the technological area of home automation networks. This involvement can be translated on investments in patent technologies, their desire for applications of these technologies and the strategic advantages or disadvantages against competitors in the technological area of home automation networks. In brief, a total of 922 assignees are detected in the retrieved patent data with more than 50% percent owning a single patent, while the assignees' mean number of granted patents is less than 7 and the largest number of granted patents is 422, owned by Samsung. This statistic indicates that the majority of assignees are not investing in many home automation network patents, being overrun by competitive technological giants, while there is also a small proportion of assignees that invest frequently in technologies that are related to home automation networks, while having the resources and market shares to do so.

By combining the information of the patent data with the technology clusters that were extracted in the previous subsection, we are able to distinguish the main of assignees of each cluster. The top three most involved assignees of each technology cluster are presented in Table .

Table 2

Main assignees of the technology clusters

Top assignees per cluster
STATE FARM MUTUAL AUTOMOBILE INSURANCE COMPANY ; 16Lab Inc ; 2Wire, Inc. (Cluster 1)
WINT WI Ltd ; 16Lab Inc ; 2Wire, Inc. (Cluster 2)
May Patents Ltd. ; 16Lab Inc ; 2Wire, Inc. (Cluster 3)
Cisco Technology, Inc. ; 16Lab Inc ; 2Wire, Inc. (Cluster 4)
University of Florida Research Foundation, Inc. ; Delos Living LLC ; 16Lab Inc (Cluster 5)
Microsoft Technology Licensing, LLC ; Canon Kabushiki Kaisha ; Hitachi, Ltd. (Cluster 6)
Intel Corporation ; 16Lab Inc ; 2Wire, Inc. (Cluster 7)
Broadcom Corporation ; 16Lab Inc ; 2Wire, Inc. (Cluster 8)
Google Inc. ; 16Lab Inc ; 2Wire, Inc. (Cluster 9)
Samsung Electronics Co., Ltd. ; Sony Corporation ; Google Inc. (Cluster 10)

By inspecting the results presented on the table above, we detect that *16Lab Inc* and *2Wire, Inc.* are involved in the majority of the extracted technology clusters. *2Wire, Inc.*[4] was a home networking equipment manufacturer that provided products to telecommunication companies while *16Lab Inc*[5] provides an integrated platform for *IOT* and wearable developers. While these two assignees are strongly related to home automation networks, we also observe some established software and hardware companies, that do not only invest in home automation networks but lead pioneering developments in other fields as well, like *Google*, *Samsung*, *Sony*, *Microsoft* and *Cisco*.

Among the different areas of interest of the assignees, we can distinguish that *farming*, *automobiles*, *information security*, *IOT*, *hardware*, *software development*, *healthcare* and *telecommunications* are the most frequent ones. These observations reveal that home automation networks are useful in many technological areas, systems, devices and applications.

5. Conclusions and Future Work

In this study, we presented and uncovered the main technological areas and assignees that are related to home automation networking by analyzing patent data derived from the USPTO. The first step in our approach was to investigate the patent objectives using the CPC assignments that characterize them. Next, by applying a dimensionality reduction technique and a clustering algorithm, we succeeded in identifying and interpreting the main technology clusters of home automation network patents, providing effective answers to **RQ_1.**

Moreover, the information regarding the assignees of each patent and the technology clusters extracted from the K-means algorithm help us in identifying the main stakeholders of the different technologies that are related to home automation networks (**RQ_2**).

Our approach unveils some interesting technological areas that are related to home automation networking which support the utilities of farming, healthcare, automobile, security and information systems. Furthermore, the applied methodology reveals the main investors in each domain that applies home networking technologies and their interest in patent granting.

The challenges that arise from the combination of ubiquitous computing in machine learning stem from the proper understanding of the field and its usage on daily life as well as the identification of proper datasets for analysis. Nevertheless, we believe that our insights prove that such challenges can be overcome when studying ubiquitous computing from an industrial perspective.

By recapping, we believe that the proposed approach provides valuable information that could be useful for companies and individuals in scoping the main competitors and potential collaborators in future projects. We expect that this information can also provide guidelines to practitioners and researchers in identifying important technologies and methodologies in the general domain of home automation

[4] https://en.wikipedia.org/wiki/2Wire

[5] https://16lab.net/about/

networking, which could be a great asset in future studies and in the forging of new inventions and applications. As far as the development and knowledge sharing on ubiquitous computing is concerned, this study is a great example of interdisciplinary research that promotes complementarity in science, a concept that should be leveraged by other researchers.

Also, we believe that the proposed methodology can provide effective results not only on patent data but also on research literature data. For instance, the CPC subgroup identifiers in the clustering approach could be replaced with the keywords of each study while the patent assignees could be replaced with the authors and publication venues of the studies. In general, this approach can be used to identify the main individuals or contributors of a dataset and further discover the main areas or topics of that are reflected on the dataset records.

Evidently, there were some limitations in our study. Firstly, the data that were used for this study are derived exclusively from the USPTO and not from all the available patent offices. Additional data from the various related industries and patent offices should be evaluated in order to generalize the results of this study. Despite the occurrence of these limitations, USPTO is characterized as a reliable source for patent analysis that reflects on the industrial technologies and interests.

Furthermore, additional experiments should be conducted in order to verify our approach as an effective alternative for general tasks and different data. Some important stages regarding the evaluation and interpretation of the clusters were conducted in a manual fashion. As a result, we believe that several alternatives in different stages of this approach should also be investigated. Despite the existence of the issues described previously, the algorithms and evaluation metrics of the proposed approach are similar to multiple existing studies that provide satisfying methodologies and outcomes.

