

Where Angels Fear to Tread:
Intelligent Privilege Review

A White Paper

By Jennifer R. Martin

Date: January 28, 2010

Where Angels Fear to Tread: Intelligent Privilege Review

A White Paper

By Jennifer R. Martin
Stroz Friedberg

Introduction.....1

A. Legal and Historical Background.....2

B. The Facciola Redgrave Framework.....6

C. Stroz Friedberg’s Intelligent Privilege Filter.....8

 1. Advanced Search Technologies.....9

 2. Identification and Classification of Potentially Privileged Documents...12

i. User-Created Search Term List.....12

ii. Categories of Potentially Privileged Documents.....12

iii. Preliminary Testing Results.....15

 3. Reporting Requirements, Indexes, and Privilege Logs.....16

Conclusion.....17

Where Angels Fear to Tread: Intelligent Privilege Review

A White Paper

By Jennifer R. Martin¹

Introduction

The race to develop new search technologies to aid in cost-effective and reliable discovery has significantly accelerated in recent years. Over the last several years, large corporations and institutions have expressed increasing alarm over the tremendous volume of data that must be preserved, processed, and reviewed in litigation, and they have pushed for new technologies to mitigate the significant associated costs. Courts also have expressed concern over the growing expenditure of time and resources, including judicial resources, associated with electronic discovery, the concomitant gamesmanship, and, most recently, the reliability of traditional search processes. As a result, law firms, judges, technology consultants, and electronic discovery vendors are hard-pressed to find affordable, defensible solutions.

This paper discusses one advanced search solution developed by Stroz Friedberg, LLC, to address the particular problems associated with conducting sensitive privilege reviews on large volumes of documents. The technical solution discussed in this paper dovetails precisely with the groundbreaking legal framework for conducting such reviews set forth by Judge John Facciola and Jonathan Redgrave, Esq., in their recent article on the subject: Hon. John M. Facciola and Jonathan M. Redgrave, *Asserting and Challenging Privilege Claims in Modern Litigation: The Facciola-Redgrave Framework*, 2009 Fed. Cts. L. Rev. 4 (November, 2009) (hereinafter “Facciola-Redgrave Framework”). As discussed more fully below, both Stroz Friedberg’s Intelligent Privilege Filter and the Facciola-Redgrave Framework focus on using advanced technology to classify potentially privileged documents into categories based on various attributes associated with privileged material, and then ranking those categories based on the relative likelihood that the documents contained therein are privileged. The application of Stroz Friedberg’s advanced text analytics and search technologies to classify documents, coupled with the cooperation of the parties as mandated by the Facciola-Redgrave Framework, will result not only in significant cost and time efficiencies, but a more defensible methodology for performing privilege reviews.

¹ Jennifer R. Martin is a Managing Director at Stroz Friedberg, an international consulting and technical services firm focusing on digital forensics, electronic discovery, cybercrime response, and investigations. She is a former Senior Counsel with the US Department of Justice’s Computer Crime and Intellectual Property Section in Washington, DC.

A. Legal and Historical Background

In 1985, David Blair and M.E. Maron conducted the first study to test the efficacy of Boolean keyword searches in the context of legal discovery. See David Blair & M.E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System*, 28 Com. A.C.M. 289 (1985). In their experiment, lawyers and paralegals were asked to develop keyword terms to search approximately 40,000 documents for materials relevant to an accident. Although the attorneys estimated that they had found approximately 75% of the relevant documents using this traditional search methodology, the study showed that, in fact, their keyword searches had identified a mere 20% of the relevant materials within the document sample.

While experts in the fields of statistical linguistics, information management, and artificial intelligence worked diligently over the next two decades to make substantial progress in the areas of text analytics and information retrieval, the legal profession largely ignored the problems associated with using traditional keyword searches in litigation. Twenty-two years after the Blair & Maron study, TREC Legal Track² conducted a similar study, and found that only 22% of relevant documents were identified out of the large sample set using Boolean keyword searches, signaling that there had been no significant improvements in legal search protocols since the 1985 findings.

