HAL open science

Patents information for humanities research

David Reymond

▶ To cite this version:

David Reymond. Patents information for humanities research. Iberoamerican Journal of Science Measurement and Communication, 2020, 1 (1), 10.47909/ijsmc.02 . hal-03010273

HAL Id: hal-03010273

https://hal.science/hal-03010273v1

Submitted on 18 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

ORIGINAL ARTICLE

Patents information for humanities research: Could there be something?

David Reymond [1]

How to cite (APA): Reymond, D. (2020). Patents information for humanities research: Could there be something? *Iberoamerican Journal of Science Measurement and Communication*; *1*(1). https://doi.org/10.47909/ijsmc.02

Received date: 2 June 2020

Accepted date: 21 September 2020

Handling editor: Carlos Luis González-Valiente

Copyright: © 2020 Reymond. This is an open access article distributed under the terms of the CC BY-NC 4.0 license which permits copying and redistributing the material in any medium or format, adapting, transforming and building upon the material as long as the license terms are followed.

ABSTRACT

Latour and co-authors proposed, in the Science and Technology Translation theory, to target the many Social and Human Science (SHS) questions addressed by social studies of sciences by considering, in complement to traditional academic matters, the complete social environment (political, economic or societal). Patents are a potential primary information source to do so. We propose to extend this considering that recent changes have evolved in our capacity to do so. We propose three preliminary steps: (a) patent documents as providing a structured information source, (b) a patent database as a technical encyclopedia, and (c) the recent expansion of the variety of uses and users in patent domains. We underline, furthermore, that minority research in the academic space does effectively use patent information, especially in SHS compared to other disciplines. We deliver an experiment to estimate the amount of data unconsidered by not questioning the huge database of the European Patent Office. By comparatively considering the terminology of the two branches of the Unesco thesaurus, namely the micro thesauri “Social and Human Sciences" and the “Information and Communication Science” branches, we evaluate a database response to the whole vocabulary. An in-depth analysis of one selected concept will complete the study to exemplify our proposal. Results show that patent information may provide several documents for a wide range of academic research questions aside from the Social Studies of Science: from strategic to state of the art, position advances, historical perspectives, cultural and market analysis, and so on. The free open-source tool is also a way to practice digital humanities expected skills in real-world corpora.

Keywords: Actor-Network Theory; Patent; Data mining; Social human science; Patent2Net

1 University of Toulon, France. Email: david.reymond@univ-tln.fr

INTRODUCTION

Patent information is reserved, traditionally, to experts, connoted to the specific world of industrial property (IP) to protect inventions and to apply laws. In academic and scientometrics studies, the role of patent information for highlighting fundamental science and applied science has been clearly stated for a long time as a primary source for the Triple Helix model (Leydesdorff, 2004). There have been many studies, from those trying to shed light on the links between science and technology (Etzkowitz & Leydesdorff, 1997; Barré & Laville, 1994; Gibbons, 1994) to those even questioning the view of science as a public good (Callon, 1994). Patent statistics feed traditional bibliometrics analysis within citations studies for instance (Li et al., 2014) or classifications (Leydesdorff, 2008) were STEM disciplines are omnipresent. In this flow, Glänzel and Meyer showed that Chemistry related subfields and Drugs and Medical patents are the most cited patents (Glänzel & Meyer, 2003), so chemists have recognized the value of patents as a source of chemical information (White, 2014); while we can also understand the absence from these studies of human science fields.

Latour *et al*. (2013), in the fundaments of the translation theory, recommend the use of patent information in social studies of science (SSS) in the actor-network theory (ANT), the anthropology of science, the ethnology of socio-technical process or usages, and even innovation studies. Encompassing a wide scope of Social and Human Science (SHS) studies, they promoted patent information to take into account the social environment (economic, political, and usage) denoted by Bourdieu and Nice (1977). Despite this recommendation, patent documents are neglected by academics from educational programs (Durand-Barthez, 2009), and are even neglected in the information search process by researchers in general (Quoniam, 2013). Patent analysis can answer many questions beyond those of mere technological intelligence (Hidalgo-Nuchera et al., 2009) that can simply address science in action (Latour, 2003) pragmatics. The scope of applications, as in bibliometrics studies, covers the impact of social structure on individual creativity (Fleming et al., 2007), the diffusion of knowledge (Singh, 2005), regional dynamics (Breschi & Lissoni, 2009), technology mapping (Leydesdorff et al., 2015; Saheb & Saheb, 2020; Valencia-Zuluaga et al., 2017), and the forecast of employment (Dou et al., 2012).

In the scientometrics field, patent analysis (named in this field as patentometrics or technometry (Polanco, 1995)) provides indicators for assessing the development of research groups and technology fields (Oppenheim, 2000) or helping research policy (Barré & Laville, 1994; Mortensen, 2011), or study innovation (Kabla & François, 1994; Suh and Jeon 2019), but also product commercialization (Wagner & Wakeman, 2016) and so on: See He and Tian (2020) for a review of literature linking institutions and innovation, see also Leydesdorff et al. (2020) for a perspective view bridging/promoting the use of quantitative and qualitative ways in human science research. Aside the tradition opposing quantitative approaches (Bowker et al., 2014; Bowker 2020), we believe that the historical context of patent information searches, -as most databases are not free of access-, (Adams, 2011), the technology-oriented information and the lack of tools for searching may be the principal causes of the disregard of this information kind. Therefore, we try hereafter to estimate the number of documents “lost” globally to human social sciences using a wide scope experiment.

The following paper is structured into two main parts. We start with basic principles, reminding about patents, then present the main freely accessible database of the European Patent Office, which we see as a technical encyclopedia. We remind then, after many works that show the possible enlargement of patent information targets across several studies, that there is a need for the expansion of users and usages that motivates the need for educational programs, besides research activity on patent mediation. The second part concerns the main

question of the experiment, situating it in terms of its global implication, the methodology developed, and the results.

A patent is described as a highly structured document covering several states (see Table 1) and recorded by the database. In short, there are the filing, publishing, granted, and expired stages. At each step, before granting and after, the document is checked by experts whose role is also to complete the content with a rich metadata set. Hence, each phase of the document may disclose different kinds of information. At the first stage of its construction, a patent document must describe the invention clearly within an abstract, title, and primary classification. Then, a background section must describe the problem to be solved and the current state of the art. Finally, claims are written in legal terminology, and the style is referred to as patentese (Lupu & Hanbury, 2013).

