International Journal of Business Analytics and Intelligence
12 (2) 2024, 18-34
www.publishingindia.com/ijbai/

An Analysis of the Patent Search Behaviour of Human Experts and Its Impact on Patent Retrieval Performance

Farshad Madani*, Charles Weber**

Abstract

Background: Patent databases are essential repositories of technical information that support innovation and technological advancements across industries. Despite their value, retrieving relevant patents from these databases remains time-consuming and often inaccurate. This challenge is exacerbated by the fact that human experts essentially perform patent retrieval, whose search approaches and skill levels vary significantly. These variations lead to inconsistent outcomes, making the retrieval process unreliable and inefficient. Understanding the factors that influence patent retrieval performance is critical for improving the utility of these databases and facilitating the discovery of enabling technologies.

Objectives: This study aims to investigate the impact of human search behaviour on patent retrieval performance. By analysing specific aspects of expert behaviour, the research seeks to identify factors contributing to variability in retrieval outcomes and propose strategies for improvement. The study focuses on keyword diversity, query complexity, and search speed metrics to measure search behaviour while evaluating reliability, efficiency, effectiveness, and judgement error. Ultimately, the objective is to assess whether current human-centred approaches are sufficient and explore potential solutions for enhancing retrieval.

Main Ideas: The study observes seven patent retrieval experts searching for, extracting, and evaluating metallurgical patents containing enabling technologies for developing and producing an innovative kitchen skillet. Through systematic analysis, the researcher identifies patterns in search behaviour and correlates them with retrieval performance. The findings reveal a strong relationship between search strategies and outcomes, with significant variations in performance across the experts. Metrics such as keyword diversity and query complexity emerge as key determinants of retrieval success. However, the results highlight that the overall performance of human experts in patent retrieval is generally low, demonstrating inefficiencies and inconsistencies in their approaches.

Conclusion: The results of this study underscore the limitations of current human-centred patent retrieval methods and the need for more effective approaches. The significant variability in expert performance suggests that relying solely on human expertise is insufficient for achieving consistent and accurate results. Applying artificial intelligence (AI) methods to patent retrieval has the potential to address these challenges by automating key aspects of the search process and improving both efficiency and accuracy. By integrating AI tools, organisations can significantly enhance patent retrieval performance and unlock the full value of patent databases.

Keywords: Patent Search, Patent Retrieval, Search Behaviour, Retrieval Performance

Introduction

Patent retrieval is utilised for many applications, such as determining the patentability of an idea, detecting infringement on prior art, conducting technology surveys, and ideation in product development (Jürgens, Hansen & Womser-Hacker, 2012; Bonino, Ciaramella & Corno,

* Department of Engineering and Technology Management, Portland State University, Oregon, United States. Email: farshad.madani@gmail.com; ORCiD: https://orcid.org/0000-0002-4451-9678
** Department of Engineering and Technology Management, Portland State University, Oregon, United States. ORCiD: https://orcid.org/0000-0002-6403-6593

2010). It is a cognitive process that currently relies heavily on interaction between human experts and machine intelligence (Chen & Dhar, 1991; Wang & Soergel, 1998). When executed improperly, patent retrieval can become a severe bottleneck in the abovementioned applications (Kumar, Tripathi & Singh, 2016; Montecchi, Russo & Liu, 2013) or result in insufficiently accurate results (Xie & Miyazaki, 2013; Noh, Jo & Lee, 2015).

Patent retrieval is an iterative process (Shalaby & Zadrozny, 2019), which is depicted in Fig. 1. An expert searches for a patent by choosing keywords based on his/her knowledge, designing a query, and extracting several patents from a patent database. The expert subsequently evaluates the extracted patents, judging whether they are relevant to the subject of the search or not. The expert revises the keywords and designs a new query according to what he/she learned from prior judgement. The iterative process ends after the expert realises there are no more relevant patents to be retrieved or there is no more time to continue patent retrieval. How human experts behave during patent searches and the errors they commit during patent evaluation can limit patent retrieval's performance (efficiency, effectiveness, and reliability). Researchers involved in patent retrieval consequently have a strong incentive to explore the behaviour of human experts who conduct patent searches and how this behaviour impacts the performance of patent retrieval (Xie & Miyazaki, 2013; Noh et al., 2015).

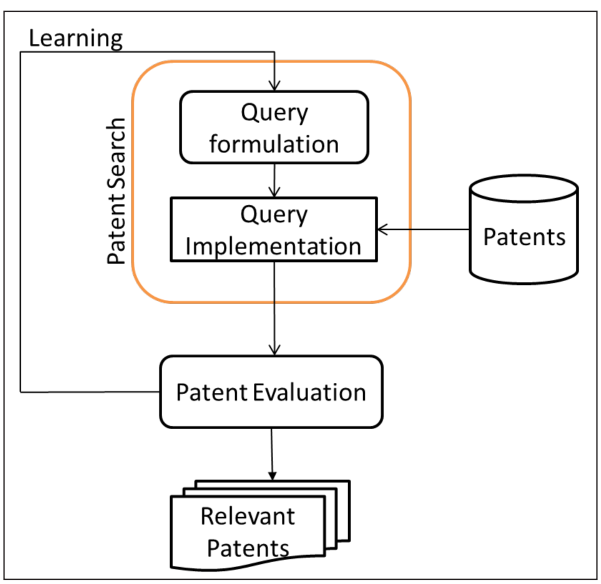

Fig. 1: The Basic Process of Patent Retrieval

Despite this strong incentive, research into the relationship between patent search behaviour and patent retrieval performance has been sparse. No study about this topic could be found in peer-reviewed journals (see Section 2 of this paper), and studies from related fields (Agichtein, Brill & Dumais, 2006; Joho, Jatowt & Blanco, 2015; Liu, Liu, Cole, Belkin & Zhang, 2012; Tamine & Chouquet, 2017; Zhang, Anghelescu & Yuan, 2005; Zhang, Liu Cole & Belkin, 2015) do not treat the topic directly. A study exploring the relationship between expert patent search behaviour and expert performance at patent retrieval is consequently warranted.

This paper describes an empirical case study that explores the abovementioned relationship. The study specifically addresses the following research question: Do the main factors that characterise patent search behaviour by human experts impact the experts' performance at patent retrieval? The case involves iron-casting technologies that could enable developing and producing a new, improved kitchen skillet. The authors of this paper have chosen this case for two reasons. Firstly, iron casting is a technology that is highly patented. Thus, a study of patents about iron casting is likely to yield statistically significant conclusions. Secondly, observing expert behaviour in a narrow context gives the authors insight into expert judgement. All searches are conducted in the US patent database.

In this study, the authors investigate the search behaviour of seven experts in iron casting and metallurgy, utilising behavioural variables suggested by the academic literature (Agichtein et al., 2006; Joho et al., 2015; Zhang, Anghelescu & Yuan, 2005; Zhang et al., 2015). The authors also determine how well these experts perform at patent retrieval according to well-known criteria for efficiency, effectiveness, and reliability (Golafshani, 2003; Machado & Davim, 2017). Finally, the authors of this paper conducted a correlation analysis of the behavioural data and performance data to provide an answer to the research question.

The study described in his paper has found a correlation between the variables that describe patent search behaviour and the performance criteria for patent retrieval. Thus, the study establishes a relationship between patent search behaviour by human experts and their performance at patent retrieval. The study also finds that the variability in patent search behaviour among experts is very high and that human experts are not efficient, effective, or reliable at patent retrieval. These findings suggest that this study could be an impetus for developing more approaches to patent search based on machine intelligence. Successful development and implementation of such approaches could perhaps alleviate the bottlenecks that currently limit the efficiency of patent search. Further study of this topic could also lead to the discovery of principles that govern the relationship between patent search behaviour and patent retrieval performance.