While the Blair & Maron and TREC studies both emphasized the failure of keyword searches to adequately identify the universe of relevant documents in a data set (“recall”), litigants were grappling with the converse problem associated with the lack of “precision.” Specifically, keyword searches were generating an enormous number of irrelevant documents that nonetheless had to be manually reviewed at great cost. In addition, a number of studies, including a 2009 study conducted by Xerox, highlighted the lack of consistency between reviewers in conducting manual reviews.³

Through a series of significant decisions brought down following the 2006 amendments to the Federal Rules of Civil Procedure relating to electronically stored information (“ESI”), courts began articulating new paradigms for resolving

² The Text REtrieval Conference is a conference series co-sponsored by the National Institute of Standards and Technology (NIST) designed to promote research into the science of information retrieval. TREC Legal Track was introduced in 2006 by Jason Baron, Director of Litigation for the US National Archives and Records Administration, specifically to evaluate search methodologies in the context of electronic discovery.

³ In the 2009 Xerox study, 5000 emails were reviewed for relevance by five different review teams. The percentage of documents deemed responsive to the review criteria ranged from 39% to 58% among the teams. In addition, the review results for each of the teams were compared to the responsiveness rate of a pre-determined “Gold Standard” for review; the correlations between the review teams’ results and the Gold Standard ranged from 72% to 79%.

the growing problems associated with electronic discovery. First, a number of courts suggested, much to the surprise of experienced litigators, that discovery is *not* part of the adversarial litigation process. Rather, citing the Federal Rules, as well as the Sedona Conference's *Cooperation Proclamation*, several courts have held that electronic discovery is a "cooperative undertaking" mandated by law. In re Seroquel Products Liability Litigation 2007 U.S. Dist. LEXIS 61287 (MD FL Aug. 21, 2007); Mancia v. Mayflower Textile Servs. Co., Civ. No. 1:08-CV-00237-CCB (D. Md. Oct. 15, 2008) (J. Grimm). See also Qualcomm Inc. v. Broadcom Corp., 2008 WL 66932 (S.D. Cal. Jan 7, 2008), vacated in part, Qualcomm, Inc. v. Broadcom Corp., 2008 WL 638108 (S.D. Cal. March 5, 2008) ("For the current 'good faith' discovery system to function in the electronic age, attorneys and clients must work together to ensure that both understand how and where electronic documents, records and emails are maintained and to determine how best to locate, review, and produce responsive documents.").

At the same time, courts also began explicitly challenging the reliability of traditional Boolean keyword searches. In three notable decisions beginning in 2008, the courts recognized the technical and linguistic complexities associated with accurately filtering large volumes of data, and they required the parties to support the defensibility of their selected search methodologies. William A. Gross Construction v. American Manufacturers Mutual Insurance Co., 256 F.R.D. 134 (S.D.N.Y. 2009); Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 260, 262 (D.Md. May 29, 2008); United States v. O'Keefe, 537 F.Supp.2d 14, 24 (D.D.C.2008). As Judge Facciola eloquently instructed in O'Keefe:

Whether search terms or "keywords" will yield the information sought is a complicated question in 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. This topic is clearly beyond the ken of a layman and requires that any such conclusion be based on evidence that, for example, meets the criteria of Rule 702 of the Federal Rules of Evidence."

O'Keefe, 537 F.Supp.2d at 24 (D.D.C.2008).

Similarly, Judge Grimm's opinion in Victor Stanley also serves as a general warning to attorneys who intend to rely primarily on keywords in conducting electronic discovery:

While keyword searches have long been recognized as appropriate and helpful for ESI [electronically stored information] search and retrieval, there are well-known limitations and risks associated with them, and proper selection and implementation obviously involves technical, if not scientific knowledge.

Victor Stanley, 250 F.R.D. at 19. See also William A. Gross Construction, 256 F.R.D. at 134 (“This Opinion should serve as a wake-up call to the Bar in this District about the need for careful thought, quality control, testing, and cooperation with opposing counsel in designing search terms or “keywords” to be used to produce emails or other electronically stored information (“ESI”).”)

Directly relevant to this paper, the Victor Stanley decision speaks to the particular perils of mechanically using keywords to search for materials protected by the attorney-client privilege and work product doctrine. In Victor Stanley, Defendants Creative Pipe *et al.* used a list of approximately seventy keywords to conduct a secondary search of otherwise discoverable electronic information to identify documents potentially protected by the attorney-client privilege and work product doctrine. After applying the secondary “privilege” keywords, defense counsel reviewed the titles of the responsive documents, as well as those of non-searchable documents, to determine if they were privileged or otherwise protected. If the title suggested a document might contain privileged information, the document was more closely reviewed. The remaining documents – those that were not responsive to the privilege keywords and those that were responsive but whose titles did not indicate privilege - were produced to Plaintiffs. During the course of the litigation, Plaintiffs identified at least 165 privileged text-searchable documents within the production provided by Defendants. Plaintiffs argued that Defendants had waived the privilege as to these documents, and there was no agreement or order under Federal Rule of Evidence 502 in place under which Defendants could claw back the privileged documents.