Phase	When	Information disclosed
Patent application	Filing date	Some metadata (min: title, IPC, abstract, applicant)
Patent application – published	18 months after the filing date	Full text and metadata
Granted patent	2-5 years or more after the filing date	Amended full text and revised metadata
Expired or ceased patent (public domain)	20 years after the filing date (or before in case of unpaid fees)	Amended full text and revised metadata

Table 1. Life-cycle of patent documents (adapted from Bonino et al., 2010).

The World Intellectual Property Organization maintains a standard hierarchical taxonomy for patent classification named International Patent Classification (IPC). Updates are provided every three months for the deepest levels and every three years for the core classification. The IPC is adopted by more than 100 nations worldwide (Bonino et al., 2010). The IPC classification constitutes a rich, controlled taxonomy to describe inventions. It is used by experts as an international and independent language (Wongel, 2005). Searching patents using the classification scheme allows the user to make abstraction of the primary language of the patent. Hence, the classification system is considered as a good alternative to full-text research and is not proposed for patent information retrieval (Stephen Adams, 2010b). To further refine a patent search, EPO and USPTO recently started (Wongel, 2005) to fuse their classification in a common classification system compatible with the IPC structure, named the Cooperative Patent Classification.

As anyone can write a patent, this checking process, from expert evaluation to expert classification means, for each step if granted, the convergence of the patent document into a new document version. This process, similar in some sense to the traditional evaluation process for academic documents, results in the patent document having a non-negligible quality and a bibliographically consistent metadata description. We believe that some patents may reflect the state of the art in research activity, such as social tendencies (political or economic) or needs.

A freely accessible patent database

Among the many freely accessible patent databases on the web (see World Intellectual Property Organization, 2009, p. 5–6), Esp@cenet is a free patent search service offered by the European Patent Office (EPO). Besides, esp@cenet also permits a text search

of European and PCT applications in English, German, and French and offers search possibilities through bibliographic data (INPADOCDB) and a legal status data set (INPADOC). The “worldwide” database currently centralizes over 89 databases and 50 patent authorities. Esp@ceNet lacks tools for analyzing data, but, in 2006, the European Patent Office introduced (Kallas, 2006) an application programming interface (API) to their databases called the Open Patent Service (OPS). The API allows a specific crawler to get access to all the content provided by the database, freely under conditions of fair-robot and fair loading of a server’s compliance. Using the API, one can download 2.5 GB per week.

The Esp@ceNet “worldwide” database has been used previously to examine patents in several fields, like for example genetic and biological material (Oldham & Cutter, 2006), the globalization of knowledge (Andersen et al., 2006) and nanotechnology (Dang et al., 2010). Kallas (2006) suggested using the OPS API to extend the usages of a patent document. The Esp@ceNet “worldwide” database is the world's largest free collection of technical information.

A patent database as a technical encyclopedia

In section II (“The uniqueness of patents as a technological resource”) of a report, the US Department of Commerce (Office of Technology Assessment and Forecast, 1977, p. 3) concludes by claiming that more than seven out of every 10 patents are not the subject of any publication in the non-patent literature and that 84% of US patents contain technology not disclosed in the non-patent literature. A patent proposal must be unique to be granted. Patent information is also typically more detailed and exhaustive than that of scientific papers (Bonino et al., 2010). The language of patents is unique and contains highly specialized or technical words not found in everyday language (Joho et al., 2010). Patent documents are often written in “patent jargon.” They are difficult to utilize and understand (Paranjpe, 2012). Finally, as claimed in their website[2], “Espacenet offers free access to more than 120 million patent documents worldwide, containing information about inventions and technical developments from 1836 to today”. One can consider this database as a huge technical encyclopedia, centralizing a worldwide collection of documents that are controlled by experts and classified by one classification taxonomy. These points make us view a patent database as a technical encyclopedia (Quoniam, 2013).

But it is an encyclopedia, which is not obvious in terms of its usage. First, there is the limitation of the tools in helping users (highly technical software, little of it free of charge, limited full-text search, the needed IPC (International Patent Classification) or CPC (Cooperative Patent Classification) skills, etc.). Secondly, there is the complexity of the data provided that may often suppose the need for interdisciplinary collaboration as well.

Patent information (mis)uses

Educational mis-usage

It is well known that patents are, in general, a misused information source (Adams, 2011). Durand-Barthez (2009) denoted the need to start educational programs in patent analysis for STEM students in France. This idea was followed by Stembridge (2010): “There is a need both for a formal education program to sit alongside the experience gained in practice and an agreed standard to provide formal recognition of the knowledge, skills, and experience vested in the competent patent information profession”.

Besides this, for scientific research Quoniam (2013) noted the need to use patents as a

research object. The author, among others, underlined that patent documents can be useful to determine the state of the art in a domain as an information source for innovation but can also be used in library services (Feng & Zhao, 2015), for instance. All the authors noted the need for a patinformatic tool to fill the gap in academic research and education as usual software in this domain is very expensive as is access to private databases. See Breitzman and Mogee (2002), and Zhang *et al*. (2015) for a comparison of several patent databases and patent analysis tools. We note that specific key points as value added by these databases are, for educational purposes, marginal.

The spread of patent uses: the variety of users

Despite their problems, for instance, Optical Caracter Recognition (OCR) that affects mainly full-text searches (Stephen Adams, 2010a), patent documents can be used to help in many technical problems. Traditionally, pure patent professionals in large companies use patent information; information specialists, private patent searchers, patent examiners, and patent lawyers are intended to be the specific users (Adams, 2011). Others have identified some examples of new kinds of users for different motivations of information needs:

- Workers at Small and Medium Enterprises (SME) (Dou, 2010; Dou et al., 2015) want to find out how their competitors have tackled a similar problem.
- Academic researchers who need to get an in-depth understanding of a research topic and need to go beyond theory and work towards technical solutions can get assistance from patent applications (Feng & Zhao, 2015; Quoniam, 2013).

All R&D people, business analysts, and managers from every company would extend this list today. Furthermore, startups, or even developing countries, are also concerned; they may need the use of patent information for “frugal innovation”. Corrêa & Gomes (2007) showed that the majority of patent documents (95%) remain in the public domain. This is due to many reasons; some expired over time or, due to nonpayment of charges due, the invention was made freely accessible. This information, which is to be encountered, could be valuable for developing countries or developing businesses. Here, we defend the use of the serendipity mode in the research activity.

