Data-Oriented Software Development: The Industrial Landscape through Patent Analysis

Konstantinos Georgiou [1], Nikolaos Mittas [2,*], Apostolos Ampatzoglou [3], Alexander Chatzigeorgiou [3] and Lefteris Angelis [1]

1   School of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
2   Department of Chemistry, International Hellenic University, 65404 Kavala, Greece
3   Department of Applied Informatics, University of Macedonia, 54636 Thessaloniki, Greece
*   Correspondence: nmittas@chem.ihu.gr

Abstract: The large amounts of information produced daily by organizations and enterprises have led to the development of specialized software that can process high volumes of data. Given that the technologies and methodologies used to develop software are constantly changing, offering significant market opportunities, organizations turn to patenting their inventions to secure their ownership as well as their commercial exploitation. In this study, we investigate the landscape of data-oriented software development via the collection and analysis of information extracted from patents. To this regard, we made use of advanced statistical and machine learning approaches, namely Latent Dirichlet Allocation and Brokerage Analysis for the identification of technological trends and thematic axes related to software development patent activity dedicated to data processing and data management processes. Our findings reveal that high-profile countries and organizations are engaging in patent granting, while the main thematic circles found in the retrieved patent data revolve around data updates, integration, version control and software deployment. The results indicate that patent grants in this technological domain are expected to continue their increasing trend in the following years, given that technologies evolve and the need for efficient data processing becomes even more present.

1. Introduction

It is an undeniable statement that one of the key aspects of the modern world and its financial, business and industrial services are data and the valuable knowledge that they provide. As society strives to become technologically adept and interconnected, the production, processing and management of data have become a focal point of businesses that adapt their operations in order to function in this data-driven environment [1–3]. Of course, these efforts would not be feasible without specialized software tools that are developed to facilitate this process. Through this perspective, Data-Oriented Software Development (DOSD) is equally important for organizations that wish to analyse significant volumes of data because the functionalities and opportunities that they offer in data processing can accelerate business growth [4–6]. Particularly due to the Industry 4.0 [7] and Industry 5.0 [8] movements, data and software are treated as two intertwined entities that complement each other [9,10]. Thus, the analysis of these entities can serve as a reflection of industrial evolution and further highlight the importance of DOSD in this area. While there are several studies that discuss this concept based on empirical data [11–14], an important, but yet neglected, source of information is industrial-granted patents.

To this regard, patents are an indispensable part of the industrial community, as they comprise a secure way of establishing the creation and ownership of a property [15]. Individuals and organizations on a global scale strive to secure a patent grant that will

allow them to own the commercial and scientific rights, enabling them to economically exploit their owned patent [16,17]. Over the years, patents have reportedly been granted from fields such as medical sciences [18–21], engineering [22,23], computer science [24–26] and many other scientific domains. The increase in patent activity is not a surprising phenomenon, as patents are research indicators that highlight research development and activity [27–30].

Due to the rapid development of new technologies, which in turn affects the methodologies and objectives of patents [31,32], the innovation potential and promising practices in any domain are constantly evolving. Thus, the early discovery of innovative technologies and the forecasting of emerging trends ensure that the research and industrial communities adapt to the ever-changing technological environment and produce high-quality results. To that end, the analysis of patent data is a potent way of uncovering technological shifts and forecasting future technologies [33–36], as their objectives and methodologies capture potential developments [37–40].

Patents are an accepted and secure way of intellectual property, with rising popularity in the industrial world in the same way that DOSD is a prominent domain of computer science as a subdomain of software engineering (SE), with a wide acceptance in the scientific community and considerable research [41,42]. Evidently, as the volume of available information increases, data generation, collection, processing and consumption is performed through software. Simultaneously, new challenges arise in integrating existing and future technologies in the evolving SE and DOSD domains [43]. To tackle these challenges and to forecast future developments, evidence from patent analysis can facilitate in the timely detection of new methodologies and highlight technological trends [33,39,44,45].

Patent activity in SE and hence in DOSD has known an ever-increasing trajectory, from the dawning of the age of computers to the modern age of information and large-scale data processing [46]. In a sense, this highly present patent activity can be a potent measure of innovation and technological advancement [47,48].

To this regard, tracking the development and research value of patents can be a challenging task, without the existence of several patent offices that contribute to the organization and storing of patent data. Simultaneously, these agencies serve as the pillar for patent applications on a global scale. Some indicative examples of popular and established patent offices are the European Patent Office (EPO, Munich, Germany), the United States Patent and Trademark Office (USPTO, Alexandria, VA, USA) and the Korean Intellectual Property Office (KIPO, Daejeon, Republic of Korea). While these offices cover regional patent applications, they also accept patents on a global scale, with academics and organizations from multiple countries filing their patents to the office of their choice.

Based on the rapid increase in patent applications of DOSD, especially during the last twenty years, as well as the growing necessity for specialized software, the main motivation of the current study is to investigate the DOSD patent landscape with the aim of identifying technological development trends and innovation dynamics covering a period from the infancy of DOSD patent activity up to the rising age of the fourth industrial revolution, known as Industry 4.0. Our findings highlight the dynamic technological shifts in DOSD patent activity, providing a roadmap of development and innovation. In addition, the empirical evidence provided serves as a point of reference for technological convergence in the DOSD domain, bridging past practices with future prospective innovations.

Currently, there are software business suites available such as Orbit Intelligence, Derwent Innovation, PatSeer, AcclaimIP and others that perform similar tasks, either focusing on the legal aspects and litigating activities of patents or by analysing patent entries. However, these tools tend to mostly focus on business indicators and industrial growth metrics, hence having a larger effect on the business and economic landscapes. To the contrary, our study serves as a methodological framework and not a software tool and as a primary investigator of the technological landscape and the innovative technologies and practices encompassed by granted patents.

The remainder of the study is organized as follows. In Section 2, we provide some indicative related work in the field of patent analysis, while in Section 3, we present our research methodology. In Section 4, we discuss the results of our analysis, while in Section 5, we present possible threats to the validity of the study. Finally, Section 6 serves as discussion points for our results and conclusions.

2. Related Work

Research activity on patents is abundant, with a plethora of methodologies for analysis observed, frequently focusing on text mining and topic modelling. In this section, we present some indicative works that have a similar scope to this paper, highlighting their main points and merits. However, as our work is the first contained effort that covers data-related patents solely based on the DOSD sector, we present similar research conducted in adjacent domains (e.g., Artificial Intelligence, Blockchain).

Several studies perform exploratory research that aims to profile the main aspects of patents and visualize the data in engaging ways. Albino et al. [49] performed a geographical and technological assessment of software used in low-carbon energy projects, highlighting the prominent contributors and leading countries. Kang et al. [50] sought essential patents in the Korean and international markets and explored the correlation between the geographical distributions and essential patents. Kim et al. [51] used clusters of patent classes in order to predict their evolution and the main objectives they represent. Moehrle et al. [52] defined "technological speciation" as the emergence of specific technologies in patents and use textual patterns in camera related software to examine its validity.

The field of artificial intelligence (AI) has known a considerable increase in later years, and multiple works study the patents under this domain, while also attempting to predict future trends and technologies. Tseng and Ting [53] conducted a quality evaluation study, focusing on patent agencies from several countries and introducing several metrics to evaluate the innovation and quality of AI patents. Fujii and Managi [54] combined patent information from several offices and grouped AI patents based on their objectives. Their analysis indicates that AI patents focus on mathematical models and knowledge extraction. Several studies delve deeper into domains that rely on AI such as nanotechnology [55] and autonomous vehicles [56], utilizing bibliometrics, citation networks and data exploration to highlight the main characteristics of patents belonging to these fields. Finally, future trends are explored in the AI sector by providing classification schemas [45], focusing on cooperation networks between the organizations that file the patents [57] and by analysing interconnections between companies and technologies [58].

Another sector that has been studied under the scope of patents is augmented reality (AR), which includes virtual devices and environments. Choi et al. [59] exploited semantic patterns in augmented reality patents in order to provide directions for further innovation. They concluded that image rendering and processing are the most promising areas that require additional research and development. Jeong et al. [60] worked in a similar spirit and extracted topics from AR patents retrieved from USPTO. Their results have a high degree of agreement with Choi et al. [59], with topics relating to display techniques being highly dominant. Finally, Evangelista et al. [61] conducted a rich exploratory analysis, dividing AR patents into five classes and uncovering geographical and organizational trends.

Similar studies have been conducted in domains relevant with security, such as blockchain, where research is focused on forecasting technologies [62]. Daim et al. [63] utilized patent classes relevant to blockchain and the Internet of Things (IoT) and produced clusters of patent objectives and types, forecasting future needs. Wustmans et al. [64] combined patent data from USPTO and microtrends data from TRENDONE and used a hybrid methodology of semantic terms and topics to predict technological developments. Zhang et al. [65] used advanced Latent Dirichlet Allocation techniques to extract patent topics over different periods in order to provide a roadmap of thematic axes and to forecast the evolution of the blockchain domain.

The last domain is IoT, of which there are plenty of opportunities for research. Several works [66,67] exploit the citation networks of IoT patent families and technologies in order to construct clusters of patent communities that contain primary characteristics of IoT sectors. Similar research was performed by Mazlumi et al. [68], exploiting social network analysis and graph metrics on patent classes. Trappey et al. [69] provided a roadmap of assignees and patent classes in order to aid the manufacturing of products related to IoT patents and the logistic procedures. In another study [70], valuable manufacturing standards are provided, depending on the country, that facilitate the validation and granting of IoT patents, in the context of Industry 4.0. Moreover, some studies shed light on innovation and trends in the IoT industry by measuring consumer satisfaction [71], exploring temporal trends [72] or using bibliometrics and graph metrics [73].

