



INFORMATION RETRIEVAL FACILITY

IRF Report 2010-00001
Date: 2010-05-01

A Survey on Patent Users Search Behavior, Search Functionality and System Requirements

Leif Azzopardi, Hideo Joho and Wim Vanderbauwhede

ABSTRACT: With a growing interest in Patent Information Retrieval, there is a need to better understand the context associated with patent users, their tasks and their needs and expectations of patent search and analysis tools, systems and applications. Patent search is known to be a complex, difficult and challenging activity, usually requiring expert Patent Information Specialists to spend a substantial amount of time sourcing (or not) documents relevant to their particular task. Information Retrieval provides a whole array of possible techniques, tools and methods which could be applied in order to ease the burden of such retrieval tasks, and also make searching patents more accessible to non-Patent Information Specialists. In this report, we outline the findings from a survey of patent users conducted to ascertain information about patent users and their requirements with respect to Information Retrieval applications.

This report was commissioned by the Information Retrieval Facility. A summary of this report has been published in [5] and a paper on these findings has been submitted.

Contents

1	Introduction	5
2	Background	6
2.1	Patent Retrieval Tasks	6
2.2	Past Surveys of Patent Users	7
3	Methods	9
4	Results and Analysis	9
4.1	Patent User Demographics	9
4.2	Domains and Search Tasks	10
4.3	Search Behavior	11
4.4	Search Functionality and Requirements	13
4.5	On the Ideal Patent System	15
5	Concluding Discussions	16

List of Figures

1	The age of patent users in years.	20
2	A slice of the demographics of Patent Users: (a) Nationality, (b) Office Location (Country), (c) Role in terms of Patent Usage, and (d) the Patent Domain according to the high level IP Classes.	21
3	Experience: Patent industry and Patent Search.	21
4	Frequency of Search Tasks performed by Patent Users for: (a) All domains, (b) Human Necessities, and (c) Chemistry / Metallurgy, where 1 indicates Mainly/Daily to 5 which indicates Rarely/Never).	22
5	Important search features (1: Strongly disagree; 5: Strongly agree)	24
6	Important techniques for query formulation (Analysts vs. Managers) . . .	26
7	Important techniques for relevance judgement (Analysts vs. Managers) . .	27
8	Other important techniques for search (Analysts vs. Managers)	28

List of Tables

1	Search Task Correlations (using Spearman rho Bold denotes statistically significant at $p \leq .05$.)	10
2	Search effort (time and quantity) expended to complete a task.	23
3	Job titles	25

With a growing interest in Patent Information Retrieval, there is a need to better understand the context associated with patent users, their tasks and their needs and expectations of patent search and analysis tools, systems and applications. Patent search is known to be a complex, difficult and challenging activity, usually requiring expert Patent Information Specialists to spend a substantial amount of time sourcing (or not) documents relevant to their particular task. Information Retrieval provides a whole array of possible techniques, tools and methods which could be applied in order to ease the burden of such retrieval tasks, and also make searching patents more accessible to non-Patent Information Specialists. In this report, we outline the findings from a survey of patent users conducted to ascertain information about patent users and their requirements with respect to Information Retrieval applications.

This report was commissioned by the Information Retrieval Facility. A summary of this report has been published in [5] and a paper on these findings have been submitted.

1 Introduction

Over the past few decades patent searching has changed dramatically from paper based access to instant online access, from library catalogue systems to internet search systems, from partially indexing to fielded and full text indexing (i.e. images, chemical structures, diagrams, tables, etc)[2, 7].

While patent search has been an important area of research for many years, this has largely been undertaken within the database community [7]. However, over the past few years several new initiatives within the IR community have been developing. Now, there are numerous large scale patent test collections available for IR research, and there has been a spate of workshops and symposium on patent search promoting research on patents¹. All these initiatives have been largely concerned with applying, researching and developing information retrieval tools, techniques and theory to address patent search problems. This has opened up an array of new search tasks that can be evaluated in IR, but while careful consideration has gone into the development of these evaluation tracks, it may be helpful to have a better overview of what patent searchers want from their information retrieval systems. This is to ensure that research and development of methods and the like fit appropriately within the patent user's search context, and that the provision of search services meet the needs and expectations of users.

Given this type of search context it is essential for developers of retrieval methods and applications to ascertain the importance of the search functionality that patent users desire and require[7]. While there have been several studies performed in the past which survey patent information specialists about specific search services (such as the British Library patent services[14] or the European Patent Office's web search service[13]), little other work has formally investigated their search requirements. To this end, we have conducted a survey of patent users in order to obtain a better picture of their search habits, preferences, and the types of functionality that they want/need and how important they perceive them to be, along with details about their search tasks (frequency, duration, etc).

In this report, we first provide a brief overview of the different patent tasks, and related surveys and studies in this area (see Section 2). Then we describe the study undertaken here in Section 3 before presenting the results of the survey and the main findings in Section 4. The points raised during the background section are considered during the analysis of results, and the implications for systems research and development is discussed, before concluding and summarizing this work in Section 5.

¹New initiatives on patent retrieval include: CLEF-IP, TREC-CHEM, MAREC, PaIR, Apsire, etc, see the IRF website <http://www.ir-facility.org/> for more details, while past initiatives on patent retrieval have been performed by NTCIR, see <http://research.nii.ac.jp/ntcir/> and TREC, see <http://trec.nist.gov>.

2 Background

In this section, we shall first provide a brief overview of the different patent retrieval tasks to provide the necessary background to the survey². Then we describe some of the previous surveys that have been conducted with patent users, prefaced by a description of the different types of patent users. Later in the discussion section we shall relate the past findings to the findings from this survey.

2.1 Patent Retrieval Tasks

According to Bonino et al[7], patent usage can be sub-divided into three main categories, search, analysis and monitoring.

1. **Search:** Within this category there are a variety of different types of search tasks which require patents and non-patent information to be retrieved in order to accomplish a particular goal/work task (see below).
2. **Analysis:** The analysis of patents can be broken into two main types: (i) micro analysis of individual patents, and macro analysis of a group of patents, such as a patent portfolio. The analysis is usually performed to evaluate and assess the Intellectual Property (IP), to map and chart the IP, to identify trends and competitors and also to identify new areas of potential to exploit.
3. **Monitoring:** Akin to filtering, the monitoring of patents is usually performed by an agent to notify users about new incoming patent information to keep users abreast of the latest developments.

The different search tasks that are routinely performed include[9]:

- **State of the Art (SOA):** identify patents for the purposes of a general review. Sometimes referred to as Landscaping or Technology survey.
- **Novelty (NOV):** identify patents and non-patents which may affect the patentability of an idea/invention (performed before writing a patent application).
- **Patentability (PAT):** given a patent application, ensure novelty.
- **Infringement (INF):** identify patents or applications which cover the proposed product or process and are still in force.
- **Opposition (OPP):** identify literature available to the public to show lack of novelty or inventive step of a granted patent. This type of search is sometimes referred to as Validity or Invalidity search.
- **Freedom to Operate (FTO):** like infringement, but also includes non-patent literature.

²We refer the reader to [12, 3, 9, 1, 2, 7] for a more comprehensive description of the field of patent search and retrieval.

- **Due Diligence (DD):** analyze strengths, weaknesses and scope of IP rights.

