Searches Without Borders

By Curtis Heckman, eDiscovery Associate,
Orrick, Herrington & Sutcliffe LLP

Even when working with native language documents, collection and production searches can be time-consuming and challenging. A multi-lingual environment introduces additional complexity to searching and therefore exposes the litigant to a greater risk of human and technology-based errors. This paper discusses the challenges of multi-lingual searches and provides simple recommendations for crafting defensible searches.

Computers were originally developed on an English-based system. Thus, the original assignment of alphabetic characters for computers was made in favor of Latin letters. Soon after, engineers developed computers using non-Latin alphabets. For a computer's purposes, a number is assigned to each alphabetic character. Strung together, these numbers are referred to as code pages. Communication between computers "speaking" languages based in different alphabets can become muddled beneath the surface as a result of their different code pages. Even different versions of the same software can vary in the interpretation and translation of non-Latin characters. Keyword searches that work in the language environment in which they were created may, despite looking identical on screen, miss documents created using a different alphabet. Accordingly, a comprehensive and accurate search may require that a litigant consider the original language environment of potentially-relevant documents.

Each custodian's software and hardware, and even the alphabet used to type a keyword, significantly impacts the accuracy of the search. If a litigant knows or anticipates that its universe of potentially-relevant documents contains information created in more than one alphabetic system, that party should consider using keyword variations in its search protocol.

The best way to ensure a defensible search is to have a quality assurance system in place to test the results of the search and collection methods at the outset.

- The Sedona Conference recommends parties evaluate the outcome of each search, using "key metrics, such as the number of included and excluded documents by keyword or filtering criteria, can be used to evaluate the outcome."[1]

- Ask questions during the initial custodian interviews to identify the language(s) used for both formal and informal communications. Also identify alternative alphabetic characters on the custodian's keyboard, and what, if any, type of software the custodian used to type in the computers non-primary language.

- Question the Information Technology department and any discovery vendor regarding the ability of proposed search engines to account for different character sets and code pages.

---

[1] Jason R. Baron et al., *The Sedona Conference: Commentary on Achieving Quality in the E-Discovery Process* (May 2009) at 15.

- If the production or review is to involve translations, determine whether the translation is to be a machine-based translation and whether the translator has the capability to differentiate mingling characters.

- Document the entire process underlying the determination and deployment of key words.

A receiving party should also be fully prepared to discuss the producing party's search obligations if multiple alphabetic systems are anticipated:

- Consider meeting with a consultant or expert who understands multi-language searches and productions prior to the initial meet and confer.

- Negotiate the parameters of the search at the meet and confer. Clearly identify and convey your production expectations.

- If no agreement is reached and the litigation moves to motions practice, have an expert provide technical affidavits regarding searching and production in multi-system environments.

Litigation imposes significant difficulties for international corporations, not the least of which is multi-lingual searches. Knowing the challenges before tackling cross-border searching and documenting each step establishes reasonableness and defensibility.

Submission by: Curtis Heckman of Orrick, Herrington & Sutcliffe, LLP.

Mr. Heckman is a member of Orrick's eDiscovery working group, which is Orrick's practice group devoted to eDiscovery. He is resident at Orrick's Global Operations Center ("GOC") in Wheeling, WV and can be reached at (304) 231-2645 or checkman@orrick.com. The GOC is Orrick's on-shore outsourcing center and home of Orrick's Document Review Services and Data Management Group.

Full biographies and description of the practice group are available at http://www.orrick.com/practices/ediscovery.