3. Methodology

3.1. Research Questions

As discussed in the introductory section, the main objective of the present study is to explore the principal thematic trends and technological ventures in DOSD, leveraging patent data. Based on the available information of granted patents, it is clear that the general SE landscape is quite broad and, for our analysis to be meaningful, should be broken down in several objectives. Thus, in order to reflect our motivations in our research methodology, and in respect to the DOSD patent landscape, we define the following research questions and objectives:

[RQ$_1$] What is the landscape of DOSD patents?

As mentioned previously, the USPTO is utilized as a repository of patent knowledge and information. Patents are organised in a concise manner, with each entry containing various metadata relevant to the patent being filed. These metadata refer to several aspects of a patent lifecycle, including temporal characteristics such as the granting year, country characteristics, as well as information about the inventors and applicants of the patent. Thus, RQ$_1$ aims to conduct an exploratory analysis on selected metadata in order to extract meaningful conclusions about DOSD patent activity over the years. In particular, we explore: (a) how patent activity evolves over time; (b) how patents are geographically distributed; and (c) which are the most active patenting organizations.

[RQ$_2$] Which thematic trends can be traced in DOSD patents?

The technologies and themes of patents vary, depending on the type of DOSD and the objectives of patents. As DOSD is a broad field which is applied to multiple other domains, it is inevitable that the patents filed and granted under this domain will be of a multifaceted nature. In RQ$_2$, we conducted a thorough analysis, employing topic modelling techniques in order to uncover the thematic areas of the technological aspects that are being patented.

[RQ$_3$] How is technological innovation portrayed in the interconnection of DOSD patents?

The literature related to patent analysis highlights that patent citations are a valid source for determining the innovation of a patent. This means that a patent, which belongs to several classes that cite (or is cited by) other patents which also belong to other classes can serve as an indicator of technological domains that intersect and are interconnected in order to reflect the objectives of a patent. Thus, an analysis of patent and patent class citations can reveal which classes and patents drive the innovations in DOSD and which patent aspects are more isolated than others. In RQ$_3$, we construct a Patent Citation Network (PCN) and a Class Citation Network (CCN) utilizing the forward and backward patent citations, and we apply Brokerage Analysis (BA) [74,75] and establish network analysis methodologies to discover influential and hence innovative patents and classes.

To answer the aforementioned questions, the process presented in Figure 1 was applied. The outlined approach consists of four phases that are (i) data collection, (ii) data preprocessing, (iii) data analysis and (iv) extraction of results accompanied with discussion of the most important findings.
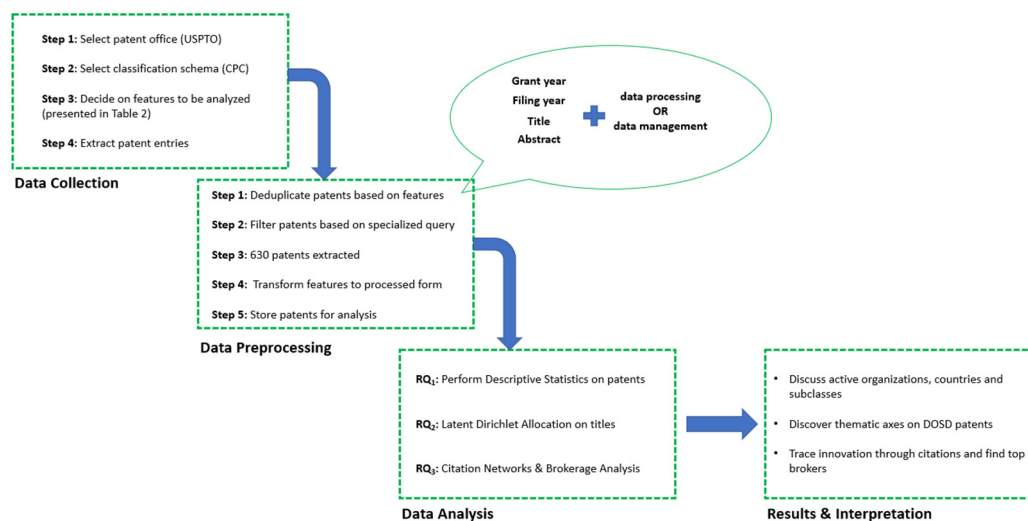
Figure 1. Methodology schema.

3.2. Patent Description

A patent is defined as "an intellectual property right granted for an invention in the technical field to a company, public organization, or individual by a national patent office, hence giving the owners the right to exclude others from the industrial exploitation of the patented invention for a defined number of years. The invention must be novel, non-obvious, adequately described, and claimed by the inventor in clear and definite terms" [49]. To this regard and in order to meet the goals of the current study, we made use of patent entries collected from a large patent office as the basic unit of analysis.

A patent entry is a semi-structured web document consisting of both textual content and metadata. In Figure 2, we provide an indicative example of a DOSD patent to demonstrate its main features and available metadata. More specifically, the title field comprises a brief description of the invention being patented and serves as an informative and short description in English. The abstract accompanies the title and is a part of the application submitted by the applicant, that gives a summary of the invention. The abstract can also contain the patent claims, as well as any helpful guidelines regarding the objectives of the patent application. The patent entry is matched by the examiners to one or more patent class, which associates the patent with a scientific domain. These classes belong to superclasses that encompass fields of technology or other disciples. Also contained in the patent entry are the inventors of the patent, which are cited by name, along with their country of origin. Similarly, the assignees of a patent, which can be either an organization, company or institution, are also included in the patent application. Finally, the granting year corresponds to the date that the patent was granted ownership by the USPTO. Each patent is characterized by a number of citations, that can be either patents that the patent cites (backward citations) during its application to the patent office or other patents where the patent is cited by (forward citations) during their applications. The citations concern not only the patents but also the classes that the patents belong to, in a sense that a patent that cites another patent also cites its corresponding classes.

3.3. Data Collection

The first phase of the approach is dedicated to the identification and retrieval of patent documents related to DOSD. To this regard, a key step is the selection of the patent office which can provide an extensive poll of patent data, thus improving the validity of our findings. To that end, the selected office for this study was the USPTO, due to the large volume of patent data stored in its databases [76]. This abundance of data can be attributed both to the global coverage of the USPTO in terms of patent applications, as well as to the leading position of the United States in the market of patent ownership and technology

forecasting [77]. Moreover, similar studies have praised the USPTO as a rich data source with minimal bias and increased patent citations [49]. Another important aspect of the USPTO is the division between the inventor of a patent, being the person that creates and develops a product, and the applicant of the patent which can either be the inventor or an organization to which the inventor is adherent to.
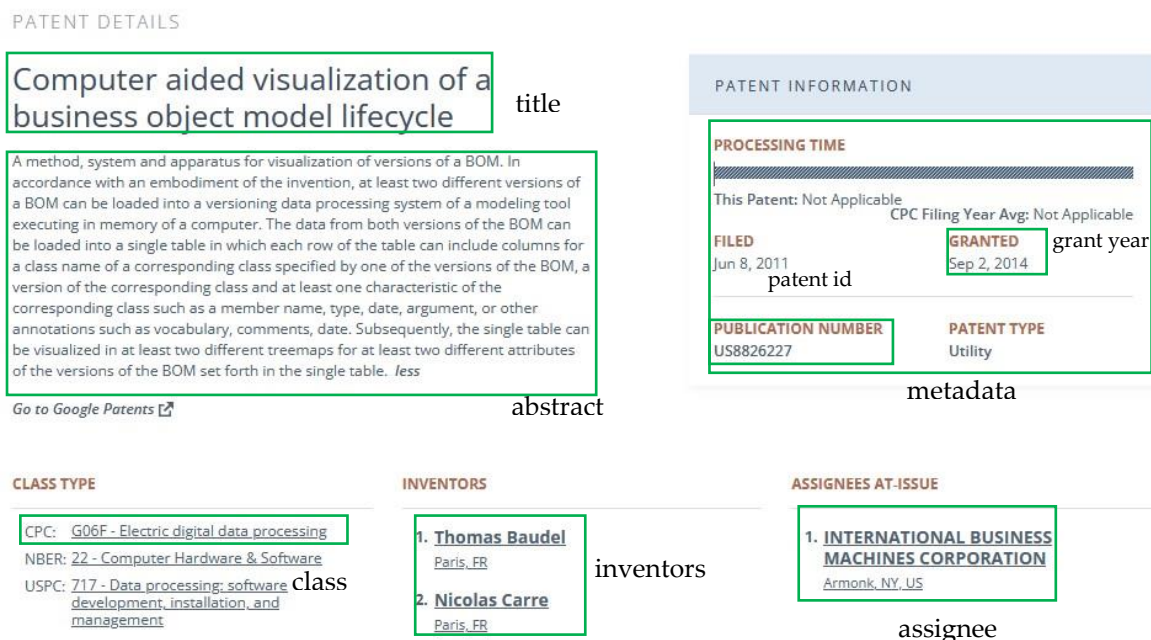

Figure 2. Example of USPTO patent entry.

