



May 20th, 11:00 AM

## Electronic Discovery: A Fool's Errand Where Angels Fear to Tread?


Milton Luoma

Assistant Professor, Metropolitan State University, milt.luoma@metrostate.edu

Vicki Luoma

Associate Professor, Minnesota State University, vicki.luoma@mnsu.edu

Follow this and additional works at: <https://commons.erau.edu/adfsl>

 Part of the [Computer Engineering Commons](#), [Computer Law Commons](#), [Electrical and Computer Engineering Commons](#), [Forensic Science and Technology Commons](#), and the [Information Security Commons](#)

---

### Scholarly Commons Citation

Luoma, Milton and Luoma, Vicki, "Electronic Discovery: A Fool's Errand Where Angels Fear to Tread?" (2010). *Annual ADFSL Conference on Digital Forensics, Security and Law*. 9.  
<https://commons.erau.edu/adfsl/2010/thursday/9>

This Peer Reviewed Paper is brought to you for free and open access by the Conferences at Scholarly Commons. It has been accepted for inclusion in Annual ADFSL Conference on Digital Forensics, Security and Law by an authorized administrator of Scholarly Commons. For more information, please contact [commons@erau.edu](mailto:commons@erau.edu).

**EMBRY-RIDDLE**  
Aeronautical University™  
SCHOLARLY COMMONS

(c)ADFSL



## **Electronic Discovery: A Fool's Errand Where Angels Fear to Tread?**

**Milton Luoma**

Assistant Professor  
Metropolitan State University  
700 East 7th Street  
St. Paul, MN 55106  
651 793-1481  
651 793-1246 fax  
Milt.Luoma@metrostate.edu

**Vicki Luoma**

Associate Professor  
Minnesota State University  
145 Morris Hall  
Mankato, MN 56001  
507 389-1916  
507 389-6497  
Vicki.Luoma@mnsu.edu

### **ABSTRACT**

Electronic discovery has transformed the discovery phase of civil litigation in recent years. The expectations of lawyers and parties were initially established in the Rowe and Zubulake cases that led to a complete revision of the electronic discovery rules contained in the Federal Rules of Civil Procedure. Subsequent cases have underscored the importance of document search methodologies and implications for attorneys, IT professionals, and digital forensics professionals. The authors review how electronic discovery has evolved thus far and offer recommendations regarding the electronic discovery process.

**Keywords:** Electronic discovery, e-discovery, keyword search, concept search,

### **INTRODUCTION**

Technology has changed the legal profession forever and civil litigation will never be the same for lawyers. Information has never been more available but at the same time so overwhelming in volume and less comprehensible to the typical attorney. By 2011 there will 1,800 exabytes of electronic data in existence, or 1.8 zettabytes. Data are growing by a factor of 10 every five years. (Mearian, 2008) In addition, electronic containers for data such as files, images and tags are presently growing 50% faster than the data are growing. (Mearian, 2008) All of the books, tapes, and other documentation contained in the Library of Congress equal only about 10 terabytes (10,000 gigabytes) of data. (Mearian L. , 2007) Today, over 94% of all information is created and stored digitally. Regardless of its source, 85% of all produced electronic data eventually passes through a corporate computer, Web site, network or asset, and then that corporation has responsibility for that information that may have to be searched in the discovery process in a lawsuit. (Mearian L. , 2007)

Further, information is rarely deleted because storage is easy and inexpensive. The devices needed to store information keep getting physically smaller but larger in capacity. For example, a thumb drive is expected to be available in 2011 with a capacity of up to terabyte of information. (Wired, 2007) This volume of stored data creates an overwhelming dilemma when litigation is commenced. A litigator

must determine what information should be sought and how to ask for the information all the while trying to determine how to find all of the information the other litigator requests.

Before the digital age attorneys searched through available hard copy documents to retrieve those documents the opposing party requested. If the opposing attorney failed to request the right documents or if they were no longer available the matter was closed. In 2007 new Federal Rules of Civil Procedure were passed to try to clarify court rules concerning digital information. One of the biggest changes is that litigants and their attorneys are required to be cooperative and to become transparent in the discovery process. This required transparency is contrary to the adversarial instinct of every trial attorney and the previous century of litigation practice. In addition, most attorneys have little training, understanding or experience in ESI. Even more alarming is that most attorneys do not know how much they do not know. Thus, discovery has become a very complex process with numerous requirements and many concern parties. Many litigants, software companies, and consultants are attempting to simplify the process at a reasonable cost but full compliance can make the process extremely expensive and time consuming.

