

Dean Mason

regarding old versions of rulebooks, for example. Most of these points have been raised with service providers already, but not all of them have been addressed as yet, unfortunately.

It will be interesting to watch the future development of these resources, as it will inevitably be linked to

new developments in IT, as well as the expectations of users.

Many thanks to everyone who completed a questionnaire, and if anyone has any feedback on the above please let me know.

My e-mail address is janice.edwards@mms.co.uk

Legal Information Management, 6 (2006), pp. 246–250

© The British and Irish Association of Law Librarians Printed in the United Kingdom

doi:10.1017/S1472669606000831

Legal Information Retrieval Study – Lexis Professional and Westlaw UK

This article by Dean Mason looks at the retrieval effectiveness of the online legal research tools Lexis Professional and Westlaw UK and is the result of research carried out for his Masters Degree in Information Science.

Introduction

The majority of legal institutions now use online databases to provide access to legal information, especially case law. There are many advantages to such systems, but the main one, it seems, is the ability of such resources to provide access to a wealth of information at the user's desktop. Information, however, is worthless without the ability to search and retrieve relevant material. How information is best retrieved has been the focus of much debate, with consideration given to how to resolve a complex range of information needs. Two common needs are a high precision and high recall search, i.e. the user requires the most relevant items or all the relevant items (Chowdhury, 2004). To resolve these needs it appears that providers of online legal databases, the main two contenders being Lexis Professional and Westlaw UK, have enabled users to search by relevance. The idea being that users are presented with a ranked list of results – the most relevant should be at the top and the least at the bottom. Therefore, when searching for case law on a



Dean Mason

particular point of law, lawyers would, in theory, have the key cases at the top of the list. Those of less importance, which also may be significant, will be further down this list. This order emphasises the need for both recall and precision measures. As a result there have been evaluations, but mainly US-based, of how effective specific systems have been in providing relevant results. The aim of this research article is to further this research to contribute to both the information retrieval and online legal research domains of study.

In a legal context, the need for a high recall and precision search is equal to the importance and nature of using case law as precedent. The doctrine of precedent can be understood in the basic sense that the English courts have to follow previous decisions, in the form of case law, within “more or less well-defined limits” (Williams and Smith, 2002, p.95). Therefore, when lawyers carry out a particular task on a point of law (e.g. advising clients, writing an article, producing know-how) they often need to see all available case law. This can be broken down further: they need to see the key authoritative cases on a point of law (thus, a high

precision search is necessary); and, if such case law does not exist, it would be important to see others that may not be as authoritative but possibly still significant (thus, a high recall search is necessary).

History

Information retrieval has come a long way in a short space of time and legal research appears to have caught up with the pace. This is evident as the key legal research texts written in the mid-to-late 1990s are now out-of-date and have been rewritten. The focus of early legal research texts is on content, e.g. what you can find, as opposed to how. The emphasis, however, has shifted with the arrival of improved technology, information retrieval evaluations and user studies. As a result, a range of tools have been developed to aid the user to find documents that meet specific information needs, e.g. high precision and recall.

There has been a range of previous research that evaluates the effectiveness of legal information retrieval. The first major evaluation was Blair and Maron's STAIRS project in 1985, which sought to evaluate the recall and precision of a full-text litigation support system comprising approximately 40,000 documents using 51 information needs (1985). The key research since has been carried out in the US by Dabney (1993) and Gerson (1999). The objective of both studies was to evaluate the retrieval effectiveness of the US versions of Lexis and Westlaw. Dabney searched a collection of in excess of one million case reports using 23 information needs and Gerson (1999) does not specify the document size of the database tested, but did so using 22 information needs.

Aims and objectives

The aim of my research was to evaluate Lexis and Westlaw to see how effectively each retrieves relevant results. The research approach adopted comprised a laboratory-style evaluation using real information needs from a legal context, i.e. a medium-sized commercial law firm based in the City of London. This research aimed to provide an indication of how effective the relevance feature is on each system when real queries are applied and, therefore, how useful the systems proved to be within an active legal context. This information could be useful to those who are considering subscribing to the featured systems; creating or updating a retrieval system; or commenting upon the effectiveness of each from a legal research perspective. The research also has the potential to pave the way for further investigation. Thus, it could be used to look more formally at the factors that affect retrieval and to aid an evaluation of users' needs when using these systems.