Thus, the USPTO was utilized as the primary source for data collection, as it has been considered as a leading authority in patent registration and granting. Similarly, to various other patent offices, the USPTO has integrated in its services initiatives that bolster the task of patent retrieval and patent search [78,79]. One such initiative is the Application Programming Interface (API) that the USPTO provides (https://patentsview.org/apis/api-endpoints/patents, accessed on 18 September 2022), where each patent is stored as a semi-structured web document that contains valuable information for our analysis. Thus, by formulating a proper search strategy, an optimal retrieval of patents would be ensured.

To that end, we decided to utilize a semi-automated approach for patent retrieval, by constructing a targeted search string that would match patents with specific keywords. The selected search strategy was used for collection of patent data based on the patent class associated with each entry. In order to identify the target class that would serve as the basis of the constructed search string, we performed a thorough study of the Cooperative Patent Classification (CPC) schema for patent categorization.

The CPC categorization schema is managed by the USPTO and comprises a straightforward and comprehensive way of describing the technological contents and objectives of patents, being comprised from general classes that are divided in subclasses, containing more specific areas. For our analysis, we focused our attention to the G06F8 (arrangements for software engineering) class. This class encompasses all patents that are related to SE and its subfields, with DOSD being one of them. It is divided into several subclasses, which are listed in Table 1, along with their description according to the CPC categorization. Patents belonging to this particular class, or its subclasses, could also be categorized to other classes, as the CPC schema is quite detailed, and some entries may be technologically adherent to other domains. However, as this study focuses on DOSD patents as a point of reference, we only used the G06F8 class in subsequent steps.

Table 1. Subclasses of G06F8 (arrangements for software engineering).

Class Number	Title
G06F8/10	Requirements analysis/Specification techniques
G06F8/20	Software Design
G06F8/30	Creation/Generation of Source Code
G06F8/40	Transformation of program code
G06F8/60	Software Deployment
G06F8/70	Software Maintenance/Management

The data collection process was carried out by constructing and passing specialized queries to the Patents Endpoint API, retrieving all SE patent activity from 1970, based on the filing year, where the first patent application related to SE is documented, up to 2019. While 1970 is the first documented filing year of SE patents in the USPTO, the granting of patents was not conducted until 1976, where the first granting of SE patents is observed. In total, the data collection phase resulted in 32,861 patents referring to the SE domain, along with all the available fields that the API provides. In addition, the collected patents were subjected to a deduplication process that removed duplicate entries based on the identification number of each patent and each abstract and title, keeping the most recently filed patent, based on the filing year or the most recently granted patent in case of a filing year match. Thus, after the deduplication process, the final number of patents in the dataset was reduced to 24,620.

Finally, to focus our analysis on DOSD, we filtered the collected patents based on a keyword search, extracting 630 patents that mentioned the words "data processing" OR "data management" in their titles and abstracts. The final search string was formulated by incorporating and testing additional terms ("data analysis", "data mining", "business intelligence", "knowledge extraction"). However, their use did not increase the number of identified patents or resulted in patents that had already been obtained. Hence, we elected to keep the two main terms for patent extraction. The final data were unified in a joint database and stored for the subsequent stages of analysis.

3.4. Data Pre-Processing

The next step of the methodology involved the detection of the most useful fields to be analysed, in accordance with the goals and objectives set by the posed RQs. The extracted patents contained a broad pool of data showcasing the purpose and general information of a patent. Similar works have focused on textual information (title), temporal trends (granting year) and longitudinal information (country). In respect to the RQs of Section 3.1 and the general focus of other works, the extracted features are presented in Table 2.

Table 2. Extracted features from patent entries.

Feature Name	Description
id	Unique identification number of the patent
granting year	The year that the patent was granted by the USPTO
country	The country of origin
assignee	The organization (company, institution) of the patent
title	The title of the patent
subclass	The G06F8 subclass to which the patent belongs
citations	The forward and backward patent citations

The majority of features were directly extracted from selected fields of the dataset. Regarding the filing and granting years, we decided to use the granting year as a reference point in the authorization of a patent. This choice was based on the fact that the granting year is a more potent indicator of patent activity because it highlights the ownership and exploitation of the patent in a more concise way [49,80,81]. In addition, most patents do not have a preassigned country or continent, as they are automatically linked to the United States of America, given that the filing organization is USPTO. Thus, to assign a country to each patent, we turned our attention to the country of its primary inventor. Our preference was to use the country of the inventor, rather than the applicant, which can be an organization, to reflect the creation of the patent, and not its ownership. The extracted countries were then parsed by a specialized Python package (https://pypi.org/project/pycountry-convert/, accessed on 18 September 2022) in order to gain the corresponding continents.

As far as the textual features are concerned (title), their unstructured nature prompted us to perform some necessary preprocessing procedures. Moreover, we used established natural language processing (NLP) techniques and removed punctuation, stopwords and any information that could generate noise, such as numbers, URLs, and symbols. Finally, all words were stemmed to their root in order to have a common representation.

3.5. Data Analysis

In order to provide answers to the posed RQs, we made the distinction between features utilized in the exploratory analysis (RQ_1) and the textual features that can serve as a baseline for the definition of thematic areas (RQ_2) and the mapping of innovation via the use of networks (RQ_3). In Table 3, we provide an overview of each RQ, the features associated with it, as well as the methodology applied for its completion.

Regarding RQ_1 and the features of the first group (granting year, country, assignee), we utilized descriptive statistics and visualization techniques in order to examine the distributions of qualitative and quantitative features. The goal of this analysis was to provide detailed patent cumulative counts and yearly distributions, in order to track the technological development of patents and indicate the various technology stages over the years. In addition, based on the methodology established by Trappey et al. [69], we utilized industrial profiling in order to draw conclusions regarding the industrial standing of technological inhibitors. Finally, the country feature was visualized in mapping software, so as to detect the most active patent granting countries and compare them with their industrial standing.

Table 3. Research goals, research questions and features on patent entries.

Research Questions	Features	Data Analysis Methods
What is the landscape of DOSD patents?		
(a) How patent activity evolves over time	granting year, subclass	Descriptive statistics
(b) How patents are geographically distributed	country	Geographical Mapping
(c) Which are the most active patenting organizations	assignee	Descriptive statistics
Which thematic trends can be traced in DOSD patents?	title	LDA
How is technological innovation reflected in DOSD patent citations?	citations	Citation Networks, Brokerage Analysis, Network Analysis

In RQ_2, our aim was to discover linguistic patterns in the preprocessed textual features that characterize each patent (title). This discovery would, in turn, allow us to extract thematic areas of granted patents that concern different technological aspects and trend in DOSD. These thematic areas are, usually, expressed by sets of words that form a clear picture of the topic to which they refer to. Thus, to obtain this representation of thematic areas and unveil topics of patents, we utilized the Latent Dirichlet Allocation (LDA) topic modelling algorithm [82] in the unified corpus of patent titles.

The most important step in the execution of an LDA model is the proper selection of the number of topics, usually expressed as K. This is a manual process, defined by the user and requires experimentation with several values in order to find the optimal value for K. In our study, after several trial executions of the LDA model, the value of K was set to eight, providing a meaningful and coherent way of extracting thematic areas from the corpus of patent titles. The selection process was evaluated by using the Coherence Score (CS) [83] for all experimentations.

In addition, by leveraging the methodologies proposed by Barua et al. [84] to assess the overall impact of each topic produced by the LDA algorithm, we utilize the share and popularity metrics, exploiting the membership value of each patent entry to each of the produced topics. These metrics are quite useful in inferring the involvement of each topic to the patent documents, with share indicating the total number of documents that are associated with a topic, while the popularity of a topic indicates the percentage of patent documents that have this topic as dominant, with the highest membership values [84]. Finally, for computing the degree of similarity (or distance) between the topics, the PyLDAVis package was utilized to project the inter-topic distances in a two-dimensional space, via the use of multidimensional scaling [85].

Finally, in RQ_3, our goal was to utilize the patent citations and construct global citation networks of interconnected patents and patent classes in order to detect influential nodes. To achieve this, we first construct the directed PCN, where two patents, denoted as p_a and p_b are connected only if p_a cites p_b. Having constructed the PCN, we then employ the HITS algorithm [86,87] to discover hubs and authorities and find influential patents that receive or provide a large number of citations. The rationale behind the use of the HITS algorithm is that the importance of a patent p in the network is not related only to the number of patents pointing to or being pointed to by p, but also to the importance of these patents. In addition, we used network analysis metrics to gain some basic insights about the structure of the network (e.g., density, modularity, etc.).

In the second step of the methodology, we construct the directed CCN. The construction of the network follows an iterative process, where for each patent of the dataset that has a distribution $c = [class_1, class_2, ..., class_n]$ of all the CPC classes that the patent belongs to and the patents of its forward and backward citations have similar distributions, a directed node is produced for each class of c to all other citation distributions, if the class is a subclass of the G06F8 class. The produced network is a directed graph of CPC classes reflecting technological objectives and connections. In order to detect valuable and innovative CPC classes comprising bridge nodes that connect other classes, we make use of the BA methodology [74] in node triads which characterize triad relationships in five different roles, which can be seen in Table 4.

Table 4. Triadic relationships of broker nodes.

Broker Role	Triadic Relationship
Coordinator	$a \rightarrow \boldsymbol{a} \rightarrow a$
Gatekeeper	$a \rightarrow \boldsymbol{b} \rightarrow a$
Representative	$a \rightarrow \boldsymbol{a} \rightarrow b$
Itinerant	$a \rightarrow \boldsymbol{b} \rightarrow b$
Liaison	$a \rightarrow \boldsymbol{b} \rightarrow c$

BA characterizes the middle node of each triad, which is referred to as the "*broker*", in one of the five roles, according to the triads in which it participates. Thus, each node receives five scores for each role, which represent the number of times that a node participates in a triad in a given role, as the broker. In the case of our paper, nodes are CPC classes, and the relations between them reveal which classes are driving innovation and which potentially restrain it.

