


April 2012

Effective Keyword Selection Requires a Mastery of Storage Technology and the Law

Daniel B. Garrie

Law & Forensics LLC, daniel@lawandforensics.com

Follow this and additional works at: <http://digitalcommons.pace.edu/plr>

 Part of the [Legal Profession Commons](#), and the [Science and Technology Law Commons](#)

Recommended Citation

Daniel B. Garrie, *Effective Keyword Selection Requires a Mastery of Storage Technology and the Law*, 32 Pace L. Rev. 400 (2012)

Available at: <http://digitalcommons.pace.edu/plr/vol32/iss2/5>

This Article is brought to you for free and open access by the School of Law at DigitalCommons@Pace. It has been accepted for inclusion in Pace Law Review by an authorized administrator of DigitalCommons@Pace. For more information, please contact cpittson@law.pace.edu.

Effective Keyword Selection Requires a Mastery of Storage Technology and the Law

By Daniel B. Garrie, Esq.*

I. Introduction

Selecting keywords for searching large volumes of electronically stored information (“ESI”) is an unavoidable, but necessary step in the process of electronic discovery.¹ The parties to a case, or the court, may choose the terms for the search.² However, an efficient alternative to both options involves a mediator, neutral, or special master with a thorough understanding of the legal elements of the case and the technology systems that will be subject to keyword search.³ This alternative can benefit both parties, as well as the court, because a “technology-aware” mediator can expedite an agreement that allows both parties to maintain oversight of the keyword selection process.⁴ This serves both parties’

* DANIEL B. GARRIE, Esq. has a B.A. and M.A. in computer science, is an e-Discovery Neutral and Special Master with Alternative Resolution Centers (www.arc4adr.com), and is a Partner at Law & Forensics LLC (www.lawandforensics.com), a legal strategy consulting firm. He can be reached at daniel@lawandforensics.com. The Author would like to thank William Spernow, Yoav Griver, and Khalid Bashjawish for their assistance with this Article.

1. *See, e.g.*, *Stern v. Shelley*, No. 4:08-cv-02753-JMC, 2010 WL 4721708, at *3-4 (D.S.C. Nov. 12, 2010); *Fisher v. Ciba Specialty Chems. Corp.*, No. 03-0566-WS-B, 2007 WL 987457 (S.D. Ala. 2007); *Treppel v. Biovail Corp.*, 233 F.R.D. 363, 374-75 (S.D.N.Y. 2006).

2. *See, e.g.*, *Mt. Hawley Ins. Co., v. Felman Prod., Inc.*, 271 F.R.D. 125 (S.D. W. Va. 2010); *William A. Gross Constr. Assocs. v. Am. Mfrs. Mut. Ins. Co.*, 256 F.R.D. 134 (S.D.N.Y. 2009); *Smith v. Life Investors Ins. Co. of Am.*, No. 2:07-cv-681, 2009 WL 2045197 (W.D. Pa. July 9, 2009); *Rhoads Indus., Inc. v. Bldg. Materials Corp. of Am.*, 254 F.R.D. 216 (E.D. Pa. 2008); *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251 (D. Md. 2008).

3. *Mancia v. Mayflower Textile Servs. Co.*, 253 F.R.D. 354, 357 (D. Md. 2008) (citing FED. R. CIV. P. 26-37 advisory committee’s notes to the 1983 amendments); Allison O. Skinner, *The Role of Mediation for ESI Disputes*, 70 THE ALABAMA LAWYER 425, 426 (2009); *see, e.g.*, *Grant St. Grp., Inc. v. Realauktion.com, LLC*, No. 2:09-cv-01407-DWA, 2010 WL 4808510 (W.D. Pa. Oct. 19, 2010).