Literature Review

Patent search is a cognitive process based on iterative trial and error (Chen & Dhar, 1991; Sutcliffe & Ennis, 1998), in which human experts try to diminish the mismatch between a search query and relevant patents (Magdy & Jones, 2011). In this process, a human expert formulates a query to identify patents that the expert believes bear relevance to a specific topic. Once applied to a patent database, the query subsequently yields a list of such patents. The expert reformulates the query to obtain a list in the next iteration, which is more focused or more comprehensive. This process continues until the expert believes that he/she has obtained a list that covers all patents relevant to the topic and excludes patents that are not. At that point, further iterations of query formulation and application will not yield additional gains (Jansen, Booth & Spink, 2009; Spink, Jansen & Cenk Ozmultu, 2000).

The approach to patent search from above is limited in two ways. Firstly, a human expert will likely possess incomplete knowledge about the topic he/she is researching. He/she may consequently design a query that misses potentially germane patents (Shiri & Revie, 2006). The second limitation is technical: If the topic under search is complex, then the query becomes complex and must be broken up into many simpler queries. If the resulting queries are too simple, the search results may contain many irrelevant patents. On the other hand, if the query is too complex, the search result may miss some relevant patents (Shiri & Revie, 2006).

Patent retrieval consists of having a computer extract the contents of the patents provided by the query and subsequently letting human experts judge these patents as relevant or irrelevant to the topic under investigation (Baeza-Yates & Ribeiro-Neto, 2011). This judgement is subject to human limitations such as errors and slow processing speed, and these limitations may vary from expert to expert, suggesting that performance at patent retrieval could vary from expert to expert (Shiri & Revie, 2006). Some studies also show that the cognitive complexity of query design has yielded query design methods that result in unsatisfactory performance in patent retrieval (Zhang, Anghelescu & Yuan, 2005; Magdy & Jones, 2011). It has been suggested that incorporating approaches from fields such as philosophy, psychology, and linguistics into query design could remedy this deficiency (Wacholder, 2013). However, a significant amount of research into patent search behaviour and retrieval performance would need to be conducted before such novel approaches become efficient, effective, and reliable.

The academic literature suggests that patent search and retrieval are complex and time-consuming. For example, Joho et al. (Joho, Azzopardi & Vanderbauwhede, 2010) conducted a patent search survey, which found that a typical search task takes 12 hours to complete and ranges from a minimum of 3 hours to a maximum of 40 hours. Bonino et al. (Bonino et al., 2010) confirmed these results by suggesting that the duration of a patent search task ranges from 6 to 24 hours. Joho et al. (Joho et al., 2010) also reported that a patent search task that is followed by a patent retrieval involves roughly 15 queries and 100 judgements to determine the relevancy of patents, where each query takes 5 minutes to formulate. Each document takes 5 minutes to judge.

The literature also notes that the patent search needs of professional users of patent databases have changed significantly since the emergence of the internet (Newton, 2000). While a fast response time is among the top expectations among most users of patent databases (McDonald-Maier, 2009), professional users like product planners have to deploy iterative search strategies to ensure they have found as many relevant patents as possible (Bonino et al., 2010; Atkinson, 2008). To professional users, missing relevant patents is unacceptable, which makes their search methods more recall-oriented (Bonino et al., 2010; Joho et al., 2010). Professional users consequently require more functionality in patent search, unlike occasional users who often prefer an easy-to-use interface and simpler commands (Bonino et al., 2010). All this suggests that the current ways in which patent searches are conducted do not meet the efficiency needs of today's professional users (Bonino et al., 2010; Atkinson, 2008).

Studies on search behaviour have been performed to predict user knowledge (Zhang et al., 2015) and to examine implicit relevance feedback on web search results (Agichtein et al., 2006). They have also been conducted to investigate search effectiveness (Zhang, Anghelescu & Yuan, 2005), temporal information searching (Joho et al., 2015), the effect of task difficulty (Liu et al., 2012), and the performance of expert users and novice users in medical information retrieval (Tamine & Chouquet, 2017). A survey of the most important studies on search behaviour follows.

Zhang et al. (Zhang et al., 2015) identified behavioural features contributing to user knowledge prediction by applying multiple regression analysis. They concentrated on three types of search behaviour among users: querying behaviour, document selection behaviour, and general task interaction behaviour. Their best regression model contained three behaviour variables that contributed to the prediction of user knowledge: the number of documents saved, the average query length, and the average ranking position of the document opened.

Agichtein et al. (Agichtein et al., 2006) examined the implicit relevance of feedback on web search results under circumstances in which user feedback can be noisy. To examine the impact of user behaviour, they developed and applied three groups of measures, which were respectively based on the amount of time spent by users on a page or URL, query characteristics such as query length keywords shared with the next query, and clicks applied by users during web browsing. The authors observed significant improvement in methods that do not consider implicit feedback.

(Zhang, Anghelescu & Yuan, 2005) tried to answer questions about the relationship between search behaviour and search effectiveness. They utilised the precision score and the total number of relevant documents identified by each participant as two measures for search effectiveness. To study the search behaviour of participants, they applied the number of queries, the average number of terms in queries, and the average number of thesaurus terms in queries. They concluded that search behaviour does not seem to be related to search effectiveness.

(Joho et al., 2015) studied the behaviour of temporal information searching in a laboratory setting. They looked at four aspects of query formulation behaviour: the average number of queries submitted in a search session, the average length of queries by terms, the total number of unique terms used in a search session, and the ratio of temporal control expressions in search vocabulary. They noticed that searching for information about the future is the most challenging and requires more support for users. On the other hand, searches for content that was generated recently were the easiest, most common, and the most successful for users.

Tamine and Chouquet (Tamine & Chouquet, 2017) studied the performance of expert and novice users in medical information retrieval. They studied the two main tasks of document search: query formulation and relevance assessment. Query formulation was evaluated using query length, specificity, and difficulty. In relevance assessment, the authors applied the assessors' relevance ratings, the degree of agreement between the assessments made by different assessors, and the difficulty of relevance assessment. The authors observed that, on average, experts formulated longer queries that contained 27% more terms than those generated by novices. Experts also applied more technical terms in their queries. The authors concluded that expert searches significantly differ from those performed by novices. Also, they figured out that the agreement in relevance assessment is low, even among experts. In addition, they concluded that both expert users and novice users had almost the same degree of difficulty in document relevance assessment. Most importantly, Tamine and Chouquet found that expert users performed tasks faster than novices.

None of the abovementioned studies on search behaviour were conducted on patent data—they were performed in other, perhaps related areas of information retrieval. Our academic literature search yielded no published papers in peer-reviewed journals about patent search behaviour and retrieval performance. Consequently, we have identified a significant gap in the scholarly literature. Given how vital patent retrieval performance is to the various applications of patent search (Jürgens et al., 2012) (Bonino et al., 2010), the case for conducting a study that explores the relationship between patent search behaviour and performance is powerful.

Hypothesis and Measures

We hypothesise that various aspects of patent search behaviour affect patent retrieval performance, and we deploy two sets of measures to test this hypothesis. The first set of measures pertains to patent search behaviour, including keyword diversity, query complexity, and search speed. The second set of measures consists of four ubiquitous performance measures tailored for patent retrieval: reliability, efficiency, effectiveness, and error rate. The number of potentially *relevant* and *judged patents* constitutes the main element in tailoring the performance measures.

Measures for Patent Search Behaviour

Measures for patent search behaviour express how an expert designs queries according to keywords selected during patent searches.