### **THE NEW FEDERAL RULES OF CIVIL PROCEDURE**

Procedural law is different from substantive law. Substantive law governs rights and obligations while procedural law specifies the methods, procedures, and practices required to be followed in civil cases, including the disclosure of electronically stored information held by the parties. Beginning with the *Rowe* (in *Rowe Entertainment Inc. v. William Morris Agency, Inc.*, 2002) and *Zubulake* (*Zubulake v. UBS Warburg LLC*, 2003) cases, the Judges Francis and Scheindlin first recognized that the court rules on discovery needed to be reinterpreted in the age of electronically stored information. Both of these cases led to the first real guidelines to litigants and their attorneys as to what would be expected of them and served as the catalyst to change.

On April 12, 2006, the U.S. Supreme Court approved amendments to the Federal Rules of Civil Procedure (FRCP) concerning discovery of electronically stored information. The amendments took effect on December 1, 2006. These revised Federal Rules put litigants and their attorneys on notice that they need to give early attention to electronic data, how and where it is stored, saved and retrieved. Rule 16 (Federal Rules of Civil Procedures, 2006) Rules 26a and f, (Federal Rules of Civil Procedure 26, 2006) require the court and the parties to give early attention to any issues concerning electronically stored information including preservation of evidence.

One of the new rules, Rule 26 (f) (Federal rules of civil procedure, 2006) requires the parties to come prepared to meet and confer to exchange information concerning the type of digital information they have, how it is stored, and how to locate the data. The major problem in this process is that attorneys are not qualified or prepared to conduct such a digital investigation.

The new rules also require that the court begin monitoring the process at the beginning of the discovery phase of litigation. The court must monitor the litigants and require them to confer as soon as possible before the first scheduling conference to discuss preservation of discoverable information and any issues relating to their discovery. (Federal rules of civil procedure, 2006)

The rules further provide that when the electronic information is not reasonably accessible, a party may not have to provide it, however, there are very few cases in which litigants have been excused from providing information, and it is much more likely that the court will sanction the non-providing party. In addition, the rules require that at the beginning of a case, each party must disclose by description, category and location of all documents and things in the disclosing party's possession, custody or control that it may use to support its claims or defenses. (Federal Rules of Civil Procedure 26, 2006).

In addition, the party requesting discovery “may specify the form or forms in which electronically stored information is to be produced.” (Federal rules of civil procedure, 2006) The producing party may then object to the requested form(s) of the information and, in such case, must “state the form or

forms it intends to use.” (Federal rules of civil procedure, 2006).

The rules further try to clarify discovery by providing that the party must produce the electronically stored information “in a form or forms in which it is ordinarily maintained or in a form or forms that are reasonably usable.” (Federal Rules of Civil Procedures, 2006) In the time since the rules were enacted the courts have consistently found there has been enough notice to potential litigants that their data must be stored and retrievable or the litigant faces sanctions.

In one such case, the court expressed displeasure with data downgrading stating, "taking an electronic document...printing it, cutting it up, and telling one's opponent to paste it back together again, when the electronic document can be produced with a keystroke, is madness. (Covad Commc'ns. Co. v. Revonet, Inc., 2009)

### **THE PRACTICAL DIFFICULTY FINDING RELEVANT INFORMATION**

The problem for litigants and their attorneys is how to find the data requested. The most common methodology has been key word searches. A keyword search is a method to locate a document by looking for certain words or terms that would be found in a record. It is a method most web browser users are comfortable with and feel competent in conducting. Some of the problems with this methodology are determining what words to search and who conducts the search.

In a recent case (Asarco, Inc. v. United States Env'tl. Prot. Agency, 2009) the plaintiff argued that the defendant's keyword search was inadequate and the court agreed, ordering an additional keyword search utilizing four additional key word terms. The court also noted that "keyword searches are no longer the favored methodology." (Asarco, Inc. v. United States Env'tl. Prot. Agency, 2009).