Nonetheless, we consider that future studies could also explore the general domain of home automation networks through additional information and purposes. The patent titles and abstracts are two very important features that describe the nature of the inventions and could possibly be analyzed for knowledge discovery through topic modelling or co-word analysis. Also, the patent granting dates, filing dates and inventors are features that are included in the patent data and provide useful information about the overall temporal growth of a technological domain and the most prolific individual inventors respectively. Finally, we believe that the citation properties of the patents contain valuable information regarding the technological value and influence of the patents that help in identifying emerging and dominant technologies and determining the strategic positioning of assignees.

6. References

[1] Gunge, V. S., & Yalagi, P. S. (2016). Smart home automation: a literature review. International Journal of Computer Applications, 975(8887-8891).

[2] Asadullah, M., & Raza, A. (2016, November). An overview of home automation systems. In 2016 2nd international conference on robotics and artificial intelligence (ICRAI) (pp. 27-31). IEEE.

[3] Withanage, C., Ashok, R., Yuen, C., & Otto, K. (2014, May). A comparison of the popular home automation technologies. In 2014 IEEE Innovative Smart Grid Technologies-Asia (ISGT ASIA) (pp. 600-605). IEEE.

[4] Toschi, G. M., Campos, L. B., & Cugnasca, C. E. (2017). Home automation networks: A survey. Computer Standards & Interfaces, 50, 42-54.

[5] Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. World Patent Information, 37, 3-13.

[6] Choi, H., Oh, S., Choi, S., & Yoon, J. (2018). Innovation topic analysis of technology: The case of augmented reality patents. IEEE Access, 6, 16119-16137.

[7] Jeong, B., & Yoon, J. (2017). Competitive intelligence analysis of augmented reality technology using patent information. Sustainability, 9(4), 497.

[8] Daim, T., Lai, K. K., Yalcin, H., Alsoubie, F., & Kumar, V. (2020). Forecasting technological positioning through technology knowledge redundancy: Patent citation analysis of IoT, cybersecurity, and Blockchain. Technological Forecasting and Social Change, 161, 120329.

[9] Trappey, A. J., Trappey, C. V., Govindarajan, U. H., & Sun, J. J. (2019). Patent value analysis using deep learning models—the case of IoT technology mining for the manufacturing industry. IEEE Transactions on Engineering Management, 68(5), 1334-1346.
[10] Nam, H. J., & An, Y. (2017). Patent, R&D and internationalization for Korean healthcare industry. Technological Forecasting and Social Change, 117, 131-137.
[11] Han, J. H., Hyun, Y. G., Chae, U., Lee, G. H., & Lee, J. Y. (2020). A Study on the Healthcare Technology Trends through Patent Data Analysis. Journal of Digital Convergence, 18(3), 179-187.
[12] Ha, S. H., Liu, W., Cho, H., & Kim, S. H. (2015). Technological advances in the fuel cell vehicle: Patent portfolio management. Technological Forecasting and Social Change, 100, 277-289.
[13] Yang, L. F., Xu, J. H., & Neuhäusler, P. (2013). Electric vehicle technology in China: An exploratory patent analysis. World Patent Information, 35(4), 305-312.
[14] Park, Y. N., Lee, Y. S., Kim, J. J., & Lee, T. S. (2018). The structure and knowledge flow of building information modeling based on patent citation network analysis. Automation in Construction, 87, 215-224.
[15] Hu, X., Rousseau, R., & Chen, J. (2012). A new approach for measuring the value of patents based on structural indicators for ego patent citation networks. Journal of the American Society for Information Science and Technology, 63(9), 1834-1842.
[16] Yang, G. C., Li, G., Li, C. Y., Zhao, Y. H., Zhang, J., Liu, T., ... & Huang, M. H. (2015). Using the comprehensive patent citation network (CPC) to evaluate patent value. Scientometrics, 105(3), 1319-1346.
[17] Song, B., & Suh, Y. (2019). Identifying convergence fields and technologies for industrial safety: LDA-based network analysis. Technological forecasting and social change, 138, 115-126.
[18] Kim, K. H., Han, Y. J., Lee, S., Cho, S. W., & Lee, C. (2019). Text mining for patent analysis to forecast emerging technologies in wireless power transfer. Sustainability, 11(22), 6240.
[19] Yoon, J., Choi, S., & Kim, K. (2011). Invention property-function network analysis of patents: a case of silicon-based thin film solar cells. Scientometrics, 86(3), 687-703.
[20] Wu, C. C., & Leu, H. J. (2014). Examining the trends of technological development in hydrogen energy using patent co-word map analysis. International Journal of Hydrogen Energy, 39(33), 19262-19269.
[21] Krestel, R., Chikkamath, R., Hewel, C., & Risch, J. (2021). A survey on deep learning for patent analysis. World Patent Information, 65, 102035.
[22] Yoon, J., & Kim, K. (2011). Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. Scientometrics, 88(1), 213-228.
[23] Kim, G., & Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. Technological Forecasting and Social Change, 117, 228-237.
[24] Kim, J., & Lee, S. (2015). Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO. Technological Forecasting and Social Change, 92, 332-345.
[25] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.
[26] Ayesha, S., Hanif, M. K., & Talib, R. (2020). Overview and comparative study of dimensionality reduction techniques for high dimensional data. Information Fusion, 59, 44-58.
[27] Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. Applied Statistics, 28, 100--108. 10.2307/2346830.
[28] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.
[29] Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. Proceedings of GSCL, 30, 31-40.