Keyword Diversity

The keyword diversity measure is defined as follows:

$$keyword\ diversity = \frac{number\ of\ distinct\ keywords\ used\ in\ patent\ searches\ of\ a\ patent\ retrieval}{number\ of\ keywords\ used\ in\ patent\ searches\ of\ a\ patent\ retrieval}$$

This measure illustrates the variety of ways through which experts utilise their knowledge to use keywords in query design. In other words, this measure compares the innovativeness of the different experts in query design. For example, an expert may use 20 keywords to design queries, but 10 are repeated. Conversely, another expert may use 50 keywords to design queries; only 10 are repeated. In this example, the second expert applied a more diversified set of keywords for his/her patent search. In other words, he/she behaved more innovatively when applying his/her knowledge of patent search.

Query Complexity

Experts exhibit different patent search behaviour depending on the number of keywords in each query. The more keywords are used in a query, the more complex the query is, the more narrow the search area is, and the fewer results the user gains (Sormunen, 2000, p. 38). Thus, query complexity in a patent search is defined as follows:

$$query\ complexity = \frac{number\ of\ keywords\ used\ in\ patent\ searches\ of\ a\ patent\ retrieval}{number\ of\ queries\ used\ in\ patent\ searches\ of\ a\ patent\ retrieval}$$

Search Speed

The speed at which human experts engage in patent searches can vary significantly (Joho et al., 2010), and search speed is an aspect of expert behaviour affecting patent searches' efficiency, effectiveness, and reliability. Efficiency is affected because human experts constitute an expensive resource. Search speed is defined as follows:

$$search\ speed = \frac{number\ of\ queries\ used\ in\ patent\ searches\ of\ a\ patent\ retrieval}{time(minutes)spent}$$

Measures for Patent Retrieval Performance

Reliability

The reliability of a research instrument concerns the extent to which the instrument yields the same results on repeated trials (Golafshani, 2003). According to this definition, we expect that when two experts use *the same keywords* in a search, they should produce *the same results*. This means that experts in patent search should find the same patents when they use the same keywords in query design in a patent search.

However, experts involved in patent retrieval may apply the same keywords in different orders, as well as design and apply different queries throughout the whole process of patent retrieval. Their attempts at patent retrieval will consequently yield different patents. Given a set of 10 keywords, for example, one user may apply two subsets of the keywords in two queries. In contrast, another user may apply different combinations of keywords in five queries. The two sets of patents retrieved in these two scenarios would be different.

Similarity is introduced to evaluate how searches utilising the same keywords in different orders can lead to slightly different results in patent retrieval (Joho et al., 2010). Two different querying strategies are deployed if the exact keywords are used in two separate searches. The two sets of keywords deployed are considered similar, and the patents retrieved from the searches with these two sets of keywords are considered similar patents. The following questions arise: How similar are keywords used in two different searches, and how similar are two sets of patents retrieved because of the two searches?

The Jaccard index (Niwattanakul, Singthongchai, Naenudorn & Wanapu, 2013) is used to calculate the similarity between two sets of keywords used in a patent search and the similarity between the two sets of patents retrieved as a result of the search. The Jaccard index compares the similarity between two sets of keywords or patents—A and B—as shown below:

$$J(A,B) = \frac{A \cap B}{A \cup B}$$

To evaluate the reliability of a patent search, the concept of reliability in patent search is extended as follows:

reliability = 1− faliure rate

= 1 – rate (retriveing dissimilar patents v using similar keywords)

= rate (retrieving similar patents v using similar keywords)

Therefore,

$$reliability = \frac{Average\ of\ similarity\ of\ patents\ judged\ in\ each\ of\ patent\ retrievals}{Average\ of\ similarity\ of\ keywords\ applied\ in\ patent\ searches\ in\ each\ patent\ retrievals}$$

The Jaccard index calculates the average similarity of the patents judged in two patent retrievals and the average similarity of keywords applied in the two patent searches.

Efficiency

Efficiency means how well resources are expended in a process (Machado & Davim, 2017). In a patent search, *time spent by experts* is the main resource used to find relevant patents. Since experts participating in patent searches are expensive, finding more patents relevant to the subject of a patent search is very important. Therefore, efficiency is defined in this research as follows:

$$Efficiency = \frac{Number\ of\ relevant\ patents\ judged\ in\ a\ patent\ retrieval}{Total\ spent\ time\ (minutes)\ in\ a\ patent\ retrieval}$$

Effectiveness

The concept of effectiveness refers to the degree of achieving possible objectives. Ideally, the most effective patent retrieval is the one in which all judged patents are relevant to the subject of the search. Therefore, the definition of effectiveness is given as follows:

$$Effectiveness = \frac{number\ of\ relevant\ patents\ judged\ in\ a\ patent\ retrieval}{number\ of\ patents\ retrived\ in\ a\ patent\ retrieval}$$

Error

Humans may frequently make errors in patent judgement, which can be classified into two types. Experts may have reviewed a patent at least two times, and their judgements agree. This error is known as a ‘duplicate’. Alternatively, the experts may have classified a patent as relevant in one judgement and irrelevant in another independent judgement. This error is known as a ‘contradiction’. The net number of patents is calculated by deducting the number of duplicates and contradictions from the number of patents judged. Finally, the net number of patents is binned as ‘relevant’ or ‘irrelevant’ (Lupu, Mayer, Tait & Trippe, 2011, Chapter 4).

The error rate of the experts is calculated as follows:

$$error\ rate = \frac{(number\ of\ duplicates + number\ of\ contradictions)\ in\ patent judgements}{number\ of\ judged\ patents\ in\ a\ patent\ retrieval}$$

Research Methods

In this study, the authors observe the patent search behaviour and evaluate the patent retrieval performance of a panel of seven experts in metallurgy. The authors have selected these experts according to two criteria: 1) academic knowledge in materials science and 2) professional experience in casting or foundry. The profiles of the participating experts are shown in Table 1.

Table 1: Profiles of the Members of the Expert Panel

Expert ID	*Degree*	*Experience*	*Job*
E1	BSc	<10 years	Product Engineer
E2	PhD	>10 years	Professor
E3	PhD	>10 years	Professor
E4	MSc	>10 years	Product Engineer Manager
E5	PhD	<10 years	Product Engineer
E6	MSc	<10 years	PhD student
E7	MSc	<10 years	PhD student

To collect data about the search behaviour of the participating experts and their performance in retrieving relevant patents, a researcher (one of the authors of this paper) organised a session with each expert under observation. In each session, the expert under observation engaged in the following activities: 1) introduction to the case, 2) examining an ontology, 3) training in patent search, and 4) executing patent searches and evaluations. The researcher observed the expert in all steps. He interacted with the expert in steps 1 through 3.

- *Introduction to the Case.* The researcher gave the expert a written description of the case, shown in Fig. 2, and a set of verbal instructions on how to proceed. The researcher answered questions regarding the case that the expert might have.

Case: A traditional skillet

As cookware company produces a traditional skillet. Here is some information:

- Length: 17.4 inches
- Width: 10.4 inches
- Depth: 5.2 inches
- Weight: 11.1 pounds

Some customers complain the skillet is too heavy. The main manufacturing process is fine grain sand casting. The engineers of the company believe they should find a way to shrink the mold to have thinner walls.

USPTO (United States Patents and Trademark Office) database is chosen as one of the sources to find some ideas. This stage is called ‘ideation’. It means we don’t expect to find a comprehensive solution for this case, but we would like to find some ideas inspiring the engineers.

In this experiment, you are supposed to do a search patent in USPTO and find relevant patents.

Fig. 2: Case Introduction Page

- *Ontology:* The researcher provided the experts with the ontology in Fig. 3 and answered questions regarding the ontology that the expert might have. The ontology was generated at the iron casting company that wants to develop and produce a new and improved skillet. The managers and engineers at that company wanted to reduce the weight of their products by utilising external technologies compatible with their existing casting technology. They used the US patent database as the source of external technologies that could identify potential opportunities for this improvement. Thus, the US patent database served as the source for patent searches and retrieval by this study's experts.