4. Results

In this section, the results of the conducted analysis are presented in accordance with the posed RQs.

[RQ$_1$] What is the landscape of DOSD patents?

This RQ aims to conduct a descriptive analysis of the collected patents by examining the temporal evolution of granted patents, mapping them geographically and pinpointing the most active organizations that seek to own a patent and possibly exploit its contents commercially.

The distribution of the granting year (Figure 3) reveals that the majority of patent-granting activity takes place after 2010 and follows an increasing trend. This showcases the necessity for DOSD in the last decade, which rose into the spotlight as the volume and types of available data rendered their manipulation from traditional software a challenging task. Patent granting in the 1990s is low, although this can be explained by the fact that patent offices had not effectively digitized their services, and thus, the stored patents were limited in number.
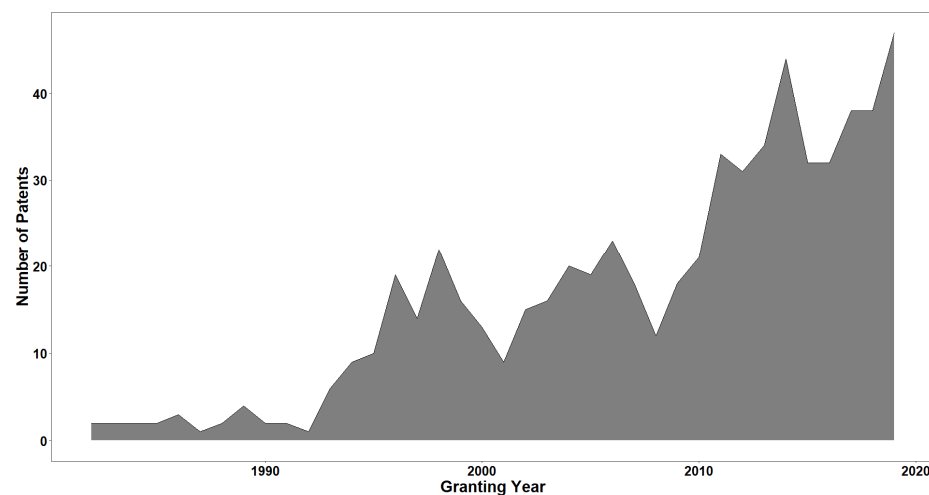

Figure 3. Distribution of DOSD patents granting year.

Moreover, the joint distribution of the patent subclasses within each decade (Table 5) showcases some interesting findings on the focus of DOSD in each chronological period. The 1980s and 1990s seem to emphasize the Transformation of Program Code (G06F8/40), while the prime class of the 2000s is Software Deployment (G06F8/60), with Transformation of Program Code (G06F8/40) and Creation/Generation of Source Code (G06F8/30) behind. Finally, Software Design (G06F8/20) and Requirements/Specifications (G06F8/10) present a steady trend thorough the examined period.

Table 5. Joint distribution of DOSD patent activity for CPC subclasses and decades.

	Decade				
Subclass	1980s	1990s	2000s	2010s	Total
G06F8/10	0 (0.0%)	1 (1.0%)	5 (3.0%)	15 (3.7%)	21 (3.0%)
G06F8/20	1 (5.9%)	3 (3.0%)	5 (3.0%)	23 (5.7%)	32 (4.6%)
G06F8/30	3 (17.6%)	19 (18.8%)	32 (19.3%)	79 (19.5%)	133 (19.3%)
G06F8/40	7 (41.2%)	31 (30.7%)	45 (27.1%)	86 (21.2%)	169 (24.5%)
G06F8/60	4 (23.5%)	22 (21.8%)	63 (38.0%)	149 (36.8%)	238 (34.5%)
G06F8/70	2 (11.8%)	25 (24.8%)	16 (9.6%)	53 (13.1%)	96 (13.9%)
Total	17 (100%)	101 (100%)	166 (100%)	405 (100%)	689 (100%)

In terms of geographical mapping, the USA is the top country, with 378 granted patents. The large number of USA patents is possibly due to the fact that some filed patents are automatically assigned to this country when filed to the USPTO. However, the USA is still a leading player in patent ownership [88], and its numbers are consequently high. European countries have a strong presence in patenting products and inventions related to this domain, with the United Kingdom (40 patents) and Germany (35 patents) having a clear advantage over the rest of the continent, with France (17 patents) and Italy (15 patents) closely following. In Asia, the top countries are Japan (58 patents), Korea (14 patents), India (8 patents) and China (4 patents), which is supported by the rapid development of their software industries in later years [89–91]. The limited number of patents belonging to Asian countries can be attributed to the fact that many Asian inventors prefer to file their patents in regional offices such as KIPO, JPO and CNIPA. Thus, the selection of USPTO, although a valuable source of information, is a minor threat to the validity of the study.

In Table 6, we present the ten organizations that have the highest number of granted patents across the entirety of our dataset. The first organization with the highest number of patents is IBM, which is hailed as one of the leading companies in computer hardware, personal computers and commercial software. IBM appears to have a very active research department that aims at owning a large number of patents, maintaining the status of the company as a torchbearer in DOSD, which has been actively happening since the 1970s (https://www.ibm.com/ibm/history/exhibits/dpd50/dpd50_intro.html, accessed on 18 September 2022). We can also observe that various companies are technology related and are active in the industries of electronics and devices (Samsung and Motorola), equipment (Siemens and Hitachi), as well as computer hardware and software products (HP and Intel). Ab Initio specializes in enterprise software facilitating data-related procedures in large companies. In terms of the countries of the top companies, the findings validate the geographical mapping, with the USA holding the lead.

Table 6. Top organizations by granted patents.

Organization	Country	# of Patents	Main Patent Subclass
International Business Machines Corporation	US	225	Software Deployment (G06F8/60)
Arm Limited	UK	23	Transformation of program code (G06F8/40)
MOTOROLA SOLUTIONS, INC.	US	11	Transformation of program code (G06F8/40)
Samsung Electronics Co., Ltd.	KR	10	Software Deployment (G06F8/60)
SIEMENS AKTIENGESELLSCHAFT	DE	10	Software Maintenance/Management (G06F8/70)
HITACHI, LTD.	JP	8	Software Maintenance/Management (G06F8/70)
Ab Initio Technology LLC	US	8	Creation/Generation of Source Code (G06F8/30)
GOOGLE LLC	US	7	Software Deployment (G06F8/60)
Hewlett-Packard Development Company, L.P.	US	7	Transformation of program code (G06F8/40)
Intel Corporation	US	7	Software Deployment (G06F8/60)

Finally, there seems to be a plethora of different DOSD areas that the top assignees are focusing on, with IBM, Samsung and Intel owning patents related to Software Deployment (G06F8/60), and companies focusing on equipment, turning their attention to Software Maintenance/Management (G06F8/70) and hardware-related companies exploiting patents related to the Transformation of Program Code (G06F8/40), possibly for communication protocols and devices. An interesting exception is the Ab Initio Technology, which focuses on patents of Creation and Generation of Source Code (G06F8/30). Given that the services of this company are linked with developing data processing application and business suites, it is apparent that the creation and delivery of high-quality code is the core of its activities.

[RQ$_2$] Which thematic trends can be traced in DOSD patents?

While RQ$_1$ aimed at performing an exploratory analysis of the identified patents, RQ$_2$ is directly leveraging the linguistic traits of the patent titles in order to trace patterns. This

process can provide insights into the targeted areas that DOSD patents revolve thematically and uncover prominent topics in patent activity.

Table 7 provides a summary of the results extracted by the LDA algorithm by setting the *K* parameter to eight, along with the share and popularity metrics for each topic. The constructed model yielded a CS of 0.59, which is an indicator of a well-rounded model that produces balanced topics [92]. Moreover, after carefully examining the key words that accompany each extracted topic in conjunction with the top five representative patents in terms of membership, we assigned a manual short title that better captures its general scope and purpose. An inspection of the topics showcases that they cover a wide range of DOSD tasks, with some of them being related to software that is used for handling memory issues (Topic 1) or being integrated in large scale systems (Topic 2) and others being closely related to dynamic frameworks that directly assist supporting business intelligence (Topic 7) or protocols that facilitate resource allocation and deployment and ensure proper knowledge transfer and data management (Topic 4). In addition, some topics cover facets of DOSD that have to do with version control and rollout of updates in software (Topic 8), while preserving software quality and the integration of processed data in interfaces and dashboards (Topic 5). Finally, two of the extracted topics are directly linked with parallel data processing, referencing the considerable amount of data produced in business and in software procedures along with specialized environments developed for this purpose (Topic 3), as well as large scale simulations of data processes, potentially for risk management and estimation (Topic 6).

In terms of the topic membership metrics, Table 7 indicates that all topics are evenly distributed across the patent documents. The most shared topics appear to be Topic 8 (Version control and software quality) and Topic 5 (Data integration, interfaces and updates). Both of these topics are directly related to the technical side of DOSD, with Topic 8 referencing the continuous need of companies to ensure that the proper versions of software are deployed in production routines and Topic 5 concerning the issues that can be raised by integrating data in different interfaces and updating software to accept new data inputs. Thus, given the importance of these issues in a business, it is more than expected that these two topics have the highest share values. Apart from that, Topic 1 (Software for memory management) has the third highest share value, highlighting the need for software that has efficient memory handling for processing large and different forms of data. In contrast, the lower share metric can be found in Topic 4 (Resource allocation and information transferring). However, this can be attributed to the coverage of similar patents by other topics that have higher share values, such as Topic 5, and its more specific nature.