Hence, one could note the extension of a variety of users, from experts to occasional users, users who handle different backgrounds and interests ranging from pure science to business (Bonino et al., 2010). These last users (occasional) and uses (casual, technical reference) refer to kinds of uses and users with objectives that are situated opposite to those usually studied (professionals of patents) in surveys or tasks of information retrieval for patents (Abbas et al., 2014; Breitzman & Mogee, 2002; Paranjpe, 2012) citing just some of them.

Do academic studies miss something?

As previously noted, the absence of academic research into patent information seems to be general, except for the case of particular fields (chemistry and drugs), especially for humanities. Traditional and cultural heritage, the noted difficulty of access to data – patent offices and paid databases – may have kept patent documents away from human science disciplines. For scientometrics, patents are the expression of the direct link between science and society (Barré & Laville, 1994) and between applied science and fundamental science or technology and science. Hence, patent documents are, in some way, a certified telltale manifestation of the links among technology, science, and society. As many applications concern utility systems for people, we suppose that human science discoveries are directly applied by patents. Clearly, in our point of view, where human beings are concerned, patents protect inventions to make business and perhaps constraint innovation but deliver precious

information about how science takes action in society.

Toward answering the previous hypothesis, many questions appear to be addressed to exploiting patent information literature, for instance, understanding how SHS knowledge is propagated pragmatically in society, or how new digital humanities artifacts or algorithms (games, tracking systems, finance prediction, criminology) transform it. Even more generally, patent documents report about the realm of technologies or applications and could be addressed in the state of the art research questions. Taking into account academic papers and patent information for such questions opens the research process to consider the problem, from one hand, with the knowledge that covers or surrounds the problem and, on the other hand, with the effective materiality and applications that link from or to the problem. As the database covers patent documents from 1836, some historical technology studies could be developed also.

From a theoretical point of view, addressing these questions this way stands for developing an in-depth analysis of the heterogeneous relations proposed by Latour's heterogenesis of the traduction-transformations, or, similarly, all the derivative theories that are centrally shaped in an information-seeking process. For instance, insights into technology dynamics, societal embedding, or societal dynamics are more generally discussed in the Constructive Technology Assessment (Schot & Rip, 1997), or even the ANT theory (Akrich et al., 2013).

To exemplify this point of view, we propose to construct an estimation of the amount of information available for SHS studies in patent information documents. To do so, we will not make a distinction in documents between filed applications and granted patents; in effect, this separation is not globally interesting, as it would insert the several granting motivations for patents as a bias against manifestations of the pragmatic propagation of science.

METHODOLOGY

We use, as a primary source, the UNESCO thesaurus. We extracted all concepts using ad-hoc python scripts, in a lexical form, in the micro thesauri "Social and Human Sciences" (SHS) and "Information and Communication Science" (ICS) (see http://skos.um.es/unescothes/). As straightforward evidence of a highly responsive capacity in indexes of single terms as opposed to those of multi-term concepts, we distinguish two sets. The first one (ST) contains the single terms (one term length description concepts), and the second (MT) is constituted by multi terms (concepts that are described by several terms). Table 2 shows a statistical description of the differently constituted corpora.

In a second step, for each term in the corpora, we build the request to the OPS worldwide database to retrieve the number of patents forming the entire patent universe attached to the concept in titles or the abstracts. This request takes expression abstractly in the CQL language using the formula 'TA="concept"', and, in a hacking expression, takes form in the API language: http://worldwide.espacenet.com/searchResults?DB=EPODOC&query=TA%3Dconcept. We used for this an inspired script from The Patent2Net (P2N) solution which offers the possibility to build up patent corpora (Reymond et al., 2018) from the worldwide database of the European Patent Organisation: gathering, cleaning, and processing data for several "distance reading" (Moretti, 2013) tools P2N is open source[3] and ready to use in educational programs (Reymond & Quoniam, 2018). The tools suite allows the practice of key instruments and skills in the field of digital humanities directly addressing the volume of data and linking. P2N is used also for building and processing the in-depth analysis section.

Micro thesauri	Single term - ST	Multi terms – MT
Social and human science	157 (SHS-ST)	256 (SHS-MT)
Information and communication science	757 (ICS-MT)	1607 (ICS-MT)

Table 2. The quantitative description of the source data sets in the number of lexical entries.

Requesting the database and imposing these words in the abstract or just the title avoids the eventual biases caused by claims (expressed in juridical terminology) or the huge descriptions (which are allowed many external references, examples, and so on). By construction, we know also that building the request this way is not the appropriate way to match the expected range of patents around the concept for many reasons. Some of this is due to ambiguity ("population policy", "target group"…), and, in other cases, is due to the terms being too generic ("intelligence", "memorization", "thinking", "behavior"). However, within these worst cases, some of the concepts may directly address the concept, and this simple query will return most of the patent universe ("adult learning", "interpersonal relations", "quality of life" and so on). Our objective is to estimate or prove the existence of the eventual missing documents in SHS studies that could be used for many purposes. We argue for some scientific research activity, position a relevant research strategy, and discover how to apply the strategy. Identifying a good request form for each concept is beyond our central interest hereafter. We are proceeding in this way to dress up the global profile of each concept addressed and, in consequence, despite obvious biases, partially cover the entire thesaurus. In a second step, an in-depth study of one relevant term extracted from these sets will complete statistical analysis. Clearly, at this point, we expect many null set responses and many large response sets. We also expect a global prominence of ICS terminology and, in general, more silence than responses in the multi-terms set.

RESULTS

As expected, some words or groups of words return a high number of responses (from ICS, "Internet", "drying", "coal", "temperature", "energy consumption", "data processing", "carbon dioxide", and from SHS, "patients", "housing", "poor", "adoption") but no multi-term universe was present in this last case. Reciprocally, null results sets were returned for theoretical concepts ("Sociologists", "antisemitism", "ethnology", "liberalism", "militarism"). Huge data set responses, as well as many of the null responses where the terms may not be relevant, are simply noisy and should not be considered. Nevertheless, some theoretical concepts do return acceptable and consistent sets (from 100 up to 1000 documents), i.e. "Philosophers", "economists", "democratization". We quantified data response sets and split them on a logarithmic scale from null sets (empty responses) to more than 100000 documents (This is the limit of OPS API, whatever the length of a response set, the system will return the number "100000").

	num=0	0<num<10	9<num<10^2	99<num<10^3	999<num<10^4	9999<num<10^5	>10^6
Single term	35	65	117	166	214	202	115
Multi term	458	370	404	359	196	65	11
SHS	85	82	103	72	45	20	6
ICS	408	353	418	453	365	247	120

Table 3. The number of words in each patent universe of length num. A logarithmic scale splits the responses sets.