4. *See Zubulake v. UBS Warburg, LLC*, 229 F.R.D. 422, 432 (S.D.N.Y. 2004)

interests because, as the *Zubulake* court noted, “[i]t might be advisable to solicit a list of search terms from the opposing party for [the purpose of preservation], so that [opposing counsel] could not later complain about which terms were used.”⁵ A poorly designed search term list guarantees that the parties will have to perform a series of subsidiary searches as gaps and problems in the original search become apparent.⁶ This can easily be mitigated with a mediator who knows the relevant law *and* technology.⁷ An effective search that results in responsive items being identified begins with the intangible creativity that forms a bond between knowledge of the law and technology.

II. Mediators Can Deliver Value to the Keyword Selection Process by Bridging the Awareness Gap Between Attorneys and Technologists

Companies and counsel faced with e-discovery have little choice but to use search terms or “keywords” in a threshold exercise to separate relevant from non-relevant information.⁸ Traditional document review techniques involving hard copies are not practical or financially feasible when reviewing a seemingly endless amount of documents in ESI form. Usually, finding relevant information in ESI form requires counsel or the court to identify search terms and apply these terms across all potential evidence in the matter.⁹ Although recent searches have become more sophisticated through the use of statistical sampling and predictive coding techniques,¹⁰ disputes over search methodology often result in the

(advising that discovery requests can be more effective if both parties “negotiate a list of search terms to be used in identifying responsive documents”); *see also* Shira Scheindlin, *We Need Help: The Increasing Use of Special Masters in Federal Court*, 58 DEPAUL L. REV. 479 (2009).

5. *Zubulake*, 229 F.R.D. at 432 n.75.

6. *See, e.g.*, *McNulty v. Ready Ice Holding Co.*, 271 F.R.D. 569 (E.D. Mich. 2011); *Dataworks, LLC v. Commlog, LLC*, No. 09-cv-00528-PAB-BNB, 2011 WL 66111 (D. Colo. Jan. 10, 2011).

7. *William A. Gross Constr. Assocs.*, 256 F.R.D. at 134.

8. *See, e.g.*, *Vioxx Prods. Liab. Litig. Steering Comm. v. Merck & Co.*, No. 06-30378, 2006 WL 1726675, at *2 n.5 (5th Cir. May 26, 2006); *Equity Analytics LLC v. Lundin*, 248 F.R.D. 331 (D.D.C. 2008).

9. *Garcia v. Tyson Foods, Inc.*, No. 06-2198-JWL-DJW, 2010 WL 5392660 (D. Kan. Dec. 21, 2010); *Smith v. Life Investors Ins. Co. of Am.*, No. 2:07-CV-681, 2009 WL 2045197 (W.D. Pa. July 9, 2009).

10. Jason R. Baron, *Law in the Age of Exabytes: Some Further Thoughts on “Information Inflation” and Current in E-Discovery Search*, 17 RICH. J.L. & TECH., vol. 3, 2011 at 1, 7. Predictive coding has been defined as:

court ultimately determining how the search will be conducted. For example, in *William A. Gross Construction Associates, Inc. v. American Manufacturers Mutual Insurance Co.*, the plaintiff's proposed keyword search was too narrow and the defendant's proposed keyword search was too broad, so the court was left in the "uncomfortable position" of crafting and imposing its own search methodology for the parties.¹¹ While the courts may be the option of last choice in resolving these matters, they also realize that even their expertise has limits and that their mandated involvement may not be the best solution.

By way of a simplified example of the ease with which poorly designed search term lists can be overlooked, a party that suggests the word "tax" or "confidential" in a finance-based litigation as a keyword is likely asking for production that is too broad and costly. A better keyword search would include an involved party's name or certain document formats (e.g., excel spreadsheets).¹² But using the wrong search terms or inappropriate document types can lead to various negative outcomes, including but not limited to overproduction, or non-production. An example of this occurred in *Nycomed U.S. Inc. v. Glenmark Generics Ltd.*¹³ In this case, the court ordered the defendant to pay one hundred thousand dollars to the plaintiff and twenty-five thousand dollars to the clerk of the court because the defendant deliberately failed to identify and search the electronic databases that were likely to contain discoverable information.¹⁴

