Healthcare information retrieval:

A survey of user needs, tasks and requirements

Tony Russell-Rose, Jon Chamberlain

UXLabs Ltd, 3000 Cathedral Hill, Guildford, GU2 7YB, UK
tgr@uxlabs.co.uk

RÉSUMÉ. *Les connaissances médicales se développent si rapidement qu'il est difficile pour les professionnels de la santé de se tenir au courant. À mesure que le volume des études publiées augmente chaque année, l'écart entre les connaissances en recherche et les pratiques professionnelles augmente de plus en plus. Les professionnels de l'information en soins de santé jouent un rôle essentiel dans la réduction de cet écart en synthétisant les résultats de la recherche biomédicale sous la forme de revues systématiques de la littérature. Ce faisant, ils abordent systématiquement certains des problèmes les plus difficiles d'extraction de l'information (IR) de n'importe quelle profession, mais les outils les outils sur lesquels ils s'appuient sont insuffisants à bien des égards. Cet article présente les résultats d'une enquête sur les pratiques et les besoins en matière de recherche d'informations concernant les systèmes et applications IR.*

ABSTRACT. *Medical knowledge is growing so rapidly that it is difficult for healthcare professionals to keep up. As the volume of published studies increases each year, the gap between research knowledge and professional practice grows ever wider. Healthcare information professionals play a vital role in closing this gap by synthesizing the findings of biomedical research in the form of systematic literature reviews. In so doing, they routinely address some of the most challenging information retrieval (IR) problems of any profession, but the tools they rely on fall short in many respects. This paper presents the results of a survey investigating their information seeking practices and needs regarding IR systems and applications.*

MOTSCLÉS : *Stratégie de recherche, examen systématique, tâche de recherche, professionnel de l'information sur la santé, processus d'examen systématique, recherche professionnelle, formulation de requêtes.*

KEYWORDS: *search strategy, systematic review, search task, healthcare information professional, systematic review process, professional search, query formulation*

1. Introduction

Medical knowledge is growing so rapidly that it is difficult for healthcare professionals to keep up. As the volume of published studies increases each year, the gap between research knowledge and professional practice grows ever wider (Bastian et al., 2012). Frontline healthcare providers responding to the immediate needs of patients may employ a Web-style search for diagnosis purposes, with Google being reported to be a useful diagnostic tool (Tang and Ng, 2006). However, the credibility of results depends on the domain (Kitchens et al., 2014). Medical staff may also perform more in-depth searches, such as rapid evidence reviews, where a concise summary of what is known about a topic or intervention is required (Hemmingway and Brereton, 2009).

However, systematic literature reviews play the primary role in closing the gap between published research and medical practice, by synthesizing the complex, incomplete and at times conflicting findings of biomedical research into a form that can readily inform healthcare decision making (Elliott et al., 2014). The systematic review process relies on the painstaking and meticulous searching of multiple literature sources using complex multi-line search strategies that often consist of hundreds of keywords, operators and ontology terms, such as:

```
1 Attention Deficit Disorder with Hyperactivity/
2 adhd
3 addh
4 adhs
5 hyperactiv$
6 hyperkin$
7 attention deficit$
8 brain dysfunction
9 or/1-8
10 Child/
11 Adolescent/
12 child$ or boy$ or girl$ or schoolchild$ or adolescen$
or teen$ or "young person$" or "young people$" or youth$
13 or/10-12
14 acupuncture therapy/or acupuncture, ear/or
electroacupuncture/
15 accupunct$
16 or/14-15
17 9 and 13 and 16
```

Performing a systematic review is a resource-intensive and time consuming undertaking, sometimes taking years to complete (Lefebvre et al., 2011). It involves a lengthy content production process whose output relies heavily on the quality of the initial search strategy, particularly in ensuring that the scope is sufficiently

exhaustive and that the review is not biased by easily accessible studies (Tsafnat et al., 2014).

