

A User-Centered Evaluation of a Web Based Patent Classification Tool

Anastasia Giachanou¹, Michail Salampasis², Maya Satratzemi¹ and Nikolaos Samaras¹

¹ Department of Applied Informatics, University of Macedonia, Greece

² Information Technology Department, ATEI Thessalonikis, Greece

Emails: agiahanou@uom.gr, msa@it.teithe.gr, maya@uom.gr, samaras@uom.gr

Abstract

This paper presents a user-centered study of a web based system that can automatically suggest classification codes with the aim to assist patent examiners on the task of patent classification. The aim of the study is two-fold. Firstly, we aim to obtain a better understanding of the search tactics patent examiners apply when they do classification search. Secondly, we examine the effect of searching at different levels of the classification scheme on classification search performance. For this user study, two conditions were tested. Both systems are web based. However, the two systems differ in their ability to allow patent examiners selecting the level from which the results will be returned. The results show that systems that allow searching at the level of subgroup are more effective for classification search.

1 Introduction

Classification schemes and metadata are heavily used in the patent domain because it is widely recognized that once the work of assigning patent documents into classification schemes is done, the search can be more efficient and language independent [LH13]. International Patent Classification (IPC) is an internationally accepted standard taxonomy for classifying, sorting, organizing, disseminating, and searching patents.

Copyright © 2014 for the individual papers by the paper's authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

In: U. Kruschwitz, F. Hopfgartner and C. Gurrin (eds.): Proceedings of the MindTheGap'14 Workshop, Berlin, Germany, 4-March-2014, published at <http://ceur-ws.org>

The IPC provides a hierarchical system of language independent symbols for the classification of patents according to the different areas of technology to which they pertain. IPC has currently about 71,000 nodes which are organized into a five-level hierarchical system which is also extended in greater levels of granularity [BG11, LH13].

Patent search tasks are challenging and often require many hours or even days to be completed. One of the tasks that patent professionals face almost every day is that of identifying relevant classification codes given a patent application. Selecting the most appropriate classification codes is a fundamental step in the patent domain as it can be used in various tasks including patent classification and prior art search. For example, in prior art search classification codes such as IPC or CPC¹ can be used as a starting point for finding relevant patents [Vij90, Ada00].

To identify all relevant IPCs of a given patent application i.e. to perform a patent classification search could be a difficult, error prone and time-consuming task, especially for a not very knowledgeable patent professional in some technical area. This is usual in small patent offices with not many patent examiners, who are usually asked to examine patents in broad technical areas. Another challenging characteristic is that patent classification is a recall-oriented task because a single missed relevant IPC and extensively a single missed relevant patent can lead to lawsuits due to patent infringement. Considering both the constantly increasing number of patent applications and the fact that IPCs are manually assigned by technical specialists, there is a need of automated or semi-automated patent classification [Smi02, LMTT11].

A number of researchers have presented new methods, algorithms and systems that improve and partially automate patent classification search [Lar99,

¹CPC is the new official classification scheme endorsed by EPO and USPTO

KZ02, GSP13, DVKB13]. Despite the various approaches that have been presented in the patent classification search literature, to the best of our knowledge, none of them has been tested in a user study. All the methods presented have been evaluated using the Cranfield paradigm. However a better understanding of how patent professionals interact with a patent classification search, what are the tactics and the strategies they use, can be much better obtained with a user study [BCST95].

In this paper, we report the results of a user-centered experiment which aimed to identify search tactics adopted during patent classification. Also, we examine the performance of the Multilayer Classification System (MCS) when it is used on classification search. The system is evaluated in comparison to Espacenet classification search system which is a widely used web based patent classification system. The two systems differ as Espacenet, at least in the version made available to the public, allows keyword-based search on a predetermined level (main group) and creates the need of searching classification codes at more specific levels using browsing and scanning strategies in contrast to MCS that allows patent examiners select the level they prefer to search.