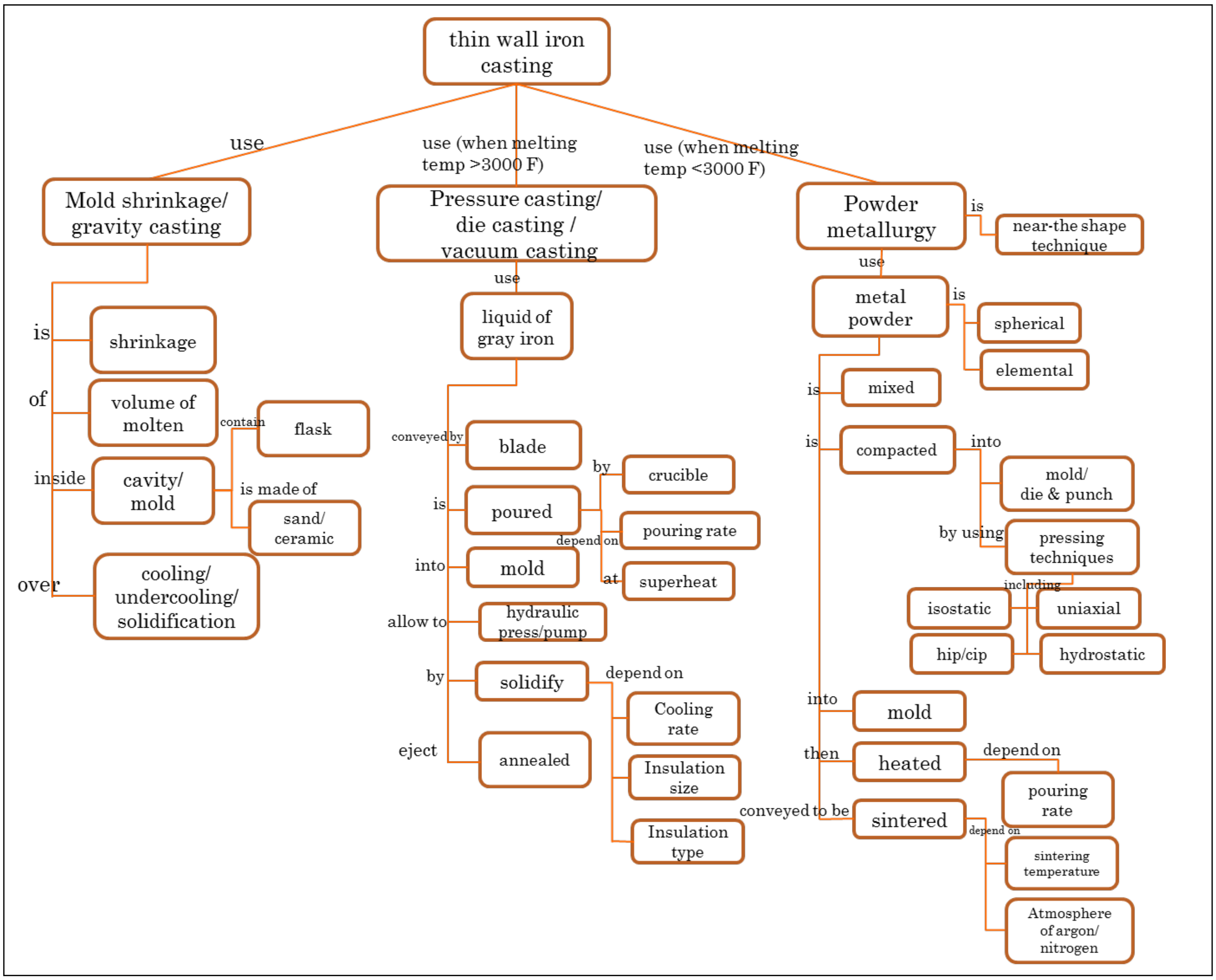

Fig. 3: The Ontology Designed for the Patent Search Experiment

- *Search Engine Training.* The researcher introduces the expert to the Google patent search engine (www.patents.google.com). The researcher spends about five minutes with the expert, practicing operating the search engine together. The researcher teaches some search tips to the expert so that he/she has enough skills to conduct a complete patent retrieval efficiently and effectively on the website.
- *Execution.* After being trained on the Google patent search engine, the experts started the patent retrieval process, as shown in Fig. 1. They aimed to generate a list of potentially relevant patents, extract a set of metallurgical patents from the US patent database, and judge which of these extracted patents contained enabling technologies for developing a new and improved kitchen skillet.

In every session, the researcher recorded data on the following variables: the behavioural variables described in Section "Measures for Patent Search Behaviour" and the well-known search performance criteria for efficiency, effectiveness, and reliability described in Section "Measures for Patent Retrieval Performance" are derived: time spent on patent search and patent retrieval (minutes), number of queries, number of keywords; the number of distinct keywords, time spent on patent retrieval; the number of patents judged; the net number of patents; the number of duplicates; the number of contradictions; the number of relevant patents; the number of irrelevant patents. Once all data were gathered, the authors performed a correlation analysis of all variables, revealing a relationship between patent search behaviour and patent retrieval performance, assuming such a relationship exists.

Results

Results Pertaining to Patent Search Behaviour

All experts' raw data about patent search behaviour are shown in Table 2.

Table 2, The experts exhibited different behaviours regarding time spent, number of queries, total number of keywords, and total number of distinct keywords. Distinct keywords refer to all keywords used in the queries without considering their repetition.

Table 2: Basic Data on the Experts' Search Behaviour

Data	*E1*	*E2*	*E3*	*E4*	*E5*	*E6*	*E7*
Number of Queries	12	16	18	13	12	12	14
Number of keywords	40	76	88	45	145	83	106
Number of distinct keywords	19	15	22	12	22	20	20

Results Pertaining to Patent Retrieval Performance

Table 3 shows the raw data for the patent retrieval performed by the experts. Experts E2 through E7 spent between 40 and 60 minutes in one session; expert E1 spent 150 minutes in two sessions. The number of patents found by the participants ranged from 13 to 60. The net number of patents ranged from 13 to 46.

Table 3: Basic Data on the Performance of the Experts in Patent Retrieval

Data	*E1*	*E2*	*E3*	*E4*	*E5*	*E6*	*E7*
Time spent on patent retrieval (minutes)	150	50	40	60	43	53	60
Number of patents judged	25	38	14	60	13	15	19
Net number of patents	24	33	14	46	13	15	18
Number of duplicates	1	5	0	14	0	0	1
Number of contradictions	0	2	0	0	0	0	0
Number of relevant patents	8	13	12	5	8	8	13
Number of irrelevant patents	16	23	2	41	5	7	5

Evaluation of Measures

The measures described in "Measures for Patent Search Behaviour" and "Measures for Patent Retrieval Performance" were calculated from the results as mentioned in Table 3.

Evaluating Measures Pertaining to Patent Search Behaviour

Determining Keyword Diversity

As shown in Fig. 4, all experts, except E1, use 15% to 25% distinct keywords in their patent search, whereas E1 uses 47.5% distinct keywords in his/her patent search. This result illustrates how expert E1 develops his queries by applying more diversified knowledge (keywords) and exploring a greater variety of technologies than the other experts.