The court agreed with Plaintiffs and held that Defendants had indeed “waived any privilege or protected status” as to the 165 documents, primarily because the Defendants had not exercised reasonable care to protect the privileged status of such documents during the discovery process. Specifically, the court held that Defendants had failed to carry their burden of demonstrating that the selected privilege keywords and review methodologies were reasonable as to any of the documents. In holding that Defendants had failed to sufficiently defend their process, Judge Grimm noted that that Defendants did not identify the keywords selected, did not provide the qualifications of the person who developed the keywords, did not perform any quality-assurance testing or sample the documents that were not responsive to the privilege keywords, and did not otherwise explain what they had done and why it was reasonable.

Judge Grimm also explicitly endorsed *The Sedona Conference Best Practices Commentary on the Use of Search & Information Retrieval Methods in E-Discovery*, noting that compliance with the Sedona Conference Best Practices “will go a long way towards convincing the court that the method chosen was reasonable and reliable.” Victor Stanley, 250 F.R.D. at 26. Because those Best Practices, although not legally binding, have influenced both legal and technological thought in this area, they are set forth below:

- Practice Point 1.* In many settings involving electronically stored information, reliance solely on a manual search process for the purpose of finding responsive documents may be infeasible or unwarranted. In such cases, the use of automated search methods should be viewed as reasonable, valuable, and even necessary.
- Practice Point 2.* Success in using any automated search method or technology will be enhanced by a well-thought-out process with substantial human input on the front end.
- Practice Point 3.* The choice of a specific search and retrieval method will be highly dependent on the specific legal context in which it is to be employed.
- Practice Point 4.* Parties should perform due diligence in choosing a particular information retrieval product or service from a vendor.
- Practice Point 5.* The use of search and information retrieval tools does not guarantee that all responsive documents will be identified in large data collections, due to characteristics of human language. Moreover, differing search methods may produce differing results, subject to a measure of statistical variation inherent in the science of information retrieval.
- Practice Point 6.* Parties should make a good faith attempt to collaborate on the use of particular search and information retrieval methods, tools and protocols (including as to keywords, concepts, and other types of search parameters).
- Practice Point 7.* Parties should expect that their choice of search methodology will need to be explained, either formally or informally, in subsequent legal contexts (including in depositions, evidentiary proceedings, and trials).
- Practice Point 8.* Parties and the courts should be alert to new and evolving search and information retrieval methods.

B. The Facciola-Redgrave Framework

Against this backdrop, and informed by their own laborious experiences working on electronic discovery matters, Judge Facciola and Jonathan Redgrave have proposed a new, groundbreaking model for parties in litigation to preserve claims of privilege without having to perform document-by-document review and logging. The Facciola-Redgrave Framework builds upon two of the primary themes running through the legal evolution described above. First, it mandates *early and regular cooperation and agreement* between the parties regarding an appropriate process for identifying potentially privileged materials. Second, it requires the parties to agree to *a defensible system to identify and classify potentially privileged documents* based on the relative likelihoods of their being protected by privilege. Based upon those classifications, the parties may then choose to automatically exclude certain categories of potentially privileged documents from production, sample other categories, or conduct a document-by-document review; the Framework also calls for the parties to agree to a system to handle challenges, which may including sampling documents for *in camera* review. If the system is defensible and followed by the parties, inadvertent disclosure of privileged material will not result in waiver.