Consequently, numerous studies have been performed in order to investigate the systematic review process and to better understand the challenges involved in strategy development. For example, Grant (2004) used a combination of a questionnaire and interviews to study researchers' experiences of searching the literature. Sampson et al. (2008) used a Web-based survey and peer review forums to investigate what elements of the search process have the most impact on the resulting evidence base. More recently, Ciapponi and Glujovsky (2012) used an online survey to study the early stages of systematic review. However, none of these studies used an approach which supports direct comparison with other professions that use complex search, such as patent law. The study reported here was designed to fill gaps in this knowledge by investigating the following questions:

- What kinds of search tasks do healthcare information professionals undertake and how does this compare with other professions?
- How do they formulate search strategies and what approaches do they use?
- How are search results evaluated and prioritised?
- What other criteria do they apply when evaluating search systems?

2. Method

The survey instrument consisted of an online questionnaire of 58 questions divided into five sections[1]. To facilitate comparisons with other professions, it was designed to align with Joho et al.'s (2010) survey of patent professionals and where possible with Geschwandtner et al.'s study of medical practitioners (Geschwandtner et al., 2011). The sections were:

- Demographics: The background and professional experience of the respondents;
- Search tasks: The types of search task that respondents perform in their work;
- Query formulation: The techniques used to formulate search strategies;
- Evaluation: How respondents assess and evaluate the results of their search tasks;
- Ideal search engine: Any other features additional to those described above.

The survey was designed to be completed in approximately 15 minutes, and was pre-tested for face validity by two health sciences librarians. Healthcare information professionals were recruited by sending an email invitation to mailing lists and interest groups that deal with systematic reviews and medical librarianship. The

[1] Available from https://isquared.wordpress.com/

survey was conducted using SurveyMonkey, and ran from July to September 2015. 218 responses were received, of which 107 were complete. Only complete surveys were examined. Responses to numeric questions were not constrained to integers as a pilot survey had shown that respondents preferred to put in approximate and/or expressive values. Text responses corresponding to numerical questions (questions 14-22 and 32-38, 16 in total) were normalised as follows:

- when the respondent specified a range (e.g. 10-20 hours), the midpoint was entered (e.g. 15 hours);
- when the respondent indicated a minimum (e.g. 10+ years), the minimum was entered (e.g. 10 years);
- when the respondent entered an approximate number (e.g. about 20), that number was entered (e.g. 20).

After normalising, 8.3% (142/1,712) responses contained no numerical data and 21.6% (370/1,712) responses were normalised.

3. Results

3.1 *Demographics*

A large proportion (86%) of the 107 respondents were female. Their ages were distributed bimodally, with peaks at 39-45 (23%) and 53-59 (18%). Most respondents worked full time (86%) and the clients they worked for were predominantly internal (73%). Respondents had considerable industry experience (median 17 years). The majority of respondents were based in the UK (51%) or the USA (27%).

3.2 *Search tasks*

The data sources most frequently searched were Medline (96%), Cochrane Library (88%) and Embase (80%). The majority of respondents (87%) used previous search strategies or templates at least sometimes. In addition, most respondents (88%) routinely share their search strategies in some form, either with colleagues in their workgroup, more broadly within their organization or with clients or as part of a published review. Table 1 shows the amount of time that respondents spend in completing search tasks, the time spent formulating search strategies, and the number of strategy lines they use. On average, it takes 4 hours to complete a search task, with a typical strategy consisting of 15 lines and taking around 60 minutes to formulate. The results also suggest that search tasks follow an iterative paradigm, with successive phases of document retrieval combined with other activities such as analysing results, exporting documents, collecting citations, etc. Task completion time is shorter than patent search (4 hours vs. 12 hours) but is longer than typical web search tasks (Broder, 2002).

	Min	Median	Max
Search task completion time (hrs)	1	4	14
Strategy formulation time (mins)	20	60	228
Number of strategy lines	5	15	59
Number of results examined	10	175	850
Time to assess relevance of a single result (mins)	1	3	10

Table 1. *Effort in completing search tasks*

3.3 *Search query formulation*

Here we examine the mechanics of the query formulation process, by asking respondents to indicate a level of agreement to various statements using a 5-point Likert scale ranging from strong disagreement (1) to strong agreement (5) – see Fig. 1.