There have been several studies conducted on this methodology to determine whether keyword searches are an adequate methodology. One of the earliest studies conducted in 1985 by David Blair and M.E. Maron found that attorneys were only 20% effective at imagining all of the different ways that document authors could refer to words, ideas, or issues in their case. (The Sedona Conference Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery, 2007) In this study, attorneys and paralegals who had special skills in a unique area of tort law (subway accidents) were asked to identify appropriate keywords to search through 40,000 documents and 350,000 pages in a case involving a subway accident. The participants were told to stop when they thought had found at least 75% of all the relevant documents. When the participants stopped they were certain that they had found at least 75% of the relevant documents but in actuality they had located only about 20% of the relevant documents. (Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery, 2007)

This study found that even experts searching other peoples' works are very poor at identifying the right key words to use in a search. (The Sedona Conference Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery, 2007) This type of study has been repeated several times and continues to be repeated. The Sedona Conference had participants review over one hundred hypothetical complaints to find relevant documents with much the same results. Further, the TREC Legal Track conducted a study in 2008 and in 2009. TREC Legal Track is a conference that researches information retrieval issues. The 2008 study used over 7 million documents available on the tobacco litigation and the 2009 used the Enron data publically available. The searchers in both of these studies were comprised of information scientists and litigators. (Oard, 2008).

There are several problems with key word searches that include determining all of the possible synonyms for a word. For example, there are approximately 120 words for the word “think.” (Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery, 2007) Another example is that there are 25 words for “early,” 15 words for “expedite,” 16 for “appointment,” and 12 for “interview.” (Rogets, 1911, amended 1991). The results of the 2008 TREC “consensus” keyword searches found that, on average, Boolean keyword searches found only 24% of the total number of responsive documents in the target data set. (Oard, 2008)

Other problems include ambiguity, acronyms, abbreviations, misspellings and typographical errors. Another keyword search methodology is Boolean searches which allow the searcher to combine words and phrases using logical operators such as the words AND, OR, NOT and to limit or increase the search. These studies along with civil cases have highlighted the problem in determining which keywords should be used. The new rules require litigants to have early meet-and-confer sessions and if the parties are going to rely on keyword searches for their ESI discovery, then they should attend with keywords in mind and with an idea of what keywords would find the documents of interest. As consultant Craig Ball states, the meet-and-confer sessions as required by Rule 26(f) consist of “two lawyers who don’t trust each other negotiating matters neither understands.” (Ball, 2006)

Judge Andrew Peck stated, “This case is just the latest example of lawyers designing keyword searches in the dark, by the seat of their pants, without adequate (indeed, here, apparently without any) discussion with those who wrote the emails.” He continued, “Moreover, where counsel are using keyword searches for retrieval of ESI, at a minimum they must carefully craft the appropriate keywords, with input from the ESI custodians as to the words and abbreviations they use. The proposed methodology must be quality-control tested to assure accuracy in retrieval and elimination of false positives. It is time that the bar – even those lawyers who did not come of age in the computer era --understand this.” (William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co. , 2009) Judge Andrew Peck further stated, “The case should be a wakeup call for them to understand the need for careful thought, quality, control, testing, and cooperation with opposing counsel in designing keywords to used to produce emails or other electronically stored information (ESI)” (William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co. , 2009)

This case points out two problems indicating that keyword searches are not the best methodology and attorneys are not prepared for ESI searches. Again, in *Equity Analytics v. Lundin* the court set the protocol for searching his former employee’s computer. The plaintiff Equity Analytics claimed that defendant, its former employee, gained illegal access to electronically stored information after he was fired. Defendant claimed that another employee gave him permission to use his username and password. The defendant by his own admission accessed his employer’s computer system at least 18 times in a 90 day period. When the defendant provided discovery, the information was obtained through keyword searching. The employer argued against the defendant’s use of these terms as inadequate because defendant had uploaded a new operating system (“Leopard”) on his home computer in 2007. Equity further argued that the files could have been compromised that were on his previous computer and further converted easily from one format to another to hide information. Equity also argued if they were not given the right to make mirror image of the hard drives that there were likely fragments of information that keyword search would not find. (*Equity Analytics , LLC v Lundin, 2008*). Judge Facciola stated that “determining whether a particular search methodology, such as keywords, will or will not be effective certainly requires knowledge beyond the ken of a lay person (and lay lawyer) and requires expert testimony that meets the requirements of Rule 702 of the Federal Rules of Evidence.” (*Equity Analytics , LLC v Lundin, 2008*). One of the most outspoken Judges on Electronic searches, Judge Facciola found in another case that the issue of appropriate searching electronic data was too complicated for lawyers and judges to address without the aid of expert testimony. The judge stated that keyword search analysis is an area of e-discovery “where angels fear to tread.” (*United States v O’Keefe, 2008*)

