

Big Data and Technology Evolution in the IoT Industry

Elona Marku¹, Maryia Zaitsava¹, Manuel Castriotta¹, Maria Chiara Di Guardo¹ & Michela Loi¹

¹ Department of Economics and Business, University of Cagliari, Cagliari, Italy

Correspondence: Elona Marku, Department of Economics and Business, University of Cagliari, Via Sant'Ignazio 74, 09123 - Cagliari, Italy. E-mail: elona.marku@unica.it

Received: July 30, 2020

Accepted: September 2, 2020

Online Published: September 21, 2020

doi:10.5539/ijbm.v15n10p94

URL: <https://doi.org/10.5539/ijbm.v15n10p94>

Abstract

The present study aims to better understand how and to what extent the different dimensions of Big Data can offer insights on technology evolution. By using a patent analytics perspective, in this paper, we introduce a novel approach based on co-words analysis using the abstracts of 170,279 European patents in the Internet of Things (IoT) field published from 2011 to 2019. In so doing, we map and visualize an industry's technology structure, development, and trends, as well as disentangle the IoT technology conceptual structure, highlighting its core and boundary concepts. This is the first study that applies a decomposition framework to clarify the determinants of IoT inventions, showing relevant changes in the focus of IoT technology overtime. By shedding light on the evolutionary dynamics of the field, this research offers a valuable contribution to the technology innovation literature.

Keywords: Big Data, patent intelligence, innovation, IoT, mapping, co-word, VOSviewer

1. Introduction

Since Big Data technologies have emerged in our networked society, a new synthesis of real-time, user-generated information and communication creates a constant flow of potential new insights for business, government, education, and social initiatives. This context has remarkably increased the interest of practitioners and scholars across various academic disciplines, including management, business, and information systems (Frizzo-Barker, Chow-White, Mozafari, & Ha, 2016; Sarica, Yan, & Luo, 2019; Ponta, Puliga, Oneto, & Manzini, 2020). Knowledge of Big Data is becoming increasingly relevant today, as firms are required to rely on Big Data to support their business activities and decision-making process (O'Donovan, Leahy, Bruton, & O'Sullivan, 2015, Wessel, 2016; Xiaofeng & Xiang, 2013; Davidson, 2019). Indeed, Big Data is used by firms in a variety of strategic ways: to create a new product, process, and business model innovations; for customer need identification; risk management and decision-making; data-driven knowledge; product and service design; quality management; and opportunity recognition and creation (Günther, Mehrizi, Huysman, & Feldberg, 2019, Erevelles, Fukawa, & Swayne, 2016, Jin, Liu, Ji, & Liu, 2016, Manyika et al., 2011).

Recent research has highlighted that patent data is particularly suitable for Big Data tools and techniques (Lee, Han, & Sohn, 2015; Ponta et al., 2020). The patent documents contained enormous and rich technical information to gain a competitive advantage for both public and private entities in their marketplace. From the technology innovation literature, we already know that patents are a meaningful instrument to measure innovation performance and quality (Ahuja & Katila, 2001; Harrigan, Di Guardo, & Marku, 2018; Hagedoorn & Cloudt, 2003; Trajtenberg, 1990). Additionally, patent analytics enable the analysis of thousands or even millions of patents, leading to insightful conclusions related to technologies and markets.

In this paper, we consider patent documents as an objective data source and the comprehensive set of all patents related to a specific topic as Big Data. In so doing, we use a patent analytics approach to understand better how and to what extent the different dimensions of Big Data can offer insight to technology evolution. Moreover, we introduce a new patent intelligence approach for mapping and visualizing an industry's profile and evolution by focusing on fine-grained patent data. Departing from extant literature, we use a novel and validated tool in management studies, the VOSviewer software (Van Eck, Waltman, Den Berg, & Kaymak, 2006; Van Eck & Waltman, 2007; Van Eck, Waltman, Dekker, & van den Berg, 2010; Waaijer, van Bochove, & Van Eck, 2010; Zupic & Čater 2015). Finally, we examine the Internet of Things (IoT) industry as it is expected to bring disruption in many other industries changing their technology connotations.

We collected and subsequently analyzed 170,279 European patent applications and the patent granted in the IoT

field, published from 2011 to 2019. By adopting a patent co-word analysis, we mapped and visualized not only the structure of the IoT technology but also the links between the different technological elements, highlighting boundary and core concepts to the field. In this way, we contribute both to patent and Big Data literature by providing new insight into the use of patents as Big Data and patent intelligence for competitive advantage.

2. Literature Background

2.1 Big Data

Big Data has been widely exploited to create a product, process, and business model innovations (Manyika et al., 2011; Jin et al., 2016; Erevelles et al., 2016) and is nowadays one of the most analyzed digital technologies in the literature (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012, Mayer-Schönberger & Cukier, 2013; Svahn, Mathiassen, & Lindgren, 2017, Nambisan, Lyytinen, Majchrzak, & Song, 2017). Big Data is commonly defined as large structured and unstructured data sets characterized by high volume and variety of data sources, and high speed of data generation, that require the use of advanced computational techniques to unveil trends and patterns to help better decision-making (Chen, Chiang, & Storey, 2012; George, Haas, & Pentland, 2014; Kwon, Lee, & Shin 2014).