The main emphasis in most of these tasks is to find all the relevant documents[7], especially for patentability and validity search tasks, where missing relevant documents is deemed unacceptable. This is because of the highly commercial nature of patents coupled with the high costs involved in creating a patent and infringing patented material. In turn, it is often required that patent searchers demonstrate that they have performed an exhaustive and comprehensive search. According to Bonino, et al.[7] published in 2010, the most common patent search tasks are: Patentability, Infringement, and Technology Survey, however in a survey[14] conducted in 2000, Novelty, State of the Art, and Infringement search were the most common tasks performed. This suggests that over time there has been a shift to Patentability. In our survey, we asked participants about how often they conducted these different search tasks.

Recent work reviewing the field of patent retrieval concluded that while patent retrieval is a specialization of information retrieval which benefits from the developments in IR, it also warrants special attention and requires the development of models that are appropriately and properly adapted to the users and their information needs[7]. It is well known that patent analysts perform a number of difficult and challenging search tasks (such as Novelty search or Infringement search)[9] and rely upon sophisticated search functionality, tools, and specialised products[4]. Furthermore, these search tasks are often performed under stringent conditions (esp. regulatory and legal requirements), and they also require different search strategies which transpire through a course of interaction to achieve the end goal (which in some cases means not actually finding documents, i.e. no “kill” document, for instance)[4, 17, 10]. These conditions and the challenges faced by patent users need to be considered when designing IR systems for such users. In this work, we asked patent users what search functionality they think is important for helping them resolve their complex and difficult information needs.

2.2 Past Surveys of Patent Users

There is a variety of different patent users and the types of search tasks that they perform vary depending on their roles. The main users of patents are Companies, Patent Analysts, Inventors, Researchers, Managers and Investors[7]. Each type requires the system for different types of search tasks. For example, Patent Analysts may be employed to determine the patentability of applications, an Inventor may wish to determine whether their ideas are novel, an investor may wish to buy out a particular area, managers may wish to identify new trends, and researchers not wanting to re-invent the wheel may need to find existing solutions or technology. According to Bonino et al[7], “professional patent searchers typically prefer more advanced functionalities, with a higher degree of control on tool capabilities and freedom in setting search parameters, while occasional users (such as managers) often require an easy to use interface and simpler commands”. This suggests that professional patent searchers prefer more advanced functionality than managers and occasional users. In the past, two major surveys of patent users have been undertaken. These were undertaken to ascertain whether users were satisfied with the

search services provided. The first was conducted by the British Library in 2000, and the second was conducted more recently in 2007 by the European Patent Office.

Newton [14] reports a survey of users of the British Library Patent Information Centre. The main finding of the survey was that the introduction of the internet has significantly impacted upon user requirements and that the services provided by the library need to adapt to the technological changes. A survey of 277 users of the Patent Information Centre was conducted. Most of those surveyed were from the south of England, predominantly from the greater London area, and were mainly patent professionals, agents and searchers. Notably, since internet search services were introduced a significant decline in library visits by those surveyed was found, while a significant increase in the usage of internet services was found. The survey revealed that most of the searches conducted were to check originality (novelty) around 40% in total, background searches (around 20%), and protection searches (validity) (20%), followed by tracking competitors, identifying trends, and investing.

McDonald-Maier (2009)[13] reported their findings from a survey conducted in 2007 based on approximately 400 esp@cenet users³. This study revealed that esp@cenet users wanted full text field search capabilities with a fast response time. The key fields that they wanted to search on were applicant, title and abstract along with keyword search of the full text, number search and patent class. This survey also revealed that users wanted a clear and easy to use interface, enabling access and navigation through the collection, and the ability to easily manipulate results. Of growing concern was the need for machine translation of results. Also, of interest was the number of non patent search specialists using the esp@cenet database - a significant rise in technical and managerial staff, as well as an increase in academics. Unfortunately, the findings report in this study are very terse and few concrete numbers or figures are provided in [13].

Our work provides a more detailed account of the demographics of patent users. More specifically, we asked patent users about their role, and the search tasks they undertook, along with a series of detailed questions about the search functionality that they required.

Fogilia [10] outlines user search strategies for the patentability search, while Tseng et al [17] report on an analysis of patent search engineers and their search tactics for patentability. The main finding from this research is that in order to complete their search tasks numerous rounds of retrieval are undertaken to ensure that they have found as many of the relevant documents as possible. This practice is also noted in [4, 7], where they state that patent searchers will often employ an iterative search strategy to complete their task and find all relevant documents. To formulate a more detailed appreciation of this aspect of the search process our survey asked patent users about the number of queries that they issue and the length of time searching and querying.

The survey we conducted is somewhat different to past work in that we focused on ascertaining the search requirements and functionality that patent users require and desire when performing their search tasks.

³Esp@cenet is a free online patent search system provided by the EPO., see <http://www.espacenet.com/>.

3 Methods

The survey instrument⁴ that we used consisted of an on-line questionnaire with 86 questions consisting of four parts:

- **Demographics:** To obtain some basic background and demographics of the subjects we canvassed the participants for details about their age, gender, nationality, language, location, education level, role, job title, and client type.
- **Domain and Search Tasks:** This section ascertained information about their patent usage and the types of search tasks that they performed and how often they performed these tasks.
- **Search Functionality Requirements:** In this part of the survey asked participants to rate the importance of different types of functionality for query formulation, search assessment and navigation, and search management.
- **Open Ended Questions:** Several open ended questions were asked to ascertain other desirable search functionality, the main functionality used to search, and what would make an ideal system.

The survey was designed to be completed in approximately 30 minutes which restricted the number of questions that we could include. To obtain a large and representative sample we sent the survey instrument out to two patent user group mailing lists:

- the Confederacy of European Patent Information User Groups (CEPIUG), and
- the international Patent Information Users Group (PIUG).

In total, these lists have over 700 members from over 27 different countries, and of these members, approximately 300 are patent information specialists. We received 81 responses in total to the survey - giving a response rate of just over 10% overall. In the following section, we shall provide a description of the results along with an analysis.

4 Results and Analysis

4.1 Patent User Demographics

We began our analysis by looking at the demography and background information of the patent user community. Of the 81 subjects, 58% were male, while 42% were female, with 60% of subjects between 39 and 59 years old (see Figure 1 for a complete breakdown). The nationality of the subjects was from 14 different countries with nationalities represented the most being Dutch, American, French, and British (see Figure 2a). While the locations of their offices were from 13 different countries with the top locations being in the USA, Netherlands, France, Belgium and the UK (see Figure 2b). Almost all subjects primarily

⁴Ethics approval was obtained from our organization.

Task Search	SOA	NOV	PAT	INF	OPP	FTO	DD
SOA	1.00	.249	.195	.113	.238	.070	.276
NOV		1.00	.657	.144	.176	.181	.163
PAT			1.00	.233	.247	.273	.231
INF				1.00	.472	.519	.436
OPP					1.00	.424	.391
FTO						1.00	.464
DD							1.00

Table 1: Search Task Correlations (using Spearman rho **Bold** denotes statistically significant at $p \leq .05$.)

searched in English (98%), with a few also using German, French and Dutch (less than 15% in total). The educational background of patent users surveyed revealed that most are highly educated, with over 65% of our participants hold a post graduate degree (PhD 23.5% and Masters 43.2%), while the remaining held degrees (28.5%) and diplomas (4.9%).

Most subjects worked full time (91%) and the clients that they worked for were predominately internal i.e. within the same organization (88%), the rest were external (22%) or both (10%). A listings of the types of positions held by the participants is shown in Table 3. The most frequent job titles found in our survey was a *patent information specialist*, followed by *patent analyst*. However, the job titles were indeed found to vary across organisations. In a post analysis of these job titles we categorized the titles into several roles: *analysts*, *managers*, *researchers*, *attorney*, and others. The roles of the subjects were as follows: Analysts (46.9%), Managers (32.1%), Researchers (12.3%), Attorneys (2.5%) and other roles (6.2%), see Figure 2c. While the proportion of patent attorneys was small in our survey, they were frequently mentioned as a collaborator within the team. These results complement the findings of the esp@cenet study, and show that over half of the patent users are non-Analysts with Managers and Researchers making up a good proportional of patent users.