We now consider the distribution of results in this scale using several kinds of log-scale sets. Table 3 shows the number of terms corresponding to each partition. We can see that 85 terms in the SHS set provided no response, and 408 terms in the ICS set provided no response. 35 single terms and 458 multi terms also provided no response and so on. As the data sets are equivalent in terms of the number of words, we cannot conclude about tendencies; we provide results just to consider the number of responses. Many terms or multi-terms return large responses. However, it appears that many terms correspond to acceptable sets (columns in the middle of Table 3).

Figure 1 shows the percentage of concepts (single terms or multi terms) in the length of response sets. SHS and ICS are also represented globally. As the original data sets (ICS and SHS) are not of the same length, we provide a percentage that normalizes the data set for our question. Curbs show the same tendencies for single or multi-term concepts whatever the thematic. 25 to 30% of words do not return anything. 4 to 15% do return too large data sets. However, 60% appears in the human acceptable/readable length set. We believe that this is probably due to patent writing rules as the effects of patents (their pragmatics application) are probably expressed with these words. MT also has the same tendency in both sets as expected response decrease in length very quickly. Considering the origins of the terminology, globally, SHS terms respond well to several results. ST and MT sets have a similar response behavior with a slightly higher rate with SHS terminology for more than 50% of the terms.

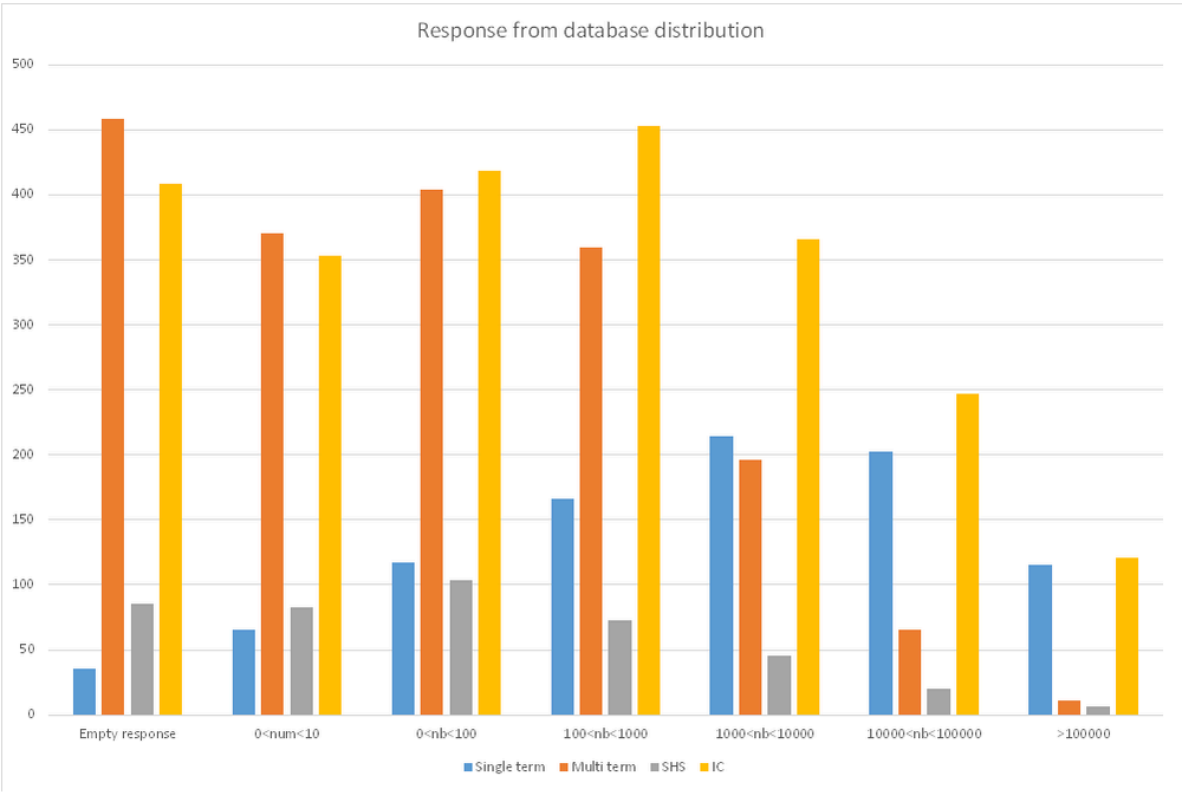

Figure 1. Distribution of the number of words corresponding to the length of data sets responses in log scale representation.

As the content of the patent universes is probably noisy, we also guess that the representation of these terms is evidence of the application of patents in human life. In other words, the

presence of these words in a patent description does not directly concern the concept, but the application of a patent may target the concept. We provide hereafter an in-depth analysis of one patent universe from the SHS corpus.

In-depth analysis: The corpus TA="Social interaction", 516 patents (26/08/2020)

As a suggestive approach to decipher the content of the corpus, we use a document-processing technique to describe the content of patent corpora built using the Patent2Net software (Reymond et al., 2018). We use the request seeking for "Social interaction*" in Title or Abstracts and get 516 patents. We then explore these documents by covering three axes: a textometry description of the contents (the micro view), a k-means classification and automatic class labeling description of classes (meso view), and, finally, a projected mind-map of the classification scheme extracted from the universe (meta-view).

Figure 2 displays the word extracted from the abstracts of the patent universe. Using the interface of the Iramuteq (Marty et al., 2013) tool we avoid the display of non-relevant words, those words that are trivial and too frequent in a patent abstract and descriptions such as "invention", "presents", "describes". In this set, we take a threshold of 15 as a minimum of occurrence to be displayed. Iramuteq, calculates, for each word, its co-occurrence and displays them at a distance (here chi² distance) corresponding to the level of cooccurrence, showing the words proportional to their apparition frequencies and their connected words in the patent abstracts. One can easily see the three main one-term topics "user", "social" and "interaction" but also the terminology set surrounding them as valuable information to describe potential relation between terms.