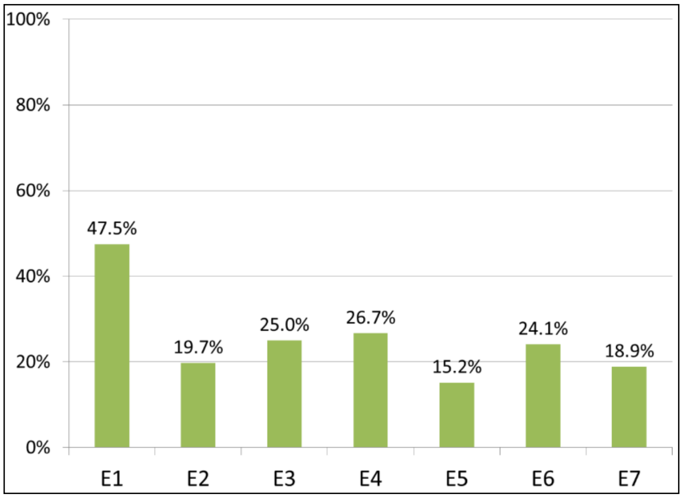

Fig. 4: Keyword Diversity (% Distinct Keywords)

Determining Query Complexity

As shown in Fig. 5, experts E1, E2, E3, and E4 used 3 to 5 keywords per query. On the other hand, experts E5, E6, and E7 used 7 to 12 keywords per query. The difference between these two groups shows that more experienced experts used fewer complex queries.

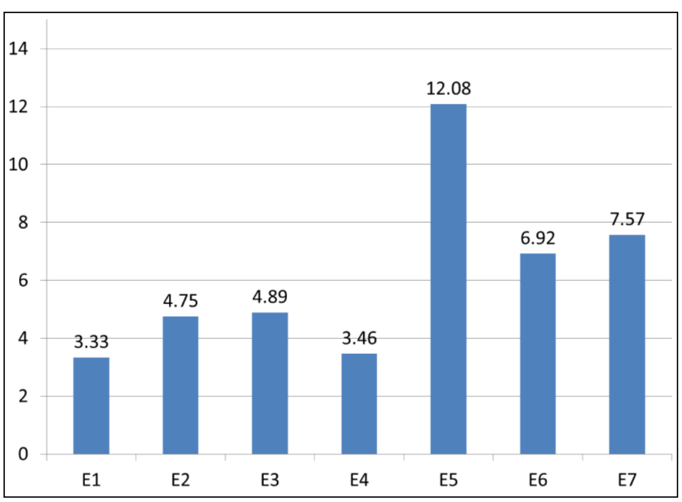

Fig. 5: Query Complexity (Keywords Per Query)

Determining Search Speed

Fig. 6 shows how fast the experts designed and implemented their queries. This result shows how patent searches can be time-consuming and consequently expensive. It also illustrates a high variability of search speed among experts. For example, expert E1, the slowest expert, spent, on average, 12.5 minutes implementing one query in each round of patent retrieval, and expert E3, the fastest expert, similarly spent 2.22 minutes to implement one query.

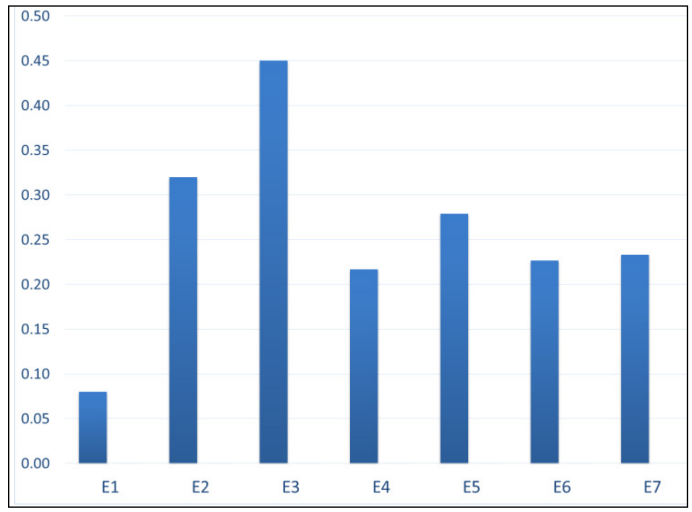

Fig. 6: Search Speed (Queries Per Minute)

Evaluating Measures for Patent Retrieval Performance

Evaluating Reliability

The similarity indexes from which the reliability of patent retrieval is determined were calculated from data collected during the patent search sessions. The results of these calculations are displayed in Tables 4 and 5. The similarity numbers shown in Table 4 indicate that, despite being presented with an ontology that provides similar knowledge about the technologies targeted in this research, the experts applied surprisingly dissimilar keywords. The similarity of the keywords varies between 5% and 35%. The similarity of the retrieved patents relevant to the ontology is less than 10%, according to Table 5.

Table 4: Similarity Indexes between Patent Searches Based on the Keywords Used

Experts	E1	E2	E3	E4	E5	E6	E7
E1	--	10%	5%	11%	14%	15%	12%
E2	10%	--	19%	29%	16%	35%	31%
E3	5%	19%	--	21%	10%	17%	24%
E4	11%	29%	21%	--	13%	23%	24%
E5	14%	16%	10%	13%	--	17%	28%
E6	15%	35%	17%	23%	17%	--	30%
E7	12%	31%	24%	24%	28%	30%	--
Ave. of similarity	**11%**	**23%**	**16%**	**20%**	**16%**	**23%**	**25%**

Table 5: Similarity Indexes between Patent Retrievals Based on the Relevant Patents Found

Expert	E1	E2	E3	E4	E5	E6	E7
E1	--	0.0%	0.0%	1.4%	0.0%	2.6%	0.0%
E2	0.0%	--	9.5%	8.3%	0.0%	2.2%	0.0%
E3	0.0%	9.5%	--	1.7%	0.0%	0.0%	0.0%
E4	1.4%	8.3%	1.7%	--	1.7%	3.4%	0.0%
E5	0.0%	0.0%	0.0%	1.7%	--	0.0%	3.3%
E6	2.6%	2.2%	0.0%	3.4%	0.0%	--	0.0%
E7	0.0%	0.0%	0.0%	0.0%	3.3%	0.0%	--
Ave. of similarity	**0.68%**	**3.34%**	**1.87%**	**2.77%**	**0.84%**	**1.37%**	**0.56%**

The average of the similarity indexes shown in Tables 4 and 5 were utilised to calculate the reliability measure. The results of these calculations are displayed in Fig. 7. They illustrate that the reliability of the patent retrieval performed by the experts is very low.

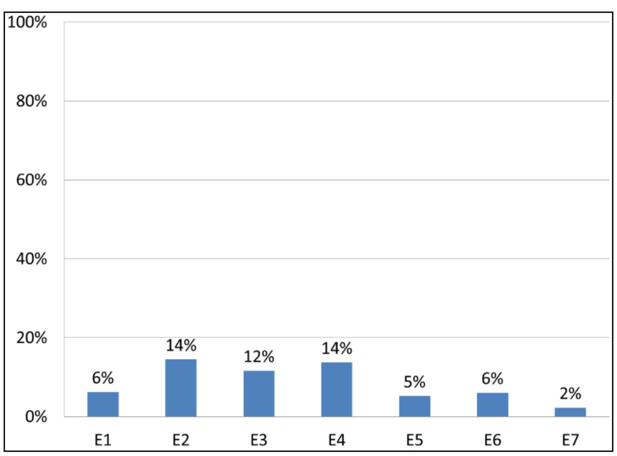

Fig. 7: Reliability of the Patent Searches (%)

Evaluating Efficiency

Fig. 8 illustrates how time-consuming it is to find a relevant patent. This study's experts required 3 to 12 minutes to find a relevant patent. Thus, the efficiency of experts varied highly. Expert E2 and Expert E3 (the professors in the expert panel) demonstrated higher efficiency than the others, and Expert E1's efficiency was the lowest.