In the search for a unified definition of Big Data, some scholars explore Big Data in terms of “what data does”, while others are focused on “what data is,” but all concepts center around defining the main characteristics and qualities of Big Data (Gandomi, & Haider, 2015). Indeed, three main characteristics introduced by Laney (2001) in his seminal work distinguish just data from Big Data. Volume, Variety, and Velocity, or the so-called 3Vs approach, reflect the three main dimensions of Big Data (Chen et al., 2012; Kwon et al., 2014, Gandomi, & Haider, 2015). Volume is the most definitive dimension of Big Data, among other characteristics. It reflects the size or the magnitude of data created from all possible sources. The ever-increasing volume of data, often measured in tera- or petabytes basically grounds the understanding of Big Data (George et al., 2014). For a long period, Big Data was perceived from the point of view of its size of datasets (George et al., 2014; Frizzo-Barker et al., 2016). However, a big volume of data does not guarantee the quality of insights (Chen et al., 2012; Sivarajah, Kamal, Irani, & Weerakkody, 2017; Yoo, 2015). Also, there is no unique threshold to meet, after which just data can be called Big Data. Not only may two data sets of the same size require different analytics technologies, but also they can be perceived differently due to their nature, exemplary, a terabyte of video data versus terabytes of text data (Gandomi & Haider, 2015). Therefore, Big Data Volume, although being the most distinguishing feature, is rather a relative dimension.

What makes Big Data more appealing for employment in organizations for decision making is the variety of sources to generate data. Modern data is generated from an increasing variety of sources from social media data to industrial and machine data; from accumulative databases of different kinds to real-time generated data; it can have diverse shapes, name a few, text, audio, video, weblogs, user-generated content. All these compose the second facet of Big Data, namely, Variety (Gandomi & Haider, 2015). It is recognized that Variety of data sources enriches Big Data and helps to generate more sophisticated insights (George et al., 2014). Moving further, Variety of Big Data has been discussed not only from the point of view of the number and different types of data of sources but also from the point of view of the so-called granularity—or the level of detail of Big Data (Günther et al., 2017; Yoo, 2015). Thus, the variety of fine-grained insights produced by Big Data reflects the distinctive nature of Big Data on the level with the variety of Big Data sources (Constantiou & Kallinikos, 2015).

However, what makes Big Data qualitatively different from a high volume of statistical and archival data is the ability to generate data in real-time and reveal hidden trends and patterns by applying computational approaches, while the mentioned above remain mainly static and historical data (George et al., 2014). Thus, the third dimension of Big Data – Velocity reflects the speed with which Big Data is generated and analyzed (Sivarajah et al., 2017). Moreover, Kitchin and McArdle (2016) emphasized that an important aspect of embracing the Velocity dimension entirely, is to understand that Big Data generation is not an occasional event but rather a continuous process. Therefore, the exponential growth of real-time and non-stop generated Big Data has triggered the need for enhanced real-time analytics (Gandomi & Haider, 2015). To respond to this challenge, scholars have added new facets of the Velocity, such as, the speed of data processing and analytics and data visualization (Kitchin & McArdle, 2016), speed of acting upon Big Data analytics (Gandomi, & Haider, 2015), and the issues related to Big Data decay or the declining value of data over time (Lee, 2017).

In addition to the three dimensions making Big Data different from other types of data, recent studies have introduced other dimensions mainly dealing with a specific type of issues of Big Data (Simsek, Vaara, Paruchuri, Nadkarni, & Shaw, 2019). Thus, the fourth dimension, namely, Veracity, reflects the ability of Big Data to be trustful (Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012). Veracity repulses issues such as

trustworthiness, accountability, and authenticity (Simsek et al., 2019). Reliable data becomes an increasingly important issue, as with the rise of various data sources, comes the challenge of trustworthy data (Gandomi & Haider, 2015; Schroeck et al., 2012). In this perspective, Big Data should possess a certain level of trust and reliability (Wamba et al., 2015). Therefore the decision-making performance can be influenced by the presence or absence of high-quality data in terms of timing and relevance (Ghasemaghaci, Ebrahimi, & Hassanein, 2018; Lee, 2017; Sukumar & Ferrell, 2013).

Value, the fifth dimension, has been introduced relatively recently, as a desirable outcome of Big Data (Gandomi & Haider, 2015; Uddin & Gupta, 2014) and already has triggered the conversation on how to analyze Big Data in a way that brings big value in managerial literature (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). In an attempt to understand what Big Data Value is, scholars apply the concept of economic or business and social value. Thus, economic and business values are those that reflect the business, income growth improved by Big Data use (McAfee et al., 2012), while social values are improvements of the level of welfare, education, health, mobility, and other social indicators caused by Big Data adoption (Wang, Kung, & Byrd, 2018; Günther et al., 2017). Some scholars employ the current view on Value from management literature, where it is defined as the input to the utility of a final product or service to end-users (Urbinati, Bogers, Chiesa, & Frattini, 2019). However, the most prominent research stream within the Value of Big Data topic is on identifying concrete types of benefits Big Data can bring to organizations. Name a few, identification of customers and users' needs for designing new products or services, improving existing product profiles (Stockstrom, Goduscheit, Lüthje, & Jørgensen, 2016; Markham, Kowolenko, & Michaelis, 2015), and innovation and digital transformation initiatives support (Urbinati, Chiaroni, Chiesa, & Frattini, 2018; Wan, Williamson, & Yin, 2015; Troilo, De Luca, & Guenzi 2017). Like all other dimensions of Big Data, Value is rather a subjective characteristic as it is linked to the context where Big Data is applied and depends on the strategic goals of an organization (Günther et al., 2017).

2.2 Patents as Big Data

Patents are intellectual property rights that grant to their owners an exclusive monopoly power over the use of an invention, and as such, they play a crucial role to encourage innovation and technology advancement (Oh, Cho, & Kim, 2014; Hall et al., 2001; Di Guardo, Harrigan, & Marku, 2019). In this paper, we consider patent documents as an objective data source and the comprehensive set of all patents related to a particular topic, as Big Data. Indeed, patent data are defined in terms of the technical information included in the patent document (Ahuja & Katila, 2001; Harrigan, Di Guardo, Marku, & Velez, 2017). A better knowledge of a firm's technology positioning within a broader scenario help is helpful to set strategies and to be prepared for competition (Marku, Castriotta, Di Guardo, & Loi, 2019).