In yet another case Judge Facciola, recommended "concept searching," -- the use of complex search engines that make use of linguistic or statistical patterning to locate responsive e-mails and electronic documents, (*Disability rights Council v Washington Metropolitan Transit Authority, 2007*) The court found that there are risks involved with conducting keyword searches. The court said it is important to determine which keywords were chosen and how they were they used to search the document population; the rationale for selecting those keywords; the qualifications of the individuals selecting such keywords to design effective searches; and whether and to what degree the results of the search were measured for reliability.

In one civil case, (*Victor Stanley, Inc. v Creative Pipe Inc, et. al, 2008*), the parties could not agree on the keywords to search through 165 documents. The court ordered the parties' computer forensic experts to meet and confer in an effort to identify a joint protocol to search and retrieve ESI responses. The computer forensic experts met and prepared a five page keyword/phrase search list. A problem arose as to how much of the material retrieved in this method inadvertently produced privileged material or work product. The defendants then came up with seventy keywords to cull out privileged material; however, the defendants were unclear as to how these 70 words were chosen, how the search was conducted, and what safeguards were employed. (*Victor Stanley, Inc. v Creative Pipe Inc, et. al, 2008*). Judge Grimm ruled that the defendants had waived the privilege because they did not employ a search expert, did not test the key word list, did not check the results and did not have a claw-back agreement. (*Victor Stanley, Inc. v Creative Pipe Inc, et. al, 2008*)

In another case the parties did have a stipulation with a claw-back agreement to return privileged documents inadvertently produced approved by the court. The stipulation provided that in case privileged material was disclosed, the document was to be returned upon written demand. If the recipient of the document wanted to challenge the privilege claim, they were required to do so in writing within five days of the demand for the document's return. As discovery progressed, plaintiffs were permitted to search certain of defendants' storage devices where they discovered several relevant documents not previously produced. Those documents were produced to plaintiffs and included one document that defendants later claimed was privileged and demanded it be returned. The plaintiff returned the document. The plaintiff later moved the court for the right to use the document to which the defendant objected on the grounds that the defendant had objected in a timely manner in accordance with the claw-back agreement. The court ruled that the defendant did not have to produce the document because the plaintiffs failed to follow the agreement made with the defendants. (*Bro-Tech Corp. v. Thermax, Inc., 2008 , 2008*)

In yet another case the plaintiff moved to compel discovery arguing that recently found e-mails and alleged irregularities in the defendant's production of documents justified additional discovery. The defendant argued that any additional e-mails were in backup storage and not reasonably accessible. In this case the court questioned the relevance of the potential evidence and noted that the timing of the request – just two weeks before the trial – made it impossible to complete the searches before the start of trial. As a result, the court granted the plaintiff limited discovery holding that the plaintiff may hire an outside expert at its own expense to search at most five of the defendant's backup tapes; the search would contain a limited number of search terms; the search would include the e-mail records of seven key employees; and the results would be subject to a confidentiality agreement. (*Kilpatrick v Berg, Inc, 2009*)

#### **RULE 702 OF THE FEDERAL RULES OF EVIDENCE**

If matters were not complicated enough for litigants, when determining appropriate sanctions, courts are now requiring litigants to justify their methods of search, the reasons for particular search words, and the expertise of the searcher. Judge Facciola held that any challenges to the search methodology in producing e-discovery must be scrutinized under Rule 702, (*Equity Analytics , LLC v Lundin, 2008*) (*United States v O'Keefe, 2008*) and decided in hearings under *Daubert v. Merrill Dow Pharmaceutical* and Rule 702. The *Daubert* case set the standard for ruling on the admissibility of expert testimony in federal courts. In *Daubert*, the court found that when scientific evidence is involved the court must determine whether the testimony underlying the reasoning or methodology is scientifically valid and properly can be applied to the facts. (*Daubert v. Merrill Dow Pharmaceuticals, 509 U.S. 579 , 1993*)

Judge Terrence McVerry in the *Smith v. Life Investors Ins. Co.* found that the defendant had a duty to demonstrate that its search methodology was reasonable, and the court suggested that proof could be established through identification of the keywords used, an explanation of the qualifications of those selecting the keywords, and proof of assurance testing. (*Smith v Life Investors Ins. Co of Ame, 2009*).