[A] combination of technologies and processes in which decisions pertaining to the responsiveness of records gathered or preserved for potential production purposes . . . are made by having reviewers examine a subset of the collection and having the decisions on those documents propagated to the rest of the collection without reviewers examining each record.

eDiscovery Institute Survey on Predictive Coding, ELEC. DISCOVERY INST., 2 (Oct. 1, 2010), <http://www.ediscoveryinstitute.org/images/uploaded/272.pdf> (internal quotation marks omitted) [hereinafter *Survey on Predictive Coding*].

11. 256 F.R.D. at 134-35.

12. *See, e.g.*, Content-Based Implicit Search Query, U.S. Patent No. 2006/0271520 (filed May 27, 2005) (published Nov. 30, 2006) (providing narrower search results by organizing the results by file type).

13. *Nycomed U.S. Inc. v. Glenmark Generics LTD.*, No. 08-CV-5023, 2010 WL 3173785 (E.D.N.Y. Aug. 11, 2010).

14. *Id.* at *12.

III. Counsel Should Select Mediators Who Not Only Understand the Law But Grasp the Nuances of Advanced Technology

Without a special master or mediator who knows the law and the technology, a poorly designed keyword search can lead to costly inefficiencies. There are several reasons why this occurs.¹⁵ First, counsel often do not fully understand the various forms of storage technology on which his or her client's information is stored. Courts have lamented this problem, which Judge Facciola addressed in a colorful section of the opinion for *United States v. O'Keefe*:

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 . . . Given this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread.¹⁶

Similarly, in *Equity Analytics, LLC v. Lundin*, the court 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 a lay lawyer) and requires expert testimony that meets the requirements of Rule 702 of the Federal Rules of Evidence."¹⁷ Second, even if counsel understands the client's technology and the law of the case, this is still not enough for purposes of an effective keyword search because e-discovery,¹⁸ for purposes of efficiency, requires that the attorneys share their understanding of the

15. See, e.g., *DeGeer v. Gillis*, 755 F. Supp. 2d 909 (N.D. Ill. 2010); *Trusz v. UBS Realty Investors LLC*, No. 3:09-CV-268(JBA), 2010 WL 3583064, at *4 (D. Conn. Sept. 7, 2010), *vacated in part on reconsideration*, 3:09-CV-268(DJS), 2011 WL 124504 (D. Conn. Jan. 13, 2011); *Spieker v. Quest Cherokee, LLC*, No. 07-1225-EFM, 2008 WL 4758604 (D. Kan. Oct. 30, 2008).

16. *United States v. O'Keefe*, 537 F. Supp. 2d 14, 24 (D.D.C. 2008) (internal citation omitted).

17. *Equity Analytics, LLC v. Lundin*, 248 F.R.D. 331, 333 (D.D.C. 2008).

18. See *Asarco, Inc. v. EPA*, No. 08-1332 (EGS/JMF), 2009 U.S. Dist. LEXIS 37182, at *3 (D.D.C. Apr. 28, 2009).

case and the technology with opposing counsel.¹⁹

Perhaps counterintuitive to some attorneys today, the origin of this principle can be traced back as far as 1947 to the case of *Hickman v. Taylor*, in which Justice Murphy provided that “Mutual knowledge of all the relevant facts gathered by both parties is essential to proper litigation.”²⁰ Such adversarial zeal inhibits any cooperative efforts and prevents the selection of keywords expeditiously. If so, a court, as in *Lundin*, can appoint a computer forensics expert at the cost of the parties to search the ESI at issue. The court may also require further affidavits from the parties as to the adequacy of proposed search methodologies.²¹ Understandably, most courts, concerned with the disclosure of facts, will usually lean in the direction of ordering additional discovery, trusting that this is the best method for extracting the truth, or encouraging the parties to try to settle their dispute.²² In such a scenario, the parties can benefit substantially by either agreeing or petitioning the judge to appoint a mediator that knows both the law and the technology, to ensure that appropriate documents are produced at a reasonable cost respective to the underlying issue.²³