Looking at patents as Big Data, the five distinct dimensions mentioned above clearly emerge. First, the Volume dimension is captured by the vast number of patent documents available worldwide; patents are massively structured data organized into searchable databases (Davidson, 2019). For instance, Derwent Innovation (DWPI) database encompasses 70 million patents. Also, the breadth of different content is worthy of mentioning. Generally, patent documents include information on assignees/applicants; inventors; technological classification(s) according to the specific authority; backward and forward citations; claims; the invention's title, abstract, detailed description, and drawings; among other insightful facets. The richness of such information is further enhanced by the construction of indicators from merging and fusing different variables. A recent study of Harrigan and colleagues (2017) combined patent classification codes and citations to operationalize patent originality.

Second, the degree of Variety of patents can be detected via different sources. Patents, as territorial rights, might be extended to several authorities (i.e., U.S. Patent Office, European Patent Office, Japanese Patent Office). Each extension provides additional meaningful information regarding the same invention. Investigating the family of inventions allows tracking a firm's technology strategy in a specific timeframe.

Third, regarding Velocity, the speed of updating databases is currently very high; for instance, DWPI keeps posted every 3/4 working days (approximately). Most data are likely to be entirely in real-time. Patent Offices provide visualization tools that foster the analysis of patent data, further increasing the speed of utilization of the information.

Fourth, when considering the Veracity dimension of patent data, once patents are granted, they have very high reliability as issued by Authorities (Patent Offices). Nevertheless, looking at the specific information included in patents, some degree of lack of Veracity can be highlighted. For example, patent backward citations are often used as a measure of a firm's knowledge flow; however, many of these citations are added by examiners (Alcácer, Gittelman, & Sampat, 2009; Jaffe & De Rassenfosse, 2019). Furthermore, Patent Offices have been found to prefer citing their patents instead of those issued by other authorities (Belenzon & Pattacconi, 2013). Patent Offices have

been observed to be understaffed. When this occurs, patent examiners tend to devote less time to each patent claim, reducing the quality and Veracity of data (Guellec & Van Pottelsberghe de la Potterie, 2007; Belenzon & Pattacconi, 2013).

Fifth, the patent's Value is one of the core dimensions that clearly emerges from patent literature. Patents are commonly conceived as the output of the R&D efforts and a tool for firms to appropriate their R&D investments and remark their technology positioning and, subsequently, their positioning in the market (Gambardella, 2013; Harrigan et al., 2018).

All these dimensions define the traits of a patent as Big Data and highlight how firms can effectively use this information in their decision-making process. In this vein, firms are continually challenged by specific business drivers, such as improving the overall competitive positioning, find technology partners, invest in promising technology. The identification of these business drivers represents the first step of patent intelligence. This phase, typically, leads to questions such as: who the competitors are, where are the new technologies being developed, or what are the emerging trends of a particular technology, among other questions. In this scenario, patent analytics is a tool able to turn these business questions into calculation and visualization of several combinations of "Who", "What", "When", "Where", or "How" (Davidson, 2019). Instead, patent intelligence allows the transformation of patent content into insightful information for informed decision-making.

In summary, patents as Big Data permit the adoption of patent analytics tools to understand better how and to what extent the different dimensions of Big Data can offer insights, for instance, regarding the evolutionary patterns of a technology field. Enablers such as analytics software and visualization tools further enhance the use of patent intelligence for both industrial and business purposes.

3. Methodology

3.1 Sample and Data

In this paper, we choose to examine the IoT field as a key emerging paradigm expected to revolutionize our society (Baldini, Botterman, Neisse, & Tallacchini, 2016; Lu, Papagiannidis, & Alamanos, 2018). The central concept behind IoT regards the generation of a smart world in which anything and anyone is connected at any time and any space. The IoT domain encompasses several industries, chemistry, life sciences, and engineering technology. This high breadth of application makes IoT particularly salient for investigation.

We used the Derwent World Patent Index database (Note 1) (DWPI) to collect patents that belong to the IoT field. More specifically, we selected all European patent applications and patents granted from 2011 to 2019; DWPI is broadly used in management studies (Harrigan et al., 2017; Di Guardo et al., 2019). To delineate the boundaries of the IoT field, we used the DWPI manual codes (see Table 1 in Appendix A). This procedure led to a sample of 170,279 patents. We used the information included in the DWPI abstracts of these patents to examine the industry's technology conceptual structure.

For a better understanding of the conceptual structure of the IoT field, we analyzed two different time-periods: the nine years window between 2011 and 2019; and the last three-year window between 2017 and 2019. This procedure allowed us to explore the industry, not in static conditions but with a dynamic approach, looking at the distinct features that were driving the whole industry throughout the long period and the acceleration of the last three years.

While focusing on the selection process of the concepts mapped, we follow several steps. Considering the largest timeframe 2011-2019, a total of 961,857 different words was used for the patent analytics process. Following prior literature, we applied a threshold of (at least) 250 links between concepts (Small, 1973). This phase aimed at reducing noise on our data and providing evidence of the most important items in the field. Our final sample of visualized items includes 1,286 items and 530,057 links.

Finally, regarding the time span from 2017 to 2019, the abstract of 73,674 patents represented the input of the patent analytics procedure. This content encompasses 568,596 different terms. Even in this case, we used a threshold of 250 links between concepts (Small, 1973). This step led to a sample of 594 concepts and 150,507 total links.

3.2 Patent Co-word Analysis

In this paper, we adopt a co-word analysis as a patent intelligence tool to investigate the structure of the IoT field. More specifically, co-word analysis is a scientometrics instrument commonly used in the management field to disentangle the conceptual profile of science (Ronda-Pupo & Guerras-Martin, 2012; Benavides-Velasco, Quintana-García, & Guzmán-Parra, 2013; Castriotta, Loi, Marku, & Naitana, 2019). When applied to patent documents, the

basic assumption of this method relies on the fact that abstracts, given their brevity, can capture core concepts included in a document (Callon, Courtial, Turner, & Bauin, 1983). Therefore, co-word analysis can be a suitable tool for the investigation of scientific constructs included in patents.