The court further found that it is the duty of the party to be aware of the literature describing the strengths and weaknesses of various methodologies. (Smith v Life Investors Ins. Co of Ame, 2009). The court also determined that if the parties' methodology is challenged they "should expect to support their position with affidavits or other equivalent information from persons with the requisite qualifications and experience, based on sufficient facts or data and using reliable principles or methodology." (Smith v Life Investors Ins. Co of Ame, 2009)

Judge McVerry also ruled that any challenges to e-discovery because of search methodology must be viewed under Federal Rules of Evidence Rule 702 and the Daubert case. Rule 702 reads:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case. (Federal Rules of Evidence, 2009)

In a later case the Supreme Court ruled that the standard outline in Daubert applies not only to testimony based on 'scientific' knowledge, but also to testimony based on 'technical' and 'other specialized' knowledge. (Kumho Tire Company, Ltd. v. Carmichael, 1999) Kumho requires that a court "exercise its gate keeping obligation so that the expert, whether relying on professional studies or personal experience, will, when testifying, employ the same level of intellectual rigor that the expert would use outside the courtroom when working in the relevant discipline." (Kumho Tire Company, Ltd. v. Carmichael, 1999)

The courts are applying the Daubert case to the e-forensic experts and their methodology. In a 2008 case involving misappropriation of trade secrets and computer fraud and abuse violations, the parties submitted competing expert testimony in the plaintiff's motion for spoliation sanctions and cross motions requesting that the opposing forensic experts be excluded under Federal Rule 702 and Daubert. (Nucor Corp. v. Bell, 2008) In this case the court found that both proposed experts met the standards under Daubert and accepted them as experts; however, the court took the unusual position of rejecting some of the plaintiff's expert testimony as not meeting the Daubert standard. Plaintiff Nucor requested sanctions claiming defendant Bell had destroyed evidence found on John Bell's laptop computer and the SanDisk device that Bell discarded. Nucor asked that the court grant a default judgment, or in the alternative, grant an adverse inference charge to the Jury. In support of the request John Jorgensen, an expert in the field of computer forensics, testified that the defendant had intentionally deleted information on its computer by wiping data. Jorgensen, the expert also testified that the missing SanDisk contained Nucor-related documents. Jorgensen also testified that defendant's continued use of the laptop resulted in the loss of relevant data, and that defendant's expert spoliated evidence because he engaged in "destructive" testing on the laptop's hard drive. Defendant's countered Nucor's arguments with the testimony of Dr. Sean McLinden, another expert in the field of computer forensics, who testified that no wiping occurred and that no relevant data was lost as a result of defendants' continued use of the compute. Further, McLinden claimed that he provided plaintiff with an exact copy of the laptop hard drive. However, Bell has admitted that he destroyed the SanDisk

After a Daubert hearing on the qualifications of the parties' computer forensics experts, the court ruled that both forensic witnesses were experts under the Daubert standard but excluded the testimony of Jorgensen on the issue of wiping. Jorgensen had testified that the laptop had large blocks of "zeroes" and therefore he theorized that the zeroes could have been created by use of a secure delete function of the computer program. Dr. McLinden testified Bell's laptop did not have a secure delete function. The court concluded that Jorgensen's methods to test wiping failed to meet the Daubert standard and

therefore barred him from testifying on that matter. (*Nucor Corp. v. Bell*, 2008)

Validation under Daubert standards when litigants use software could be extremely difficult. If the person using the software is not a forensic expert it is doubtful that they would be qualified to testify as to the validity of the methodology. Furthermore, search engine software applications are proprietary. It is not reasonable or appropriate to expect the designers and proprietors of such software to testify in court to explain the proprietary methods and source code of their programs, and the actual methodology or other valuable information to satisfy the Daubert requirements. An expert could testify that he or she has validated a program by testing it in the lab obtaining the same results as the results obtained by a program known to be valid.

If litigants are required to validate their methods, they should be prepared to discuss, how many false positives were discovered, how many false positives have been found, what methods were developed to measure what is missing, and the statistical sampling methods used. Attorneys need to be aware that they may be required to defend the methodology they used in their search. Required proof may include providing quality assurance testing and providing measurement protocols to support their search methodology. Rule 34(a) of the Federal Rules of Civil Procedure has mentioned sampling as an accepted methodology for producing information so it could be argued as a method of validation.