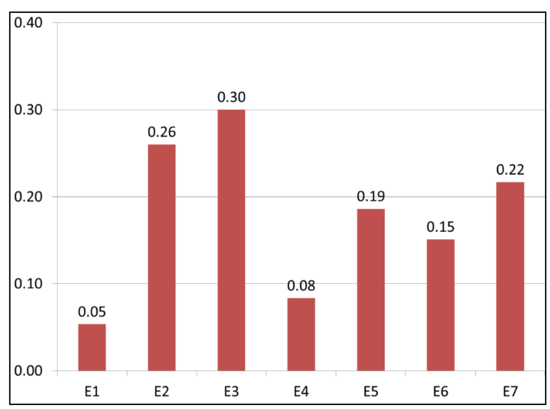

Fig. 8: Patent Retrieval Efficiency (Patents Retrieved Per Minute)

Evaluating Effectiveness

The effectiveness measures of the patent searches are shown in Fig. 9. Despite exhibiting low efficiency in patent retrieval, almost half of the participants demonstrate a relatively high degree of effectiveness in finding relevant patents.

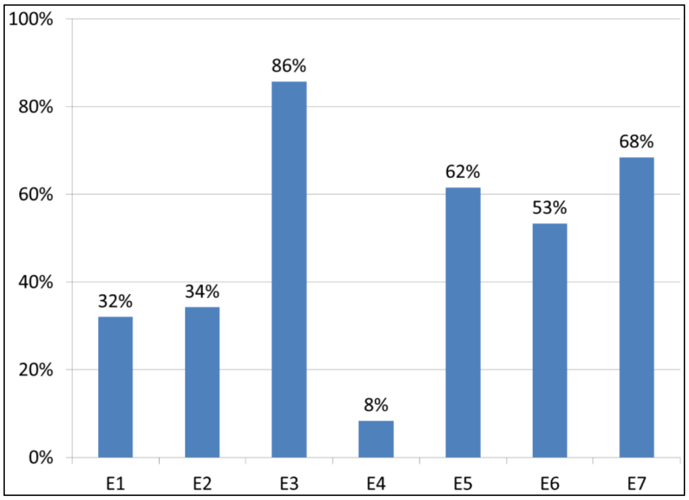

Fig. 9: Patent Retrieval Effectiveness (% Relevant Patents Found of Patents Judged)

Error

Retrieving duplicate and misjudging patents constitute the main sources of error in patent retrieval. As shown in Fig. 10, some experts committed errors that adversely affected their performance, and the error rate varied greatly. Expert E4 had an error rate of 23%, whereas experts E3, E5, and E6 made no errors.

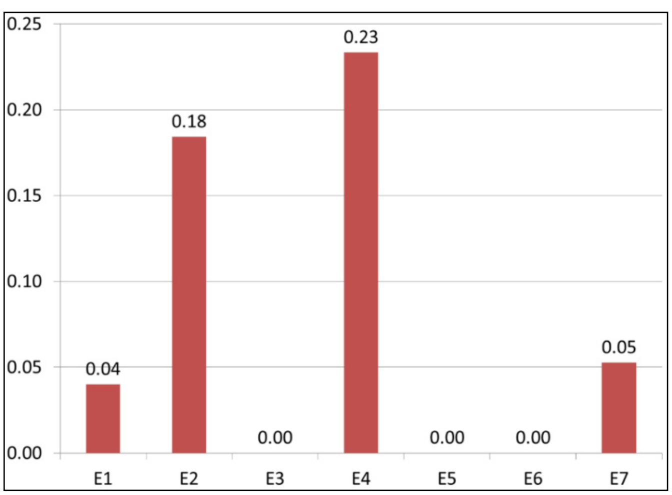

Fig. 10: Error Rates in Patent Retrievals

Analysis and Discussion

The patent search results showed that the experts performed poorly in patent retrieval. Their performance can be summarised as follows:

- *Reliability:* The patent searches are between 2% and 14.45%. These numbers show that the reliability is *very low*. The average of similarities between keywords (shown in Table 4) and the average of similarities between relevant patents found in the patent retrievals (shown in Table 5) illustrate that the experts used very few similar keywords and consequently found very few similar relevant patents despite having the ontology at their disposal.
- *Efficiency:* The *efficiency* of patent retrievals varies between 0.05 and 0.30, which is 0.18 on average. An expert who wants to find 100 *relevant* patents would have to spend 9.26 hours if he/she were to perform at a rate that reflects the average of these efficiencies. Considering cases with thousands of relevant patents, finding them would take a long time. Thus, the efficiency of the experts is *too low*.
- *Effectiveness:* The *effectiveness* of the patent retrieval ranges from 8%, which is very low, to 86%, which is fairly high. Half of the experts have more than 50% effectiveness, and half have less than 50%.
- *Error:* Some experts created duplicates in their patent retrieval and contradictions in their judgements. These errors, especially the contradictions, negatively affect effectiveness because the experts have misjudged some patents. Thus, they consider some irrelevant patents relevant and some relevant patents irrelevant.

The correlation between the measures is analysed by calculating the Pearson correlation coefficient in SPSS to gain insights into how patent search behaviour affects patent retrieval performance. The results of the correlation analysis are displayed in Table 6 and Fig. 11. The correlations between patent search measures and patent retrieval performance measures can be taken as *causation* relations because the search behaviour of the experts (as expressed by keyword diversity, query complexity, and search speed) influences their ability to retrieve patents and judge them as relevant or irrelevant. The explanations of the correlations between patent search behaviour measures and patent retrieval performance measures are as follows:

- *Keyword Diversity*: A moderately negative correlation exists between keyword diversity and efficiency (-0.65). The correlation is moderately significant; its p-value is 0.11. When an expert applies more diversified keywords, it means he/she performs a search on a greater variety of subjects, so he/she retrieves more patents and must spend more time evaluating and judging them. Therefore, a more diverse set of keywords negatively impacts the efficiency of his/her patent search. Also, keyword diversity has a moderate negative correlation with query complexity and a weak negative correlation with search speed.

Table 6: Correlation Analysis between the Measures

		Keyword Diversity	Query Complexity	Search Speed	Efficieny	Effectiveness	Reliabelity	Error
Keyword Diversity	Pearson Correlation	1.00	-0.67	-0.62	-0.66	-0.38	0.01	-0.01
	Sig. (2-tailed)		10%	14%	11%	40%	99%	99%
Query Complexity	Pearson Correlation	-0.67	1.00	0.17	0.28	0.50	-0.58	-0.52
	Sig. (2-tailed)	10%		72%	55%	26%	17%	23%
Search Speed	Pearson Correlation	-0.62	0.17	1.00	0.89	0.59	0.44	-0.10
	Sig. (2-tailed)	14%	72%		1%	16%	33%	83%
Efficieny	Pearson Correlation	-0.66	0.28	0.89	1.00	0.72	0.16	-0.21
	Sig. (2-tailed)	11%	55%	1%		7%	74%	65%
Effectiveness	Pearson Correlation	-0.38	0.50	0.59	0.72	1.00	-0.42	-0.80
	Sig. (2-tailed)	40%	26%	16%	7%		35%	3%
Reliabelity	Pearson Correlation	0.01	-0.58	0.44	0.16	-0.42	1.00	0.68
	Sig. (2-tailed)	99%	17%	33%	74%	35%		9%
Error	Pearson Correlation	-0.01	-0.52	-0.10	-0.21	-0.80	0.68	1.00
	Sig. (2-tailed)	99%	23%	83%	65%	3%	9%	

- *Query Complexity:* There is a weak negative correlation (-0.58) between query complexity and reliability. The correlation is moderately significant, with a p-value of 0.16. To explain this correlation, imagine two experts: expert A and expert B. Both experts perform a patent search for a similar concept, so they select their keywords from the same set of keywords to build up their queries. Now, imagine expert A develops more complex queries by putting more keywords in his/her queries. Despite using similar sets of keywords in developing queries and looking for a similar concept, expert A gains fewer patents than expert B because expert A applied more complex queries, narrowing down his/her domain search. The two imaginary experts come up with fewer similar patents in their search results, suggesting that more complex queries lead to less reliable patent searches.