Methodology

The initial phase of the research was to investigate and choose the online systems that would be tested. The aim was to find two or more systems that were the least dissimilar in terms of functionality and content, so that they could be compared within a reasonable framework, and the decision was Lexis and Westlaw. The next important factor was to ensure that the source to be tested on each system was the least dissimilar. As with the choice of system, a number of conditions had to be the same, or similar, to make the test as fair as possible, and to be applicable to the scope of the project. The type of source chosen on each system was the Incorporated Council of Law Reporting for England & Wales (ICLR) reports of cases (i.e. The *Law Reports*). The *Law Reports* are made available electronically by Lexis and Westlaw and are therefore likely to provide an exact set of documents. An analysis of a small set of the same cases from the *Law Reports* source on each system revealed that there are no visible editorial differences, i.e. extra indexing, summaries, etc. It was important to eliminate any inconsistencies between the sources to ensure that it was the search facilities being tested and not the content.

The systems were tested with real information needs selected from request sheets that have been previously filled out by information professionals in the previously mentioned legal context. Fifty of these information needs, on a range of subjects, were chosen from a two-year archive. The needs were originally recorded using a "pre-search interview" (as discussed in Chowdhury, 2004) to determine the user's exact information requirement. The above information needs were formalised into a query statement using a specific strategy. This was necessary to explore the concepts and to refine and modify the query to ensure the need was expressed sufficiently to retrieve the maximum amount of relevant results. The systems were tested with 50 queries each and only the first 10 results from each search (a total of 1000 documents) were judged. The queries were submitted to each system in turn within a similar period of time, so that changes in content on the systems did not influence the validity of the results.

The evaluation policy that was used to judge each result for relevance was the stated information need, e.g. "cases on applications for summary judgments in respect of copyright infringement". The cases returned were only considered relevant if the subject criteria as a whole were met, e.g. if a case was returned that only considered "copyright" or only "summary judgment" – this was not considered relevant, but a case displaying a relationship between the concepts would be judged relevant. A general policy of two diagnostic measures was used for the whole evaluation to state what action to take should documents be duplicated or not returned. Any duplicates found were considered as not relevant, as these were considered a deficiency on the part of the system.

Likewise, any documents not retrieved, i.e. because of broken links or other technical issues, were also considered not relevant.

The TREC measures of performance evaluation were used in the assessment of the data (Chowdhury, 2004). The specific measures employed were: binary relevance judgements - to decide that either a document is relevant or not relevant; precision at five and 10 documents retrieved (P@5/P@10) - the number of relevant documents retrieved in the first five and 10 results; and, estimated average precision (EAP) – to provide an indication of how well each system provides relevant results against the total number relevant (recall). It is impossible to know the total number of documents on each system that were relevant to a specific information need as this would require every document to be judged for relevance.

Key findings

The findings indicate that both systems performed effectively in providing and ranking relevant results. Westlaw was the most effective providing higher results for all the precision measures, which are displayed in the table below. It is noticeable that at five and ten results, Westlaw is respectively 6% and 7% more accurate than Lexis which, if rounded to the closest whole number, represents one extra relevant document for both measures. The total EAP results were also high for both systems. The figures, expressed as percentages, show that Lexis was successful in ranking 67% of relevant results and for Westlaw the figure was 77%. In terms of documents, the findings show that Lexis ranks five out of ten relevant documents successfully, while Westlaw ranks six. Thus, the majority of relevant results, as identified by the P@5 and P@10 performance measures, have been successfully ranked by both systems. Interestingly, the performance measures follow the same pattern on both systems: P@5 providing the highest result; P@10 the second; and, EAP the third. The diagnostic measures were low or zero and were dismissed as not having any impact. The overall high precision results and subsequent correlations are conclusive in terms of the effectiveness and behaviour of both systems.

Analysis

The findings of the research established that both systems are effective in providing high-precision results.

