

## Tip of the Month



### **Developing and Testing Search Terms in E-Discovery**

#### **Scenario**

A global pharmaceutical company is sued in the United States. As discovery begins, in-house counsel is concerned about the cost and time associated with collecting, reviewing and producing the electronically stored information (“ESI”), which likely will result in the collection of hundreds, if not thousands, of gigabytes of data. In-house counsel assumes that search terms should be used to cull this massive amount of data to a more reasonable size, but wants to be sure that any decisions regarding the use of search terms are defensible and cost-effective.

#### **The Purpose of Search Terms**

Keyword searches are among the most common automated tools for fulfilling the obligation to conduct a diligent search for documents potentially responsive to a request for production. As technology in the e-discovery space continues its rapid development, there has been a great deal of discussion about the most effective, and most defensible, uses of keyword searches. Cooperation with an opposing party, of course, is one way to develop and structure keyword searches. In addition, there are a number of techniques or methodologies that may be useful in validating the effectiveness of keyword searches.

The key to developing and properly testing effective search terms is having a clear purpose for their use. If the purpose for the keyword search is to identify the “hot” documents for an investigation (i.e., the needle in the haystack), then the keyword search should focus narrowly on specific topics. But if the purpose is to create a cost-effective and defensible review population, then the keyword search should focus on capturing the relevant and material information from the larger volume of data collected. And if the purpose is an initial collection, it may be appropriate to use broad search terms in the first instance, and then employ narrower search terms later in the process to target documents that may be responsive to specific request’s.

In addition, designing a keyword search requires constant balancing of the materiality of issues in a matter and the risk of being under- or over-inclusive. For example, a date range or time period restriction will often be an appropriate way to exclude per se irrelevant information and avoid over-inclusion. Similarly, limiting the review population to information collected from particular company employees (“custodians” in e-discovery jargon) will also often be appropriate. But any judgment call must be made with an eye to maintaining the proportion between the cost of discovery and the information’s importance to the case. The standard for discovery is reasonableness—not perfection—and the same is true for the use of keyword searches in e-discovery.

#### **The Risks and Costs of Using Search Terms**

- Search Technology: Not all search technologies operate in the same manner or have the same functionalities. Performing effective searches depends on understanding the available technology and tailoring any keyword searches to the strengths, and weaknesses, of that technology. For example, the technology available on an organization's internal business systems may not have the same search functionality as the technology available through an e-discovery vendor. Conducting an effective search—and being able to explain the nature and sufficiency of that search if required—depends on possessing a full understanding of the search technology being used when designing and implementing the keyword search.
- Failure to Identify and Capture Relevant Documents: Relying solely on search terms may well result in certain relevant documents not being captured. No single search term set is going to be perfect. However, as long as the process of developing and testing search terms is reasonable, the use of search terms is defensible even if it is hard to show that the search terms captured every relevant document. The risk associated with missing relevant documents is higher when search terms are not negotiated with the requesting party (i.e., the requesting party does not assume the risk of missing relevant documents), when the documents not captured by the search terms may be subject to deletion or destruction (and thus not available for further search), or when the search terms are not adequately developed and tested.
- Capturing Non-Relevant Documents: Use of search terms alone will almost certainly result in the capture of non-relevant documents. If there is no attempt to identify and remove such documents, the use of search terms will incur additional costs (e.g., processing and document review costs), particularly if the search terms are broad and over-inclusive. Obviously, the larger the data set, the greater the costs associated with over-inclusiveness. In some circumstances, there also can be risks associated with over-inclusiveness (e.g., collection of non-relevant personal information protected by various data privacy laws and regulations). Appropriate development and testing of search terms should provide, at a minimum, some transparency about the trade-offs between the costs and risks of using certain search terms. Such analysis also provides some metrics that can be used in negotiations with the requesting party, or in defense of the process before a court (e.g., a 10 percent responsiveness rate in a data set resulting from a requesting party's search terms).

## **Developing Search Terms**

An organization may want to consider the following steps when developing a keyword search:

1. Develop the terms using a well thought-out process, with input from knowledgeable parties and/or a review of relevant documentation;
2. Test and refine the terms in an iterative fashion;
3. Apply the terms to a real-world subset of the documents under consideration; and
4. Where appropriate, review a sample of the population of documents that do not appear in the search results to ensure that the search is not missing important documents. If important documents do not appear in the search results, iterate, as suggested in (2) above (called "null set testing").

Keep in mind that the timing of any of these steps may depend on the purpose for the keyword search and the technology being used. For example, when using search terms for collection, it may not be practical to test the population of documents that do not hit on the search terms or to test or refine the search terms at the outset of the process. What is important is to remember that developing keyword searches is an iterative process and that search terms may need to be expanded, or contracted, over time as the legal matter develops and more information becomes

available.

### **Testing Search Terms**

After an initial set of search terms has been developed, the next step is to consider whether to test the selected terms for under- and over-inclusiveness. This testing allows the parties to create new or revised search terms (e.g., combining terms with connectors such as “and” or “within 10”) or expand the pre-existing search terms based upon key words appearing in those documents (e.g., adding wildcards). When confronted with large sets of data, consideration should be given to the use of statistical sampling in the testing process, which can provide guidance on the characteristics of the larger data set (e.g., the percentage of responsiveness in the larger data set). Whether it is appropriate to test search terms for over- or under-inclusiveness may depend, *inter alia*, on the circumstances of the case, the purpose of the search terms, the timing of the production, the volume of the data at issue and the negotiations among the parties.