The results suggest two observations in common with patent search. Firstly, the average of all but one of the features is above 3 (neutral), which suggests a willingness to adopt a wide range of search functionality to complete search tasks. This represents a marked contrast to the behaviour of typical web searchers who rarely, if ever, use any advanced search functionality (Spink et al., 2001). Secondly, the use of Boolean logic was shown to be the most important feature (4.97), most likely reflecting the need for transparency and repeatability within the systematic review process. This was closely followed by the use of synonyms and related terms (4.79). A number of other syntactic features, notably proximity operators, truncation and wildcarding, all scored highly, reflecting a need for fine control over search strategies. By contrast, weighting of results (where results are ordered by relevance citeria) was seen as least important, which suggests that the benefit provided is outweighed by a potential lack of transparency and repeatability in the retrieval process that is required for systematic reviews.

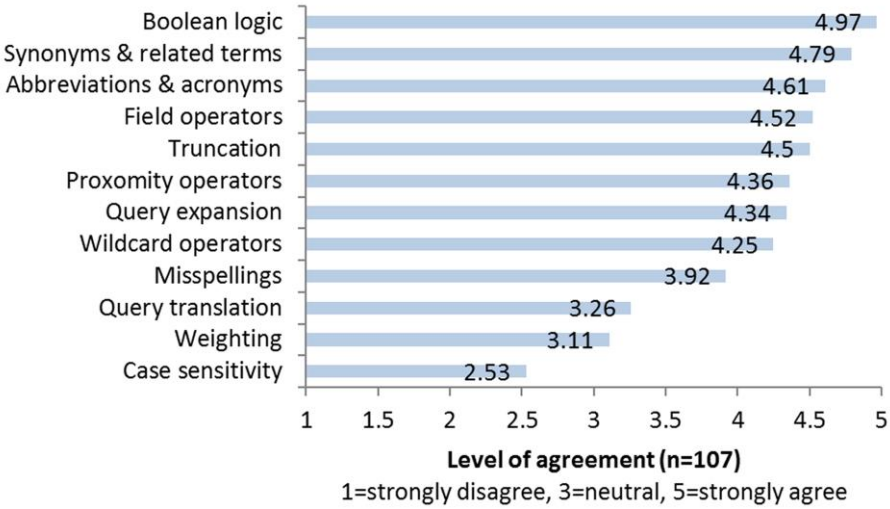

Figure 1. *Importance of query formulation functionality*

3.4 *Results evaluation*

In this section we examine the results evaluation process, as shown in Table 1. The number of results examined is higher than that of patent searchers (175 vs. 100) and higher than their ideal number of results returned (100). The time to assess each result is shorter than patent searchers (3 vs. 5 mins). However, the search process in systematic review is only the first stage of a work flow in which the retrieved documents are exposed to further phases of analysis and evaluation, often involving other colleagues and reviewers.

We then asked respondents to indicate how frequently they use search limits and other criteria to narrow down results. Publication date was the most important (3.44), followed by publication language (3.36). These choices are consistent with the criteria applied by medical professionals in other surveys (Geschwandtner et al., 2011). We also examined respondents' strategies for interacting with results sets. The most popular approaches were to start with the result that looked most relevant (54%) or simply select the first result (23%). No respondent suggested selecting the most trustworthy source, which contrasts sharply with the approach used in other studies of the healthcare profession, e.g. Geschwandtner et al. (2011). This most likely reflects the difference between the curated databases referred to in this study and the relatively uncontrolled web resources used in Geschwandtner's study.

We also investigated other features related to search management, organisation and history. Combining search queries and combining search results were rated as the most important (4.82 and 4.80 respectively). This reflects the line-by-line strategy building approach offered by most native databases. Participants also rated the ability to export search queries (histories) highly, reflecting their need to publish completed search strategies as part of their professional practice.

4. Discussion & Conclusions

This paper summarises the results of a survey of health information professionals in a form that allows direct comparison with other professions. In this section we discuss the findings with verbatim quotes from respondents shown in italics where applicable.