The framework, in a nutshell, proposes the following sequence of events:

- (1) *Meet and Confer*: Rule 26(f) of the Federal Rules of Civil Procedure requires the parties to “meet and confer” early in the discovery process to discuss preservation of ESI, form of production, and “any issues relating to disclosure or discovery of ESI.” The Facciola-Redgrave Framework requires the parties to discuss issues relating to privilege claims at the “meet and confer” conferences. The parties should discuss, among other things, the types of privilege claims that may be asserted and the process by which such claims will be identified and adjudicated, as well as how inadvertent disclosure and waiver will be handled. The parties’ agreement on these subjects will be incorporated into a court order pursuant to Federal Rule of Evidence 502.
- (2) *Classification and Identification*: Pursuant to Rule 26(f), the parties also must meet and confer regarding the classification of potentially privileged documents. Such discussions should focus on and result in --
 - a. Agreement as to search terms, concepts, or “other methodolog(ies)” which will be used to identify and segregate potentially privileged documents and ESI.
 - b. Agreement as to which documents and ESI may be excluded from production and privilege logs or other indexes due to the “high likelihood” that they are privileged or otherwise protected.

- c. Agreement as to what type of information will be supplied with respect to the various classes of withheld documents. Essentially, this body of information will be used to describe the number and types of documents, including, for example, authors and/or recipients, to understand the basis for potential claims of privilege without having to conduct “document-by-document” reviews.
- d. Agreement as to categories for classifying various documents based on any number of criteria including, for example, their subject matter, date range, document type, or author, sender, or recipient. The object of creating such categories is “to create a set of natural differentiations among documents so the parties can say, once again with confidence, what is true of items within the category is true of the whole.” Notably, documents may reside in more than one category.
- e. Agreement as to what categories of documents should be indexed and what fields to include in such an index. The authors of the Framework note that such an index is not intended to be a privilege log but, rather, “it is a catalog of information regarding withheld documents” “likely to be sufficient to justify many privilege claims.”
- f. Agreement as to which documents, *if any*, should be logged pursuant to document-by-document privilege log standards.

The Framework also includes requirements for the identification and indexing of email strings, email attachments, and duplicate and near duplicate documents. Generally, these relationships should be reflected in the indexes and logs. Such steps ensure that the parties are following the same standards for searching and identifying protected documents and ESI.

- (3) *Court Order*: All agreements reached by the parties will be memorialized in a Court Order. The Order should make clear that (i) neither information shared in the meet and confers, nor actions taken by the parties to meet their obligations, will operate as a waiver of the attorney-client privilege, work product doctrine, or any other agreed-to protections; and (ii) intentional or inadvertent inclusion of privileged information in the indexes and logs will not result in waiver. The Framework authors also “advise” parties to include a specific “clawback” protocol in the event of inadvertent disclosure of a protected document.

The Facciola-Redgrave Framework also proposes protocols for handling the inevitable challenges that will arise. The Framework foresees at least two types of challenges: (i) challenges to the premise that documents within a certain category are privileged; and (ii) challenges to the status of the documents within a category. Although the Framework leaves it to the discretion of the court to determine the best way to handle challenges, the Framework authors propose a sampling process, rather than a document-by-document *in camera* review, to test

the validity of privilege claims within categories. Based on this sampling, the court can then determine whether to rule on privilege challenges on a document-by-document basis, by category, or by subsets of documents within a category according to distinguishing attributes.

Significantly, in proposing a sampling process for *in camera* review Judge Facciola's and Mr. Redgrave's skepticism of the traditional document-by-document approach is unmistakable:

[W]e note that the lawyers' notion that only document-by-document review will suffice is *flatly wrong*. [emphasis added] Studies have established that manual document-by-document review alone may be one of the poorest ways to find what one is looking for in a large data set. Accordingly, when we propose a different method, we have no concern that we are displacing a system that already works well.

The Facciola-Redgrave Framework, 2009 Fed. Cts. L. Rev. at 51-52. That is, the Facciola-Redgrave Framework clearly recognizes that the current mechanisms used in eDiscovery are broken, and it signals that parties must begin to adopt new, defensible paradigms based in part on advancements in search and review technologies and knowledge.

C. Stroz Friedberg's Intelligent Privilege Filter

Stroz Friedberg's Intelligent Privilege Filter (the "Privilege Filter") is remarkably compatible with the practical and conceptual ideas described by Judge Facciola and Jonathan Redgrave. Consistent with the Facciola-Redgrave Framework, the Privilege Filter uses two advanced search technologies to identify and classify potentially privileged ESI based on various attributes and then ranks such classifications based on the relative strengths of the attributes used to identify such documents as privileged. Specifically, consistent with the requirements of Step 2 of the Framework, Classification and Identification, the Privilege Filter provides a powerful tool for litigators to agree on a defensible methodology (a) to identify and segregate potentially privileged documents from the larger document set; (b) to aid in deciding which categories of documents can be automatically excluded from review because of their "high likelihood" of containing privileged documents; (c) to provide detailed information about the various categories, subcategories, and individual documents within those categories for easy sampling; (d) "to create a set of natural differentiations among documents so the parties can say, once again with confidence, what is true of items within the category is true of the whole"; and (e) to automatically export relevant information into spreadsheets to be used as an index or privilege log.

