EDD Basics: The Three Parts of a Keyword Search, Part 2

Keyword searching is one of the workhorses of eDiscovery. The **first part of this series** from **QDiscovery** covered responsiveness review, privilege review, targeted issue searches and more. The article picks up with the second part of keyword searches, starting with grammar.

2) Grammar — The second part of a good keyword search is constructing the search string by making the most effective use of operators, wildcards, and search parameters.

QD EDD Basics Keyword Searches Part 2

— Boolean operators — AND, OR, and AND NOT;
— Wildcards — * (multi-character expander) and ! (single character expander) at the beginning or end of a term;
— Nested searches — parentheses to "nest" or connect search terms, i.e., "(this OR this) AND that";
— Fuzziness — an instruction to the search engine that some of the characters can differ from the term as written, which is useful in finding proper names with variant spellings as well as technical terms and other hard to spell words;
— Proximity parameters — within sentence, within paragraph, or within a certain number of words.

Search strings run the gamut from simple "this term or that term" searches to complex searches that draw on the full battery of technical options. While the basic "grammar" of a keyword search will be familiar to lawyers from searching in case law databases, it's sound practice to work with an experienced project manager or eDiscovery consultant to build a complex search. An eDiscovery support professional can:

— Suggest keyword variants;
— Give advice on when and how to use operators, etc.;

— Collate multiple keyword lists to eliminate duplicates (overlap isn't always obvious with complicated terms);
— Reconcile inconsistencies within the search string.

3) Validation — Just as no battle plan survives first contact with the enemy, no keyword search (should) survive first contact with the ESI. It's important to review, analyze, and refine the keyword search in light of the results. Even more than in the second stage of building the search, a good project manager or consultant can provide invaluable assistance at the final validation stage.

a) Review a "hit count" report — The hit count report will provide the total number of documents that have hits on the search terms and also break out the number of hits per term. The total volume must be considered in light of whether it's feasible to review the search results given discovery deadlines, staffing availability, and the budget for the review. If the data volume is still inordinately high after the keyword search is run, search terms with high hit counts should be evaluated to see if it's possible to narrow them, such as by removing wildcards or adding a proximity search term.

In addition, search terms with unexpectedly low hit counts should also be scrutinized. Particular terms may need to be broadened by removing search limitations or reducing the fuzziness percentage. Overall low hit counts may indicate that the keyword list needs to be augmented.

b) Sample the search results — Analyzing the hit count report is most effective in conjunction with sampling the search results. For the terms with high hit counts, review a small sample of the results to determine if there are significant false positives. Draw the sample from a range of data sources (e.g., custodians, network folders) for a more accurate picture of the search results. Assigning unique color-coding to terms or concept groups of terms when the search is

initially run will expedite this review.

c) Refine the terms — The final step is to refine the keyword search by adding, removing, narrowing, or broadening terms based on the review and analysis of the hit count report and sampling. It may be necessary to repeat the cycle several times before finalizing the search string. Any keyword search agreement made with opposing counsel should to take this validation process into account by including discretion to modify the search string or providing for a mechanism to revisit the agreed terms.