In the O'Keefe case Judge Facciola found that when testing search engines against the Daubert test it might make a difference if the expert is attempting to prove something is present versus attempting to prove that something is not present. The problem is that typically the requirement is to find everything.

In a Texas appellate case the court found the expert's use of Guidance Software's EnCase forensic tools to create a bit-stream forensic image of defendant's hard drive to search for child pornography was appropriate. Since the purpose was to find whether something existed and it was found, the court accepted the evidence as meeting the Daubert standard. (*Sanders v. Texas*, 2006) However, in O'Keefe the challenge was that the information was not complete. It is difficult under the Daubert standard to prove that nothing was missed. Successful Daubert challenges can totally destroy the opponent's case so litigators and forensic experts need to prepare for the eventual challenge.

Since there are better search methods than keyword methods, it would be difficult for an attorney to defend this methodology as the sole search method used after applying the foregoing standards. Litigants must use other search methodologies, such as concept searches or pattern recognition searches. However, even with other search methodologies attorneys and their experts must be able to defend the methodology used.

## **SEARCH METHODOLOGIES OTHER THAN KEYWORDS**

The primary problem in selecting a search methodology is that the method chosen must include the ability to search, to cluster data, to classify, to filter, to analyze social network information, and finally, to extract information. In the Lehman Brothers bankruptcy case there were over 3.2 billion e-mail and instant message records that had to be reviewed. It became clear to the investigators when they realized that keyword searching would not be adequate so they had to rely on other methods including concept searching and visual mapping. (*Lehman Brothers Holdings Inc.*, 2008) In cases like the Lehman Brothers case and others attorneys had to learn to use other search methodologies that should include fuzzy searches, semantic searches as well as concept searches.

### **1. Fuzzy Searches**

Fuzzy searches find misspellings in words so they are particularly effective in finding results where words are typically misspelled. Spelling errors are fairly common in cases involving technical terms. The degree of "fuzziness" is normally adjusted via a numeric value. Many search engines will use the degree of fuzziness to determine the number of letters that can be wrong in the misspelling. For instance, a fuzzy value of 1 would mean that only one letter can be wrong in the word and a value of 3 means three letters can be wrong. For example, using a search term of "rake" and a fuzzy degree of 1



will find the term “fake.” Fuzzy searching is a very valuable ability when dealing with sophisticated terms, where the proper spelling is not widely known.

## **2. Concept Searches**

Concept searching takes the input term and returns results that are related in meaning. This is best explained in this example. If the term ‘car’ was used, then results of ‘automobile’ and ‘vehicle’ would be in the returned hits. One of the most well known of the concept search engines is Attenex. The Attenex engine is very powerful, but not for the faint of heart. The cost is very high and is typically used in large, high profile cases where there is a large volume of electronic information. The key issue to be aware of with conceptual search engines is that they are only as good as the programming logic.

## **3. Bayesian Network Searches**

Essentially, a Bayes network is a simple probabilistic classification method based on applying Thomas Bayes’ theorem regarding conditional probabilities. A Bayes network can be used to answer questions such as, “I am Caucasian male, 42 years old, and am overweight. What is the likelihood that I am diabetic?” This method has been effective in classifying documents for retrieval based on the content of the documents and the concepts that are contained in the document population. For example, based on the content of all of the documents in a given population of documents, what is the likelihood a particular document is relevant to a particular concept or search term? A Bayesian network will retrieve documents that have a high probability of being relevant to a particular class or category. As documents are added to the population of documents, the probabilities are updated. This methodology has proven to be very useful in information retrieval and data mining. The advantage of Bayesian networks in data mining is that the probabilities are updated continually during the search process, and thus, it can connect information with causal relationships. (Heckerman, 1997)

## **CONCLUSION**

In conclusion, whatever the chosen search methodology used, attorneys must obtain the assistance needed whether it is from a linguist, a computer forensics expert, or an IT expert to provide the information requested and to demand the information through discovery needed to properly litigate. Courts have come to recognize the problem with keyword searches before litigants have. Judge Facciola stated, “Whether search terms or ‘keywords’ will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics.” (United States v O’Keefe, 2008).