On the other hand, the popularity metric indicates topics that are dominant in the distribution of patent documents. With this in mind, the most popular topic in patent objectives appears to be the version control and software quality in data-related products (Topic 8) along with the integration of data in interfaces and the proper updates (Topic 5). The high popularity values of these topics correspond with their high share values and prove that integration and software quality procedures are the pillars of efficient DOSD. In contrast, the topics with the lowest popularity scores are the creation of automated software to be used in complex systems (Topic 2) and the exploitation of parallel processing and specialized programming environments (Topic 3). However, their restrained popularity values can be explained by the more technical and domain-specific aspects of Topic 2 and the fact that many patents that reference parallel processing may be focused on other primary objectives and may thus belong to other topics.

In addition, Figure 4 serves as a visualization of the distances between the topics by projecting them onto a two-dimensional axis, utilizing the multidimensional scaling technique. In this figure, the circles correspond to the presence of each topic in the corpus of patent titles, while the circles are positioned based on the inter-topic distance. The exploration of Figure 4 indicates a well-defined LDA model, since there are no overlapping circles, while topics represented by circles are located in every quadrant. In addition, the topics are well-distributed over the corpus of patent documents, as there is no clear

dominant topic. This finding proves that DOSD patents express multiple equally important objectives. Furthermore, topics closer to one another are thematically adherent, focusing on similar technological objectives. Topic 1 (Software for memory management) and Topic 2 (Automated software for large scale systems) seem to refer to the management of memory, which can be expanded in large scale systems. In addition, Topic 2 appears to be the farthest away from other topics, while the small radius of its circle indicates that it is dominant in the least number of patent entries. However, this is not surprising, as its objectives are very specific and are tackled by field experts. Topic 7 (Dynamic frameworks and business environments) is also in a close distance and thus similar with Topic 1 (Software for memory management), which can be explained by the fact that dynamic interfaces and business environments usually handle advanced visualizations and need to efficiently distribute memory. Other distinct groups are Topic 5 (Data integration, interfaces and updates) and Topic 8 (Version control and software quality), which complement each other, as data integration and updates in software also require version control and quality routines to be deployed. These topics have the highest dominance score, as their circles are larger. Finally, in the lower right quadrant, the parallel processing architectures (Topic 3) are close to both simulations for advanced data processing (Topic 6) and the handling of resource allocation tasks (Topic 4).
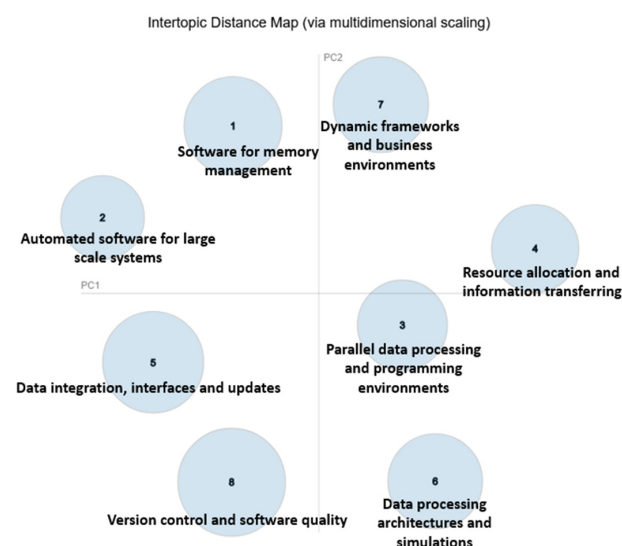

Figure 4. Inter-topic distance map.

Table 7. Extracted topics and metrics.

Topic Description	Key Words	Share %	Popularity %
Topic 1: Software for memory management	memory, operation, product, patch, service, content, update, device, enterprise	20.3	13.8
Topic 2: Automated software for large scale systems	configure, service, automate, source, efficient, transform, aircraft, device, server, function	17.4	8.3
Topic 3: Parallel data processing and programming environments	develop, environment, base, object, perform, platform, processor, parallel, structure, format	16.7	9.5
Topic 4: Resource allocation and information transferring	resource, deploy, network, correct, model, microcode, analytics, multimedia, platform, error	15.4	10.1
Topic 5: Data integration, interfaces and updates	integrate, interface, type, firmware, update, control, user, link, display, feature	21.7	15.3
Topic 6: Data processing architectures and simulations	instruct, file, compile, circuit, associate, stream, communicate, synchronize, vector, simulate	18.7	11.9
Topic 7: Dynamic frameworks and business environments	framework, upgrade, dynamic, virtual, network, distribute, automate, business, flow	18.2	13.5
Topic 8: Version control and software quality	control, install, distribution, dynamic, medium, storage, version, digital, set, language	26.2	17.2

[RQ$_3$] How is technological innovation portrayed in the interconnection of DOSD patents?

The creation of the PCN and CCN directed networks reveals some very interesting findings about influential patents and patent classes that drive innovation and serve as guidelines that other patents follow when formulating their objectives and purposes. The top hub and authority patents extracted from the application of the HITS algorithm in the PCN network are presented in Table 8. The PCN nodes tend to form communities of patents that cite each other, with a modularity score of 0.91, which is expected, given that each patent has its own set of forward and backward citations, even if some patents may cite the same patents.

Table 8. Top authorities and hubs in PCN.

Top Authorities		
Patent Title	**Granting Year**	**Assignee**
Data integration by object management	1997	Wang Laboratories
Object oriented programming based global registry system, method, and article of manufacture	1998	Object Technology Licensing Corporation
Method for managing globally distributed software components	1999	Novell, Inc.
Method for forming a reusable and modifiable database interface object	1996	POWERSOFT S.P.A.
System and method for completing an electronic form	1996	Wright Strategies, Inc.
Selecting screens in a GUI using events generated by a set of view controllers	2007	International Business Machines Corporation
Top Hubs		
Patent Title	**Granting Year**	**Assignee**
Method and apparatus in a data-processing system for the issuance and delivery of lightweight requests to concurrent and multiple service providers	2005	International Business Machines Corporation
Method and apparatus in a data-processing system for providing an interface for non-intrusive observable debugging, tracing, and logging data from execution of an application	2005	International Business Machines Corporation
Controlling presentation of a GUI, using view controllers created by an application mediator, by identifying a destination to access a target to retrieve data	2005	International Business Machines Corporation
Method and apparatus in a data-processing system for the controlling and sequencing of graphical user interface components and mediating access to system services for those components	2004	International Business Machines Corporation

The identified authorities are patents that, when present in the PCN, receive a large number of incoming edges by hubs, which essentially means that they are highly cited by other influential patents that shape the objectives of subsequent patents. It is apparent that the identified authorities concern patents with highly valuable objectives for DOSD, with the top authority being relevant to data integration. This finding is in line with the topics extracted in RQ2 and proves that proper integration of data is crucial in organizations, as is its commercial exploitation. Other authorities are relevant to object-oriented programming architectures, distributed software and databases, which are all aspects of developing software primarily targeted for data manipulation and management.

In contrast, the hubs of the PCN represent patents that have a large number of outgoing edges to authorities, hence being patents that highly cite other important patents. This

fact indicates patents that are directly referencing other technological fields and may combine objectives and methodologies from different patents, thus creating an innovative result [93,94]. An interesting finding is that IBM is the sole assignee that has top hubs, which compliments the fact that it ranks first when it comes to the highest number of granted patents. Among the top hubs, there are some quite promising objectives of GUI handling, concurrent and parallel processing, as well as debugging and processing data in applications.

The other facet of RQ$_3$ was the identification of bridge nodes (or CPC classes) that drive or control innovation and transfer of knowledge in patent objectives. In Table 9, we present the top brokers of the CCN network for each brokerage role. Given that the patents of the CCN network contained only the citations of classes that were subclasses of G06F8, there were no itinerants detected. However, we present the top brokers for the remaining triad roles.

Table 9. Top brokers in each role.

Coordinators	Gatekeepers	Representatives	Liaisons
Compilation (G06F8/41)	Software Deployment (G06F8/60)	Software Deployment (G06F8/60)	Installation (G06F8/61)
Software Deployment (G06F8/60)	Updates (G06F8/65)	Installation (G06F8/61)	Updates (G06F8/65)
Parallelism (G06F8/45)	Graphical or Visual Programming (G06F8/34)	Updates (G06F8/65)	Software Deployment (G06F8/60)
Graphical or Visual Programming (G06F8/34)	Installation (G06F8/61)	Graphical or Visual Programming (G06F8/34)	Graphical or Visual Programming (G06F8/34)
Updates (G06F8/65)	Version Control (G06F8/71)	Software Design (G06F8/20)	Software Maintenance/Management (G06F8/70)

According to Gould et al. [74], each brokerage role reveals different stages of innovation and knowledge transfer for the participating classes. Of course, a node (or class) can have multiple brokerage roles. In the case of the constructed CCN, coordinator classes facilitate the connection between internal classes of the same superclass, thus allowing knowledge and patent objectives to be transferred directly and without limitations. These classes essentially serve as "stopping points" for other classes that utilize them to reach other similar subclasses and are generally classes that define DOSD. Among them, we can find Compilation (G06F8/41), Software Deployment and Updates (G06F8/60, G06F8/65), Graphical Programming (G06F8/34) and Parallelism (G06F8/45). We can see that the coordinator classes also correspond to the identified topics of RQ$_2$, further validating the prominent fields of DOSD.

