

## BUSINESS MATTERS

# Misspellings can create...

# 'e-diskcovery' issues

By Correy E. Stephenson  
Staff writer

**Y**ou might have thought that electronic discovery always requires accuracy. But the latest advice from e-discovery experts is that misspelled words and names are just as essential to the process as accurate ones.

That's because misspellings and abbreviations are common in both texts and e-mails, where people tend to be more casual with their language and keep their messages short.

"This is a huge issue," said Michael R. Arkfeld, an attorney at Arkfeld & Associates in Phoenix and the author of the treatise, "Electronic Discovery and Evidence." "E-mail accounts for almost 60 percent of all [electronically stored information], with text messaging as another huge source."

In a recent order, U.S. District Court Judge Blanche M. Manning of the Northern District of Illinois ordered the defendant in *Northington v. H&M International, Inc.* to "include misspellings of [the] plaintiff's first name as well as other key search terms reasonably related to each of the topics set forth" in the production request.

Josh Gilliland, an e-discovery practitioner in San Jose, Calif. and author of the *Bow Tie Law* blog on e-discovery, said it was the first time he had seen a court order parties to include misspellings.

"The order reminds us that there is no easy way to go through the thousands or sometimes millions of documents to cull through and find what is truly responsive to a search in a particular case," said Andrew Cosgrove, a partner at Redgrave LLP in Minneapolis.

Cosgrove, whose information law practice focuses on e-discovery, information management, privacy and data protection, said the order also shows "the danger of using too strict a keyword search approach, where one missing or transposed letter in a word in a key e-mail may [mean it is] excluded from a search."

Due to these problems, some practitioners are now turning to alternative ESI search methodologies, such as conceptual searching, he said, where the search is less tied to specific words.

### How many ways to spell a name?

While the plaintiff in the *Northington* case — a Title VII sex and race discrimination suit — has an unusual first name (Ehnae), lawyers should include alternate spellings even for common names.

For example, with a party named William, the lawyers should include alternatives like Bill, Billy, Will and Willy, Gilliland said, adding that he knows a family friend who spells Bill with one "l." And in some cases, people go by a completely different name, a middle name or a nickname.

In addition to individual names, names of corporations or products, or in a pharmaceutical case the name of a drug, might also be terms for which to consider misspellings.

Social networking sites, including Twitter accounts, will also be a source of abbreviations and misspellings, Gilliland noted.

"Anything where people are typing on a smartphone in a moving vehicle increases the chance of error," he said.

Documents that involve more formal work will typically have correct spelling and many programs — like Microsoft Word, for example — have validation protocols that will highlight words spelled incorrectly or unusual spellings, Arkfeld noted.



Lawyers should address the issue early in the e-discovery process and discuss it with opposing counsel at the Rule 26(f) meet and confer, Gilliland suggested.

Arkfeld said a lawyer's approach may depend upon whether he or she is the requesting or producing party.

"A requesting party wants the other side to search for any kind of iteration of words and put in any and all spellings," he said. "For a producing party, it's a double-edged sword."

While a longer list of search terms will increase the defense's obligation to preserve all relevant ESI, not including misspellings could skew the search results, Arkfeld said.

"The best thing [the defense] can do is sit down and agree with the other side on what search terms to use," he explained.



Attorney Andrew Cosgrove

That way, if the agreed-upon search terms are used and don't result in a smoking gun e-mail or document, the defense can point to the parties' agreement and avoid sanctions or a second collection.

"Especially if you are the producing party, get a search protocol in place to protect you," Arkfeld said.

### A new type of searching

The misspellings issue aside, Arkfeld noted that many federal court judges are unhappy with keyword searching itself for failing to bring back all the data expected in a search.

In fact, U.S. Magistrate Judge John M. Faciola in the District of Columbia, a well-respected jurist in the world of e-discovery, has said that in some cases, the use of as certain search methodology requires an expert opinion.

In a 2008 case, he wrote that "[w]hether search terms will yield the information sought is a complicated question involving the sciences of computer technology, statistics and linguistics. ... For lawyers and judges to dare opine that a certain search term would be more likely to produce information is truly to go where angels fear to tread. This topic is clearly beyond the ken of a layman," he wrote. (*U.S. v. O'Keefe*, 537 F. Supp. 2d 14 (D.D.C. 2008).)

Ken Withers, director of Judicial Education and Content at The Sedona Conference, a non-profit organization in Phoenix that works to advance law and policy in areas like electronic discovery, noted that the issue of misspelling words or alternative spellings can be addressed with different types of searching.

"Computer scientists and others have developed methods of training computers to be much smarter than using brute force word searching," he said. "Sophisticated

search and information retrieval now involves criteria that go way beyond character strings that appear in a text database, and now include relationships between various communicants, having the computer identify whole concepts, not just words, and involving the proximity of words and concepts to each other."

For example, in a case about gemstones and searching for the word diamond, the computer could exclude discussions of baseball or wedding planning. Using mathematical probability, the computer could also look for things that don't exist in the database — such as at what point in time people stop talking about an issue, which could be very important in a case, Withers said.

The new methodologies allow litigants to maximize two variables: recall and precision, he explained. While the use of alternate spellings and misspellings may expand recall, it will decrease precision, he said, which is why courts and litigants are increasingly turning to new types of searching.

For the best results, Withers said attorneys should take a small sample of the ESI collection and run a search with a few keywords or data concepts in collaboration with the opposing party.

Then, both sides should examine the results and determine how successful the first run was, he said, before refining the search parameters to increase accuracy.

"The parties might do this two or three times before they press the button and do the final search of the complete ESI collection. They will have a much narrower, much better-defined search that will maximize precision and recall," Withers said.

Questions or comments can be directed to the writer at: [correy.stephenson@lawyersusaonline.com](mailto:correy.stephenson@lawyersusaonline.com)