In some instances, search term testing may occur during the review process (i.e., by evaluating, based on a manual review of a sample of the documents, whether the search terms are returning a high or low number of responsive documents). In other instances, it may be necessary to run the search terms against a set of documents reflecting the characteristics sought by the data set, as well as those documents not captured by the collective search term set, and to conduct a formal measurement of the results. While certain types of measurements may be used to assess the effectiveness of search terms (commonly referred to as “precision” and “recall” testing), it is commonly more important for a party to develop a robust and effective process directed at the goals of the particular case, rather than to be focused on achieving any particular metric.

### **Strategies for Developing, Testing and Using Keyword Searches**

- Effective Collaboration: Lawyers, especially at the outset of a legal matter, may not have sufficient information to identify a full set of potentially useful search terms. Clients know the subject matter, the words used to convey key concepts and nuances in the language employed. Collaboration with lawyers, clients, e-discovery experts and, if appropriate, the requesting party in developing search terms is critical to managing the risks and costs associated with a large document review.
- Understand the Search Process and Technology: Developing effective search terms takes time, effort, diligence, creativity, flexibility and, above all, sound judgment. It is important to work with individuals knowledgeable in the use of search terms in order to understand how best to test and refine the search terms and how to use the available technology to its greatest advantage. Having a well thought out, clear process for designing and testing a search that takes into account the available technology may provide an important strategic advantage if the search is challenged.
- Test and Reevaluate: Remember that the reasonableness of any set of search terms often depends on testing and evaluation of those terms. Developing keyword searches is an iterative process, and the initial set of search terms may need to be expanded, or contracted, over time as the legal matter develops and more information becomes available. Be prepared both to consider modifications to any set of keyword terms and to defend against attempts to expand or contract search terms where necessary.
- Consider Sampling: If circumstances allow for it, use sampling on large data sets to help better define the review population and test search terms. Experts and technology (usually available at e-discovery vendors) may help you develop appropriate sample sizes and random sample selections.
- Consistency in Approach: Consistency and documentation are key elements to defending

any keyword search process. Parties should record the search terms they ultimately select for use in the overall document population and their reasons for choosing them.

For inquiries related to this Tip of the Month, please contact Anthony Diana at [adiana@mayerbrown.com](mailto:adiana@mayerbrown.com), Therese Craparo at [tcraparo@mayerbrown.com](mailto:tcraparo@mayerbrown.com) or Daniel Fisher at [dfisher@mayerbrown.com](mailto:dfisher@mayerbrown.com).

Learn more about Mayer Brown's [Electronic Discovery & Records Management](#) practice or contact Anthony J. Diana at [adiana@mayerbrown.com](mailto:adiana@mayerbrown.com), Eric Evans at [eevans@mayerbrown.com](mailto:eevans@mayerbrown.com), Michael Lackey at [mlackey@mayerbrown.com](mailto:mlackey@mayerbrown.com) or Edmund Sautter at [esautter@mayerbrown.com](mailto:esautter@mayerbrown.com).

Please visit us at [www.mayerbrown.com](http://www.mayerbrown.com).

---

If you would like to be informed of legal developments and Mayer Brown events that would be of interest to you please fill out our [new subscription form](#).

Mayer Brown is a global legal services provider comprising legal practices that are separate entities (the "Mayer Brown Practices"). The Mayer Brown Practices are: Mayer Brown LLP and Mayer Brown Europe – Brussels LLP, both limited liability partnerships established in Illinois USA; Mayer Brown International LLP, a limited liability partnership incorporated in England and Wales (authorized and regulated by the Solicitors Regulation Authority and registered in England and Wales number OC 303359); Mayer Brown, a SELAS established in France; Mayer Brown JSM, a Hong Kong partnership and its associated entities in Asia; and Tauil & Chequer Advogados, a Brazilian law partnership with which Mayer Brown is associated. "Mayer Brown" and the Mayer Brown logo are the trademarks of the Mayer Brown Practices in their respective jurisdictions.

IRS CIRCULAR 230 NOTICE. Any tax advice expressed above by Mayer Brown LLP was not intended or written to be used, and cannot be used, by any taxpayer to avoid U.S. federal tax penalties. If such advice was written or used to support the promotion or marketing of the matter addressed above, then each offeree should seek advice from an independent tax advisor.

This email and any files transmitted with it are intended solely for the use of the individual or entity to whom they are addressed. If you have received this email in error please notify the system manager. If you are not the named addressee you should not disseminate, distribute or copy this e-mail.

Mayer Brown LLP, 71 S. Wacker Drive, Chicago IL, 60606, Tel: +1 312 782 0600

© 2013 The Mayer Brown Practices. All rights reserved. This publication provides information and comments on legal issues and developments of interest to our clients and friends. The foregoing is not a comprehensive treatment of the subject matter covered and is not intended to provide legal advice. Readers should seek legal advice before taking any action with respect to the matters discussed herein.

[See our privacy policy and important regulatory information.](#)