In particular, the co-word technique combines science mapping and performance analysis (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011). The technology conceptual structure emerges by linking different concepts (Small, 1999). These links allow us to disentangle the connections not only between items but also between groups, signaling the centrality of items within and between clusters (Noyons, Moed, & Van Raan, 1999; Castriotta & Di Guardo 2016).

Finally, the degree of connections between items is assessed using the co-occurrences of two words in the same abstract (Cambrosio, Limoges, Courtial, & Laville, 1993). A high number of co-occurrences captures the degree of similarity within contents applied to a specific technology field (Ding, Chowdhury, & Foo, 2011).

3.3 Patent Mapping and Visualization Software

This study introduces a new tool for patent intelligence to investigate the emergence and evolution of a technology field, the VOSviewer software. The latter is a validated tool in management studies used to map and visualize science (Zupic & Čater, 2015; Castriotta et al., 2019). It combines two mapping techniques, namely, clustering and multidimensional scaling (MDS). For this reason, it is widely considered a valuable alternative to other statistical software (Waltman, Van Eck, & Noyons, 2010). In particular, cluster analysis is commonly used to identify groups and subgroups in a science or technology domain, while MDS generates a map in a low-dimensional space and finds a structure in a set of proximity measures between items (Kruskal, 1977). Concepts with the highest degree of similarity will be positioned closer in the map (Leydesdorff & Vaughan 2006).

Moreover, the VOSviewer algorithm is used to compute the distance between the different concepts included in the patents' abstracts according to the degree of similarity between nodes. In this way, it allows a more accurate result investigation (Van Eck & Waltman 2007). More specifically, VOSviewer methodology accounts for the association strength s_{ij} between items (concepts) that can be synthesized as follows:

$$s_{ij} = \frac{c_{ij}}{w_i w_j} \quad (1)$$

where c_{ij} refers to the number of co-occurrences of items i and j , while w_i and w_j concern to either the total number of occurrences of items i and j or to the total number of co-occurrences of these items. Finally, VOSviewer has been adopted as a network visualization tool to identify the links between items and clusters better and, therefore, to enhance the interpretation of the relationship between technology concepts and groups (Di Stefano, Gambardella, & Verona, 2012; Van Eck & Waltman, 2017).

4. Discussion of Results

4.1 Technology Conceptual Structure: 2011-2019

Figure 1 depicts the conceptual structure of the IoT field from 2011-2019 and the groups that emerge from the cluster analysis. More specifically, nodes are represented by the technological concepts (words) used in the patent document; core technology-related concepts are identified by nodes with a big size. Additionally, thick ties highlight the existence of a strong connection in terms of co-occurrences of the two words within a patent's abstract. Nodes are grouped into different clusters according to their degree of similarity.

Also, the positioning of the clusters on the maps is logic-wise. A cluster positioned closer to the center of the map plays the focal role in the IoT field. On the analogy, clusters surrounding the central cluster as on the map either have supportive nature or play a more independent role in the IoT industry.

Cluster analysis allows the identification of four groups that encompass the IoT industry patents' structure for the nine years' time-period:

- Cluster 1 - Core Technologies (Yellow);
- Cluster 2 - Supportive Infrastructure (Red);
- Cluster 3 - Applications (Blue);
- Cluster 4 - Communication and Transmission (Green).

Figure 1 introduces the overall overview of the industry characterized by four well-defined clusters with a few overlaps between concepts evidencing a distinct role of each group. However, it is also evident that two single items, namely, "text" and "phone," form central nodes in the very heart of the map. Core Technologies cluster is placed at the very center of the map. Even though the cluster is not large in terms of the number of items presented (23 concepts), it has co-occurrences with all the other three clusters. Moreover, these other clusters are linked with

each other through the Core Technologies cluster, which works as a connection hub. The map has a prolate form. The distortion is mostly caused by the Transmission/Communication cluster (366 items). Thus, the right side of the map is totally dedicated to the Transmission and Communication cluster. In contrast with Applications and Supportive Infrastructure clusters that are more compact, the Transmission/Communication group is stretched out on the side of the map. The cluster Applications (410 items) is placed on the left upper side of the map. The items within the cluster are equally spread and have quite equal numbers of links between each other, with some items placed on the very periphery of the cluster. The bottom downside of the map is presented by the Supportive Infrastructure cluster (486 items), which has small nodes connected more strongly with each other than other items in the very center.

The large-scale map introduces the main technological drivers of the industry or the so-called core technologies driving the whole IoT industry, as well as niche technologies influencing more the IoT industry performance and further development. Thus, although being dramatically smaller in terms of items included (23 over more than 360 items in other clusters), the Core Technologies cluster is positioned in the core of the map. It has a logical explanation; all items listed there were at the origins of the industry and triggered industry emergence. Indeed, “Bluetooth”, “Personal Assistant Device”, “GPS”, “cell phone” were the enablers of the industry, while three other surrounding clusters played either a supportive role or are related to the concrete application areas of IoT.

Applications cluster deals with niche applications of IoT and devices serving different applications. Specifically, items put closer to the center are related to serving devices, such as “control apparatus”, “speaker”, “field device”, “sensor”, while receding from the center of the cluster, and the map are positioned functions that IoT could perform, namely “control”, “testing”, “analyze”. Moreover, items placed on the periphery of the cluster and the whole map are related to the specific use in different areas and industries, such as transport (“vehicle”, “movement”, “truck”, “bus”), energy (“energy”), environment (“temperature”, “water”), medicine (“hospital”, “patient”) and other.

In addition, the Supportive Infrastructure group includes items that are strongly related to technologies and devices needed to support the functioning of the IoT networks in general. These technologies, being not vital for the whole IoT industry, however, play an important role in enabling the efficient performance of IoT networks disregarding applications. Namely, “computing system”, “database”, “storage”, “file”, “management system” are at the core of the cluster or closer to the map center, while items positioned near to the periphery sides of the cluster are mostly niche-related items (i.e., “sms”, “content delivery network”, “credit card”, “kiosk”, “disk”).