2 Related Work

Improving information retrieval performance does not necessarily mean that the user performance will be also improved [ACL05]. Understanding the interaction between the user and the search system is vital for the evaluation of a system [Kel09]. Over the last thirty years, many researchers focused in understanding the information seeking process with the aim to assist in building systems that can satisfy the user needs. Marchionini [Mar97] focused on new system designs and user interfaces which could support multiple information seeking strategies. Issues such as information seeking strategies [Bat89, Bat90, BCST95] and user behavior [Wil97] were also examined.

User studies are of high importance in understanding information seeking process and search tactics. The need for user studies is more intensive for systems used by professionals working within complex information workplace such as patent domain. User studies not only give the opportunity to explore cognitive aspects of the entire process of information seeking but also give a better understanding of a search process and how a specific tool can attain a specific objective of this process. For example, Lupu and Hanbury [LH13] in a recent review of patent retrieval present a typical prior art search use case, analysed in different sub-processes, performed by a patent examiner (pp. 15) to model and better understand prior art search

[LH13].

Some efforts in the field of information seeking and modelling search process on patent domain were already made. Hansen [HJ00] presented a study performed at the Swedish patent office which aimed to examine the relationship between the users work-task and the information seeking and retrieval process. A study by Becks [BGWH10] examined different impact factors on information seeking and discussed how these can be integrated on the user interface of a system. More recently, Joho et al. [JAV10] published the results of a survey on patent users in which they analyzed the relationship between the patent search tasks and the functionalities of patent retrieval systems.

A number of studies also tried to identify patent search strategies. The study presented by Foglia [Fog07] summarized search strategies adopted by patent examiners for the patentability search. The search tactics during a patentability search were also examined by Tseng and Wu [TW08] with the aim to improve the patent search systems. 18 of the 43 patent searches that were involved in the study were also observed of their search behaviour. This study showed that the patent examiners follow an iterative search process to identify the relevant patents.

3 Patent Classification with MCS

In this paper we are concerned with web based systems that can assist patent examiners on the task of patent classification. We have developed a web-based patent classification search system and performed a user-centered experiment to evaluate its performance and the utility of a key design feature, i.e. selection of classification level to search. The tool is compared with another classification search system, which is part of the Espacenet service.

The first system that was used for our study is Espacenet's classification search², a search service offered by the European Patent Office (EPO). Espacenet classification search system is a widely used web based patent classification system and can be used by patent professionals as a supplementary tool during a patent search.

The second system is the MCS patent classification system. MCS is a web-based system that aims to assist patent professionals identify relevant IPC classification codes. The system can automatically suggest IPCs given a query and be used as a classification search tool. MCS implementation is based on the multilayer collection selection algorithm [GSSS13].

The multilayer algorithm is a collection selection method that uses multiple evidence utilizing (when calculating the relevant IPCs/collections) the ranking

²<http://worldwide.espacenet.com/classification>

of ancestors collections in higher level of the classification hierarchy [GSSS13]. Clusters of patents are created based on their manually assigned IPC codes. The patents are first divided using the subclass, group and subgroup level of IPC and then they are allocated to the sub-collections. The system can retrieve relevant classification codes on three different IPC levels: subclass, main group and subgroup.

The basic characteristic of the MCS system is that it allows patent examiners to choose the level they prefer to perform their search. This characteristic differentiates the system from Espacenet which, at least in the version made available to the public, allows keyword-based search on a predetermined level (main group) and imposes the need of searching classification codes at more specific levels using browsing and scanning strategies. Using MCS, patent examiners can view results at a narrow level (subgroup) while using Espacenet they have to manually browse and scan the classification codes. Also, MCS allows users to view the ancestors and their descriptions of the retrieved IPC codes.

4 User Experiment

4.1 Aim

The general aim of this user study is to increase the understanding of the tactics and strategies that patent professionals use when they perform classification search. The second aim is to evaluate the performance of a proposed classification method and algorithm (Multilayer-Condition 2) which utilizes the hierarchical nature of the classification scheme. This method has been tested in previous studies using a number of system-oriented evaluations [GSSS13, GSP13]. In this experiment we aim to evaluate the performance of the method in comparison to a widely used web-based patent classification search system (Espacenet, Condition 1) and also to study the use of one key feature of the method which is the selection of the IPC level to search.