Figure 3 shows the participants’ search experience as well as industry experience in the patent domain. As can be seen from the graphs, most participants have several years experience in patent search, with an average of around 10 years.

4.2 Domains and Search Tasks

To gain a deeper insight into the patent user community, we analyzed the data across domains and search tasks. Figure 4 shows the result of the analysis along with the data of all domains. The tasks such as *Freedom to operate* search, *Novelty* search, and *Patentability* search were frequently performed by our participants, while the *Opposition* search and *Due diligence* search were rarely performed. These findings lend weight to

the assertion made in [7], which claimed that Patentability was one of the main search tasks, and re-affirms the finding in [14] that Novelty search is also a main search task.

In Figure 2d, we report the breakdown of the respondents into the patent domains they work in predominantly. The domains correspond to the highest IP class assigned to Patent Documents. It can be seen from the pie chart that most respondents worked in the area of Chemistry/Metallurgy (35.8%) and Human Necessities (34.6%), followed by Mechanical Engineering (12.3%) and then all the other class domains. We then examined the effect of the two most popular domains on the frequency of search tasks (see Figures 4b and 4c). The breakdown across domains suggests that the frequency of search tasks was reasonably similar across the subject domains. However, the main difference between domains was that in the Chemical/Metallurgy domain the Freedom to Operate and State of the Art search task was more frequently performed than within the Human Necessities domain.

To appreciate the relationship between the search tasks that participants performed in common we measured the correlation between the search tasks undertaken by participants. The results of the pair-wise comparison of search tasks are shown in Table 1. As can be seen, a reasonably high significant correlation was found between the Novelty search and Patentability search (as noted earlier novelty is usually performed given an idea, and patentability is performed given the patent application). The Freedom To Operate search and Infringement search also showed a significant positive correlation. As the description of these search types indicates (see Section 2.1), there are common elements in these search tasks too. The Freedom to Operate search is conducted to determine if a product is not generating an infringement, while an Infringement search is conducted to determine whether some one does not have full freedom to operate. Essentially, these search tasks are two sides of the same coin. It also suggests that it is relatively common for patent searchers to perform multiple types of tasks with a similar level of frequency.

4.3 Search Behavior

In the previous subsection we obtained a macro level view of the search contexts of patent users. This subsection focuses on ascertaining specific details about the user search behavior when performing the tasks in particular domains and under various conditions. The results for this subsection were obtained by asking participants to indicate the maximum, the average, and the minimum amount of time that they spent completing their most frequently performed search task, formulating individual (single) queries, and evaluating or judging individual documents, along with the number of queries they submitted and the number of documents that they examined. The results of this analysis are shown in Table 2. The values reported are the median of each category as the variance of data was large.

As can be seen in the rows for all patent users (referred to as Overall in the Table), the effort required to complete a patent search task are quite substantial. On average a typical search task will taken 12 hours to complete and range from a minimum of 3 hours to a maximum of 40 hours (based on the median values). This is substantially longer than the typical web search[8], as even a simple task can take hours to complete. These

findings confirm that patent searching largely follows an iterative search paradigm [7], and shows precisely the breakdown of the different interactions required to complete the task. Note that because we have asked participants for an estimate of the time taken, the number of hours taken to complete a search task is not exactly or strictly equal to the time taken to formulate a query multiplied by the number of queries submitted plus the time taken to assess documents multiplied by the number of documents examined. This is because the search task also includes other duties like recording documents and collecting citations, but also due to the imprecision of the answers possible through a survey instrument. However, we believe the estimates are reasonable and provide a good indication of the amount of effort expended during the search process.

We further examined how the duration and quantity of the different factors varied according to search task, domain, academic level of education, and client type. We only used the average data for this part of the analysis and so values in the lower part of the table should be compared to the Overall Average. Table 2 reports the median values for the Search Task Completion Time (in hours), the time to formulate a each single query (in minutes), the time to evaluate and assess a document (in minutes), the number of queries submitted, and the number of documents assessed. We can make a number of observations based on this analysis.

First, we can see that the Novelty search and Patentability search require a similar amount of effort on behalf of the patent searcher. So does the Infringement search and Freedom to Operate search. These similarities echo our earlier examination of the task type correlation (See Table 1).

Second, the subject domains were found to have some effect on the query formulation process. For example, in the Chemistry domain substantially more time is required to formulate a query and more queries are submitted when compared to the human necessities domain. This is perhaps not surprising since in the Chemistry domain, queries often take the form of complex chemical structures which have numerous variations. On the other hand, the time taken to assess the relevance of documents is very similar, as is the number of documents examined.

Third, the level of education of respondents appears to have a noticeable effect on their patent search behavior. Our results suggest that those with a PhD tend to take longer to formulate a query than the other groups. However, perhaps compensating for this, they also tended to issue fewer numbers of queries than the other groups. It may be that holders of a PhD might have a more advanced knowledge of the subject which enables them to formulate more effective queries, and thus need less of them. Alternatively, it is maybe that non-PhD holders are more cautious preferring to issue more queries, or because of a greater anomalous state of knowledge[6] require more queries in order to refine and reduce their uncertainty. We can not provide a definitive answer given the data, but this finding warrants further investigation into the differences in higher education levels on searching and search behavior.

Finally, our results suggest that more efforts are required for external clients than internal clients (See the bottom rows of Table 2). We can speculate that there can be a higher degree of uncertainty in capturing the information needs of external clients coupled with the need to ensure that the searcher has been diligent and exhaustive i.e.

earning the trust and respect of external clients usually requires going the extra mile (especially if being paid by the hour). It may be that the criticality of the search task performed for external clients often demands a more thorough and exhaustive search to be carried out.

4.4 Search Functionality and Requirements

So far we have gained a greater insight into the search behavior of patent users across several dimensions. However, in this section we examine the requirements and generic search functionality that patent users believe is important in helping them complete their patent retrieval search tasks. To achieve this goal, we asked our participants to rate the importance of a number of retrieval application and search features. We used a 5-point Likert scale where participants were asked to indicate a level of agreement to a statement such as “Boolean Operators are important to formulate effective queries”, “Patent family information is important to accurately assess the relevance of retrieved documents”, and “Combining multiple search results is important for my work” and we fielded responses of strong disagreement (1) to strong agreement (5). As mentioned in Section 3, the system features were roughly grouped into (a) query formulation, (b) results assessment and results navigation, and (c) search management, organization and history. The results of this analysis are shown in Figure 5.

The most consistent pattern in the results was the scale; the median value of almost all 40 system features rarely goes below the point 3. This suggests that the patent community can appreciate and desire a wide range of search functionality during the course of undertaking patent search tasks. This also suggests that they are also willing to adopt and leverage search functionality to complete their task. This is in stark contrast to the behavior of typical web searchers who prefer simple minimalist search functionality and rarely, if ever, use any advanced search functionality[16].