We notice also the "interaction" pole represent patents dealing with sensors, games, screen or robots (left side) while the term "social" is connected to a variety of concepts from individual (physical, real or virtual) identification, embodiment (person) or age (child) or activity. The term 'oligosaccharide' shows also that noise (regarding pure "social interaction" in the technologic world) may come from the use of both terms in the abstract as an instance, not as an application (in this case it medicine to improve "social interaction" in behavioral point of view (see Garssen et al., 2020; or Yongde 2011, for instance).

Aside from this terminology description, Figure 3 represents the results of the classification of patent abstracts using the k-mean algorithm (see Noh et al. (2015) for a discussion on the interest of such a method) on the same data set. Helped by a previous projection, we expected nine classes to be discovered by the classification tool Carrot2 (Osiński & Weiss, 2005). The software displayed 25 classes; the most important covered games, personal data, and messages, videos, and media, (at the center of the figure) separating them into virtual worlds, devices, and services platforms and matching algorithms. The tool provides hence an interface to documents in each subset class, offering a way to explore specific subsets.

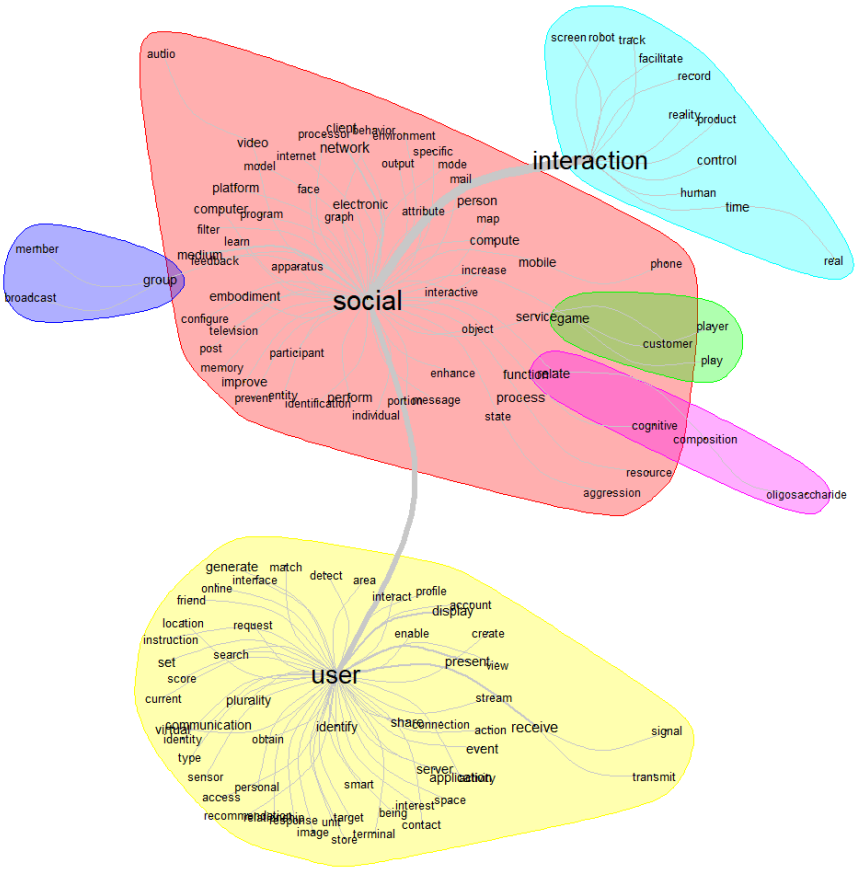

Figure 2. Similitude textometry treatment and representation with Iramuteq. This figure shows the terminology relationship that can help to appreciate concepts interrelations in the technological domain.

We will terminate this characterization by displaying, in Figure 4, the mind-map projection of IPC (International Patent Classification) codes retrieved from the patent universe. Each code is replaced in the diagram by a humanly readable description. The IPC[4] is also a normalized thesaurus, a hierarchical system of language independent symbols. As in precedent projections, the font police size is proportional to frequency. We retrieve here the assertion previously proposed: three big predominant classes (H04L: transmission of digital information (100 patents), A63F: games (36 patents), and G06F: digital data processing (296 patents)). The patent in these fields shows how technology provides user social interactions (essentially in data transmission and games) and how to do technology stores or process it (the data processing field).

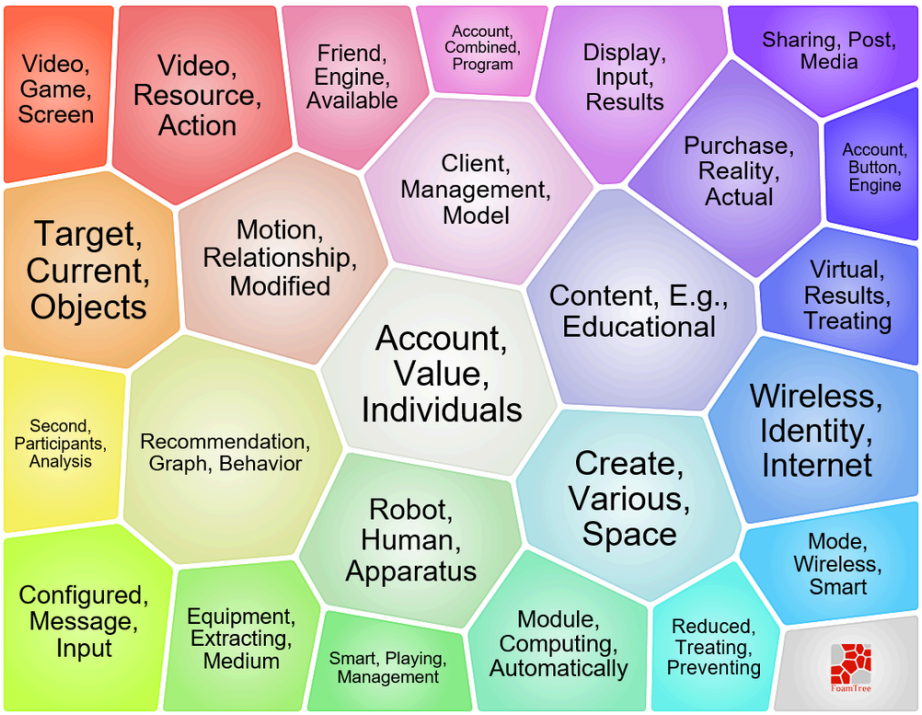

Figure 3. The patent abstract classes created using carrot2 software, k-mean algorithm, 25 classes. Showing quite clearly the main components of social interactions in the technological point of view.