Transmission and Communication cluster groups items connected with information transmission, and it is stretched out on the side of the map. On the one hand, this “horizontal” dynamics trend of distinct from each other items signalize about the immature nature of the cluster, when technologies are variable, niche, and not standardized. Exemplarily, items that are positioned on the extreme periphery of the Transmission and Communication cluster are related to specific niche transaction or communication protocols and are strongly case- and technology-related. On the other hand, as long as there is no standard yet in transmitting information, there are market opportunities for newcomers willing to enter this field of the IoT industry.

Finally, specific items, namely, “phone” and “text”, related to the two different clusters, Supporting Infrastructure and Applications, respectively, are positioned at the very heart of the map, visually forming a prominent focus-zone. It is explained by the number of co-occurrences of the “phone” and “text” not only within their neighboring clusters but also with items within other clusters. It can signal the exceptional importance of the two single items for the whole industry. Indeed, “phone” and “text” were the key features of IoT network functioning and information transmission.

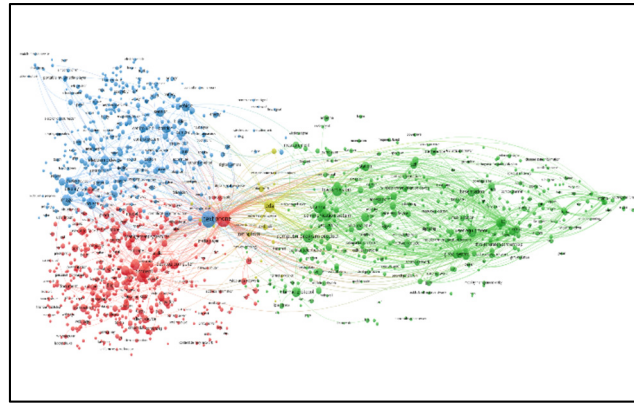


Figure 1. The conceptual structure of the IoT domain 2011-2019

4.2 Technology Conceptual Structure: 2017-2019

The map depicted in Figure 2 visualizes the technological concepts included in patent documents during the last time span 2017 – 2019. All items are grouped into three clusters listed below:

- Cluster 1: User Supportive Infrastructure and Functions (Red)
- Cluster 2: Solutions and Devices (Blue)
- Cluster 3: Transmission and Communication (Green)

The map 2017-2019 is characterized by three well-defined clusters. Three items form very prominent nodes on the map and play the role of technology core, namely, “text”, “user”, “personal digital assistant”. These three nodes serve as hubs grouping the number of items around them and are very strongly positioned within their neighboring clusters, and “text” also with links with all three clusters and is positioned therefore at the very core of the map. The left upper side of the map is represented by Solutions and Devices cluster (132 items) which has several focus-points of items positioned closer to the center of the map and one group of items placed on the very periphery of the cluster and map. The left bottom side of the map is presented by the User Supportive Infrastructure and Functions cluster (249 items), which has two nodes’ hubs distinct from each other, one placed at the core of the cluster, another placed in the core of the map. Finally, the Transmission and Communication (213 items) cluster is positioned on the right side of the map, and as on the 2011-2019 map is stretched on the side, causing the prolate form distortion in the shape of the map. This cluster is sub-divided into two visual zones. One is closer to the center, while the other is closer to the right side of the map. However, these two groups are interconnected.

Also, the conceptual structure (Figure 2) reflects the major changes happening in the IoT industry recently. Firstly, the main changes are connected with the new shape of the IoT industry map. Thus, while no key emerging technology themes appear on the map 2011-2019, the 2017-2019 map representing a three-years scenario has a well-pronounced consolidated or nodes structure. This can be explained in terms of the maturity of IoT technology. The emerging key themes presented in terms of new nodes within clusters signalize the more developed nature of the industry, where there are the key technologies or solutions that drive the IoT industry. Moreover, the emerging themes, although being presented on the general ten-years landscape map, are prominent only in the three-years scenario map. It is explained in terms of the emerging nature of the key themes. However, if the predominance of the emerging themes continues, their role in the next ten-years scenario will grow. Additionally, we found that the clusters are focused on users and not purely on the technologies *per se* as on the 2011-2019 map.

Thus, items from existing previously Core Technologies cluster (Figure 1) have migrated to the new Solutions and Devices cluster. The role of the core technologies previously forming the whole industry has been reduced. There are now several prominent groups of items within the core group around the “personal digital assistant” item. The close look at the Solutions and Devices cluster allowed us to distinguish four main characteristics. First, single disorganized technologies have transformed into solutions – ready-to-use devices to manage the new IoT era (i.e., robot, speaker, workstation, smartwatch, personal digital assistant). Second, the focus within the cluster moved from technologies towards devices that are now the new core of the IoT industry (“personal digital assistant”, “smartphone”, “sensor”, “electronic device”). Third, there are more items that relate to information gathering and processing, and transaction commands recipients (i.e., sensor, camera, microphone, portable media player, refrigerator, smart T.V., television). Finally, it is quite surprising that the cluster has only several items reflecting

the connection with a specific context (“housing”, “air”, “train”). This dynamic might be connected with the fact that the industry is focused more now on standardized products, rather than single use-cases solutions. Instead, the Solutions and Devices cluster now lists general-purpose devices with “electronic devices” being the prominent item. Overall, the renewed cluster signalizes the maturation process happening in the industry, as well as the emerging focus on developing ready-to-use solutions instead of just technologies. Moreover, this cluster directly indicates the rapid digitalization happening in all sectors of humans’ life, where all devices can be connected and used to gather data.