4.2 Participants

Twelve participants voluntarily participated in this study. The participants, seven male (58%) and 5 female (42%), are patent examiners in the Greek Patent Office. The age of the participants ranged from 34 to 54 with an average of 41.8, a standard deviation of 7.7 and a median of 39 years. Additionally, all the participants had very good skills in using the internet in order to look for information and very good skills in understanding English. The skill in understanding English was included in the questionnaire about demographics because the patent documents that were given to the

participants for examination were in English.

4.3 Method

A very short presentation (2-5 minutes) was given to the participants before the test, to help them gain an overview of the MCS system. No training was provided for the Espacenet since it is used very frequently from the patent examiners. With the training session we wanted to ensure the participant had understood the nature of the task that s/he will be asked to perform.

The participants were tested individually. Each participant was randomly given one patent from which the IPC codes had been removed. The task that participants were asked to perform was to examine this patent and find as many IPC codes as possible in 10 minutes. Before executing the task with the patent that was assigned to them, each subject confirmed that had a fair knowledge about the subject of the patent. If the subject declared that had no knowledge about the technical domain or is extremely familiar with it, another patent was selected. Participants were told to perform the tasks, as they would normally do for themselves.

Each participant performed the search task two times using each time one of the systems for classification search: Espacenet and MCS. The participants have experience using the Espacenet system, but they have no experience with the MCS system. Although the two patents given to subjects were coming from a different technical domain, to counterbalance potential effect of prior knowledge the order that participants were asked to use the different systems changed in a round-robin fashion.

Each search session was logged and the data were analyzed. Judged recall and judged precision were calculated to compare the performance of the two systems. *Judged recall* is the proportion of relevant IPCs which are actually retrieved while *judged precision* is the proportion of retrieved IPCs which are actually relevant. The only difference from the traditional recall and precision metrics is that judged recall and judged precision refer to the judgment list that participants produced during the experiment.

5 Results

5.1 Performance of the Systems

Table 1 shows the results of the system in terms of judged recall and judged precision. We have calculated the recall and the precision for the level of main group and subgroup.

Before discussing the results, we would like to make some observations. The evaluation of the results has some difficulties on the level of subgroup which are

Table 1: Effectiveness results of subjects using the Espacenet (condition 1) and the MCS (condition 2)

	Judged Recall		Judged Precision	
	Main Group	Sub-group	Main Group	Sub-group
Condition1	0.28	0.19	0.24	0.15
Condition2	0.45	0.37	0.41	0.32

related to the hierarchical structure of patent collections. All the participants were asked to find classification codes for a patent preferably at the subgroup level but several times subjects were identifying relevant IPCs at a different level because they were not certain. Consequently, there are cases that the participant reached a different level from this that is reported in the patent application. Therefore, we made the following assumption: if a classification code given by a participant was included in the patent application but at least one code from the previous or the next level was included, then the classification code was judged as correct.

The results show that MCS had the best performance. This result may suggest that the key feature of the system to allow participants searching at the level of subclass contributed to its performance. Additionally, it is worth to mention that all the participants performed at least one search at the level of subgroup when they used MCS.

5.2 Search Tactics

The most frequent approach that was used to initiate a classification search using the MCS system is the use of keywords in contrast to the use of classification codes that was not often adopted by the participants. Another search tactic adopted by the participants was to reform the query using synonyms of the terms. Also, use of "AND" and "OR" for conjunction in search vocabulary and nested Boolean to combine Boolean Operators was adopted by some participants. This tactic is also adopted by patent professionals during patentability search [TW08].

A very interesting observation is the extensive use of searching at the level of subclass. We should mention that searching using the level of subclass was not only used in cases when participants were certain for the level of group, but also when submitting initial queries. Although participants had the chance to manually scan the codes using the respective Espacenet webpage, they preferred searching at the level of subgroup at once.

