

developed to meet their requirements. As a firm we may consider extending the content the system searches to emails. It would also be useful to be able to search our CRM database. In the Information Services team we would be interested in exploring federated searching once again.

### Conclusion

The implementation of KnowledgeSearch has improved the firm's effectiveness at sharing knowledge. It is a system which matches our original requirements, was straightforward to implement and has proved easy to use.

### Biography

Victoria Jannetta is currently Head of Information Services at Field Fisher Waterhouse. She has worked in both legal and accountancy information services. Her previous positions have included working at KPMG and Charles Russell. She is an active member of BIALL and was Chair in 2004–2005.

*Legal Information Management*, 8 (2008), pp. 277–279

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

doi:10.1017/S1472669608000674

# Federated Searching – the Process and the Problems

**Abstract:** This paper, taken from a presentation given by Melanie Farquharson, 3Kites Consulting, to the 39<sup>th</sup> Annual BIALL Conference held in Dublin in June 2008, discusses recent developments in the provision of federated search systems and how they can ease the legal research process. It also considers the attitude of legal publishers to such systems.

**Keywords:** online information; law firms; federated search; knowledge management

### Introduction – too much information

We are now producing more information annually than it is possible to store<sup>1</sup>. There is so much information, and things change so quickly, that it is a real struggle to keep up. New sources of information crop up all the time, and new ways of mining that information become available. Knowing where to look and how to use the sources is challenging. For lawyers the challenge is particularly acute – they have to provide the right answer, not just a good answer, and clients expect quick responses.

Librarians in law firms are experiencing greater demands to provide a 24-hour service. In my practice as a lawyer I have many painful memories of what I call 'the Sunday afternoon syndrome'. Working on something that has to be on the client's desk on Monday morning, but looking out of the window and wishing I could be doing

something else, I would diligently produce the best quality advice I could. However, when a point of law needed checking, or some factual information needed to be confirmed, there was no-one around in the office to help me. I was on my own. As a junior lawyer I would have used the available information sources constantly to carry out research, and I was very familiar with them. As I gained experience, and became a less frequent user, the sources changed and became more sophisticated. Trying to find one's training notes from the last release of a legal publisher's online service on a Sunday afternoon, when all you want is a quick answer to a straightforward question, is very frustrating and it has left its scar on me.

So in my mind it is clear that an important part of the role of the knowledge and information team in a law firm is to work out how to make the right information accessible, easily, when lawyers need it, without them necessarily having to call the library for help. Federated searching has a role to play here.

## What is federated searching?

Before going any further, let us be clear what we mean by federated searching. It is the process of using a single search term to query multiple sources. The search engine will take the user's query and in effect drop it into the search box of each of the selected sources and, if necessary, translating the syntax of the query to match the required syntax of the source. It will also ensure that any authentication requirements, such as user names and passwords, are dealt with behind the scenes. The sources themselves carry out the searches, so the results are dependent on the quality and speed of the underlying search functions of the sources. Results will normally be returned grouped by source – since the search engine has no reliable way of knowing how the different sources calculate relevance, it cannot tell that perhaps the first result from source A is less relevant to the query than the first ten results from source B.

As described by [Peter Jacso \(2004\)](#), federated searching consists of (1) Transforming a query and broadcasting it to a group of disparate databases with the appropriate syntax, (2) Merging the results collected from the databases, (3) Presenting them in a succinct and unified format with minimal duplication, and (4) Providing a means, performed either automatically or by the portal user, to sort the merged result set.

In contrast, a search engine could take all the information from a set of sources, including the related metadata, and create its own index, so that the user actually searched that index, rather than the original sources (though when results were returned, the user could be linked through to the original source to view each result). This is known as indexed searching. It requires a certain amount of server space to store the indexes, and the engine needs to check each source periodically to see whether anything new has been added. Consequently, this approach is often inappropriate for significant sources of legal information, such as the legal publishers' online services. Moreover, many information providers will not allow search engines to index their content. Indeed, some are very reluctant to allow federated searching of their sites, but we will return to that issue.

The advantage of indexed searching is that the metadata associated with the underlying content can be used to help the user work with the results of a search, by refining them by document type, date, industry sector, etc. Since the metadata returned with federated search results is unlikely to be in a form that will enable mapping across different sources, the potential for manipulating the results is much less.

In law firms to date, federated searching has generally been used for external sources. It helps to solve the 'know where' problem. Whilst librarians are familiar with

the areas of strength and weakness of the different legal sources, many practising lawyers will not be. Since there are so many sources, and since each has its own interface and requires a slightly different approach to searching and navigating, there is a danger that lawyers will fail to try all the relevant sources when seeking to answer a query. They may give up after trying two or three sources, without realising that the most up-to-date source is one that they failed to look at. This is a risk issue for law firms, which pay very large sums of money to ensure that they have the most authoritative sources available, but may not actually be using them to their full potential.