Health information professionals display a number of search characteristics that differentiate their behaviour from web search (Lupu et al., 2014), such as lengthy search sessions, different notions of relevance, different sources searched separately, and the use of specific domain knowledge. They use complex search strategies, and actively cultivate skills in the formulation of such expressions. Previous studies have shown that as many as 90% of published strategies contained an error, including spelling errors, truncation errors, logical operator errors, etc. (Sampson and McGowan, 2006). A number of respondents suggested that healthcare information professionals need better query formulation support: "…automate checking of parentheses, operators and field codes…". They work across multiple platforms and share a need for greater standardisation and consistency between suppliers: "A service that could map search strategy between databases would save a lot of time." They routinely share their strategies suggesting a need for improved facilities for managing and sharing strategies such as: "…being able to download, share, remix, transfer and translate search strategies". There is also a need for translating strategies between terminologies, ideally with "one universal thesaurus of medical terminology for all databases" and advanced strategy editing functionality: "…move search lines up and down the history…", "…add tags or descriptions to search strategies, sort by name, topic or date".

These findings have important consequences for the IR community and the assumptions underlying many of its research priorities. For example, much academic research continues to assume that searches are formulated as natural language queries, but this study shows that many professions prefer to formulate their queries as Boolean expressions (Tait, 2014). In closing, we would hope that these findings will inform the development of future information retrieval systems and applications, not only for those undertaking systematic reviews but also for professionals performing other healthcare-related search tasks.

5. References

Bastian, H., Glasziou, P. and Chalmers, I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Medicine* 7.9:1112, 2012.

Broder, A. A taxonomy of web search. *ACM SIGIR Forum*, 36(2):3-10, 2002.

Ciapponi, A. and Glujovsky, D. Survey among Cochrane authors about early stages of systematic reviews. In: *Proceedings of the 20th Cochrane Colloquium*, 2012.

Elliott, J. H., Turner, T., Clavisi, O., Thomas, J., Higgins, J., Mavergames, C. and Gruen, R. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Medicine*, 11(2)2:e1001603, 2014.

Geschwandtner, M., Kritz, M. and Boyer, C. D8.1.2: Requirements of the health profes-sional search. *Technical report*, Khresmoi Project, 2011.

Grant, M. J. How does your searching grow? A survey of search preferences and the use of optimal search strategies in the identification of qualitative research. *Health Information and Libraries Journal*, 21.1:21-32, 2004.

Hemingway, P. and Brereton, N. What is a systematic review? *Hayward Medical Communications* 2ed, 2009.

Joho, H., Azzopardi, L. and Vanderbauwhede, W. A survey on patent users: An analysis of tasks, behavior, search functionality and system requirements. In: *Proceedings of the 3rd Symposium on Information Interaction in Context (IIiX 2010)*, 2010.

Kitchens, B., Harle, C. and Li, S. Quality of health-related online search results. *Decision Support Systems* 57:454 – 462, 2014.

Lefebvre, C., Manheimer, E. and Glanville, J. Searching for Studies. In: Higgins, J. and Green, S. (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.* The Cochrane Collaboration, 2011.

Lupu, M., Salampasis, M. and Hanbury, A. Domain specific search. *Professional Search in the Modern World*. Springer International Publishing, 2014.

Sampson, M. and McGowan, J. Errors in search strategies were identified by type and frequency. *Journal of Clinical Epidemiology*. 59(10):1057–63, 2006.

Sampson, M., McGowan, J., Lefebvre, C., Moher, D., and Grimshaw, J. PRESS: Peer review of electronic search strategies. Ottawa: *Canadian Agency for Drugs and Technologies in Health*, 2008.

Spink, A., Wolfram, D., Jansen, M. and Saracevic, T. Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3):226–234, 2001.

Tait, J. An introduction to professional search. *Professional Search in the Modern World. Springer International Publishing*, 2014.

Tang, H. and Ng, J. Googling for a diagnosis—use of Google as a diagnostic aid: internet based study. *British Medical Journal* 333:1143, 2006.

Tsafnat, G., Glasziou, P., Choong, M. K., Dunn, A., Galgani, F. and Coiera, E. Systematic review automation technologies. *Systematic Reviews* 3(1):74, 2014.