Gatekeeper classes are quite different from coordinators, as they have increased authority. Essentially, the classes that belong to this category are "guarding" subclasses of the same superclass and decide whether to allow or deny access to them from other classes. Gatekeeper classes can be defined as well-defined and robust aspects of DOSD that influence the objectives of a large number of patents in the network while simultaneously being intermediaries between different technological fields. Novel classes in this category are Version Control (G06F8/71) and Installation (G06F8/61), while the other classes have been described in the previous role.

Representatives are the exact opposite role of gatekeepers. Where a gatekeeper class would control access in a class of the same group, representative classes are trying to communicate with other classes and transfer knowledge. Representative nodes of the CCN are classes that actively cite other classes and are used in interdisciplinary patents of DOSD. An interesting class of this category is Software Design (G06F8/20), with the top representative class being Software Deployment (G06F8/60).

Finally, liaison classes are patent classes that link other classes that are unrelated to each other, in the sense that neither node belongs to the same class group. Liaison nodes act as mediators between technological fields and can be used to bridge different ideas and objectives of patents in elegant software solutions. Installation (G06F8/61) is the top liaison class, with Updates (G06F8/65) and Software Deployment (G06F8/60) closely following. An interesting addition in this category is Software Maintenance/Management (G0F8/70), which was also present in the main subclasses of the top organizations.

It is apparent that Software Deployment is an important broker, holding both promoting (Coordinator, Representative), authoritative (Gatekeeper) and neutral (Liaison) roles. This is an excellent indicator of the importance of proper software deployment architectures for data processing that need to be carefully developed and have potential to be applied in multiple fields. Another major class is Updates, which is another key aspect of DOSD while Version Control is a prominent gatekeeper, possibly due to the more technical nature of patents filed under this class. Finally, Software Design is a class that is actively used to promote innovation and define patent objectives with a Representative role and Software Maintenance/Management acts as a mediator and necessary procedure for the development of software and the granting of patents that belong to other classes.

In addition, the results from the network analysis on the CCN are presented (Table 10), where the most important nodes ranked by their centralities can be seen. Overall, the CCN has a more abstract community structure, with a modularity score of 0.39.

Table 10. Network analysis of CCN.

Highest Nodes by		
Degree Centrality	**Betweenness Centrality**	**Closeness Centrality**
Software Deployment (G06F8/60)	Software Deployment (G06F8/60)	Installation (G06F8/61)
Installation (G06F8/61)	Installation (G06F8/61)	Software Deployment (G06F8/60)
Updates (G06F8/65)	Updates (G06F8/65)	Updates (G06F8/65)
Graphical or Visual Programming (G06F8/34)	Graphical or Visual Programming (G06F8/34)	Software Design (G06F8/20)
Version Control (G06F8/71)	Version Control (G06F8/71)	Requirements Analysis/Specifications (G06F8/10)

As far as node centralities are concerned, the nodes that have a larger number of external and internal edges as citations (Degree Centrality) and the nodes that act as immediate connections between node paths (Betweenness Centrality) are similar to the results of the BA, with Software Deployment and Updates occupying the top spots. A more interesting finding lies in the nodes that are closer to every other node in the network (Closeness Centrality), thus being immediate or intermediate citations of other classes, with Requirements Analysis/Specifications (G06F8/10) and Software Design (G06F8/20) being present, indicating that proper requirement definition and design of software before the implementation are very important factors in patent objectives and innovation.

5. Threats to Validity

In this section, we discuss some existing threats to the validity of our study while also presenting the mitigating actions taken to limit their effect.

Regarding the internal validity of the study, a principal threat is identified in the data collection and patent selection process. The collection of patent data was meticulously carried out and involved the identification of a relevant SE CPC class in the upper level of collection and the leveraging of keywords relevant to DOSD at a secondary level. However, due to the multivariate nature of patent data, the threat of omitting or missing patent entries that may not belong to this specific CPC class or may not correspond to the utilized keywords is possible. We deem this event to not reflect the typical state of the collected data, however, since most SE-related patents are naturally assigned to the G06F8 class, and the extracted keywords underwent expert judgement and reiterations so as to better capture

and accumulate the largest possible number of DOSD patents. In addition, although the selection of a single source of data collection, namely the USPTO, is adequately justified, the application of the methodological framework to other patent offices would certainly enhance the credibility of the current study.

In the data analysis phase, the application of the LDA algorithm posed a problem, as the appropriate selection of the number of topics is a crucial part of a proper execution and different algorithms setups can significantly alter the produced latent topics. To mitigate this threat, multiple experiments were deployed, evaluated and cross-validated by experts of the field, ensuring that the produced topics fully captured the different thematic axes of the collected DOSD patents. Although manual and human interpretation is always required when applying LDA and errors in judgement can be detected, we believe that our validation process is robust, and hence, the produced topics are credible.

Regarding the external validity of the study, a limitation of our methodological framework is its application on one patent office. Of course, USPTO has been proven to be the most well-known and established patent office on a global scale, but the extension of the study to other patent offices (EPO, JPO, KIPO) would certainly offer opportunities for a more concise and solid presentation of our results on a collective scale and a proper generalization of our findings. However, despite the choice of a single data source, we still consider the practical implications to stakeholders and policymakers to overcome the restrictions of the data collection. Finally, in regard to the country and organization profiling, while the primary investors in DOSD patent granting are large countries and organizations, we recognize that innovation in this domain can have multiple forms, besides patents, such as research papers, startup ventures and funded projects, and can originate from smaller countries or companies. Hence, the investigation of other innovation forms could be beneficial for a more complete profiling of the countries and assignees involved in our study. However, the goals of this study, which emphasize patents and their value in the technological landscape, as well as the absence, to the best of our knowledge, of an organized data source that could provide detailed information on other innovation forms, prevented us from applying this type of analysis.

6. Conclusions

The industrial landscape of patents related to DOSD is constantly growing, as the need for software that can handle large volumes of data and perform complicated tasks is crucial for business services. In this increasing trend, patents stand as a reliable way of securing and exploiting an invention, while also promoting innovative technologies. The findings show that multiple countries and organizations around the globe are interested in patent grants in this field, and in the last decade, patent grants have been on the rise. Described in more detail, the geographical analysis of the assignees showcased that countries with an established "patent culture" such as the USA, Germany or the United Kingdom gain an advantage over smaller countries that may not be so active in patent grants. The top organizations that invest in DOSD patents are all high profile and established with IBM, Google and other large-scale companies, having a large presence in our dataset. Finally, the analysis of CPC in a temporal scale indicates that Software Deployment (G06F8/60) and Transformation of Program Code (G06F8/40) present the highest rise in each decade, while Software Design (G06F8/20) and Requirements Analysis (G06F8/10) increase at a much slower rate.

In addition, the analysis of topics reveals that DOSD patents mainly revolve around data integration, updates, software quality and development environments and the results of the advanced network analysis validate this statement, with Software Deployment (G06F8/60) and Transformation of Program Code (G06F8/40) being once again the most influential patent classes that mediate between the knowledge transfer of other classes. Finally, in terms of patent citations that dictate the most influential patents, our findings indicate that data integration, data interfaces and large data-processing systems are the core of DOSD patent applications.

The results of this study can yield multiple practical implications to stakeholders, policymakers, technology investors and practitioners or researchers, by not only highlighting the most active and growing organizations and countries but also by further highlighting the innovation prospects of patents. The thematic analysis clearly showcases the dominant technological domains that DOSD focuses on, prompting decision makers and business sectors to gain a perspective in the technological convergence of the domain and adjust their business strategies related to the development of similar software while encouraging them to pursue additional patent grants. Finally, the identification of prominent topics, influential CPC classes and technological objectives facilitates the conduction of other relevant studies in the field, providing comprehensive guidelines to practitioners and researchers that wish to further examine and profile DOSD patents or other forms of innovation in the field.

Given the tremendous rates of data production and the rapid advancements in technology and software, we expect this rise of patent grants and objectives to be even more impressive in the future, bolstering the standing of software enterprises and contributing to the diffusion of innovation across multiple domains.

Author Contributions: Conceptualization: all authors; Data Curation: K.G.; Formal Analysis: K.G. and N.M.; Methodology: all authors; Software: K.G. and N.M.; Visualization: K.G. and N.M.; Writing—original draft: all authors; Writing—review and editing: all Authors. All authors have read and agreed to the published version of the manuscript.