Special attention attracts the User Supportive Infrastructure and Functions cluster. It has the “user” item as the core. This new item placed in the center for the cluster drives the new dynamics of the industry. It is under the technology diffusion theory that technologies shift their attention from technology advancement towards market needs and users at a more mature degree of development. Thus, the User Supportive Infrastructure and Functions cluster gathers items that are responsible for performing or supporting specific functions of IoT and are strongly connected with “user”, and its needs in save transaction (“Security”, “attacks”, “certificate”, “public key”, “encryption key” items), comfortable storage (“storage”, “file”, “document”, “server device” items), system maintenance (“maintenance”, “identification information” items), “presentation”, “calculation”, “user interface”, and other.

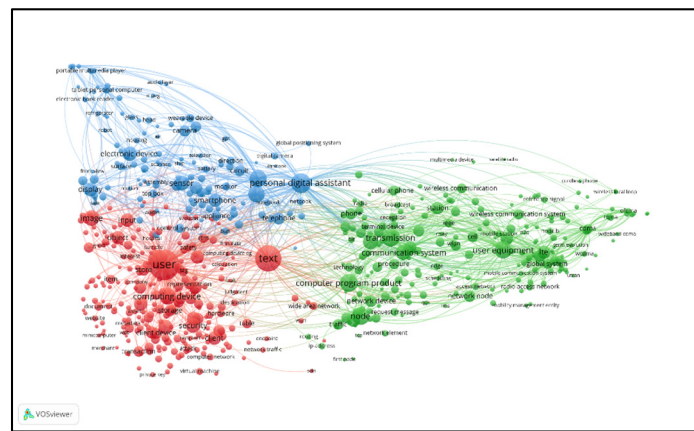


Figure 2. The conceptual structure of the IoT domain 2017-2019

The Transmission and Communication is the only remained cluster from the 2011-2019 map (Figure 1); however, it has changed. Previously single transmission items form now two prominent focus-points on the map. First, well-developed transmission technologies and general items (i.e., “access point”, “broadcast”, “wireless device”, “phone”, etc.) are centralized around two key items of the cluster, namely, “transmission” and “communication system”. Second, item “user equipment” is grouped with complex solutions aimed at ease information transmission on long distances, globally (i.e., “global system”, “universal territorial radio”, “node”, “base station”, “mobile station”, “worldwide interoperability”, “data rate”). Finally, niche technologies are still placed at the very periphery of the cluster. The newly discovered logics of items positioning within the Transmission and Communication cluster shows that the user is at the core of the transmission technologies too. The focus from advancing technologies is shifting towards building solutions that will solve concrete needs to ease the use, standardize information transmission globally.

5. Conclusion

In this paper, we adopted a patent analytics approach to better understand how and to what extent the different dimensions of Big Data can offer insights on technology evolution. We explored the conceptual structure of an emerging technology such as the IoT and introduced a new patent intelligence tool for mapping and visualizing an industry’s structure, evolution, and trends. Our co-words analysis of 170,279 European patents allows us to disentangle the IoT industry and highlight boundary and core concepts to the field.

Our results showed that during the timeframe 2011-2019, the IoT conceptual structure consists of four groups: one core cluster and three surrounding clusters reflecting the IoT industry’s developing nature during these years. Indeed, the situation when there are only a few key basic technologies forming the industry and plenty of niche technologies is characteristic for young growth industries when the prime focus is on technology and its advancements. In the last window, from 2017 to 2019, we found that the IoT domain has experienced a shift from

technology advancements and technology per se towards market needs and ready-to-use solutions. Thus, although the 2017-2019 map does not have a cluster at the center of the map, it has two vibrant focus-items, namely, “user” and “personal digital assistant” belonging to the User Supportive Infrastructure and Functions and Solutions and Devices clusters, respectively.

The present work contributes to the existing literature in several ways. First, it is the earliest study that applies a decomposition framework to clarify the determinants of IoT technology invention, providing new insight into the use of patents as Big Data. Second, this paper offers a valuable contribution to the technology innovation literature by shedding light on the evolutionary dynamics of an emerging paradigm; this can be applied to various other technology industries serving as a starting point for more general models. Third, by introducing a novel tool for mapping the co-occurrences of technology concepts included in patents’ abstracts (the VOSviewer software), this research offers a methodological contribution that might serve as a basis for researchers and practitioners for detecting core and emerging technologies using patent as Big Data.

Managerial implications can also be highlighted. This study mapped and visualized the conceptual structure of the IoT domain, providing practitioners with critical insights about the core and frontier technologies and the degree of centrality of items in a more recent time window. This can be of help, especially when comparing an industry landscape with the firms one, to investigate a firm’s competitive positioning in the broader scenario.

Despite the contributions mentioned above, this work presents several limitations. This study examines patents to map a field’s conceptual structure. Nevertheless, not all inventions are patented as firms might choose to rely on secrecy. To mitigate this effect, we examined an industry with a high intensity of patenting activity. Finally, we built the co-occurrences matrix using the general abstracts of patents. Future research might focus on additional specific facets, such as the abstract of technology novelty, activity, use, advantage, and technology focus.

References

- Ahuja, G., & Katila, R. (2001). Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study. *Strategic Management Journal*, 22(3), 197-220.
- Alcácer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in U.S. patents: An overview and analysis. *Research Policy*, 38(2), 415-427.
- Baldini, G., Botterman, M., Neisse, R., & Tallacchini, M. (2018). Ethical design in the internet of things. *Science and Engineering Ethics*, 24(3), 905-925.
- Belzon, S., & Pataconi, A. (2013). Innovation and firm value: An investigation of the changing role of patents, 1985–2007. *Research Policy*, 42(8), 1496-1510
- Benavides-Velasco, C. A., Quintana-García, C., & Guzmán-Parra, V. F. (2013). Trends in family business research. *Small Business Economics*, 40, 41-57.
- Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22, 191-225.
- Cambrosio, A., Limoges, C., Courtial, J. P., & Laville, F. (1993). Historical scientometrics? Mapping over 70 years of biological safety research with cword analysis. *Scientometrics*, 27, 119-143.
- Castriotta, M., & Di Guardo, M. C. (2016). Disentangling the automotive technology structure: a patent co-citation analysis. *Scientometrics*, 107(2), 819-837.
- Castriotta, M., Loi, M., Marku, E., & Naitana, L. (2019). What’s in a name? Exploring the conceptual structure of emerging organizations. *Scientometrics*, 118(2), 407-437.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165-1188.
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Infometrics*, 5, 146-166.
- Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, 30(1), 44-57.
- Davidson, J. M. (2019). Improved Business Decisions Through Patent Intelligence. *Advanced Materials & Processes*, 177(8), 16-21.
- Di Guardo, M. C., Harrigan, K. R., & Marku, E. (2019). M&A and diversification strategies: what effect on quality