Another search tactic that was used by a number

of participants was to alphabetically sort the results. Using this tactic the participants could easily find if a specific classification code was retrieved and if other codes that belong to the same branch were retrieved.

Also, in some cases when participants viewed results at the level of subclass, they were uncertain for the specific technical field of the classification code. It is not always possible for patent examiners to recall all the descriptions of the classification codes, especially when they are not specialists in this technical field, due to the large number of classification codes. Therefore, a usual tactic was to view the descriptions of higher levels in order to become more certain about the branch this classification code belongs to.

Figure 1 shows the number of participants that adopted each search tactic.

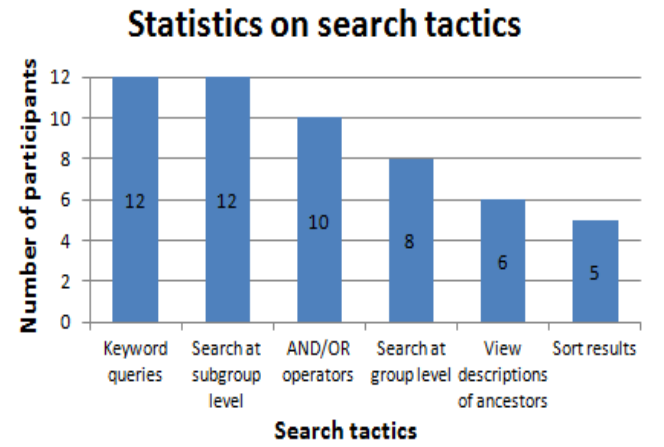


Figure 1: Number of participants used each search tactic

5.3 Search Process

To sum up the above tactics in the search process, patent examiners normally adopt a sequential and multi-directional process, as shown in Figure 2.

Patent examiners submit the query after selecting the desired IPC level. Then they examine the results and search for relevant IPCs. Viewing the description of an IPC code increased the confidence of deciding if the IPC was relevant or not. The step of viewing the results in alphabetical order assisted patent examiners to confirm if a certain IPC code is included in the returned results. Patent examiners choose to create new queries or change the IPC level when they are unable to identify relevant IPCs in the returned results. For example, they choose to search in a higher level so they can first identify the IPCs representing broader topics.

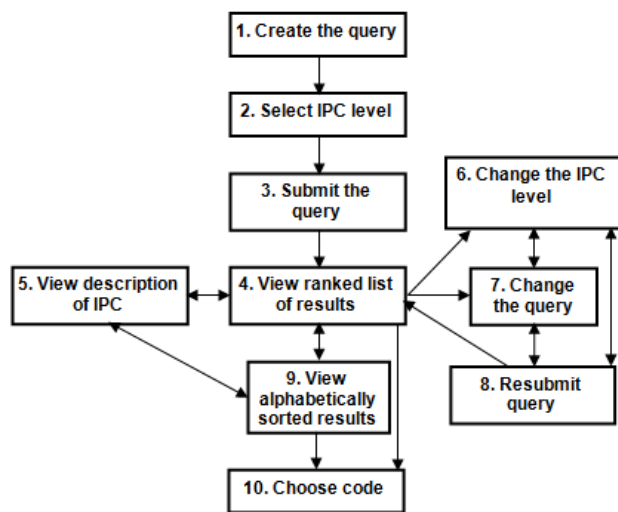


Figure 2: Patent search process for patent classification using the MCS system

6 Conclusions

In this paper, we have presented the results of a user-study that was performed with the aim to identify the search tactics adopted by patent professionals when they perform classification search. The second aim of this paper was to evaluate the performance of a web based system which can be used as a supplementary tool during classification search.

The results have shown that the MCS system performed better. A potential contribution of this study is the observation that patent professionals not only preferred searching at the level of subgroup but also this tactic may be related to the ability to identify more relevant classification codes.

We believe that the discussion and the experiment presented in this paper are also useful to the designers of patent search systems. However, we believe that more and larger experiments are required before we can reach a general conclusion.