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

References

1. Chen, H.; Chiang, R.H.; Storey, V.C. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Q.* **2012**, *36*, 1165. [CrossRef]
2. Choi, T.-M.; Chan, H.K.; Yue, X. Recent development in Big Data Analytics for Business Operations and Risk Management. *IEEE Trans. Cybern.* **2017**, *47*, 81–92. [CrossRef] [PubMed]
3. Fan, S.; Lau, R.Y.K.; Zhao, J.L. Demystifying big data analytics for business intelligence through the lens of Marketing Mix. *Big Data Res.* **2015**, *2*, 28–32. [CrossRef]
4. Singh, S.K.; El-Kassar, A.-N. Role of big data analytics in developing sustainable capabilities. *J. Clean. Prod.* **2019**, *213*, 1264–1273. [CrossRef]
5. Alsghaier, H. The importance of Big Data Analytics in Business: A Case Study. *Am. J. Softw. Eng. Appl.* **2017**, *6*, 111. [CrossRef]
6. Ghimire, A.; Thapa, S.; Jha, A.K.; Adhikari, S.; Kumar, A. Accelerating business growth with Big Data and artificial intelligence. In Proceedings of the 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 7–9 October 2020.
7. Lasi, H.; Fettke, P.; Kemper, H.-G.; Feld, T.; Hoffmann, M. Industry 4.0. *Bus. Inf. Syst. Eng.* **2014**, *6*, 239–242. [CrossRef]
8. Xu, X.; Lu, Y.; Vogel-Heuser, B.; Wang, L. Industry 4.0 and industry 5.0—Inception, conception and perception. *J. Manuf. Syst.* **2021**, *61*, 530–535. [CrossRef]
9. Axmann, B.; Harmoko, H. Industry 4.0 readiness assessment. *Teh. Glas.* **2020**, *14*, 212–217. [CrossRef]
10. Dalmarco, G.; Ramalho, F.R.; Barros, A.C.; Soares, A.L. Providing industry 4.0 technologies: The case of a production technology cluster. *J. High Technol. Manag. Res.* **2019**, *30*, 100355. [CrossRef]
11. Subramanian, G.H.; Pendharkar, P.C.; Wallace, M. An empirical study of the effect of complexity, platform, and program type on software development effort of Business Applications. *Empir. Softw. Eng.* **2006**, *11*, 541–553. [CrossRef]
12. Woods, M.; Paulus, T.; Atkins, D.P.; Macklin, R. Advancing qualitative research using qualitative data analysis software (QDAS)? reviewing potential versus practice in published studies using atlas.ti and NVIVO, 1994–2013. *Soc. Sci. Comput. Rev.* **2016**, *34*, 597–617. [CrossRef]
13. Moral-Munoz, J.A.; López-Herrera, A.G.; Herrera-Viedma, E.; Cobo, M.J. Science Mapping Analysis Software Tools: A Review. In *Springer Handbook of Science and Technology Indicators*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 159–185.
14. Abdellatif, T.M.; Capretz, L.F.; Ho, D. Software analytics to software practice: A Systematic Literature Review. In Proceedings of the 2015 IEEE/ACM 1st International Workshop on Big Data Software Engineering, Florence, Italy, 23 May 2015.
15. Odaki, K. Legitimacy of employer ownership. In *The Right to Employee Inventions in Patent Law: Debunking the Myth of Incentive Theory*; HeinOnline, 2018.
16. Merges, R.P. Commercial success and patent standards: Economic perspectives on innovation. *Calif. Law Rev.* **1988**, *76*, 803. [CrossRef]
17. Ernst, H.; Conley, J.; Omland, N. How to create commercial value from patents: The Role of Patent Management. *R&D Manag.* **2016**, *46*, 677–690.

18. Grabowski, H. Patents, innovation and access to New Pharmaceuticals. *J. Int. Econ. Law* **2002**, *5*, 849–860. [CrossRef]
19. Danzon, P.M.; Towse, A. Differential pricing for pharmaceuticals: Reconciling Access, R&D and patents. *Int. J. Health Care Financ. Econ.* **2003**, *3*, 183–205.
20. Gilchrist, D.S. Patents as a spur to subsequent innovation? evidence from pharmaceuticals. *Am. Econ. J. Appl. Econ.* **2016**, *8*, 189–221. [CrossRef]
21. Dai, R.; Watal, J. Product patents and access to innovative medicines. *Soc. Sci. Med.* **2021**, *291*, 114479. [CrossRef]
22. OuYang, K.; Weng, C.S. A new comprehensive patent analysis approach for new product design in Mechanical Engineering. *Technol. Forecast. Soc. Chang.* **2011**, *78*, 1183–1199. [CrossRef]
23. Hunter, E.M.; Perry, S.J.; Currall, S.C. Inside multi-disciplinary science and engineering research centers: The impact of organizational climate on invention disclosures and patents. *Res. Policy* **2011**, *40*, 1226–1239. [CrossRef]
24. Kwon, O.; An, Y.; Kim, M.; Lee, C. Anticipating technology-driven industry convergence: Evidence from large-scale patent analysis. *Technol. Anal. Strateg. Manag.* **2019**, *32*, 363–378. [CrossRef]
25. Curran, C.-S.; Leker, J. Patent indicators for monitoring convergence—Examples from NFF and ICT. *Technol. Forecast. Soc. Chang.* **2011**, *78*, 256–273. [CrossRef]
26. Geum, Y. Technological convergence of it and BT: Evidence from patent analysis. *ETRI J.* **2012**, *34*, 439–449. [CrossRef]
27. Buerger, M.; Broekel, T.; Coad, A. Regional Dynamics of Innovation: Investigating the co-evolution of Patents, research and development (R&D), and Employment. *Reg. Stud.* **2012**, *46*, 565–582.
28. Mueller, D.C. Patents, research and development, and the measurement of inventive activity. *J. Ind. Econ.* **1966**, *15*, 26. [CrossRef]
29. Jemala, M. Long-term research on technology innovation in the form of new technology patents. *Int. J. Innov. Stud.* **2021**, *5*, 148–160. [CrossRef]
30. Hall, B.H. Patents, innovation, and development. *Int. Rev. Appl. Econ.* **2022**, *36*, 1–26. [CrossRef]
31. Ferreira, M.; Oliveira, B.M.P.M.; Pinto, A.A. Patents in New Technologies. *J. Differ. Equ. Appl.* **2009**, *15*, 1135–1149. [CrossRef]
32. Elfenbein, D.W. Publications, patents, and the market for University Inventions. *J. Econ. Behav. Organ.* **2007**, *63*, 688–715. [CrossRef]
33. Qiu, Z.; Wang, Z. Technology forecasting based on semantic and citation analysis of patents: A case of robotics domain. *IEEE Trans. Eng. Manag.* **2022**, *69*, 1216–1236. [CrossRef]
34. Kim, M.; Park, Y.; Yoon, J. Generating patent development maps for technology monitoring using semantic patent-topic analysis. *Comput. Ind. Eng.* **2016**, *98*, 289–299. [CrossRef]
35. Erzurumlu, S.S.; Pachamanova, D. Topic modeling and technology forecasting for assessing the commercial viability of healthcare innovations. *Technol. Forecast. Soc. Chang.* **2020**, *156*, 120041. [CrossRef]
36. Bamakan, S.M.; Babaei Bondarti, A.; Babaei Bondarti, P.; Qu, Q. Blockchain technology forecasting by patent analytics and text mining. *Blockchain Res. Appl.* **2021**, *2*, 100019. [CrossRef]
37. Schiff, E. *Industrialization without National Patents: The Netherlands, 1869–1912; Switzerland, 1850–1907*; 2015.
38. Ernst, H. Industrial Research as a source of important patents. *Res. Policy* **1998**, *27*, 1–15. [CrossRef]
39. Basberg, B.L. Patents and the measurement of Technological Change: A Survey of the literature. *Res. Policy* **1987**, *16*, 131–141. [CrossRef]
40. Giarratana, M.S.; Mariani, M.; Weller, I. Rewards for patents and inventor behaviors in industrial research and development. *Acad. Manag. J.* **2018**, *61*, 264–292. [CrossRef]
41. Kitchenham, B.; Pearl Brereton, O.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. Systematic literature reviews in software engineering—A systematic literature review. *Inf. Softw. Technol.* **2009**, *51*, 7–15. [CrossRef]
42. Beecham, S.; Baddoo, N.; Hall, T.; Robinson, H.; Sharp, H. Motivation in software engineering: A systematic literature review. *Inf. Softw. Technol.* **2008**, *50*, 860–878. [CrossRef]
43. Hoda, R.; Salleh, N.; Grundy, J. The rise and evolution of Agile Software Development. *IEEE Softw.* **2018**, *35*, 58–63. [CrossRef]
44. Saheb, T.; Saheb, T. Understanding the development trends of Big Data Technologies: An analysis of patents and the cited scholarly works. *J. Big Data* **2020**, *7*, 12. [CrossRef]
45. Habibollahi Najaf Abadi, H.; Pecht, M. Artificial Intelligence Trends based on the patents granted by the United States Patent and Trademark Office. *IEEE Access* **2020**, *8*, 81633–81643. [CrossRef]
46. Nichols, K. The age of software patents. *Computer* **1999**, *32*, 25–31. [CrossRef]
47. Lee, S.; Yoon, B.; Lee, C.; Park, J. Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping. *Technol. Forecast. Soc. Chang.* **2009**, *76*, 769–786. [CrossRef]
48. Geum, Y.; Kim, M. How to identify promising chances for technological innovation: Keygraph-based patent analysis. *Adv. Eng. Inform.* **2020**, *46*, 101155. [CrossRef]
49. Albino, V.; Ardito, L.; Dangelico, R.M.; Messeni Petruzzelli, A. Understanding the development trends of low-carbon energy technologies: A patent analysis. *Appl. Energy* **2014**, *135*, 836–854. [CrossRef]
50. Kang, B.; Huo, D.; Motohashi, K. Comparison of Chinese and Korean companies in ICT Global Standardization: Essential Patent Analysis. *Telecommun. Policy* **2014**, *38*, 902–913. [CrossRef]
51. Kim, G.; Bae, J. A novel approach to forecast promising technology through patent analysis. *Technol. Forecast. Soc. Chang.* **2017**, *117*, 228–237. [CrossRef]