Within the three different groups of system features, there were some very noticeable differences. Within the query formulation group, the results show that Boolean operators were very important to almost all respondents. Proximity, truncation, wildcards and field operators were important to most respondents (80-91%). On the other hand, Expansion and Translation were seen as important by only around half the respondents, while a large proportion of respondents were impartial to these features. This may be because they have not used such functionality to form an opinion, or because such functionality is not required as most searched primarily in English. Finally, the Weighting of terms in the query obtains a very mixed response where most respondents are impartial while the rest are split between it being important and not important. This may be because of the difficulties associated with weighting query terms manually. It appears that the features that introduce some uncertainty in the process (i.e. Weighting, Expansion and Translation) are not considered to be as important as the other very precise operators, which can be controlled and have a precise semantic and interpretation. This is perhaps due to the fact that Patent Analysts are often required to fulfill strict and stringent practices given the legal and regulatory requirements.

In the results assessment and navigation section, the most important features were

highlighting of keywords to facilitate navigation within documents as well as features that enabled the results set to be grouped by patent family, status, and date, and the document to be viewed. Of slightly lesser importance was functionality that sorted or grouped by relevance, classification and class codes, access to citations and maximum size of the result list. The utility of the different features is likely to impact upon the perceived importance of these features, it is an open question as to how and when these features (and the others feature) could be engaged to increase the efficiency or effectiveness of retrieval.

Perhaps more interestingly, within the search management group of features most were designated as very important. This suggests that crucial to most patent search tasks is the recording and tracking of the search activities (i.e. ensure accountability and audit mechanisms are available to demonstrate due diligence in performing a search task). This makes sense given the nature of the tasks, but also because the tasks often span several days. If the previous search paths or results were re-examined this would add extra time to the task. Or worse still, is that some documents might be missed or skipped if a searcher is not meticulous in recording where they have been. Missing documents could lead potentially disastrous consequences.

In [7], it was hypothesized that patent analysts would prefer more fine grained and precise control over their search, while managers and other users would prefer simpler querying functionality[7]. To examine this hypothesis, we further investigated the effect of roles on the perceptions of search feature importance and compared the perceptions of the analyst group to the manager group. The results of this analysis are shown in Figure 6, 7, and 8, respectively. Our results suggest that the two groups generally had a similar perception on the search functionality. This provides evidence against the hypothesize that managers prefer simpler search functionality when querying. However, there were some noticeable differences between the two groups for particular search functionality. For example, in query formulation, the analyst group tended to appreciate the query expansion and weighting function stronger than the manager group (See Figure 6). While, in results assessment and navigation section, the analyst group tended to appreciate more features than the manager group (See Figure 7). A similar trend was found in the search management section (See Figure 8). These findings tend to suggest that it is the navigation, organization and management of search results and the search which are more important to analysts than managers, as oppose to simpler querying functionality. This implies that facilitating, tracking and supporting long term interaction is a critical requirement for analysts.

In the open ended questions, we asked respondents to list other search functionality which they also felt was very important. Other features which we did not ask about yet were mentioned as important by our respondents were image search, statistical analysis, reporting tools, claim progress mapping (e.g., filed to granted status), and citation traversal functions. Most of these types of functionality again suggest that it is post retrieval interaction that is important and should provided by a patent search tool/system.

Furthermore, these findings suggest that research and development of models, methods and systems for patent search needs to consider the search and interaction requirements of the different types of patent users. It would appear that patent users prefer search

functionality which provides a high degree of control and precision for accomplishing their search tasks, and that supporting post-search interaction, navigation, organization and management of results is perceived to be crucial in accomplishing their tasks successfully. Since patent searchers are willing to spend a lot of time and effort in constructing requests and examining documents it is important that the tools and systems developed support their needs and requirements.

4.5 On the Ideal Patent System

As part of the survey we included an open ended question asking what the users thought would be in an ideal patent system.

Approximately 17 responses referred to existing patent databases (both commercial and freely available systems), as ideal or close to ideal if they provided some extra functionality or were combined with other existing systems. On the other hand, 11 responses indicated that an ideal patent retrieval system would not be possible. Three technically correct responses provided answers like “Fast accurate retrieval of only relevant document”, and “One that retrieves all of the relevant references and none of the irrelevant references”. These responses are very close to the text book definition of an ideal retrieval system[18], but are not very helpful in this context. Many respondents indicated the need for speed was paramount, but perhaps Simmons[15] sums up best many of the features of an ideal patent retrieval system which many of our respondents touched upon in their descriptions:

the ideal patent information ... would have universal coverage of all patents from all countries, with immediate updating. It would provide maximum recall with minimum false drop. Chemical structures would be searchable from specific and generic queries, with retrieval of all records with any amount of overlap between the query structure and the structures disclosed in the patent. Mechanical and electrical drawings would be searchable. All inventive concepts would be indexed for retrieval, and all prior art disclosed in the patent would be retrievable and distinguishable from the technology claimed as new. Search strategy building would be intuitive. There would be easy access to the original documents. And the system would be inexpensive and accessible by both large companies and SMEs.

The point regarding accessibility is particular important and was emphasized by a number of respondents, who had great cause for concern with respect to coverage and access with numerous responses stating something to the effect of, “access to all information in one language of choice”. Interestingly, while within the descriptions of ideal systems many pointed out the need for machine translation, overall most respondents indicated that machine translation was not as important as other features. This might be because current machine translation techniques are not adequate, and that this presents one of the major challenges for patent information retrieval (a sentiment expressed in both [2] and [7]).

5 Concluding Discussions

This report reported the results of a survey of patent users. We aimed to elicit the general requirements of search system functionality required by patent users. In contrast to previous work, we have performed a more detailed and specific survey of such users. Since we were aware of the importance of the user context [11] (e.g. work tasks, roles, educational background, and interaction) in the design of IR systems, we also reported the search behavior and the effect of several factors. However, providing an accurate picture of a large international community like the patent users is challenging, and this report can only offer a part of it. Nonetheless, the descriptive statistics and analysis reported here should serve as an excellent starting point for further research of patent search and development of patent search tools and systems. Some of the highlights of our findings are as follows.

- Our results indicate that patent search tasks are inherently interactive – requiring multiple iterations, support for the combination, organization and management of results, while searching needs to be considered in context of the larger task at hand. An environment where a long-term interaction history is recorded and effectively incorporated to subsequent interactions appears to be very desirable.
- It appears that the effect of domain and user is likely to affect the design of query formulation and specification. However, there might be a common requirement on the presentation of retrieved documents for relevance judgements.
- A similar set of functionality might support multiple types of search by understanding the characteristics of search tasks. For example, Novelty search and Patentability search are likely to require a similar functionality. The same might apply to Freedom to Operate search and Infringement search. The design of search interfaces and search system can take these trends into consideration.
- Our results also indicate that patent analysts and managers both demand search functionality and features which provide control over the searches being undertaken. In particular, our results indicate a strong preference towards functionality that provides fine grained control over the search process using operators with a clear semantic and interpretation. For example, in query formulation, the operators to control the *matching* tended to be more appreciated and perceived as more important than the *weighting* or translation of query terms. This suggests that a real time preview of the sample texts that match the current query might be a useful feature to predict the performance of search.
- We still have a long way to go to achieve the ideal patent system. However, these descriptions of the perfect system demonstrated the professionalism and enthusiasm of the patent searcher community. In the last few years, the IR community arguably has focused over much on web users[8]. Our survey suggests that the

patent user community can drive the advancement of search technologies for complex and difficult search tasks which transpire over days and consume a significant amount of effort.

These findings motivate further research into understanding more deeply the importance of each type of functionality within the different patent search tasks along with a better understanding of the context in which the search is undertaken. The design of Information Retrieval methods, systems and applications should be tailored to their needs and behavior: and this study provides useful insights into the needs and requirements of patent users, which can be used to direct this further research in a more directed way. Finally, these findings also impact upon the development of test collections and tasks for evaluation that should be created – where the focus is certainly on provisioning for interaction, organization and management; and this presents a significant challenge for future research.