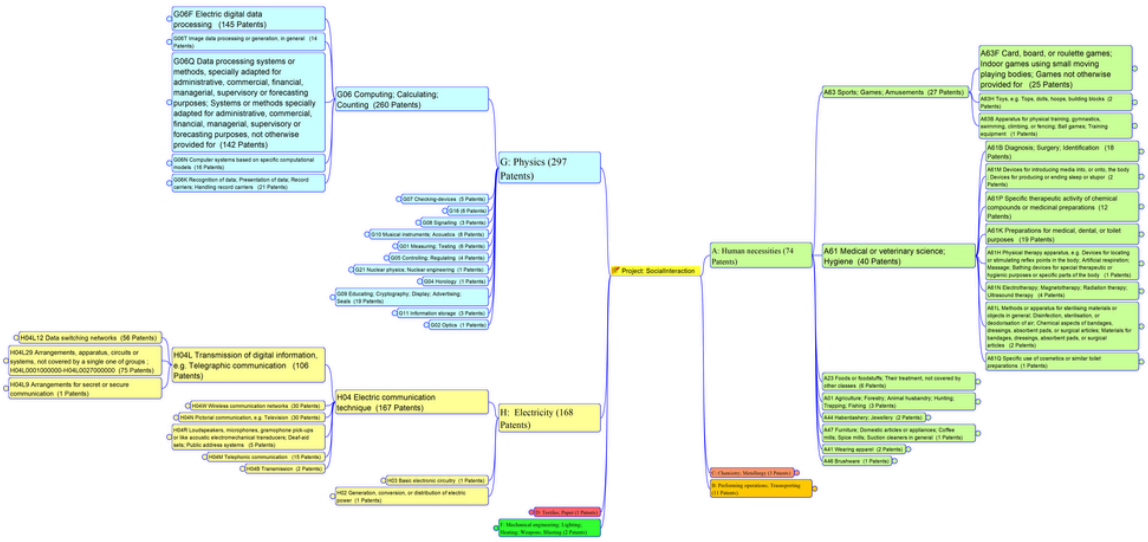

Figure 4. Patent classification mind map of the patent universe.

One can easily project that deeper analysis can be done also at this step: the previous “distance reading” three tools helps in exploring and classifying the corpus and allow also to represent Latour’s translation of the Actor-Network Theory, hence offering a way to illustrate

society and technology interrelations and then focus on the specific actor (documents, countries or applicants or even words) that leads to relevant discussions.

DISCUSSION AND CONCLUDING REMARKS

Promoting the use of patent information in more educational programs to identically cover STEM and human science disciplines, we follow Callon and Latour's recommendations in the ANT approach by stating that the non-usage of patent information may result in a lack of many SHS questions. The experiment showed statistically that SHS researchers might find valuable data extracted from patent databases for different research purposes. Many terms from the Unesco thesaurus refer to quite large data patent sets. We show that this information is valuable using a hybrid exploration methodology, from distance reading to in-depth analysis (with a real preparation of a query for a patent search purpose). SHS terminology has proven a better response rate as expected: as patents are solutions to human problems, the technological description of an invention is contextualized in social words.

Hence, in-depth analyses for our case: "social interaction showed by major patents documents subsets that are concerned how we support, store and technically process social interactions and games, which appears to be the primary tools in social interactions. To generalize this position, in a complement of traditional academic state of the art methods, we claim that addressing the patent literature stands for developing an in-depth analysis of the heterogeneous associations" (Cambrosio et al., 2020) that constitute (hyper)networks (Latour's heterogenesis of the translation-transformations) as well as insights into technology dynamics, societal embedding, and societal dynamics; texts can be analyzed "as assemblages of terms leading to specific ways of problematizing issues by defining the roles assigned to relevant entities (researchers, substances, tools, and technologies)". Underuse of patent documents may suppose historical consideration, perhaps ideology, but also a lack of tools to search databases.

The Patent2Net (P2N) offers an up to date and state of the art tool and open-source suite for patent analysis[5] and ready to use in research for establishing corpora but also for educational programs (Reymond & Quoniam, 2018): the suite allows the practice of key instruments and skills in the field of digital humanities directly addressing the volume of data and linking. Thus, associated words, document mining, elementary statistics, mapping, network exploration, and concept maps to communicate, share and explore the collected domains (Reymond et al., 2018) can be summoned covering major digital humanities skills (Mayer-Schönberger & Cukier, 2013). Further study may allow the extension of this study addressing the eventual usefulness of patent information here calling for a hybrid approach: quantitative and qualitative studies with interviews of the SHS researchers at work to confirm or not an effective interest of patent information and eventual modes of usage.

Notes

[2] See: https://worldwide.espacenet.com/patent/
[3] Available at https://github.com/Patent2net/P2N-v3
[4] See: https://www.wipo.int/classifications/ipc/en/faq/
5 Available at https://github.com/Patent2net/P2N-v3

Conflict of interest statement

The author declares that there is no conflict of interest.

Statement of data consent

The data generated during the development of this study has been published as supplementary material.

Acknowledgements

Patent2Net development originally started with the collaboration of Professor Luc Quoniam. His advice has contributed to its improvement from the beginning. The author thanks also the anonymous evaluators of the paper and their useful comments.

REFERENCES

Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. World Patent Information, 37, 3-13. http://dx.doi.org/10.1016/j.wpi.2013.12.006

Adams, S. (2010). The text, the full text and nothing but the text: Part 1–Standards for creating textual information in patent documents and general search implications. World Patent Information, 32(1), 22-29. https://doi.org/10.1016/j.wpi.2009.06.001

Adams, S. (2010). The text, the full text and nothing but the text: Part 2–the main specification, searching challenges and survey of availability. World Patent Information, 32(2), 120-128. https://doi.org/10.1016/j.wpi.2009.06.002

Adams, S. 2011. Information Sources in Patents. Guides to Information Sources. De Gruyter Saur. http://books.google.fr/books?id=I_dI0FsL2a4C

Akrich, M., Callon, M., & Latour, B. (2013). Sociologie de la traduction: Textes fondateurs. Paris: Presses des Mines.