- of inventive activity? *Journal of Management and Governance*, 23(3), 669-692.
- Di Stefano, G., Gambardella, A., & Verona, G. (2012). Technology push and demand pull perspectives in innovation studies: Current findings and future research directions. *Research Policy*, 41, 1283-1295.
- Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management*, 37, 817-842.
- Engelsman, E. C., & van Raan, A. F. (1994). A patent-based cartography of technology. *Research Policy*, 23(1), 1-26.
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897-904.
- Frizzo-Barker, J., Chow-White, P. A., Mozafari, M., & Ha, D. (2016). An empirical study of the rise of big data in business scholarship. *International Journal of Information Management*, 36(3), 403-413.
- Gambardella, A. (2013). The economic value of patented inventions: Thoughts and some open questions. *International Journal of Industrial Organization*, 31(5), 626-633.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
- George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, 57(2), 321-326.
- Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems*, 27(1), 101-113.
- Guellec, D., & Van Pottelsberghe de La Potterie, B. V. P. (2007). *The economics of the European patent system: I.P. policy for innovation and competition*. Oxford University Press on Demand.
- Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, 26(3), 191-209.
- Hagedoorn, J., & Cloudt, M. (2003). Measuring innovative performance: is there an advantage in using multiple indicators?. *Research Policy*, 32(8), 1365-1379.
- Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools (No. w8498). *National Bureau of Economic Research*. Retrieved from <https://www.nber.org/papers/w8498.pdf>
- Harrigan, K. R., Di Guardo, M. C., & Marku, E. (2018). Patent value and the Tobin's q ratio in media services. *The Journal of Technology Transfer*, 43(1), 1-19.
- Harrigan, K. R., Di Guardo, M. C., Marku, E., & Velez, B. N. (2017). Using a distance measure to operationalise patent originality. *Technology Analysis & Strategic Management*, 29(9), 988-1001.
- Jaffe, A. B., & De Rassenfosse, G. (2019). Patent citation data in social science research: Overview and best practices. In *Research Handbook on the Economics of Intellectual Property Law* (pp. 20-46). Edward Elgar Publishing.
- Jin, J., Liu, Y., Ji, P., & Liu, H. (2016). Understanding big consumer opinion data for market-driven product design. *International Journal of Production Research*, 54(10), 3019-3041.
- Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1). <https://doi.org/10.1177/2053951716631130>
- Kruskal, J. (1977). The relationship between multidimensional scaling & clustering. In *Classification and Clustering: Proceedings of an Advanced Seminar Conducted by the Mathematics Research Center* (pp. 17-44). Wisconsin, Madison.
- Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387-394.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6(70), 1.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 21-32.
- Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293-303.

- Lee, W. S., Han, E. J., & Sohn, S. Y. (2015). Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. *Technological Forecasting and Social Change*, *100*, 317-329.
- Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: extending ACA to the web environment. *Journal of the American Society for Information Science and Technology*, *57*, 1616-1628.
- Lu, Y., Papagiannidis, S., & Alamanos, E. (2018). Internet of Things: A systematic review of the business literature from the user and organisational perspectives. *Technological Forecasting and Social Change*, *136*, 285-297.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition*. Washington, DC: McKinsey Global Institute.
- Markham, S. K., Kowolenko, M., & Michaelis, T. L. (2015). Unstructured text analytics to support new product development decisions. *Research-Technology Management*, *58*(2), 30-39.
- Marku, E., Castriotta, M., Di Guardo, M. C., & Loi, M. (2019). Mapping Innovation in the Digital Transformation Era: The Role of Technology Convergence. In *Business Transformations in the Era of Digitalization* (pp. 160-178). IGI Global.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60-68.
- Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing innovation management research in a digital world. *MIS Quarterly*, *41*(1), 223-238.
- Noyons, E., Moed, H., & Van Raan, A. (1999). Integrating research performance analysis and science mapping. *Scientometrics*, *46*(3), 591-604.
- O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. (2015). Big data in manufacturing: a systematic mapping study. *Journal of Big Data*, *2*(1), 20. <https://doi.org/10.1186/s40537-015-0028-x>
- Ponta, L., Puliga, G., Oneto, L., & Manzini, R. (2019). Innovation Capability of Firms: A Big Data Approach with Patents. In *INNS Big Data and Deep Learning conference* (pp. 169-179). Springer, Cham.
- Ronda-Pupo, G. A., & Guerras-Martin, L. Á. (2012). Dynamics of the evolution of the strategy concept 1962–2008: A co-word analysis. *Strategic Management Journal*, *33*, 162-188.
- Sarica, S., Yan, B., & Luo, J. (2019). Data-Driven Intelligence on Innovation and Competition: Patent Overlay Network Visualization and Analytics. *Information Systems Management*, *37*(3), 198-212.
- Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: The real-world use of big data. *IBM Global Business Services*, *12*, 1-20.
- Simsek, Z., Vaara, E., Paruchuri, S., Nadkarni, S., & Shaw, J.D. (2019). New Ways of Seeing Big Data. From the Editors. *Academy of Management Journal*, *62* (4), 971-978.
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263-286.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265-269.
- Small, H. (1999). Visualizing science by citation mapping. *Journal of the Association for Information Science and Technology*, *50*, 799-813.
- Stockstrom, C. S., Goduscheit, R. C., Lüthje, C., & Jørgensen, J. H. (2016). Identifying valuable users as informants for innovation processes: Comparing the search efficiency of pyramiding and screening. *Research Policy*, *45*(2), 507-516.
- Sukumar, S. R., & Ferrell, R. K. (2013). 'Big Data' collaboration: Exploring, recording and sharing enterprise knowledge. *Information Services & Use*, *33*(3-4), 257-270.
- Svahn, F., Mathiassen, L., & Lindgren, R. (2017). Embracing Digital Innovation in Incumbent Firms: How Volvo Cars Managed Competing Concerns. *MIS Quarterly*, *41*(1), 239-254.
- Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, 172-187.