Litigants will need to defend their methodology and it may be necessary to employ more than one methodology to convince the court that the search can be defended. In a typical case there can be more than one hundred thousand documents that must be searched to find all of the necessary and relevant documents without providing irrelevant or privileged documents. If the search methodology seeks to be too precise, that is, a search that results in a high percentage of retrieved documents being considered relevant, there is a risk that some relevant documents may be overlooked. If the search methodology seeks to have a high level of recall, that is, to find every possible relevant document, there will likely be a high percentage of false positives, or documents that are not relevant. There is necessarily a degree of tradeoff between precision and recall that is involved in developing any search methodology. Despite the safe harbor rules in place in the Federal Civil Rules of Procedures, the courts have had little tolerance for litigants’ failure to provide the necessary discovery requests. Noncompliant litigants can be sanctioned by courts imposing fines, attorney fees, discovery costs, adverse inferences, and dismissal of the case in its entirety.

All attorneys must work with the opposing attorneys to determine the method and amount of searching that will be conducted. The new rules and subsequent cases demand that attorneys cooperate with one another and that the discovery process be transparent. However, even if attorneys agree to limit their ESI search there is always the chance of a malpractice lawsuit looming if the demand was inadequate. Attorneys must employ computer forensics experts throughout the discovery phase of the litigation

process. Further, attorneys must understand that searching methodologies require more knowledge than how to Google. Attorneys must take classes in computer forensics as well as information retrieval and electronic discovery. Universities need to develop more programs in computer forensics to prepare specialists who can advise counsel on such matters and who are prepared to take testify in court.

In this digital world attorneys must understand how to discover, produce, preserve and justify electronic data. Attorneys and their experts must be prepared for the eventual challenges to their method and experts. The cases in which the judges have required justification of the methodology has put litigators on notice that choosing a search method and expert may not be an easy matter.

“While it is now possible to store enormous amounts of reported decisions in electronic databases, the retrieval of relevant cases remains an extremely difficult task.” (Smith, 1998) It will continue to be a difficult task as data storage continues to multiply and data searching and retrieving continues to be scrutinized.

Regardless of the methodology chosen, attorneys need to have a plan to determine the best way to find, retrieve and seek electronic data. Attorneys need to know what experts are needed and engage them early in the process. Attorneys need to be prepared to argue what is reasonable to find and produce and what is reasonable to request.

#### **REFERENCES**

- Asarco, Inc. v. United States Env'tl. Prot. Agency, WL 1138830 (D.D.C. April 28, 2009).
- Ball, C. (2006). Musings on Meet and Confer. Retrieved December 12, 2009, from Craig Ball, LTD: [http://www.craigball.com/Musings\\_on\\_Meet\\_and\\_Confer.pdf](http://www.craigball.com/Musings_on_Meet_and_Confer.pdf)
- Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery. (2007). Sedona Conference Journal , 191-217.
- Bro-Tech Corp. v. Thermax, Inc., 2008 , WL 5210346 (E.D.PA December 11, 2008).
- ClearOne Commc'ns, Inc. v. Chiang, WL 920336 (D.Utah April 1, 2008).
- Covad Commc'ns. Co. v. Revonet, Inc., WL 2595257 (D.D.C. August 25, 2009).
- Daubert v. Merrill Dow Pharmaceuticals, 509 U.S. 579 , 509 U.S. 579 (Court of Appeals 9th circuit 1993).
- Di Sciuillo, A. (2005). Natural language asymmetry and internet infrastructures. International Journal of Electronic Business , 3(3,4), 328-338.
- Disability Rights Council v. Washington Metropolitan Transit Authority, 242 F.R.D. 139 (D.D.C. 2007).
- Equity Analytics, LLC v. Lundin, WL 615528 (D.D.C. Mar 7, 2008).
- Fattahi, R. W. (2008). An alternative approach to natural language query expansion in search engines: Text analysis of non-topical terms in Web documents. . Information Processing & Management , 44(4), 1503.
- Federal rules of civil procedure, 16(b) (2006).
- Federal Rules of Civil Procedure, 37 (2006).
- Federal Rules of Civil Procedure, Rule 26(b)(5) (2007).
- Federal Rules of Civil Procedure, Rule 34 (2007).
- Federal Rules of Civil Procedure, 37(f) (December 2007).
- Federal Rules of Civil Procedure, Rule 702 (2009).

