



CBA Cincinnati Bar  
ASSOCIATION

# Report

August 2009

## Staring into the Abyss

Of Electronic Discovery

# The Four Hard Drives of the Apocalypse Staring into the Abyss



By Richard L. Moore

**E**lectronic discovery is a threat to our very way of life. The volume of electronically stored information (“ESI”) that needs to be reviewed in even a moderately sized case is making litigation cost prohibitive. As a result, litigants are being denied access to the courts.

These are not the ramblings of a madman crying in the electronic wilderness, but are the findings of no less than the American College of Trial Lawyers.<sup>1</sup> The College’s final report concludes that in many jurisdictions, some deserving cases are not brought because litigation is cost prohibitive.<sup>2</sup> In addition, meritless and smaller cases are settled rather than tried

storage. For example, in 1990 a gigabyte of storage, approximately 80,000 pages, cost about \$20,000. Today, it costs less than \$1.<sup>5</sup> As a result, companies are less motivated to assign limited and expensive human resources to the task of purging their systems of unnecessary ESI. Instead, companies are likely to opt to keep the ESI and simply add more inexpensive storage capacity. This certainly makes economic sense — that is until litigation arises. Then, that gigabyte of data that cost \$1 to store may cost over \$30,000 to review.<sup>6</sup>

It is unlikely that the volume of ESI subject to discovery will decrease.

the amount of ESI that requires human review. Second, counsel and the courts must approach discovery differently. Instead of continuing to rely on an adversarial model, litigants must take a much more cooperative approach.

## Searching for Salvation

*U.S. v. Philip Morris* was a RICO lawsuit brought by the U.S. government in 1999 against multiple tobacco companies.<sup>7</sup> The defendants filed 1,726 requests for production against 30 government agencies, which included requests for relevant e-mails.<sup>8</sup> The government was able to winnow down the 20 million Clinton-era White House e-mails that had been preserved to about 1% of the total, or 200,000 relevant e-mails, by unilaterally applying keyword searches.<sup>9</sup> These remaining e-mails then required a manual review to determine relevance. To complete this task, the government had to put in place a team of 25 lawyers, law clerks, and archivists, working full time for a period of six months.<sup>10</sup> This “boots on the ground” approach, however, becomes impractical if the universe of e-mails requiring human review becomes larger.

*Philip Morris* involved 20 million e-mails generated during the Clinton Administration. The outgoing Bush II administration generated 200 million e-mails. Estimates are that if the Obama administration runs for two terms, it will generate over 1 billion e-mails.<sup>11</sup> To review 1 billion e-mails it would take 100 people, working 10 hours a day, seven days a week, 52 weeks a year, over 54 years to complete at a cost of over \$2 billion.<sup>12</sup>

## Parties must increasingly rely on electronic search and retrieval technology to help separate responsive ESI from the gigabytes of irrelevant electronic chaff.

because litigation is too expensive.<sup>3</sup> The Final Report cautions that “it is up to counsel and the judiciary to ensure that e-discovery does not place the courtroom out of the reach of parties seeking a fair adjudication of their disputes.”<sup>4</sup>

### The Rise of the Machines

One of the principal causes of the rapid escalation of discovery costs is the expedient growth in the amount of ESI generated by normal business operations. Adding to this problem is the tendency of companies to hold on to ESI longer than they have to, which is facilitated by the rapid reduction in the cost of electronic

storage. Instead, with the continued use of e-mail and the increased use of new technologies like blogs, Twitter, and texting, the volume of ESI will grow even more rapidly. As a result, the cost of litigation will also continue to rise. Without a change in the way attorneys and the courts approach discovery, there is a significant risk that the cost of discovery alone will put the courtroom out of reach of all but the most well funded litigants.

There are two important changes that need to take place to move us away from the abyss. First, attorneys and the courts must take advantage of search and retrieval technology to help manage

As an alternative to throwing the population of Iowa at a document review, parties must increasingly rely on electronic search and retrieval technology to help separate responsive ESI from the gigabytes of irrelevant electronic chaff. The use of search technology can significantly reduce the cost of discovery by reducing the amount of ESI that requires expensive human review. To intelligently take advantage of available search technologies, however, counsel must first be aware of the different search technologies available and the strengths and weaknesses of each.

## Recent studies have shown that these new search technologies may not outperform keyword searches.

### Keywords Can Be Fuzzy

Keyword searches are the most familiar and widely used search methodology. Most attorneys practicing today are comfortable using such searches to do online research or to troll the Internet. Boolean operators are a common search tool used to expand on simple keyword searches. These familiar searches utilize operators like “and,” “or,” “and not,” and “but not,” along with keywords to attempt to better target relevant documents.