Although agreeing on a mediator that is skilled in both the law and technology will more than likely lead to efficiency, one note of caution is required.²⁴ Often, the court-appointed mediator knows the particular business area in dispute but has no more technological education or experience than the parties or the court. For example, if the parties are in an insurance-related dispute, organizations such as ARIAS have a stable of potential mediators and arbitrators with years of impressive, insurance-related experience available for choosing. Few of them,

19. *E.g.*, *Rexall Sundown, Inc. v. Perrigo Co.*, 651 F. Supp. 2d 9, 12 (2009); *Clearone Commc'ns, Inc. v. Chiang*, No. 2:07 CV 37 TC, 2008 U.S. Dist. LEXIS 27617, at *3 (D. Utah Apr. 1, 2008); *Elliott v. U.S. Att'y Gen.*, No. 06-1128 (JDB), 2006 U.S. Dist. LEXIS 80204, at *7-9 (D.D.C. Nov. 2, 2006); *J.C. Assocs. v. Fid. & Guar. Ins. Co.*, No. 01-2437 (RJM/JMF), 2006 U.S. Dist. LEXIS 32919, at *1-4 (D.D.C. May 25, 2006).

20. *Hickman v. Taylor*, 329 U.S. 495, 507 (1947).

21. *See Lundin*, 248 F.R.D. at 333.

22. *See Carrie Lonetti, When the Emperor Has No Clothes: A Proposal for Defensive Summary Judgment in Criminal Cases*, 84 S. CAL. L. REV. 661, 709 (2011).

23. *See, e.g.*, *Hammann v. 800 Ideas, Inc.*, No. 2:08-cv-00886-LDG-GWF, 2010 WL 4943391 (D. Nev. Nov. 22, 2010); *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 269 F.R.D. 497 (D. Md. 2010); *Pension Comm. of the Univ. of Montreal Pension Plan v. Banc of Am. Secs., LLC*, 685 F. Supp. 2d 456 (S.D.N.Y. 2010); *Zubulake v. UBS Warburg LLC*, 220 F.R.D. 212 (S.D.N.Y. 2003).

24. *See also* MODEL STANDARDS OF CONDUCT FOR MEDIATORS 2 (2005); Amy Cook, *ADR Is a-OK*, 22 CBA REC. 6 (2008).

however, would likely know the differences involved in recovering data from an IBM mainframe computer system as opposed to one from Hewlett Packard. Few lawyers can appreciate the subtle nuances of Boolean “proximity searches,”²⁵ “stemming,” and “fuzzy logic.”²⁶ Although anyone who has used an Internet search engine may have some familiarity with Boolean logic, when it comes to metadata, compression algorithms, artifacts data fragments, entropy tests, sub-OS level searches, and even the visual examination of free space areas of the drive, these subjects will be beyond the understanding of the business-oriented mediator. Demonstrably, a neutral mediator or special master appointed for discovery purposes needs both a firm grasp of the business field as well as firm grasp of theory and application around the field of electronic search methodology, including an understanding of the different algorithms by which forensic software searches for information.²⁷

Since most mediators are not technologists, this lack of specialized knowledge unfortunately works against both parties and negates most or all of the value that a mediator can deliver.²⁸ If a mediator does not understand the technology, the litigants will still have to provide independent, technical, expert reports supporting any objection to the scope of discovery,²⁹ or see the mediator retain a consulting technical expert of its own. As a result, this additional consultant will serve at the parties’ expense. Thus, unless a proper mediator is chosen, the parties once again will find themselves in a situation where the cost of production far outweighs the limited value resulting from the execution of a poorly designed discovery search.³⁰ One example of this is the case

25. Kevin Shay, *Google API Proximity Search (GAPS)*, STAGGER [NATION], <http://www.staggernation.com/gaps/readme.php> (last visited Feb. 21, 2012).