- *Search Speed*: Unsurprisingly, search speed and efficiency have a strong, positive correlation. The faster an expert designs and implements queries in patent searches, the more patents he/she finds in a patent retrieval, and more of these patents might be judged as relevant. On the other hand, search speed is negatively correlated to effectiveness; however, this correlation is weak. Speeding up query design and implementation may negatively affect the number of relevant patents found in a patent retrieval process.

In addition to the abovementioned correlations, which explain how search behaviour impacts patent retrieval performance, there are some correlations between the performance measures for patent retrieval. The correlation between the error rate and effectiveness is highly negative. Since contradictions and duplications cause experts to retrieve fewer relevant patents, the error rate impacts effectiveness adversely. Also, the error rate has a moderate positive correlation with reliability. This correlation implies that experts exhibiting a high error rate in their judgement of patents may have applied more similar keywords and thus come up with more similar patents, and vice versa.

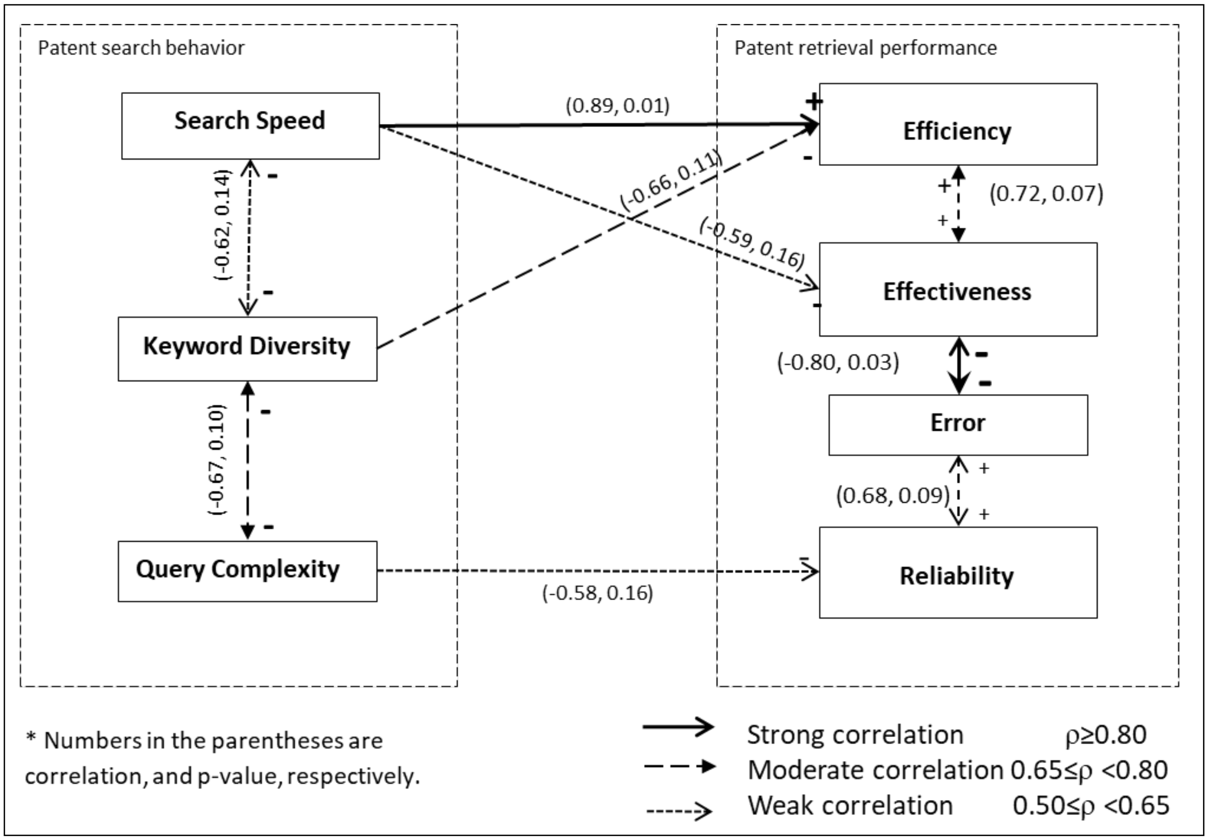

Fig. 11: Correlations between Patent Search Behaviour Measures and Patent Retrieval Performance Measures

Summary and Conclusion

The purpose of the research described in this paper was to explore the relationship between patent search behaviour by human experts and their performance at patent retrieval. To achieve this purpose, the authors of this paper conducted an empirical case study in which a researcher observes seven experts in patent retrieval as they search for, extract, and judge metallurgical patents that could contain enabling technologies for developing and producing a new and improved kitchen skillet. The study utilised two sets of measures: one evaluated search behaviour, and the other evaluated the patent retrieval performance of the participating experts. The study found correlations between these two sets of measures, establishing a relationship between patent search behaviour and retrieval performance. The study concludes that the success of keyword-based patent retrieval depends on search behaviour in query formulation and implementation. The study has also found that the performance of human experts in patent retrieval varies greatly from expert to expert. Human experts are *highly unreliable*, *highly inefficient*, and *relatively ineffective* at patent retrieval, even when similar knowledge regarding the subject of a specific patent retrieval is provided to them.

Implications

By establishing a direct link between patent search behaviour and retrieval performance, the study described in this paper contributes to the field of patent retrieval. Prior studies on search behaviour (e.g., Agichtein et al., 2006; Joho et al., 2015; Liu et al., 2012; Tamine

& Chouquet, 2017; Zhang, Anghelescu & Yuan, 2005; Zhang et al., 2015) were conducted in other, perhaps related areas of information retrieval. To the best of our knowledge, other than Joho et al. (Joho et al., 2010), no prior study on search behaviour had been performed on patent data. However, Joho et al. (Joho et al., 2010) did not draw any significant conclusions about search behaviour's impact on patent retrieval performance. Thus, the effect of patent search behaviour on patent retrieval constituted a subject that had hitherto remained unexplored.

The findings of the research described in this paper align with many prior studies in patent research and related fields, but they do not support the findings of some prior studies. For example, the conclusion that the success of keyword-based patent retrieval depends on search behaviour in query formulation and implementation is in alignment with Tannebaum and Rauber (Tannebaum & Rauber, 2015). In addition, this paper's observation that human experts are inefficient at patent retrieval is supported by the findings of Joho et al. (Joho et al., 2010), whose survey of patent users suggests that experts at patent retrieval must spend many hours to find patents that are relevant to the topic of their search. Furthermore, the high variability in patent retrieval performance that has been observed in this study is in alignment with Zhang et al**.** (Zhang, Anghelescu & Yuan, 2005) and Tamine & Chouquet (Tamine & Chouquet, 2017), who determined that agreement in relevance assessment is low, even among experts. However, the research of Zhang et al**.** (Zhang, Anghelescu & Yuan, 2005) and Tamine and Chouquet (Tamine & Chouquet, 2017) suggests that search behaviour is unrelated to search effectiveness. This finding is not supported by this paper's discovery of a moderate, negative correlation between search speed and effectiveness. In conjunction, all these findings lead to the conclusion that much about patent analysis remains unknown, suggesting that a significant amount of further research is required before the principles that govern the relationship between patent search behaviour and patent retrieval performance are discovered.