52. Moehrle, M.G.; Caferoglu, H. Technological speciation as a source for emerging technologies. using semantic patent analysis for the case of Camera Technology. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 776–784. [CrossRef]

53. Tseng, C.-Y.; Ting, P.-H. Patent analysis for technology development of Artificial Intelligence: A country-level comparative study. *Innovation* **2013**, *15*, 463–475. [CrossRef]

54. Fujii, H.; Managi, S. Trends and priority shifts in Artificial Intelligence Technology Invention: A global patent analysis. *Econ. Anal. Policy* **2018**, *58*, 60–69. [CrossRef]

55. Wu, L.; Zhu, H.; Chen, H.; Roco, M.C. Comparing nanotechnology landscapes in the US and China: A patent analysis perspective. *J. Nanoparticle Res.* **2019**, *21*, 180. [CrossRef]

56. Li, S.; Garces, E.; Daim, T. Technology forecasting by analogy-based on social network analysis: The case of autonomous vehicles. *Technol. Forecast. Soc. Chang.* **2019**, *148*, 119731. [CrossRef]

57. Tsay, M.-Y.; Liu, Z.-W. Analysis of the patent cooperation network in Global Artificial Intelligence Technologies based on the assignees. *World Pat. Inf.* **2020**, *63*, 102000. [CrossRef]

58. Liu, N.; Shapira, P.; Yue, X.; Guan, J. Mapping Technological Innovation Dynamics in artificial intelligence domains: Evidence from a global patent analysis. *PLoS ONE* **2021**, *16*, e0262050. [CrossRef] [PubMed]

59. Choi, H.; Oh, S.; Choi, S.; Yoon, J. Innovation Topic Analysis of Technology: The case of augmented reality patents. *IEEE Access* **2018**, *6*, 16119–16137. [CrossRef]

60. Jeong, B.; Yoon, J. Competitive Intelligence Analysis of augmented reality technology using patent information. *Sustainability* **2017**, *9*, 497. [CrossRef]

61. Evangelista, A.; Ardito, L.; Boccaccio, A.; Fiorentino, M.; Messeni Petruzzelli, A.; Uva, A.E. Unveiling the technological trends of Augmented Reality: A Patent Analysis. *Comput. Ind.* **2020**, *118*, 103221. [CrossRef]

62. Janavi, E.; Emami, M. A co-citation study of Information Security Patents in the USPTO database. *Libr. Hi Tech* **2020**, *39*, 936–950. [CrossRef]

63. Daim, T.; Lai, K.K.; Yalcin, H.; Alsoubie, F.; Kumar, V. Forecasting technological positioning through technology knowledge redundancy: Patent citation analysis of IOT, cybersecurity, and Blockchain. *Technol. Forecast. Soc. Chang.* **2020**, *161*, 120329. [CrossRef]

64. Wustmans, M.; Haubold, T.; Bruens, B. Bridging trends and patents: Combining different data sources for the evaluation of Innovation Fields in Blockchain technology. *IEEE Trans. Eng. Manag.* **2022**, *69*, 825–837. [CrossRef]

65. Zhang, H.; Daim, T.; Zhang, Y.P. Integrating patent analysis into technology roadmapping: A latent Dirichlet allocation based technology assessment and roadmapping in the field of Blockchain. *Technol. Forecast. Soc. Chang.* **2021**, *167*, 120729. [CrossRef]

66. Takano, Y.; Mejia, C.; Kajikawa, Y. Unconnected Component Inclusion Technique for Patent Network Analysis: Case Study of Internet of things-related technologies. *J. Informetr.* **2016**, *10*, 967–980. [CrossRef]

67. Lei, L.; Qi, J.; Zheng, K. Patent analytics based on feature vector space model: A case of iot. *IEEE Access* **2019**, *7*, 45705–45715. [CrossRef]

68. Mazlumi, S.H.; Agha Mohammadali Kermani, M. Investigating the structure of the internet of things patent network using social network analysis. *IEEE Internet Things J.* **2022**, *9*, 13458–13469. [CrossRef]

69. Trappey, A.J.; Trappey, C.V.; Fan, C.-Y.; Hsu, A.P.; Li, X.-K.; Lee, I.J. Iot patent roadmap for smart logistic service provision in the context of industry 4.0. *J. Chin. Inst. Eng.* **2017**, *40*, 593–602. [CrossRef]

70. Trappey, A.J.C.; Trappey, C.V.; Hareesh Govindarajan, U.; Chuang, A.C.; Sun, J.J. A review of essential standards and patent landscapes for the internet of things: A key enabler for industry 4.0. *Adv. Eng. Inform.* **2017**, *33*, 208–229. [CrossRef]

71. Wang, Y.-H.; Hsieh, C.-C. Explore technology innovation and intelligence for IOT (internet of things) based Eyewear Technology. *Technol. Forecast. Soc. Chang.* **2018**, *127*, 281–290. [CrossRef]

72. Ardito, L.; D'Adda, D.; Messeni Petruzzelli, A. Mapping innovation dynamics in the internet of things domain: Evidence from patent analysis. *Technol. Forecast. Soc. Chang.* **2018**, *136*, 317–330. [CrossRef]

73. Li, X.; Pak, C.; Bi, K. Analysis of the development trends and innovation characteristics of internet of things technology—Based on patentometrics and Bibliometrics. *Technol. Anal. Strateg. Manag.* **2019**, *32*, 104–118. [CrossRef]

74. Gould, R.V.; Fernandez, R.M. Structures of mediation: A formal approach to brokerage in Transaction Networks. *Sociol. Methodol.* **1989**, *19*, 89. [CrossRef]

75. Park, Y.-N.; Lee, Y.-S.; Kim, J.-J.; Lee, T.S. The structure and knowledge flow of building information modeling based on Patent Citation Network Analysis. *Autom. Constr.* **2018**, *87*, 215–224. [CrossRef]

76. Huang, M.-H.; Chang, H.-W.; Chen, D.-Z. The trend of concentration in scientific research and Technological Innovation: A reduction of the predominant role of the U.S. in World Research & Technology. *J. Informetr.* **2012**, *6*, 457–468.

77. Michel, J.; Bettels, B. Patent citation analysis. A closer look at the basic input data from patent search reports. *Scientometrics* **2001**, *51*, 185–201. [CrossRef]

78. Krestel, R.; Chikkamath, R.; Hewel, C.; Risch, J. A survey on Deep Learning for patent analysis. *World Patent Inf.* **2021**, *65*, 102035. [CrossRef]

79. Tseng, Y.-H.; Lin, C.-J.; Lin, Y.-I. Text mining techniques for patent analysis. *Inf. Process. Manag.* **2007**, *43*, 1216–1247. [CrossRef]

80. Bessen, J. Estimates of patent rents from firm market value. *Res. Policy* **2009**, *38*, 1604–1616. [CrossRef]

81. Hall, B.; Jaffe, A.; Trajtenberg, M. Market Value and Patent Citations: A First Look. *Rand J. Econ.* **2000**, *36*, 16–38.

82. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

83. Röder, M.; Both, A.; Hinneburg, A. Exploring the space of topic coherence measures. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, 2–6 February 2015.

84. Barua, A.; Thomas, S.W.; Hassan, A.E. What are developers talking about? an analysis of topics and trends in stack overflow. *Empir. Softw. Eng.* **2012**, *19*, 619–654. [CrossRef]

85. Cox, M.A.; Cox, T.F. Multidimensional scaling. In Handbook of Data Visualization; 2008; pp. 315–347.

86. Kleinberg, J.M.; Kumar, R.; Raghavan, P.; Rajagopalan, S.; Tomkins, A.S. The web as a graph: Measurements, models, and methods. In Proceedings of the International Computing and Combinatorics Conference, Tokyo, Japan, 26–28 July 1999; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 1999; pp. 1–17.

87. Kleinberg, J.M. Authoritative sources in a hyperlinked environment. *J. ACM* **1999**, *46*, 604–632. [CrossRef]

88. Cohen, W.M.; Goto, A.; Nagata, A.; Nelson, R.R.; Walsh, J.P. R&D spillovers, patents and the incentives to innovate in Japan and the United States. *Res. Policy* **2002**, *31*, 1349–1367.

89. Zhao, L.; Wang, X.; Wu, S. The total factor productivity of China's software industry and its promotion path. *IEEE Access* **2021**, *9*, 96039–96055. [CrossRef]

90. Iyer, A. Moving from industry 2.0 to industry 4.0: A case study from India on leapfrogging in Smart Manufacturing. *Procedia Manuf.* **2018**, *21*, 663–670. [CrossRef]

91. Prause, M. Challenges of Industry 4.0 technology adoption for smes: The case of Japan. *Sustainability* **2019**, *11*, 5807. [CrossRef]

92. Beyer, S.; Macho, C.; Di Penta, M.; Pinzger, M. What kind of questions do developers ask on stack overflow? A comparison of automated approaches to classify posts into question categories. *Empir. Softw. Eng.* **2019**, *25*, 2258–2301. [CrossRef]

93. Ji, Y.; Yu, X.; Sun, M.; Zhang, B. Exploring the evolution and determinants of open innovation: A perspective from patent citations. *Sustainability* **2022**, *14*, 1618. [CrossRef]

94. Duguet, E.; MacGarvie, M. How well do patent citations measure flows of technology? evidence from french innovation surveys. *Econ. Innov. New Technol.* **2005**, *14*, 375–393. [CrossRef]