Table 1. TREC Measures of Performance Evaluation

	P@5	P@10	EAP	Duplicates	Not retrieved
Lexis	0.78	0.74	0.67	2	0
Westlaw	0.84	0.81	0.77	2	0

In comparison with Dabney’s (1993) and Gerson’s (1999) overall precision for Westlaw and Lexis (see table below) these results share two commonalities. Firstly, that Westlaw was the most effective system in all three studies and, secondly, that both systems provided a similar result in each of the studies. Therefore, the current research confirms an established pattern, i.e. that both systems work to provide a similar level of precision and that Westlaw is slightly more successful. However, the current results differ from Dabney and Gerson’s findings, as they are significantly higher. There are a number of possible factors that could have contributed to this difference. One is that Lexis and Westlaw have improved their probabilistic processes, which are now more effective in providing relevant results. Another could be Dabney’s (1993) and Gerson’s (1999) rigorous laboratory approach, i.e. case reports are judged relevant only if they meet those on a pre-defined list, and Dabney’s use of Boolean operators in the testing, which is a more exact search. In the present study, cases were judged relevant if they satisfied a real information need, which is a less stringent criterion. Dabney’s study (1993) notes that “full-text retrieval systems are flexible enough to allow the user to control the trade-off between recall and precision” (p.109). This certainty appears to be the case in each of the evaluations listed in the table below. Dabney’s exacting approach was geared toward providing high recall at the expense of precision, whereas the present study and the STAIRS project (Blair and Maron, 1985) can be seen to have used, but indirectly, searches that provided high precision. Therefore, while the studies are working at different levels, the systems appear to share a common area of activity, regardless of the queries submitted – a plausible factor being the systems’ probabilistic ranking of documents.

In Dabney’s analysis of legal information retrieval databases (1986), he decided – with all the evidence available at that time – that litigation support databases (i.e. STAIRS) could be compared with commercial systems (i.e. Lexis and Westlaw). The findings in the present evaluation appear to confirm this theory. If we establish a total overall result for the present study by averaging the Lexis and Westlaw figures in the above table, the answer is 78%, which is almost identical to Blair and Maron’s overall precision level of 79% in the STAIRS project. It should be noted that the present study and the STAIRS project share some commonalities, i.e. they both used real needs and tested a similar-sized source: 47,779

Table 2. Figures for Precision

Study	Precision	
STAIRS	79%	
Dabney	12.4% (W)	11.5% (L)
Gerson	37% (W)	31% (L)
Present study	81% (W)	74% (L)

W=Westlaw/L=Lexis

documents in the present study, (Jansen op de Haar, 2005); and, just under 40,000 in the STAIRS project (Blair and Maron, 1985). It could be said that these commonalities contributed to the similar results reported for both studies. However, this would be speculation in the case of query types as there is no direct research on the effect of this on retrieval effectiveness. In terms of source size, research has challenged the assumption that source size has an impact on precision (Hawking and Robertson, 2001). Therefore, while parallels in the findings can be drawn between the present study and that of Dabney and Gerson, it has not been possible to make any links to the STAIRS findings.

The current research has provided another perspective on the key research on legal information retrieval that has been published to date, i.e. the aspect of using real information needs from an active legal context. Thus, in summary, the picture so far is that the STAIRS project (Blair and Maron, 1985) shows that a full-text system can provide lawyers with high-precision results at the expense of recall. Dabney's (1993) study shows poor results when recall is achieved at the expense of precision; Gerson's findings are also low, but a greater precision can be achieved via a natural language search; and, the current study also shows that high precision and ranking can be achieved for lawyers in an operational context. The above studies have different and important applications, but the present study and the STAIRS project are the only ones that can be realistically applied to an operational legal context. These two studies, when compared, have distinct advantages and disadvantages. Two lawyers generated the information needs in the STAIRS project (Blair and Maron, 1985), whereas the present study used real needs from a number of lawyers involved in a specific law firm's commercial activity. Therefore, the present study could be considered more representative of how a legal information retrieval system behaves in an operational environment. However, the STAIRS project could be considered a better representation of how useful the documents retrieved were as the study measured 'utility', i.e. how useful the retrieved documents were to a particular activity (Blair, 1996), as opposed to the current study, which only considers the topical content of the documents. If we consider both studies together, accepting that they can be generalised, then the picture is that legal information retrieval in general is working well in providing high precision in a commercial legal environment.

Burson (1987) and Gerson (1999) argue that recall is not necessary in legal information retrieval, as there are other means to facilitate this type of search and that therefore all lawyers actually need is high precision. If we subscribe to this ideology then the findings of the current study show that Lexis and Westlaw provide the service lawyers require. It is the author's view that recall as a concept is essential, but it still remains extremely difficult to calculate, which is the reason why it was excluded

from the present study. Recall is necessary because other tools for obtaining recall are sufficient (e.g. the *Case Locator* index on Westlaw) in dealing with clearly defined points of law, but not for finding less well-defined points of law, e.g. the "obiter dictum" ("chance remark") that may be valued by a judge (Williams and Taylor, 2002, p.105). It is not feasible to expect editors of such tools to index every possible facet of every case – it is usually their policy to pick out and index the central legal concepts. The concept of recall, however, is to provide all the relevant cases and, therefore, the case where a "chance remark" is made could be vital. Thus, recall should be evaluated (if possible) in order for the concept to be realised.