Keyword searches work best in situations where the use of language in documents is relatively predictable and consistent — legal research, for instance. On the other hand, keyword searches fall short in many real world situations. For example, keyword searches have trouble with synonymy — multiple words meaning the same thing (trip, vacation, getaway). Keyword searches also get tripped up when dealing with words that have multiple meanings — polysemy. Another challenge for keyword searches are situations involving the use of code words, abbreviations, or new words — texting for example — OMG!

Another shortcoming of keyword searches is their inability to pick up on misspellings, alternate spellings, or different derivations of words. Fuzzy search technology attempts to address these shortcomings by locating relevant documents even when the keywords are not an exact match. Search engines that

allow for fuzzy searches allow the query to be adjusted to pick up keywords that may not be an exact match but have a designated number of matching letters.

The biggest problem with keyword searching is that it is completely dependent on the person formulating the search to be able to identify the correct keywords. Unfortunately, attorneys significantly overestimate their capability to do this. The most cited study of the effectiveness of humans in locating responsive documents using keywords is the Blair and Maron Study completed in 1985.<sup>13</sup> That study found that humans

are less than 25% accurate and complete in searching for and retrieving information from a set of documents. Although, the reviewers were convinced that they were at least 75% effective.<sup>14</sup> The principle problem faced by the reviewers was the innumerable ways that authors of the documents could come up with to describe the particular accident, calamity, deed, episode, incident, mishap, occasion, or situation.<sup>15</sup>

Despite its shortcomings, keyword searches remain the default option for most attorneys. A number of recent court opinions, however, have cited the limitations of keyword searches and chastised attorneys who used keyword searches without having a firm grasp on their limitations.<sup>16</sup> One recent case went so far as to say that keyword searches are now disfavored.<sup>17</sup>

### Searching Beyond Keywords

More sophisticated search tools have been, and are being developed to improve on keyword searching. One of these sophisticated search technologies are Bayesian classifiers. This process identifies relevant documents by formulating concepts using a starter set of representative relevant documents.<sup>18</sup> The classifier then uses the information gleaned from the smaller set to identify other responsive documents in the larger group of potentially responsive documents. Similarly, clustering technology

uses statistics to recognize what category certain information belongs to and attempts to group together documents with similar content. Clustering systems do not require human intervention to organize documents and, therefore, can serve as an economical and effective first pass at organizing ESI.<sup>19</sup>

One of the most recent technological innovations is concept searching. Concept searching attempts to locate relevant documents without relying on keywords.<sup>20</sup> Instead, as the name implies, concept search technologies attempt to find documents that relate to certain concepts. For example, using concept search technology to identify documents related to Alaska might yield documents dealing with snow machines, polar bears, or pipelines.<sup>21</sup>

### Are We Better Off?

Despite some of the advances in search and retrieval technology that have taken place over the last two decades, recent studies have shown that these new search technologies may not outperform keyword searches. In 2006, the Text Retrieval Conference (“TREC”) designed an independent search project to compare the effectiveness of a variety of search technologies. The test compared Boolean keyword searches with 31 different automatic search methodologies. The results determined that Boolean searches located 57% of the known relevant documents out of the test set of 7 million documents. None of the alternative search methodologies reliably performed any better.<sup>22</sup>

Importantly, the Boolean searches were formulated by different groups of TREC coordinators — some serving in the role of the requesting party and others as the responding party. These groups worked together to determine the appropriate searches to run.<sup>23</sup> This cooperative approach appeared to outperform more recent technological advances and represents the path forward in handling e-discovery.

### Man vs. Machine

The outcome of the TREC study supports the conclusion that the use of search and retrieval technology is most



effective when the search methodology is used in a cooperative and iterative manner. Instead of one party unilaterally attempting to identify the most appropriate search to run, both parties should discuss and agree on how best to construct the search. In addition, the parties should continue to meet and confer during the course of discovery to review the results, and adjust course as appropriate.<sup>24</sup>


Recently, Magistrate Andrew J. Peck of the United States District Court in the Southern District of New York sent out a wake up call to lawyers in his district with respect to cooperating with e-discovery generally and in developing search strategies in particular.<sup>25</sup> Because the parties could not come to an agreement on which tools to use or even which keywords should be applied, the court had to intervene. The court was not amused. Magistrate Judge Peck cautioned that:

*Electronic discovery requires cooperation between opposing counsel and transparency in all aspects of preservation and production of ESI. Moreover, where counsel are using keyword searches for retrieval of ESI, they at a minimum, must carefully craft the appropriate keyword, with input from the ESI's custodians as to the words and abbreviations they use, and the proposed methodology must be quality control tested to assure accuracy in retrieval and*

*elimination of false positive. It is time that the Bar – even those lawyers who did not come of age in the computer era – understand this.*<sup>26</sup>

In this same vein, the Sedona Conference has recently issued its Cooperation Proclamation.<sup>27</sup> The Proclamation has already been endorsed by a number of courts across the country, including several in Ohio.<sup>28</sup> The Proclamation cautions that over-contentious discovery in the era of ESI is a cost that has outstripped any real or perceived advantage, is not in anyone's interest, and is a waste of resources.<sup>29</sup> The Proclamation declares that lawyers have twin duties of loyalty — they are obligated to be zealous advocates for their client, but also bear a professional obligation to conduct discovery in a diligent and candid manner.<sup>30</sup> The Proclamation argues that cooperation does not conflict with the advancement of a client's interest — it enhances it.<sup>31</sup>

## Conclusion

Automated search tools will become increasingly necessary to enable litigants to manage the huge volumes of ESI subject to discovery. However, no automated search tool, or combination of tools, can guarantee that discovery will be completed in a cost efficient manner. The solution to the e-discovery problem lies in the lawyers' ability to work cooperatively in applying these technological solutions. Only by working cooperatively can counsel secure for their clients "the just, speedy, and inexpensive determination of every action and proceeding."<sup>32</sup> 

*Richard L. Moore is a partner in the Cincinnati office of Vorys, Sater, Seymour and Pease LLP, and a member of the litigation group and the firm's e-discovery core team. He has a general litigation practice with specific expertise in the areas of employment discrimination, workers' compensation, construction, medical malpractice, and personal injury.*

- 1 Final Report on the Joint Project of The American College of Trial Lawyers. Task force on Discovery and the Institute for the Advancement of the American Legal System, March 11, 2009 (Rev. March 20, 2009) ("Final Report").
- 2 Id. at 2.
- 3 Id.
- 4 Id. at 15 (quoting B. Sells and T.J. Adhietty, *E-Discovery, you can't always get what you want*, International Litigation News, Sept. 2008, pp. 35-36.).
- 5 The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery ("Best Practices"), Vol. VIII, Fall 2007, 189, 192.
- 6 See Id. at 192.
- 7 United States v. Philip Morris USA, 449 F. Supp. 2d. 1 (D.D.C. 2006).
- 8 George L. Paul and Jason R. Baron, *Information Inflation: Can the Legal System Adapt*, Richmond Journal of Law and Technology, Vol. XIII, Issue 3, p. 11 (2007).
- 9 Jason Krause, *In Search of the Perfect Search*, ABA Journal, April 2009.
- 10 Information Inflation at 12.
- 11 Ralph Losey, *Jason Baron on Search – How Do You Find Anything When You Have a Billion E-mails* ("Baron on Search"), <http://ralphlosey.wordpress.com/2009/03/04/jason-baron-on-search-how-do-you-find-anything-when-you-have-a-billion-emails/>.
- 12 Information Inflation at 13.
- 13 Best Practices at 199.
- 14 Id. at 206.
- 15 See Id.
- 16 See *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 260, 262 (D.Md. May 29, 2008) (Grimm, M.J.); *United States v. O'Keefe*, 537 F. Supp. 2d 14, 24 (D.D.C. 2008) (Facciola, M.J.).
- 17 *Asarco, Inc. v. W.S. E.P.A.*, 2009 WL 1138830, \*2 (D.D.C.) (April 28, 2009).
- 18 Best Practices at 218.
- 19 Id. at 219.
- 20 Id. at 202.
- 21 Id.
- 22 H. Christopher Boehning, and Daniel J. Toal, *Assessing Alternative Search Methodologies*, New York Law Journal April 22, 2008, at Col. 1).
- 23 Id. at 202.
- 24 Information Inflation, at 32-36.
- 25 *William A. Gross Construction Assoc., Inc. v. American Mfgs. Mutual Ins. Co.*, 256 F.R.D. 134 (S.D.N.Y.) (2009).
- 26 Id. at 136.
- 27 The Sedona Conference Cooperation Proclamation, at 5 (available at [www.TheSedonaConference.org](http://www.TheSedonaConference.org)).
- 28 Id.
- 29 Id. at 1.
- 30 Id.
- 31 Id.
- 32 See Fed. Civ. R. P. I.