- Federal Rules of Civil Procedure 26, 26(a)(1) (2006).
- Federal Rules of Civil Procedures, 16 (2006).
- Fukumoto., T. (2006). An analysis of image retrieval behavior for metadata type image database. *Information Processing & Management* , 42(3)723-728.
- Golden G. Richard III, & V. (2006). Next-generation digital forensics. *Association for Computing Machinery Communications of the ACM* , 49(2), 76.
- Gunter, D. (2010). Semantic Search. *Bulletin of the American Society for Information Science and Technology* , 36-37.
- Hansen, K. (1998). The Effect of Presearch Experience on the Success of Naive (End-User) Searches. *Journal of the American Society for Information Science* , 37(5), 315.
- Heckerman, D. (1997). Bayesian Networks for Data Mining . *Data Mining and Knowledge* .
- Holt, J. &. (2009). An Evolution of Search. *Bulletin of the American Society for Information Science and Technology* , 11-15.
- In re Zurn Pex Plumbing Prods. Liab. Litig., WL 1606653 (D.Minn June 5, 2009).
- in Rowe Entertainment Inc. v. William Morris Agency, Inc., 205 F.R.D. 421, 423 (S.D.N.Y 2002).
- Kilpatrick v. Breg, Inc, WL 1764829 (S.D. Fla June 22, 2009).
- Kumho Tire Company, Ltd. v. Carmichael, 526 U.S. 137 (U.S. Supreme Court 1999).
- Lehman Brothers Holdings Inc., No. 08-13555 (US Bankruptcy Court, Southern District of New York September 15, 2008).
- Mearian, L. (2007, February 23). Computerworld. Retrieved January 31, 2007, from From bits to yottabytes How much data is that?: <http://blogs.computerworld.com/node/5054>
- Mearian, L. (2008, March 11). Study:Digital Universe and its impact is greater than we thought. Retrieved January 31, 2010, from Computer World: [http://www.computerworld.com/s/9067639/Study\\_Digital](http://www.computerworld.com/s/9067639/Study_Digital)
- Nucor Corp. v. Bell, WL 4442571 (D.S.C. Jan 11, 2008).
- Oard, D. H. (2008, April 13). Overview of the TREC 2008 Legal Track. Retrieved January 21, 2010, from Trec Legal Track: <http://trec.nist.gov/pubs/trec17/papers/LEGAL.OVERVIEW08.pdf>
- Procter & Gamble Co. v. Haugen, ., 179 F.R.D. 622 (D.Utah 1998), rev'd on other grounds, 222 F.3d 1262 (10th Cir. 2000) (10 Cir 1998).
- Rogets, P. (1911, amended 1991). Roget's Thesaurus of English Words. Chicago: University of chicago.
- Rothstein, R. H. (2007). Managing Discovery of Electronic Information. . *A Pocket Guide for Judges* , 3.
- Sanders v. Texas, 191 S.W.3d 272 (Ct.App 2006).
- Smith v Life Investors Ins. Co of Ame, WL2045197 (W.D. PA July 9, 2009).
- Smith, J. C. (1998). Machine Intelligence and Legal Reasoning. *CHI-Kent Rev* , 334-335.
- State ex rel. Gehl v. Connors, WL 3024436 (Wisconsin Ct. App October 18, 2007).
- The Sedona Conference Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery. (2007). *The Sedona Conference Journal* , 189-217.
- Toussie v. County of Suffolk, WL 4565160 (E.D.N.Y Dec 21, 2007).

*ADFSL Conference on Digital Forensics, Security and Law, 2010*

United States v O'Keefe, 537 F. Supp. 2d at 22 (D.D.C February 18, 2008).

Victor Stanley, Inc. v Creative Pipe Inc, et. al, FRD 251 MJG-062662 (District Court of Maryland 2008).

William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co. , WL 724954 (S.D.N.Y March 19, 2009).

Wired. (2007, October 26). Terabyte Thumb Drives Made Possible by Nanotech Memory. Retrieved January 31, 2010, from [http://www.wired.com/gadgets/miscellaneous/news/2007/10/ion\\_memory](http://www.wired.com/gadgets/miscellaneous/news/2007/10/ion_memory):  
[http://www.wired.com/gadgets/miscellaneous/news/2007/10/ion\\_memory](http://www.wired.com/gadgets/miscellaneous/news/2007/10/ion_memory)

Zubulake v. UBS Warburg LLC, 217 F.R.D. 309 (S.D.N.Y 2003).