The present study also employed the precision-relation measure EAP for each search, i.e. a measure to indicate how well the systems have retrieved results against the estimated recall. EAP is a relatively new addition to the range of tools available to evaluate information retrieval. This addition appears to respond to the emergence of the complex ranking algorithms that now feature on databases like Lexis and Westlaw. This measure has had an important function in the present study, which is to indicate how effective the systems have been in ranking documents. The findings show that Westlaw is the most effective, but overall both systems performed to a high standard. As this measure is relatively new, no other legal information retrieval research is available for comparative purposes.

Conclusion

With the arrival of the first computer-assisted legal research products in the 1970s, law firms were faced with a choice: to invest in such a product to keep pace with developments; or, rely on traditional methods of research in the form of hardcopy materials. Several years later, with an array of computer-assisted legal research products available (now reclassified as online legal research resources), the choice is which products to choose and not whether to have them in the first instance. This choice is not easily made, as a range of criteria usually have to be met, e.g. the firm's needs; the effectiveness of the product; cost; usability; support from vendor; integration with other products, etc. The main advance in computer-assisted legal research is the availability of tools that address case finding. Without such facilities, or the benefit of colleagues' wisdom, using hardcopy materials to find relevant case law can be an arduous experience. However, just because such tools exist does not mean that they are effective in providing lawyers with relevant case reports. With this problem in mind, the present study has carried out an evaluation of the two leading online research tools Lexis and Westlaw, from the perspective of a city law firm, to see how effective they actually prove to be.

Dean Mason

Overall, the present study shows that Lexis and Westlaw are highly effective when real information needs are used from an operational legal environment but, in terms of resolving the case finding problem, Westlaw is marginally ahead. The debate on how information retrieval systems should be evaluated is one that will not go away in the near future, whereas the commercial development of such resources currently seems to be more focused on content and usability. However, if we place the current study in the context of the previous

research, it is evident that one fact remains undisputed, which is that Westlaw has been the more effective of the two systems over a significant period of time.

Acknowledgement

Many thanks to my supervisor at City University, Andy MacFarlane, who supported me throughout my research with guidance, comments and support.

Bibliography

- Blair, D. C. & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. *Communications of the ACM* 28(3) 289–299
- Blair, D. C. (1996). STAIRS redux: thoughts on the STAIRS evaluation, ten years after. *Journal of the American Society for Information Science* 47(1) 4–22
- Burson, S. (1987). A reconstruction of Thamus: comments on the evaluation of legal information retrieval systems. *79 Law Library Journal* 134
- Chowdhury, G. G. (2004). 2nd edition. *Introduction to modern information retrieval*. Facet Publishing: London
- Dabney, D. (1993). *Statistical modelling of relevance judgments for probabilistic retrieval of American Case Law*. University of California Ph.D: Berkeley
- Dabney, D. (1986). The curse of Thamus: an analysis of full-text legal document retrieval. *78 Law Library Journal* 5
- Gerson, K. (1999). Evaluating legal information retrieval systems: how do the ranked-retrieval methods of WESTLAW and LEXIS measure up? *Legal Reference Services Quarterly* 17(4) 53–67
- Hawking, D. & Robertson, S. (2003). On collection size and retrieval effectiveness. *Information Retrieval* 6 99–150
- Jansen op de Haar, B. (12 July 2005). RE: ICLR Query. Justis [Email]
- Williams, G. & Smith, A. T. H. (2002). *Learning the law*. Sweet & Maxwell: London

Biography

Dean Mason holds a masters degree in Information Science that he gained from City University in 2006. He began his career in the Library & Information Services department at Freshfields Bruckhaus Deringer, where he was awarded sponsorship by the firm to participate in the course in 2001. In December 2004 he joined the Library & Research Centre at Lewis Silkin LLP, where he is now a Research Librarian. Further information can be found on the Business Services pages at www.lewissilkin.com

Contact: dean.mason@lewissilkin.com

Dean Mason
8 Four Acres
East Malling
Kent
ME19 6AY