26. See David C. Blair & M. E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System*, 28 COMM. OF THE OF THE ACM 289 (1985); *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 8 SEDONA CONF. J. 189, 192 (2007) [hereinafter *Best Practices Commentary*].

27. See Scheindlin, *supra* note 5, at 481 (stating that primary considerations for a court selecting a special master are “(1) time commitment; (2) knowledge and expertise; (3) resources; and (4) neutrality.”).

28. See *Mt. Hawley Ins. Co. v. Felman Prod., Inc.*, 271 F.R.D. 125 (2010).

29. See *In re Bankers Trust Co.*, 61 F.3d 465, 469 (6th Cir. 1995); *Searock v. Stripling*, 736 F.2d 650, 653 (11th Cir. 1984); *McCoo v. Denny’s Inc.*, 192 F.R.D. 675, 692 (D. Kan. 2000); *In re Kuntz*, 124 S.W.3d 179, 181 (Tex. 2003).

30. See, e.g., *FSP Stallion 1, LLC v. Luce*, No. 2:08-CV-01155-PMP-PA, 2009 WL 2177107, at *5 (D. Nev. July 21, 2009); *Young v. Pleasant Valley Sch. Dist.*, No. 3:07cv865, 2008 WL 2857912, at *2 (M.D. Pa. July 21, 2008); *Advante Int’l Corp. v.*

of *Ross v. Abercrombie & Fitch Co.*, in which the court found that a 123 keyword search returning 1.3 million documents produced largely irrelevant results and the costs of review for relevance and privilege outweighed the benefit of ordering production.³¹ Therefore, like a good chess move, opposing counsel can make two moves in one by selecting a mediator that is knowledgeable of both the law and technology.

IV. Conclusion

In conclusion, the management of keyword selection by a skilled mediator offers both parties to a suit the opportunity to navigate the immense volumes of ESI that have come to characterize traditional discovery as “e-discovery” and that threaten to expose counsel to inadvertent disclosure, misconduct, spoliation, and worse. The guiding light that a skilled lawyer, or a retired judge with IT expertise, brings to the keyword selection process benefits everyone including the parties themselves, their attorneys, the court, and even third-parties and nonparties who may be custodians of ESI. Even with emphasis on keyword searches, e-discovery may be entering the next phase of legal technology, a phase that succeeds the use of keyword searches. In the case of *Asarco, Inc. v. EPA*, the court granted a request for additional search terms but noted that keyword searches are no longer the preferred methodology.³² The *Cache La Poudre Feeds* court, interpreting *Zubulake V*, denied a sanctions request that was based on a “perceived obligation” to conduct keyword searches.³³

A mediator, neutral, referee, or special master may be as important in conducting the selection and implementation of keywords as in consulting the parties and the court on when this methodology is non-productive and at odds with the search for the truth.

Mintel Learning Tech., No. C 05-01022 JW (RS), 2006 WL 1806151, at *1 (N.D. Cal. June 29, 2006); *In re* 3817 W. West End, 321 F. Supp. 2d 953, 958 (N.D. Ill. 2004); *Bray & Gillespie Mgmt. LLC v. Lexington Ins. Co.*, 259 F.R.D. 568, 590 (M.D. Fla. 2009); *F&M Expressions Unlimited, Inc. v. O’Connell*, No. C-240-04 (N.J. Sup. Ct. Sept. 10, 2004).

31. *Ross v. Abercrombie & Fitch Co.*, Nos. 2:05-cv-0819, 2008 WL 4758678, at *2 (S.D. Ohio Oct. 27, 2008).

32. *Asarco, Inc. v. EPA*, No. 08-1332 (EGS/JMF), 2009 U.S. Dist. LEXIS 37182 (D.D.C. Apr. 28, 2009).

33. *Id.*; *see also* *Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth.*, 242 F.R.D. 139 (D.D.C. 2007).