Andersen, P. H., Drejer, I., & Waldstrøm, C. (2006). In the eye of the storm-Knowledge appropriation and the globalization of regional competencies. The Proceedings of the Annual Meeting of the Academy of Management.

Barré, R., & Laville, F. (1994). La bibliométrie des brevets: une mesure de l'activité technologique. Economie et statistique, 275(1), 71-81.

Bonino, D., Ciaramella, A., & Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. World Patent Information, 32(1), 30-38. https://doi.org/10.1016/j.wpi.2009.05.008

Bourdieu, P., & Nice, R. (1977). Outline of a Theory of Practice. Cambridge: Cambridge University Press.

Bowker, G. C. (2020). Numbers or no numbers in science studies. Quantitative Science Studies, 927-929. https://doi.org/10.1162/qss_a_00054

Bowker, G., S.L. Star, L. Gasser, and W. Turner. 2014. Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide. New York: Psychology Press.

Breitzman, A. F., & Mogee, M. E. (2002). The many applications of patent analysis. Journal of information science, 28(3), 187-205. https://doi.org/10.1177/016555150202800302

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. Journal of economic geography, 9(4), 439-468. https://doi.org/10.1093/jeg/lbp008

Callon, M. (1994). L'innovation technologique et ses mythes. Gérer et comprendre, 34, 5-17.

Cambrosio, A., Cointet, J. P., & Abdo, A. H. (2020). Beyond networks: Aligning qualitative and computational science studies. Quantitative Science Studies, 1(3), 1017–1024. https://doi.org/10.1162/qss_a_00055

Corrêa, F.C., & Gomes, S.L.R. (2007). A Patente Na Universidade: Sigilo, Transparência e Direito à Informação. In Encontros Nacionais de Pesquisa Em Ciência Da Informação, vol. 8. Salvador, Brazil.

Dang, Y., Zhang, Y., Fan, L., Chen, H., & Roco, M. C. (2010). Trends in worldwide nanotechnology patent applications: 1991 to 2008. Journal of nanoparticle research, 12(3), 687-706. https://doi.org/10.1007/s11051-009-9831-7

Dou, H. (2010). Information Brevet 2.0, Transfert de Technologies, Valorisation Des Ressources. In L. Quoniam, & A. Lucien (Eds.), Intelligence Compétitive 2.0 (181–199). Hermès Science.

Dou, H., Haudeville, B., & Wolff, D. (2015). L'analyse de l'information brevet comme catalyseur de l'innovation et du développement des entreprises. Vie sciences de l'entreprise, (1), 72-91. https://doi.org/10.3917/vse.199.0072

Dou, H., Mohellebi, D., & Kister, J. (2012). L'importance Du Traitement Bibliométrique Des Brevets Pour Développer l'activité Industrielle. Exemple Des Bitumes En Algérie. RIST (Revue Scientifique et Technique).

Durand-Barthez, M. (2013). Former à l'information brevets dans l'enseignement supérieur. Revue internationale d'intelligence economique, 5(1), 25-38. https://doi.org/10.3166/R2IE.5.25-38

Etzkowitz, H., & Leydesdorff, L. (1997). Universities and the Global Knowledge Economy: A Triple Helix of University-Industry-Government Relations. London: Printer.

Feng, J., & Zhao, N. (2015). A New Role of Chinese Academic Librarians—The Development of Embedded Patent Information Services at Nanjing Technology University Library, China. The Journal of Academic Librarianship, 41(3), 292-300. http://dx.doi.org/10.1016/j.acalib.2015.03.010

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. Administrative science quarterly, 52(3), 443-475. https://doi.org/10.2189%2Fasqu.52.3.443

François, J. P., & Kabla, I. (1994). Présentation générale. les enjeux économiques du brevet. Economie et statistique, 275(1), 49-53.

Garssen, J., Van Elburg, R., Bartke, N., & Kraneveld, A. D. (2018). U.S. Patent No. 9,974,816. Washington, DC: U.S. Patent and Trademark Office.

Gibbson, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). The new production of knowledge: the dynamics of science and research in contemporary societies. London: SAGE Publications.

Glänzel, W., & Meyer, M. (2003). Patents cited in the scientific literature: An exploratory study of'reverse'citation relations. Scientometrics, 58(2), 415-428. https://doi.org/10.1023/A:1026248929668

He, J., & Tian, X. (2020, forthcoming). Institutions and Innovation: A Review of Recent Literature. Annual Review of Financial Economics. https://doi.org/10.2139/ssrn.3530214

Hidalgo-Nuchera, A., Iglesias-Pradas, S., & Hernández-García, Á. (2009). Utilización de las bases de datos de patentes como instrumento de vigilancia tecnológica. El profesional de la información, 18(5), 511-519. https://doi.org/10.3145/epi.2009.sep.04

Joho, H., Azzopardi, L. A., & Vanderbauwhede, W. (2010, August). A survey of patent users: an analysis of tasks, behavior, search functionality and system requirements. In Proceedings of the third symposium on Information interaction in context (pp. 13-24). https://doi.org/10.1145/1840784.1840789

Kallas, P. (2006). Open Patent Services. World Patent Information, 28(4), 296–304. https://doi.org/10.1016/j.wpi.2006.03.007

Latour, B. (2003). Science in Action: How to Follow Scientists and Engineers through Society. Cambridge, Mass: Harvard University Press.

Leydesdorff, L. (2004). The university–industry knowledge relationship: Analyzing patents and the science base of technologies. Journal of the American Society for Information Science and Technology, 55(11), 991-1001. https://doi.org/10.1002/asi.20045

Leydesdorff, L., Alkemade, F., Heimeriks, G., & Hoekstra, R. (2015). Patents as instruments for exploring innovation dynamics: geographic and technological perspectives on “photovoltaic cells”. Scientometrics, 102(1), 629-651. https://doi.org/10.1007/s11192-014-1447-8

Leydesdorff, L., Ràfols, I., & Milojević, S. (2020). Bridging the Divide between Qualitative and Quantitative Science Studies. Quantitative Science Studies, 1(3), 918–26. https://doi.org/10.1162/qss_e_00061

Li, R., Chambers, T., Ding, Y., Zhang, G., & Meng, L. (2014). Patent citation analysis: Calculating science linkage based on citing motivation. Journal of the Association for Information Science and Technology, 65(5), 1007-1017. https://doi.org/10.1002/asi.23054

Lupu, M., & Hanbury, A. (2013). Patent retrieval. Foundations and Trends in Information Retrieval, 7(1), 1-97. https://doi.org/10.1561/1500000027

Marty, E., Marchand, P., & Ratinaud, P. (2013). Les médias et l'opinion-Eléments théoriques et méthodologiques pour une analyse du débat sur l'identité nationale. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 117(1), 46-60. https://doi.org/10.1177%2F0759106312465550

Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. New York, USA: Houghton Mifflin Harcourt.