References

- [1] S. Adams. 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, 2010.
- [2] S. Adams. 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, 2010.
- [3] N. J. Akers. The european patent system: an introduction for patent searchers. *World Patent Information*, 21(3):135 – 163, 1999.
- [4] K. H. Atkinson. Toward a more rational patent search paradigm. In *Proceeding of Patent Information Retrieval Workshop at CIKM '08*, pages 37–40. ACM, 2008.
- [5] L. Azzopardi, W. Vanderbauwhede, and H. Joho. Search system requirements of patent analysts. In *To appear at SIGIR '10*, 2010.
- [6] N. J. Belkin, R. N. Oddy, and H. M. Brooks. Ask for information retrieval: part i: background and theory; part ii: results of a design study. *Journal of Documentation*, 38(2) 61-71 and 38(3) 145-164, 1982.
- [7] D. Bonino, A. Ciaramella, and F. Corno. Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, 32(1):30 – 38, 2010.
- [8] A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.
- [9] M. R. D. Hunt, L. Nguyen. *Patent Searching*. 2007.
- [10] P. Foglia. Patentability search strategies and the reformed ipc: A patent office perspective. *World Patent Information*, 29(1):33 – 53, 2007.
- [11] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, Secaucus, NJ, USA, 2005.
- [12] E. W. Kitch. The nature and function of the patent system. *Journal of Law and Economics*, 20(2):265 – 290, 1977.
- [13] L. McDonald-Maier. esp@cenetš: Survey reveals new information about users. *World Patent Information*, 31(2):142 – 143, 2009.
- [14] D. Newton. A survey of users of the new british library patent information centre. *World Patent Information*, 22(4):317 – 323, 2000.
- [15] E. S. Simmons. Patent databases and gresham’s law. *World Patent Information*, 28(4):291 – 293, 2006.

- [16] A. Spink, D. Wolfram, M. B. J. Jansen, and T. Saracevic. Searching the Web: The Public and Their Queries. *Journal of the American Society for Information Science and Technology*, 52(3):226–234, 2001.
- [17] Y.-H. Tseng and Y.-J. Wu. A study of search tactics for patentability search: a case study on patent engineers. In *Proceeding of Patent Information Retrieval Workshop at CIKM '08*, pages 33–36. ACM, 2008.
- [18] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 1979.

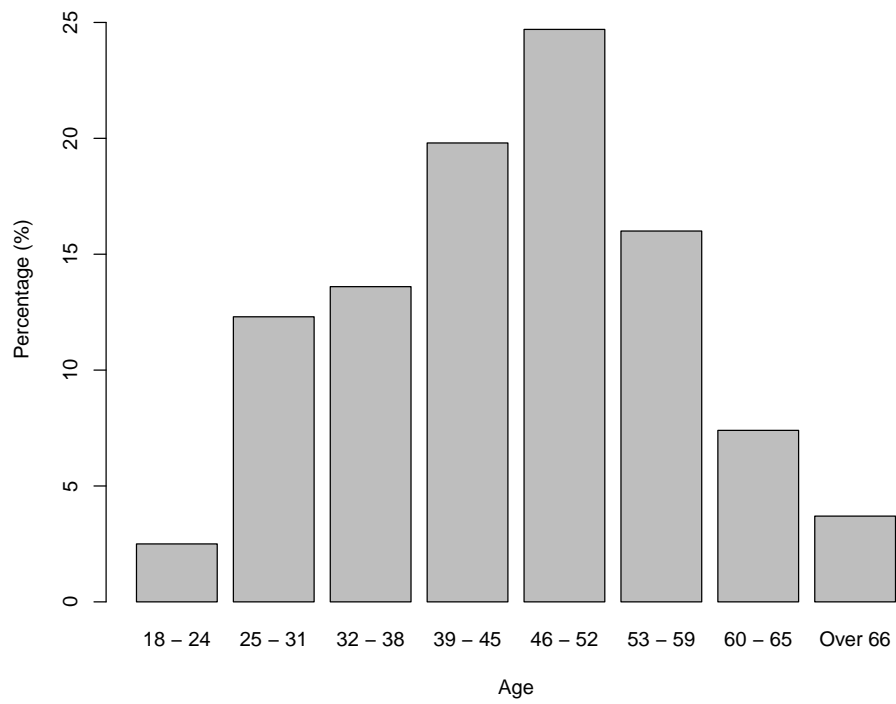


Figure 1: The age of patent users in years.

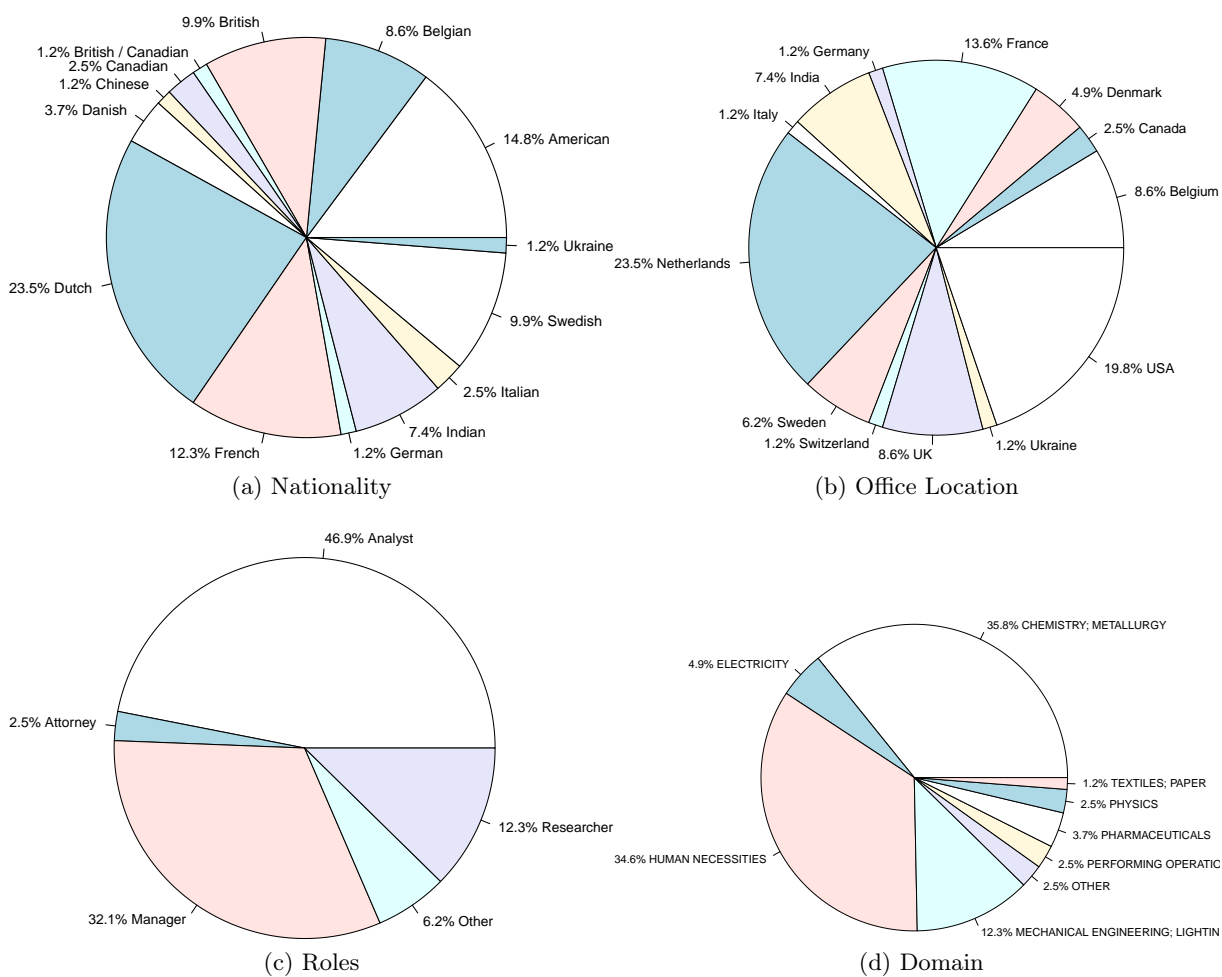


Figure 2: A slice of the demographics of Patent Users: (a) Nationality, (b) Office Location (Country), (c) Role in terms of Patent Usage, and (d) the Patent Domain according to the high level IP Classes.

