EDD Basics: The Three Parts of a Keyword Search, Part 1

Keyword searching is one of the workhorses of eDiscovery, reports **QDiscovery** in an article on its website. The article continues:

Keyword searching sees heavy use in the areas of:

Responsiveness review — Substantive keywords are used to identify likely responsive documents in the collection dataset before the documents are given to the legal team for review or produced to the other side. Alternatively, keywords may be used in conjunction with Technology Assisted Review ("TAR") to test the accuracy of responsiveness predictions or to flag key documents for manual review.

Privilege review — Although keyword searching has been losing ground to TAR as the go-to solution for responsiveness review, it remains the primary means for identifying privileged communications and work product. Privilege searches primarily target names (i.e., law firms, lawyers, paralegals, legal support staff, consultants, associated email domains). The secondary focus is on legal terms like plaintiff and defendant; click here to download a starter list of non-name privilege keywords.

Targeted issue searches — Keyword searching is useful throughout the life of the case for finding documents that relate to particular people and issues. For example, keyword searching is helpful in preparing for depositions by focusing on a deponent's documents and communications about key topics.

Use of keyword searching in general is driven by two interlocking goals. The first is to improve review efficiency by narrowing the scope of review to documents most likely to be relevant. The second goal is to improve feasibility and

cost-effectiveness of reviewing large volumes of ESI by culling non-relevant documents prior to the costly attorney review stage. While the first point applies equally to each of the three review areas, the second is most important for responsiveness review. The average volume of ESI has increased to the point that it is now impractical to make a linear (document-by-document) review of the full collection dataset even in smaller eDiscovery cases.

The three parts of a good keyword search are the keywords themselves, the "grammar" or construction of the search string, and validating the results.

1) Keywords — The three main sources for substantive keyword terms and phrases are the pleadings, the client, and opposing counsel. For illustration I'll draw on my prior litigation experience in patent infringement defense work.

a) Complaint and Answer — A starter list of keywords is found in the pleadings. Between the allegations on their face and the lawyer's experience with the legal subject matter, a list can be compiled of relevant names, fact allegations, and typical key terms. In a patent infringement case, obvious keywords are the:

Names of the parties;
Inventors' names;
Other names listed in the patent, such as assignees and patent lawyers;
Patent number and last three digits of the patent number (e.g., "the '574 patent"), the common shorthand in the field;
Distinctive words and phrases used in the patent to describe the invention;
Product names of the alleged infringing products.
b) Client — In most cases it is the client that is most knowledgeable about the subject matter of the case. As such, the client is an invaluable resource for keywords like:

Product names of competitors' products;
Project and product names used internally at the company during R&D and pre-launch marketing planning;
Industry acronyms and shorthand;
Product code numbers.
In many cases, the client will already have identified a small number of critical documents based on the allegations in the pleadings. Like the pleadings, these documents can be mined for additional keywords.

c) Opposing counsel — Whether to include opposing counsel in developing the keyword list is a strategic decision. The potential benefit is that an agreed list heads off later disputes over the search terms, as discovery fights tend to be both costly and a time-consuming distraction from the substantive issues. However, keyword negotiations may be fruitless or even create unnecessary work if opposing counsel is either unreasonable or unsophisticated about search methodologies or eDiscovery processes in general.

Assuming the proposed list compiled from the first two categories is already reasonably comprehensive, opposing counsel is most likely to add terms going to legal issues; for example, knowledge and state of mind.

Read Part 2.