Moretti, Franco. 2013. Distant Reading: The Formation of an Unorthodox Literary Critic. London, New York: Verso.

Mortensen, P.S. (2011). Patentometrics as Performance Indicators for Allocating Research Funding to Universities'. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.7855&rep=rep1&type=pdf

Noh, H., Jo, Y., & Lee, S. (2015). Keyword selection and processing strategy for applying text mining to patent analysis. Expert Systems with Applications, 42(9), 4348-4360. http://dx.doi.org/10.1016/j.eswa.2015.01.050

Office of Technology Assessment and Forecast. (1977). Eighth Technology Assessment and Forecast Report 8. United States: U.S. Department of Commerce, Patent and Trademark Office. Retrieved from https://ia801601.us.archive.org/5/items/technologyas00unit/technologyas00unit.pdf

Oldham, P., & Cutter, A. M. (2006). Mapping global status and trends in patent activity for biological and genetic material. Genomics, Society and Policy, 2(2), 62. https://doi.org/10.1186/1746-5354-2-2-62

Oppenheim, C. (2000). Do Patent Citations Count. In B. Cronin, & B. H. Atkins (Eds.), The Web of Knowledge: A Festschrift in Honor of Eugene Garfield (pp. 405–432). Medford, NJ: AIS Monograph Series.

Osiński S., Weiss D. (2005) Carrot2: Design of a Flexible and Efficient Web Information Retrieval Framework. In Szczepaniak P.S., Kacprzyk J., Niewiadomski A. (eds), Advances in Web Intelligence. AWIC 2005. Lecture Notes in Computer Science, vol. 3528. Berlin, Heidelberg : Springer. https://doi.org/10.1007/11495772_68

Paranjpe, P. P. (2012). Patent Information and Search. DESIDOC Journal of Library & Information Technology, 32(3).

Polanco, X. (1995). Aux sources de la scientométrie. Solaris, 2, 13-78.

Quoniam, L., Reymond, D., & Rey, C. (2014). Frugal Innovation, Innovative Conception, KDD. Journée des Utilisateurs Intellixir. Montpellier, France.

Quoniam, Luc. 2013. 'Le Brevet : Objet de Recherche En Sciences de l'Information et de La Communication'. In Papy, F. (Ed.) Recherches Ouvertes Sur Le Numérique: Approches Pratiques En Information-Communication (pp. 95–114). Paris : Hermès Science Publications.

Reymond, D., & Quoniam, L. (2018). Patent Documents in STEM and PhD Education. In Proceedings of the IEEE EDUCON2018 International Conference, Santa Cruz de Tenerife, Canarias Islands, Spain.

Reymond, D., Faga, R., Molt, A., Vanderlei, C.A., & Quoniam, L. (2018). A Free Open Source Patent Corpus Generator. In Proceedings of the 9th International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction, CLEF 2018. Avignon, France.

Saheb, T., & Saheb, T. (2020). Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. Journal of Big Data, 7(1), 1-26. https://doi.org/10.1186/s40537-020-00287-9

Schot, J., & Rip, A. (1997). The past and future of constructive technology assessment. Technological forecasting and social change, 54(2-3), 251-268. https://doi.org/10.1016/S0040-1625(96)00180-1

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. Management science, 51(5), 756-770. https://doi.org/10.1287/mnsc.1040.0349

Stembridge, R.A. (2010). Education and Certification of Patent Information Professionals in Europe. In A. Twiss-Brooks (Ed.), Special Topics in Intellectual Property (87–93). Washington, DC: American Chemical Society. http://pubs.acs.org/doi/abs/10.1021/bk-2010-1055.ch010

Suh, Y., & Jeon, J. (2019). Monitoring patterns of open innovation using the patent-based brokerage analysis. Technological Forecasting and Social Change, 146, 595-605. https://doi.org/10.1016/j.techfore.2018.01.037

Valencia-Zuluaga, T., Rivera-Rodriguez, S. R., & Sánchez-Ortíz, N. (2017). Potencial en el uso de la consulta de patentes para determinar el estado de la tecnica. Analisis en microredes con energías renovables. Ingeniería Investigación y Desarrollo, 17(2), 16-22. https://doi.org/10.19053/1900771X.v17.n2.2017.7180

Wagner, S., & Wakeman, S. (2016). What do patent-based measures tell us about product commercialization? Evidence from the pharmaceutical industry. Research Policy, 45(5), 1091-1102. https://doi.org/10.1016/j.respol.2016.02.006

White, M.J. (2014). Chemical Patents. In J. Currano, & D. Roth (Eds.), Chemical Information for Chemists (pp. 53–90). Cambridge: Royal Society of Chemistry. https://doi.org/10.1039/9781782620655-00053

Wongel, H. (2005). The Reform of the IPC—Consequences for the Users. World Patent Information, 27(3), 227–31. http://dx.doi.org/10.1016/j.wpi.2005.02.002

World Intellectual Property Organization. (2009). Guide to Technology Databases. L434/11E. Geneva, Switzerland: WIPO Publication. Retrieved from http://www.wipo.int/edocs/pubdocs/en/patents/434/wipo_pub_l434_11.pdf

Yongde, G. (2011). Chinese herbal preparation for adjusting blood fat and controlling fatty liver and hepatocirrhosis. CN102228663. Retrieved from. https://worldwide.espacenet.com/patent/search/family/044841295/publication/CN102228663A?q=cn102228663

Zhang, L., Li, L., & Li, T. (2015). Patent mining: a survey. ACM Sigkdd Explorations Newsletter, 16(2), 1-19. https://doi.org/10.1145/2783702.2783704