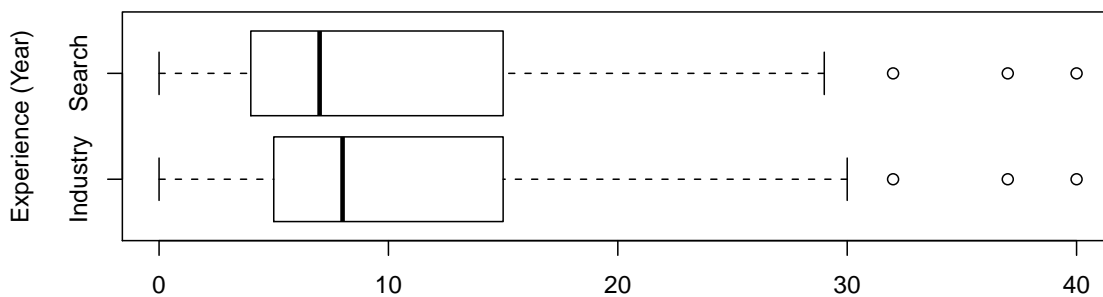
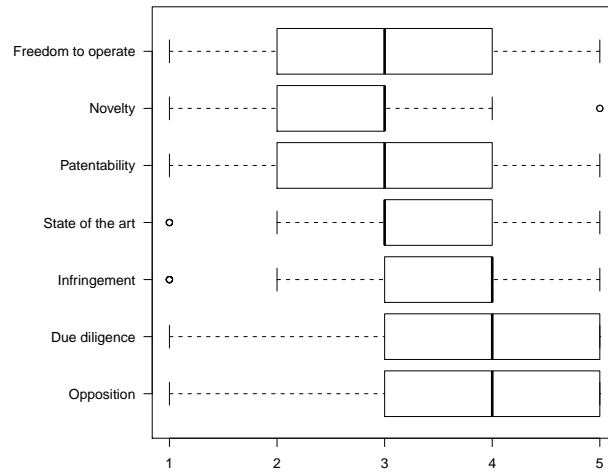
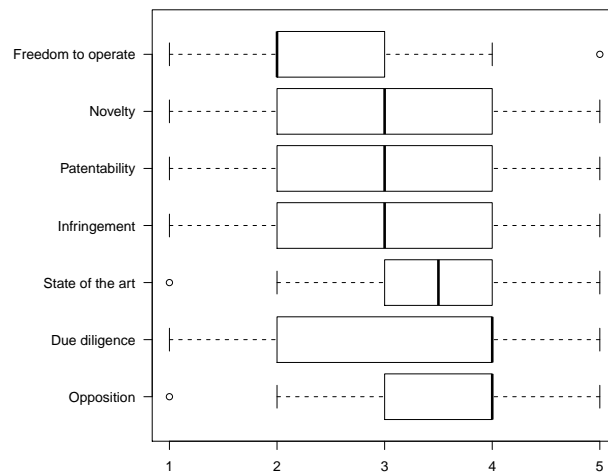


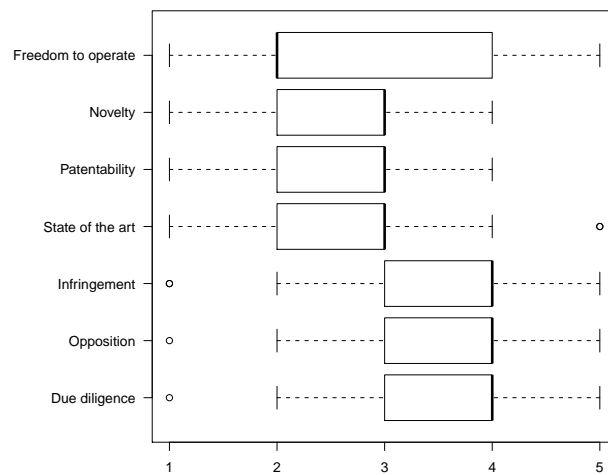
Figure 3: Experience: Patent industry and Patent Search.



(a) All domains



(b) Human Necessities Domain

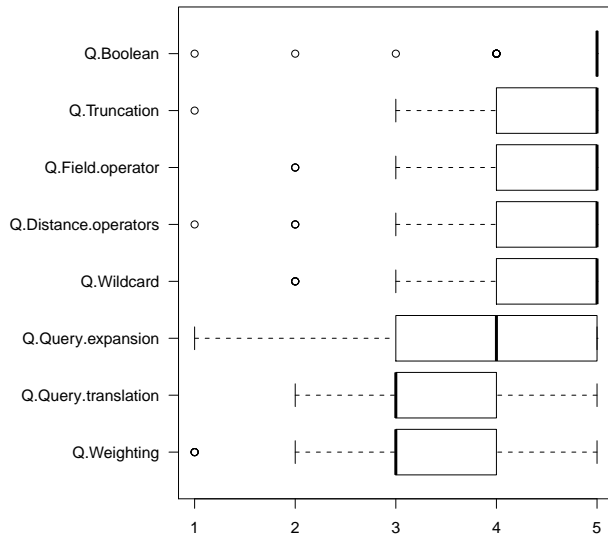


(c) Chemistry / Metallurgy Domain

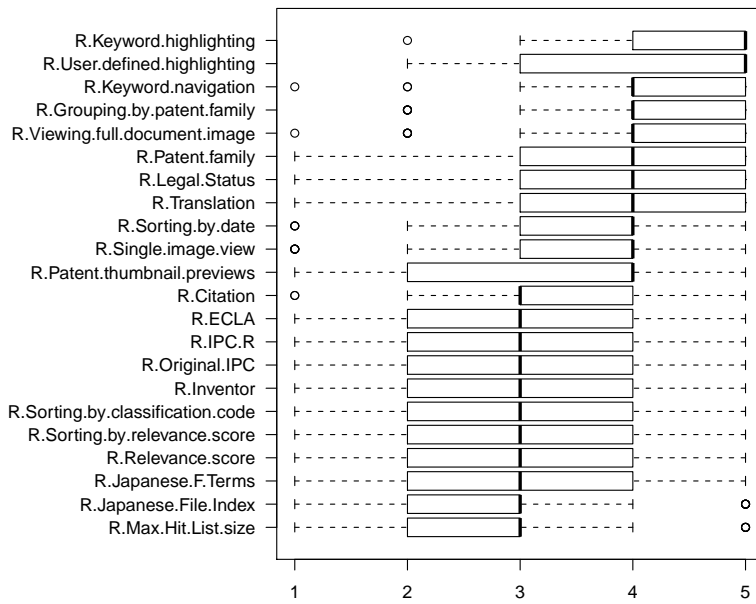
Figure 4: Frequency of Search Tasks performed by Patent Users for: (a) All domains, (b) Human Necessities, and (c) Chemistry / Metallurgy, where 1 indicates Mainly/Daily to 5 which indicates Rarely/Never).