## The Google generation

In our work with law firms, where we are often looking at the ways in which information and know how can be made more readily accessible to lawyers, we will often start by asking a cross-section of the firm how they would go about finding the answer to different types of query. We stress that it is important that they tell us what they really do, rather than what they think they ought to do. It is startling how often lawyers will say that they use Google as a starting point for all kinds of research, including legal research. This is because they find the Google interface very easy, and they generally find useful material within the first two or three pages of results. Google is of course an excellent tool and can be very useful in a number of contexts. However, its page ranking algorithm depends on the popularity of web pages, based on links to those pages from other pages on the web. It may not therefore necessarily be the best place to look for obscure or esoteric points. More crucially, it will not provide access to information held in subscription sources, such as the legal publishers' sites, and these may well be the most authoritative and up to date sources for legal research.

The purist may say that the answer is to provide more training for lawyers on using the different sources. However, in my view this is a case of swimming against the tide. People are now accustomed to using information sources on the web without any training at all. The BBC website, often high on the list of most visited sites in a law firm, is easy to use, well laid out, and requires no training. Similarly, Amazon, Tesco, ebay and other sites that people use regularly are sufficiently intuitive not to require training. True, it is hard to provide such ease of use when the underlying content is essentially complex. Nevertheless, we have to recognise that users' tolerance for complexity in the access to information is diminishing all the time.

## Pitfalls of federated searching

Whilst federated searching of key legal sources may appear to be an answer to these issues, it is not problem-free. The user will lose out on the sophistication of the user interface of the original source, which may provide rich

functionality to enable results to be refined and worked with. Some see this as a danger of ‘dumbing down’ the sources. This may be avoided by providing the user with a way to get to the results within the source’s own interface, once the search has been performed. At this point the user has already seen that a particular source has returned some useful results, and will be willing to spend some time working with those results to find the answer they need.

If the original source does not have a single search box into which the query can be passed, this can mean that the results returned will be poor. One way of getting around this issue is to set up the federated search engine with different options for different types of search (e.g. case law searches and legislation searches), so that the query can be passed to the appropriate part of the source. Some firms have found this very helpful, although there can be a temptation to provide too many options, thereby recreating the complexity that the tool is intended to avoid.

### Reactions of the legal publishers

It is fair to say that the trend towards federated searching of their material has not been welcomed by all the legal publishers. There are a number of reasons for this.

With new search engines coming onto the market all the time, the publishers fear that they could devote a great deal of time and effort to working with all the different suppliers of those tools and much of that time will be wasted, since the legal market will not support a large number of search products in the long term. Whilst this issue could be overcome by the publishers providing standard ways in to their material for federated searching (APIs), there remains a concern that the results will be presented in a way that loses the added value the publishers seek to provide. Their results may be mixed in with other sources and may appear less useful than they would be if viewed through the publisher’s own interface. These are legitimate concerns, and will need to be overcome by

agreeing standards as to the way in which results are presented to the user. In addition, ensuring that the user always has the ability to click through to the results in the original resource will help to address these concerns.

For some time there has been little progress on these issues and, with the exception of Solcara, search vendors have struggled to set up reliable federated search functions to the legal publishers’ services (though they have in many cases managed to federate searches to the websites of government departments and regulators, which has proved helpful for lawyers). In recent months there appears to be some movement on the part of the publishers, which is to be welcomed. Ultimately, making material more easily accessible through federated searching may bring to the publishers’ sites those users who might otherwise have relied on Google.

### The way forward

I hope that it will become easier for firms to set up federated searches covering a range of authoritative sources, since this is likely to be a real help to their lawyers.

In some jurisdictions (such as the Netherlands), third parties have set up what is in effect a shared index of legal sources. Access to this overcomes the limitations of federated searching in terms of the manipulation of the results, but without requiring each law firm to hold its own index of the content, with the associated overhead in terms of space and administration. I refer to this as a kind of ‘legal Google’ or ‘Loogle’.

Whether this will happen in the UK or Ireland remains to be seen. Perhaps, however, there will in the future be a standard way in which the metadata associated with the results of federated searches is returned, so that the user gets a better opportunity to work with the results.

In any event, those law firms that manage to crack the nut of providing easy access to the right legal sources for their lawyers, even on a Sunday afternoon, will reap the benefits in terms of efficiency and risk management.

### Footnote

<sup>1</sup>Reported e.g. by Mark Lewis of EMC Corporation.

### Biography

Melanie Farquharson has been a consultant with 3Kites Consulting, [www.3kites.com](http://www.3kites.com), since 2007. She was previously a partner at Simmons & Simmons from 1994 onwards and in 2001 she moved into a central role in the firm with responsibility for knowledge management and information services.