References

- [1] S. Adams. Using the International Patent Classification in an online environment. *World Patent Information*, 22(4):291–300, Dec. 2000.
- [2] J. Allan, B. Carterette, and J. Lewis. When will information retrieval be good enough? In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 433–440, 2005.
- [3] M. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424, 1989.
- [4] M. Bates. Where should the person stop and the information search interface start? *Information Processing & Management*, 26(5), 1990.
- [5] D. Becks, M. Görtz, and C. Womser-Hacker. Understanding information seeking in the patent domain and its impact on the interface design of IR systems. In *Proceedings of the HCIR*. Citeseer, 2010.
- [6] N. Belkin, C. Cool, A. Stein, and U. Thiel. Cases, scripts, and information-seeking strategies: on the design of interactive information retrieval systems. *Expert systems with applications*, 9:379–395, 1995.
- [7] K. Benzineb and J. Guyot. Automated patent classification. In *Current Challenges in Patent Information Retrieval*. Springer Berlin Heidelberg, 2011.
- [8] E. D’hondt, S. Verberne, C. H. A. Koster, and L. Boves. Text Representations for Patent Classification. *Computational Linguistics*, 39(3):755–775, 2013.
- [9] P. Foglia. Patentability search strategies and the reformed IPC: A patent office perspective. *World Patent Information*, 29(1):33–53, Mar. 2007.
- [10] A. Giachanou, M. Salampasis, and G. Paltoglou. Multilayer Collection Selection and Search of Topically Organized Patents. In *Integrating IR Technologies for Professional Search*, 2013.
- [11] A. Giachanou, M. Salampasis, M. Satratzemi, and N. Samaras. Report on the CLEF-IP 2013 Experiments : Multilayer Collection Selection on Topically Organized Patents. In *CLEF (Online Working Notes/Labs/Workshop)*, 2013.
- [12] P. Hansen and K. Järvelin. The information seeking and retrieval process at the Swedish Patent and Registration Office. In *Proceedings of the ACM SIGIR 2000 Workshop on Patent Retrieval*, 2000.
- [13] H. Joho, L. A. Azzopardi, and W. Vanderbauwhede. A survey of patent users: an analysis of tasks, behavior, search functionality and system requirements. In *Proceeding of the third symposium on Information interaction in context - IiX ’10*, page 13, New York, USA, 2010. ACM Press.

- [14] D. Kelly. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends in Information Retrieval*, 3(12):1–224, 2009.
- [15] M. Krier and F. Zaccà. Automatic categorisation applications at the European patent office. *World Patent Information*, 24(3):187–196, Sept. 2002.
- [16] L. S. Larkey. A patent search and classification system. In *Proceedings of the fourth ACM conference on Digital libraries*, pages 179–187, Berkeley, California, USA, Aug. 1999. ACM New York, NY, USA.
- [17] M. Lupu and A. Hanbury. Patent Retrieval. *Foundations and Trends in Information Retrieval*, 7(1):1–97, 2013.
- [18] M. Lupu, K. Mayer, J. Tait, and A. J. Trippe. *Current Challenges in Patent Information Retrieval*, volume 29 of *The Information Retrieval Series*. Springer, 2011.
- [19] G. Marchionini. *Information seeking in electronic environments*, volume 30 of *Cambridge Series on Human-Computer Interaction*. Cambridge University Press, 1997.
- [20] H. Smith. Automation of patent classification. *World Patent Information*, 24(4):269–271, Dec. 2002.
- [21] Y. Tseng and Y. Wu. A study of search tactics for patentability search: a case study on patent engineers. In *Proceedings of the 1st ACM workshop on Patent Information Retrieval, PaIR 2008*, pages 33–36, 2008.
- [22] W. Vijvers. The international patent classification as a search tool. *World Patent Information*, 12(1):26–30, Jan. 1990.
- [23] T. D. Wilson. Information behaviour : an interdisciplinary perspective. 33(4):551–572, 1997.