Category	Category	Search Task Completion Time (Hour)	Single Query Formulation Time (Mins)	Single Document Relevance Judgement Time (Mins)	# Queries Submitted	# Documents Examined
Overall	Minimum	3	1	1	5	10
	Average	12	5	5	15	100
	Maximum	40	20	25	40	600
Search Task	SOA	13.5	7.5	4.5	10.0	75
	NOV	11.0	5.5	6.0	15.5	75
	PAT	12.2	6.5	9.0	10.0	50
	INF	15.0	3.0	5.0	20.0	300
	OPP	20.0	10.0	5.0	20.0	300
	FTO	15.0	5.5	5.0	20.0	200
	DD	20.0	5.0	3.0	10.0	150
Domain	Human Necessities	11.0	2.0	4.5	15.0	100
	Chemistry Metallurgy	12.5	10.0	5.0	20.0	100
Education	Bachelor	15.0	5.0	7.0	15	100
	Masters	15.0	3.5	5.0	15	100
	Phd	12.0	10.0	5.0	10	200
Clients	Internal	12.5	5.0	5.0	15	100
	External	17.5	10.0	4.5	20	150

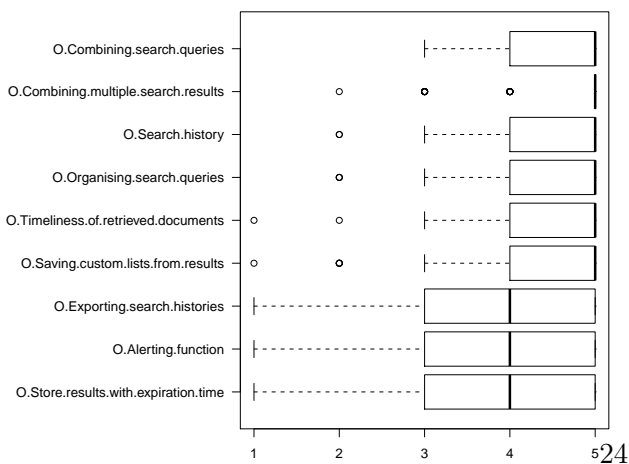
Table 2: Search effort (time and quantity) expended to complete a task.



(a) Query formulation



(b) Result assessment and navigation



(c) Search management, organization and history

Figure 5: Important search features (1: Strongly disagree; 5: Strongly agree)

#	Job Title
13	Patent Information Specialist
5	Patent Analyst
4	Patent Engineer, Patent Information Analyst, Research Engineer
3	Consultant, Information Scientist, Manager IP
2	Information Manager, IP Analyst, IP Information Specialist, Patent Attorney, Patent Information Manager
1	Documentation Manager, Information Analyst, Information Professional, Information Researcher, Information Specialist, IP Director, IP Information Scientist, IP Manager, Litigation Research Analyst, Managing Member, Patent Administrator, Patent Agent, Patent Information Scientist, Patent Searching Manager, Principal Scientist, Quality Manager, Reference Librarian, Scientific Advisor, Scientific Information Researcher, Scientist, Searcher, Specialist, etc.

Table 3: Job titles

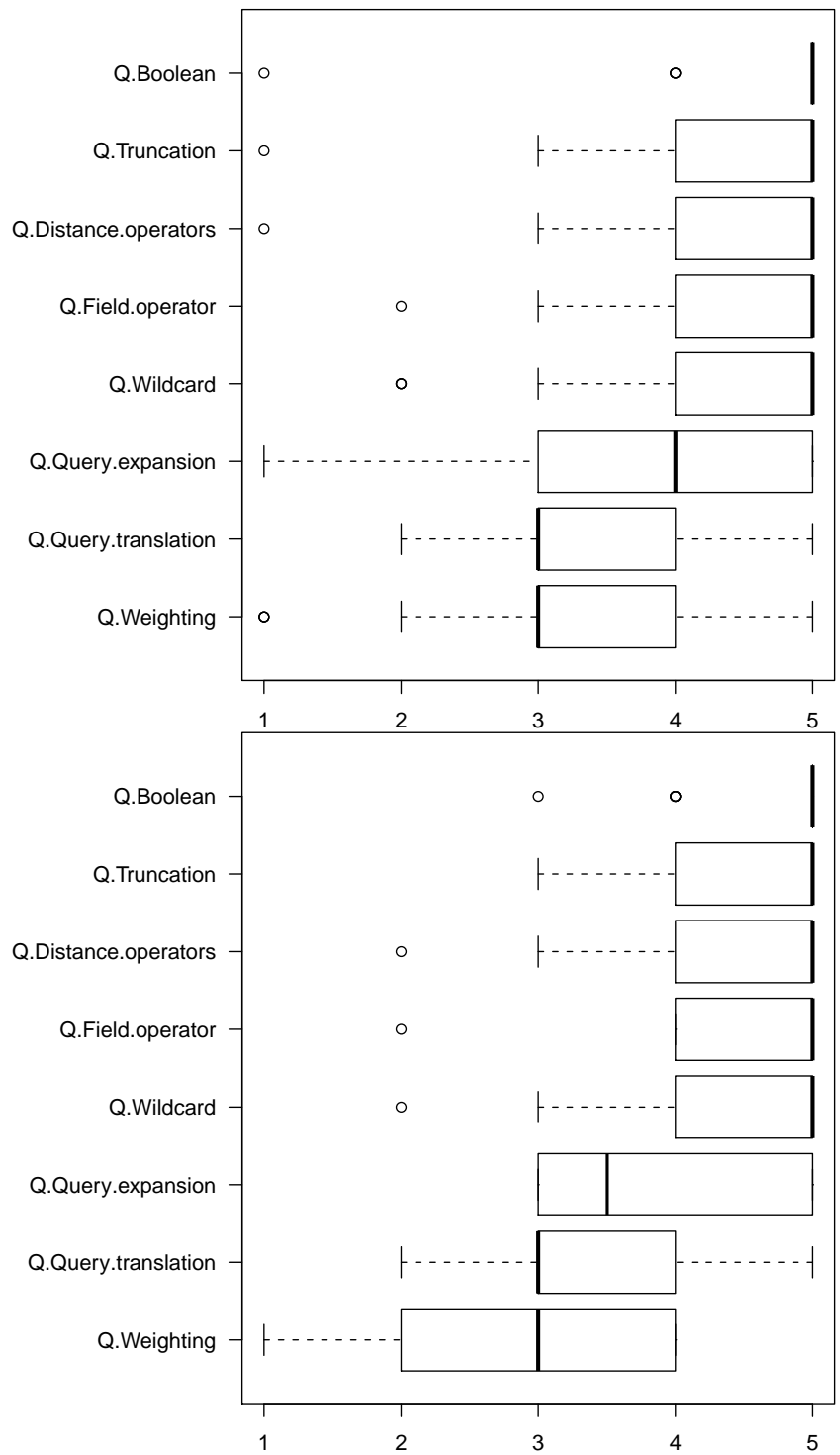


Figure 6: Important techniques for query formulation (Analysts vs. Managers)

