

Keyword searching – is there a better option?



Andrew King looks at the problems with keyword searches

In today's digital world, the exponential growth in the sources and volumes of electronic information is making the discovery process more challenging. Recent reports estimate that 70 per cent of discovery costs are incurred at the document review stage. The only way to significantly reduce the cost of discovery is to limit the number of documents for lawyers to review.

The new High Court discovery rules expect parties to come up with strategies to conduct a reasonable and proportionate search. Keyword searches can be one of the most common methods to target relevant documents, but are also one of the most problematic.

Keyword searches frequently do not produce the results expected, and, more importantly, parties are often unaware of the limitations of keyword searches. Unless a thorough thought-out strategy is devised, a keyword search approach may be little more than a blind search for documents.

Problems with keywords

A keyword search is a search carried out by using specific words/terms to isolate documents to review. In theory, what you search for is what you get, but therein lies the problem. Often what you search for and what you want from the search are two different things.

It has been estimated by industry experts that a typical keyword search approach can 'miss' approximately 80 per cent of relevant documents.

Keywords create a black or white scenario – *a document is either in or out* – a document either contains the selected word or it doesn't. Running search terms across a set of documents may reduce the volume of documents to review, but it could mean that you are missing many relevant documents because the keyword you have used has not 'hit' in a document.

Basic keyword searches can be inaccurate and inefficient as they regularly return over-inclusive

or under-inclusive results.

Over inclusive

Keyword searches often result in a high volume of documents that are irrelevant to the case. Many of the keywords may return 'false positives', where the keyword appears in a document, but it may have a different meaning to that being sought. This can result in large volumes of documents being reviewed that do not need to be, ultimately increasing the time and cost to review the documents.

Under inclusive

Keyword searches can miss many documents because a particular word was not contained within the text of the documents. Critical documents may not ever be reviewed or discovered.

Parties are often unaware of what has not been

The technology does not replace lawyers, but instead equips lawyers with the tools to carry out litigation more effectively in a world complicated by the increasing volumes and sources of electronic information.

picked up by the keywords. The danger is that certain information is carved off and is never sighted again. Most litigation evolves, and it is imperative that you have the ability to refine search parameters accordingly.

Technical issues can also restrict a keyword-search approach. Many run their keyword search terms presuming the keywords will pick up all documents that contain the keywords. Keyword searches rarely work with scanned paper as the content of all of those documents will not be 100 per cent accurate – with the ability only to recall some words at best.

As not all software searches in the same way, it is also important to understand your search tool and that used by other parties to ensure they are capable of picking up the keywords. Common issues include not all the documents being indexed (searchable), and other parties using a search tool that may not search the same way as yours.

How to make more effective keyword searches

One of the main reasons for the inaccuracies of keyword searches is the lack of understanding around how they can be deployed effectively. The problems inherent with keyword searches can be reduced by spending more time planning your search strategy and then cooperating with the other parties over the search.

The following steps may assist in making a keyword approach more effective:

1. **Construct initial keyword search terms:** Use the knowledge of individuals involved in the matter to identify any specific language, terminology, or abbreviations that may be specific to the matter. Too often, search terms are devised only by lawyers, without direct knowledge of potential words and usually devised too early before any documentation has been looked at.
2. **Test the initial search results:** Test and sample searches to find if they have produced the results you were seeking. Confirm the results are not over inclusive or under inclusive. Ensure the keywords are comprehensively searching all file types. If one party has scanned paper documents that are not 100 per cent searchable, their search terms will not capture the documents you are seeking. Confirm the tool(s) being used to search for documents are capable of running the searches you require.
3. **Refine search terms:** Keep refining search terms, until all parties can agree on the keywords that will be effective in locating the information you require.
4. **Consider if alternative search and review**

strategies may be more appropriate: Spending sufficient time devising effective keyword searches may reduce the volume of documents to review, but often these documents are then just reviewed in the same way as paper – essentially conducting a page-turn exercise, but on screen. This can still lead to a time-consuming and expensive process. There are now more effective methods to carry out the document review, using the full benefits of technology to reduce the cost and improve the accuracy of the review.

Alternatives to keyword searching

A document review which groups similar documents together is a more effective method to search and review documents. This enables lawyers to prioritise what is reviewed, decide upon the order to review, and who reviews particular documents.

Alternative search and review strategies

- › **Near Duplicate** technology identifies files that have small differences, although they are not exact duplicates. The technology groups near duplicates files together so they can be reviewed at the same time. The advantage is that it allows the reviewer to quickly focus on the small differences and move through the documents more quickly and accurately.
- › **Email Threading** technology identifies all related emails in a thread/chain. This technology can identify the email 'endpoint' and determine the unique emails in a thread. Parties can focus on the entirety of the email conversation at the same time, instead of coming across many duplicated parts of the same email conversation throughout the document review.
- › **Concept searching** finds the true context of the words by returning documents with similar meaning. It adds additional information to the very basic keywords as it evaluates both words and the context in which they appear.
- › **Predictive Coding** analyses the decisions of a human review of a sample set of documents. The technology then applies the review calls it has learnt from the sample set of documents to the remainder of documents. The documents can then be ranked or prioritised by grouping similar documents to assist the review. To improve accuracy, the initial sample set of documents is best reviewed by someone with the most knowledge about the matter.

These methods of grouping similar documents together facilitate quicker access to the most

important information. It easily isolates the irrelevant information, allowing more time to concentrate on the important documents.

Not only do these tools allow for a more cost-effective review, they enable a more consistent and accurate document review. They can identify important documents and their relevance that may be missed in a traditional document review. By grouping similar documents together, the workflow is able to be managed more effectively by allocating documents to reviewers based on their importance and complexity.

These tools search and analyse information in a way humans never could. The technology does not replace lawyers, but instead equips lawyers with the tools to carry out litigation more effectively in a world complicated by the increasing volumes and sources of electronic information.

All these methods are not new, as the Discovery Checklist now encourages parties to consider these other automated searches.

Conclusion

As the volumes of information increase, it is important to come up with an effective search and review strategy to minimise the burden and cost of discovery.

The new discovery rules require technology to be used efficiently and effectively to facilitate a more proportionate and cost-effective discovery process. To achieve this, all practitioners and litigants will be required to have a greater understanding of the benefits and limitations of any approach and the technology available. Every matter is different, so it is important to cooperate with the other parties to devise an appropriate search and review strategy.

Keyword searches are just one of many options to identify relevant documents. It is important to be fully aware of the limitations of running keyword searches and spend enough time devising a search strategy to run more effective searches.

Largely due to the ineffectiveness of keyword searching, more sophisticated alternatives of grouping similar documents together are proving more effective in getting to the important information more quickly, cost-effectively, and accurately.

If you are unsure of your search and review strategy, then seek expert guidance to assist you with your discovery process. 

Andrew King is a litigation support consultant at E-Discovery Consulting (www.e-discovery.co.nz), where he advises on strategies and tools to simplify the discovery process. Andrew can be contacted on 027 247 2011 or andrew.king@e-discovery.co.nz.