The study in this paper is based on the analysis of a single technology. This condition demonstrates a relationship between patent search behaviour and patent retrieval performance, but it limits the generalisability of the study's findings. Once again, further research would counteract this shortcoming. The authors of this paper suggest that this study should be repeated in multiple contexts and that a diverse set of technologies should be examined. Differences between the findings of these studies could then provide insight into the principles that govern the relationship between patent search behaviour and patent retrieval performance.

The research described in this paper suggests that patent retrieval performance could be improved by applying intelligent tools that increase search speed, avoiding the application of many diversified keywords, and designing less complex queries. Intelligent tools, in particular, facilitate more efficient, more effective, and more reliable patent retrieval in many ways (Madani, 2018). Specifically, recommender systems effectively provide keywords that are more relevant to the subject of a search. Also, machine learning (more specifically, classification methods) can assist an expert in automating patent judgement, help avoid errors in judgement, and speed up the patent retrieval process. Intelligent tools could thus relieve bottlenecks (Kumar et al., 2016; Montecchi et al., 2013) that plague the multitude of applications (Jürgens et al., 2012; Bonino et al., 2010) for patent retrieval and yield sufficiently accurate results to contribute significantly to the performance of patent retrieval by practitioners (Jürgens et al., 2012; Bonino et al., 2010).

Conflicts of Interest

The authors confirm that no financial, professional, or personal conflicts of interest could have influenced the work reported in this paper.

References

Agichtein, E., Brill, E., & Dumais, S. (2006). Improving web search ranking by incorporating user behavior information. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '06* (p. 19). New York, USA: ACM Press.

Atkinson, K. H. (2008). Toward a more rational patent search paradigm. In *Proceeding of the 1st ACM Workshop on Patent Information Retrieval - PaIR '08* (p. 37). New York, USA: ACM Press.

Baeza-Yates, R. (Ricardo), & Ribeiro-Neto, B. (2011). *Modern information retrieval: The concepts and technology behind the search*. Addison Wesley.

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

Chen, H., & Dhar, V. (1991). Cognitive process as a basis for intelligent retrieval systems design. *Information Processing & Management*, *27*(5), 405-432. doi:https://doi.org/10.1016/0306-4573(91)90060-Y

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*. doi:https://doi.org/10.46743/2160-3715/2003.1870

Jansen, B. J., Booth, D. L., & Spink, A. (2009). Patterns of query reformulation during Web searching. *Journal of the American Society for Information Science and Technology*, *60*(7), 1358-1371. doi:https://doi.org/10.1002/asi.21071

Joho, H., Azzopardi, L. A., & Vanderbauwhede, W. (2010). 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 - IIiX '10* (p. 13). New York, USA: ACM Press.

Joho, H., Jatowt, A., & Blanco, R. (2015). Temporal information searching behaviour and strategies. *Information Processing & Management*, *51*(6), 834-850.

Jürgens, J. J., Hansen, P., & Womser-Hacker, C. (2012). Going beyond CLEF-IP: The 'reality' for patent searchers? In *Third International Conference of the CLEF Initiative* (pp. 30-35). Rome, Italy: Springer, Berlin, Heidelberg. doi:https://doi.org/10.1007/978-3-642-33247-0_4

Kumar, R., Tripathi, R. C., & Singh, V. (2016). Keyword-based search and its limitations in the patent document are used to secure the idea from infringement. *Procedia Computer Science*, *78*, 439-446. doi:https://doi.org/10.1016/J.PROCS.2016.02.086

Liu, C., Liu, J., Cole, M., Belkin, N. J., & Zhang, X. (2012). Task difficulty and domain knowledge effects on information search behaviors. In *Proceedings of the American Society for Information Science and Technology* (vol. 49, pp. 1-10). Wiley-Blackwell.

Lupu, M., Mayer, K., Tait, J., & Trippe, A. J. (Eds.). (2011). *Current challenges in patent information retrieval* (vol. 29). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:https://doi.org/10.1007/978-3-642-19231-9

Machado, C., & Davim, P. (2017). *Productivity and organizational management*. De Gruyter.

Madani, F. (2018). *Opportunity identification for new product planning: Ontological semantic patent classification*. Portland State University.

Magdy, W., & Jones, G. J. F. (2011). A study on query expansion methods for patent retrieval. In *Proceedings of the 4th Workshop on Patent Information Retrieval - PaIR '11* (p. 19). New York, USA: ACM Press. doi:https://doi.org/10.1145/2064975.2064982

McDonald-Maier, L. (2009). esp@cenet®: Survey reveals new information about users. *World Patent Information*, *31*(2), 142-143.

Montecchi, T., Russo, D., & Liu, Y. (2013). Searching in cooperative patent classification: Comparison between keyword and concept-based search. *Advanced Engineering Informatics*, *27*(3), 335-345. doi:https://doi.org/10.1016/J.AEI.2013.02.002

Newton, D. (2000). A survey of new British library patent information centre users. *World Patent Information*, *22*(4), 317-323.

Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013). Using of Jaccard coefficient for keywords similarity. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*. Hong Kong.

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. doi:https://doi.org/10.1016/J.ESWA.2015.01.050

Shalaby, W., & Zadrozny, W. (2019). Patent retrieval: A literature review. *Knowledge and Information Systems*, 1-11. doi:https://doi.org/doi:https://doi.org/10.1007/s10115-018-1322-7

Shiri, A., & Revie, C. (2006). Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation. *Journal of the American Society for Information Science and Technology*, *57*(4), 462-478. doi:https://doi.org/10.1002/asi.20319

Sormunen, E. (2000). *A method for measuring wide-range performance of Boolean queries in full-text databases*. University of Tampere.

Spink, A., Jansen, B. J., & Cenk Ozmultu, H. (2000). Use of query reformulation and relevance feedback by Excite users. *Internet Research*, *10*(4), 317-328. doi:https://doi.org/10.1108/10662240010342621

Sutcliffe, A., & Ennis, M. (1998). Towards a cognitive theory of information retrieval. *Interacting with Computers*, *10*(3), 321-351. doi:https://doi.org/10.1016/S0953-5438(98)00013-7

Tamine, L., & Chouquet, C. (2017). On the impact of domain expertise on query formulation, relevance assessment, and retrieval performance in clinical settings. *Information Processing & Management*, *53*(2), 332-350.

Tannebaum, W., & Rauber, A. (2015). Learning keyword phrases from query logs of USPTO patent examiners for automatic query scope limitation in patent searching. *World Patent Information*, *41*, 15-22. doi:https://doi.org/10.1016/J.WPI.2015.02.005

Wacholder, N. (2013). Interactive query formulation. In *Annual Review of Information Science and Technology* (pp. 157-196). Wiley-Blackwell. doi:https://doi.org/10.1002/aris.2011.1440450111

Wang, P., & Soergel, D. (1998). A cognitive model of document use during a research project. Study I. Document selection. *Journal of the American Society for Information Science*, *49*(2), 115-133. doi:https://doi.org/10.1002/(SICI)1097-4571(199802)49:2<115::AID-ASI3>3.0.CO;2-T

Xie, Z., & Miyazaki, K. (2013). Evaluating the effectiveness of keyword search strategy for patent identification. *World Patent Information*, *35*(1), 20-30. doi:https://doi.org/10.1016/J.WPI.2012.10.005

Zhang, X., Anghelescu, H. G. B., & Yuan, X. (2005). Domain knowledge, search behaviour, and search effectiveness of engineering and science students: An exploratory study. *Information Research: An International Electronic Journal*, *10*(2).

Zhang, X., Liu, J., Cole, M., & Belkin, N. (2015). Predicting users' domain knowledge in information retrieval using multiple regression analysis of search behaviors. *Journal of the Association for Information Science and Technology*, *66*(5), 980-1000.