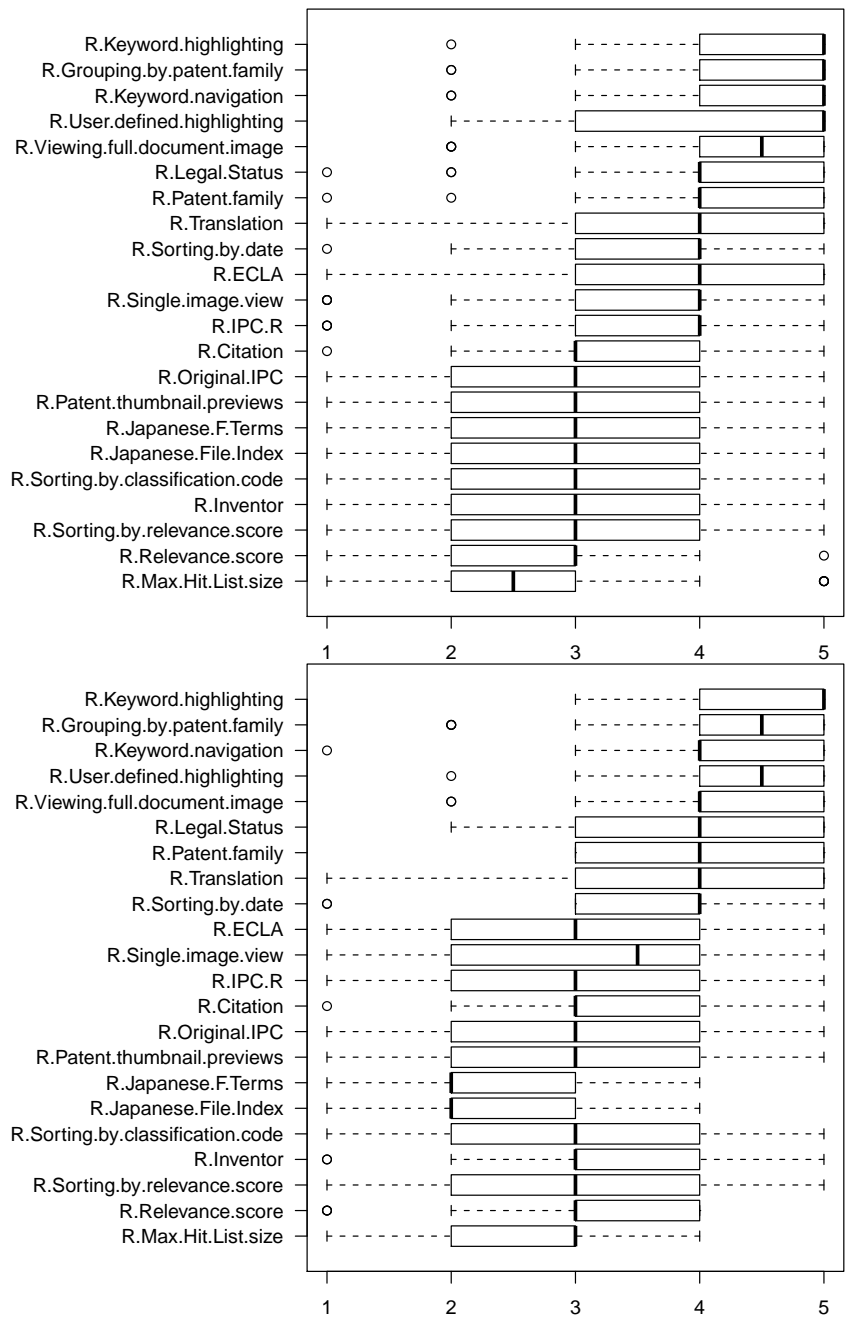


Figure 7: Important techniques for relevance judgement (Analysts vs. Managers)

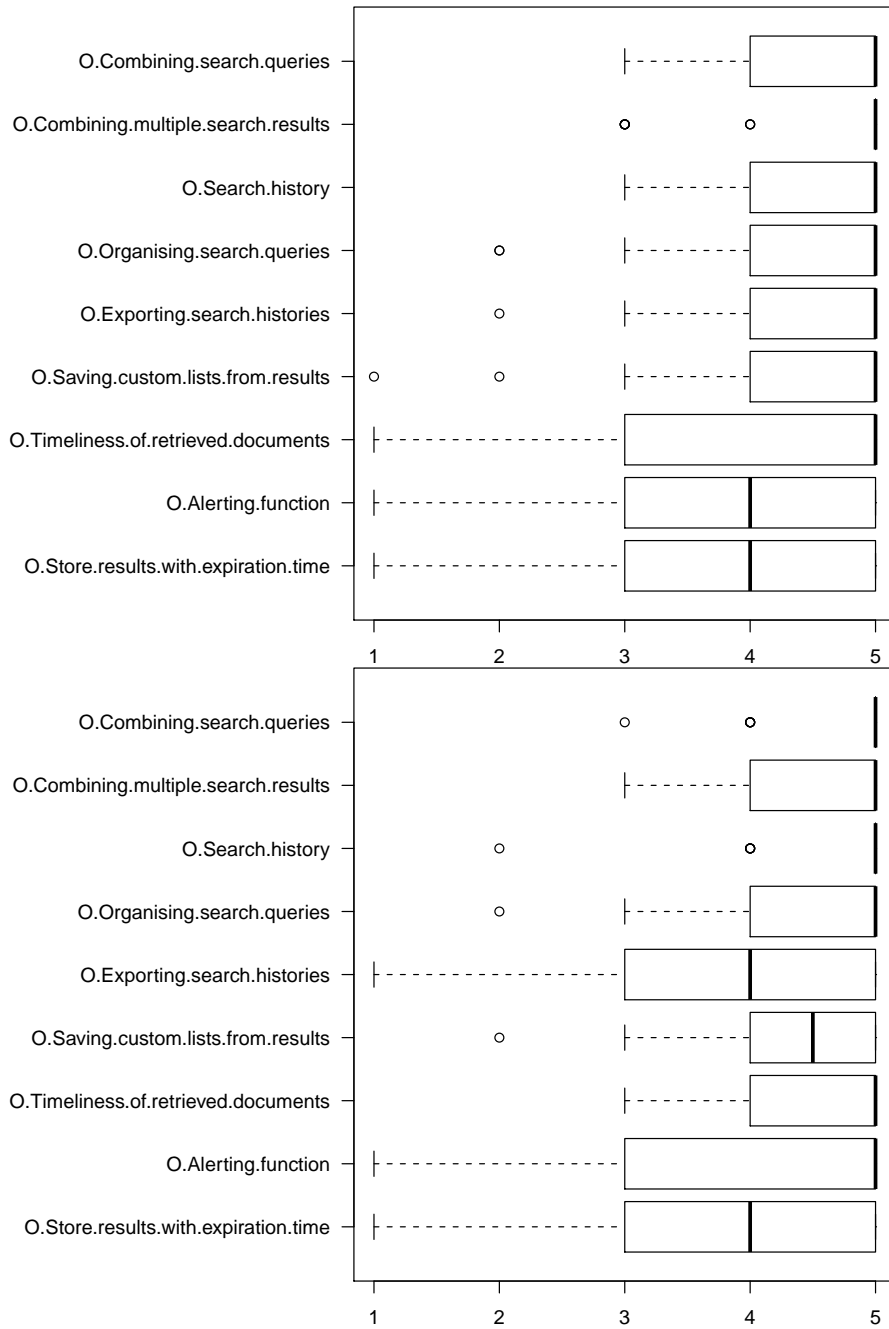


Figure 8: Other important techniques for search (Analysts vs. Managers)