- Troilo, G., De Luca, L. M., & Guenzi, P. (2017). Linking data-rich environments with service innovation in incumbent firms: A conceptual framework and research propositions. *Journal of Product Innovation Management, 34*(5), 617-639.
- Uddin, M. F., & Gupta, N. (2014). Seven V's of Big Data understanding Big Data to extract value. In *Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education* (pp. 1-5).
- Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. *Technovation, 84*, 21-36.
- Urbinati, A., Chiaroni, D., Chiesa, V., & Frattini, F. (2018). The role of digital technologies in open innovation processes: an exploratory multiple case study analysis. *R&D Management, 50*(1), 136-160.
- Van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNet Explorer and VOSviewer. *Scientometrics, 111*, 1053-1070.
- Van Eck, N. J., & Waltman, L. (2007). Bibliometric mapping of the computational intelligence field. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15*(05), 625-645.
- Van Eck, N. J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology, 61*(12), 2405-2416.
- Van Eck, N. J., Waltman, L., Den Berg, J., & Kaymak, U. (2006). Visualizing the computational intelligence field [Application Notes]. *Computational Intelligence Magazine, IEEE, 1*(4), 6-10.
- Waaaijer, C. J., van Bochove, C. A., & Van Eck, N. J. (2010). Journal editorials give indication of driving science issues. *Nature, 463*(7278), 157-157.
- Waltman, L., van Eck, N. J., & Noyons, E. C. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics, 4*, 629-635.
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356-365.
- Wan, F., Williamson, P. J., & Yin, E. (2015). Antecedents and implications of disruptive innovation: Evidence from China. *Technovation, 39*, 94-104.
- Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change, 126*, 3-13.
- Wessel, M. (2016). You Don't Need Big Data—You Need the Right Data. Harvard Business Review. Retrieved from <https://hbr.org/2016/11/you-dont-need-big-data-you-need-the-right-data>
- Xiaofeng, M., & Xiang, C. (2013). Big data management: concepts, techniques and challenges. *Journal of Computer Research and Development, 1*(98), 146-169.
- Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429-472.

Note

- Note 1. "Derwent World Patents Index™ is the world's most comprehensive database of enhanced patent information. Experts from Derwent's editorial team analyze abstracts and manually index patent records.

Appendix A

Table 1. List of DWPI Manual codes of the IoT domain used for patent search.

1.	Connectivity (T01-N)	
A.	Data Transmission	T01-N01D T01-N01D1 T01-N01D1A T01-N01D1B T01-N01D2 T01-N01D3 T01-N01D3A W05-D06E1 W05-D06E1A W05-D06E1C
B.	Network Control	T01-N02A T01-N02A1 T01-N02A1A T01-N02A1B T01-N02A2 T01-N02A2A T01-N02A2B T01-N02A2C T01-N02A2D T01-N02A2E T01-N02A2X T01-N02A3 T01-N02A3A T01-N02A3B T01-N02A3C T01-N02B T01-N02B1
2.	General (T01-M02A1)	T01-M02A1A T01-M02A1B T01-M02A1C T01-N01F W05-D06E1 W05-D07
3.	Home	
A.	Home Automation – General	
•	Control	X27-V
•	Data Transmission	W05-D07A
•	Tv Equipment	W03-A16C5K
B.	Kitchen Appliances (X27-B)	
•	Intelligent Refrigerator	X27-F05
C.	Lighting	
•	Remote-Controlled Switching	X26-C03C
D.	Robots – Domestic Assistance (X27-U)	
E.	Smart Grids X12-H08	
F.	Utility Meter Reading / Measurement	S01-B01 X12-H03E7 X12-H04 X12-H04A

	X12-H04B
	W05-D07G
4. INDUSTRIAL	
A. Agricultural/Farming (W05-D07N)	
B. Building Control (W05-D07C)	
C. Earth Drilling and Well Logging (W05-D07H)	
D. Factory Automation (W05-D07B)	
E. Office Automation (W05-D07A)	
F. Power Generation and Distribution (W05-D07F)	
G. Smart Grids (X12-H08)	
H. Utility Meter Reading / Measurement	S01-B01
	X12-H03E7
	X12-H04
	X12-H04A
	X12-H04B
	W05-D07G
5. Medical	
A. Medical Systems/Equipment	
• Information Systems	T01-N01E1
• Data Transmission	W05-D07M
• Telediagnosis	S05-D06A
6. Network Security	
A. Data Transmission	W05-D05B5A
	W05-D05B5C
	W05-D05B5E
B. System/Network Monitoring	T01-N02B2
	T01-N02B2A
	T01-N02B2B
	T01-N02B2C
	T01-N02B3
C. User Authentication/Control	T01-N02B1A
	T01-N02B1B
	T01-N02B1C
	T01-N02B1D
	T01-N02B1E
	T01-N02B1F
	T01-N02B1G
	T01-N02B1H
7. Vehicles	
A. Control	X21-A01L
	X22-P15
B. Data Transmission	W05-D07D
	X21-K
	X22-K08

Source: Derwent World Patent Index (DWPI)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).