Although ideal for meeting the requirements of the Facciola-Redgrave Framework, the Privilege Filter also results in enormous cost and time efficiencies during traditional privilege reviews and provides an audit trail and defensible methodology if challenges do arise. In addition, it provides a means for courts and appointed third parties to perform or adopt a neutral methodology to expedite the review process in a range of situations.

1. Advanced Search Technologies

Stroz Friedberg's Intelligent Privilege Filter identifies and classifies potentially privileged documents using two text-analytic technologies which are far more sophisticated than applying Boolean keyword searches. First, documents are processed using Stroz Friedberg's proprietary "Stroz Extract" technology to extract key information about a document, including, for example, its authors, recipients, title, and document type from the text of the document. Second, key terms are located within the document and evaluated based on their placement within the text of the document and the surrounding context. Each of these advanced search technologies is discussed more fully below.

Stroz Friedberg's proprietary targeted data extraction program, Stroz Extract, is unique to the Stroz Friedberg eDiscovery solution. It resolves many problems resulting from inaccurate or unavailable metadata commonly relied upon when querying a typical database during the review process. The tool analyzes the face of an electronic file or scanned document by targeting language patterns and specific areas of the document to programmatically determine key information about the document, including the document type, author, date, recipient, cc, bcc, and title. It then extracts this objective bibliographic information from the document and imports it into fields within a searchable database. Based on our internal testing, we estimate 85% accuracy of extraction of key information across multiple document types for electronic documents, with a near 100% accuracy for emails because they are so highly structured.

Stroz Extract also can be used to extract more specific information such as proper names, email addresses, dates, keywords, job titles and other customizable descriptors. By extracting such information during the processing phase, users can efficiently query across a structured database and easily group documents based on reliable, text-based attributes and terms. Stroz Extract works for both emails and other types of electronic documents, as well as for scanned paper-based documents and ESI that do not have metadata associated with them. The following is an example of the information that would be extracted from the text of a document using Stroz Extract, as viewed on Stroz Friedberg's review platform:

DOCDATE INFO
 DOCDATE
 20160390
 T_DOCDATE
 20110620
DOCTYPE INFO
 DOCTYPE
 EDJC
 T_DOCTYPE
 correspondence
TDE SUBJECTIVE CODING
 T_STOCK_TRANSFER
 [STOCK_TRANSFER]non_contrac@Kenneth L. Lay Chairman of th...Corp. to Com
TITLE INFO
 EXH089-00927.PDF
 T_TITLE
 Sale of 76,491 shares of Enton Corp. to Company for Repayment of \$4,000,000 L
AUTHORS
 T_AUTHORNAME
 T_AUTHORORG
 Enton Corp.
RECIPIENTS
 T_RECIPIENTNAME
 Emery, Rocky
 T_RECIPIENTORG
 PaineWebber, Inc.
 Private Client Group
CCs
 T_CCNAME
 Butler, Pam
 Covert, Joanne
 Gibson, Sherrin
 Dwyer, J. Elaine

Keneth L. Lay
 Enton Corp.
 P.O. Box 1288
 Houston, TX 77255-1188
 281-491-8779
 Fax: 281-491-8722
 keneth.lay@enton.com

May 30, 2011


Rocky Emery
 Paine Webber, Inc.
 Private Client Group
 108 N. Post Oak Lane, Suite 150
 Houston, TX 77024

RE: Sale of 76,491 shares of Enton Corp. to Company for Repayment of \$4,000,000 LDC plus accrued interest through May 23, 2011

Dear Rocky:

This letter is your authority to transfer the remaining 6,892 shares of Enton Corp. stock from the stock option award, Grant #3454, with an exercise date of 2/16/08, 58,587 shares of Enton Corp. stock from the stock option award, Grant #3718, with an exercise date of 3/13/08, and 12,801 shares of Enton Corp. stock from the stock option award, Grant #3112, with an exercise date of 1/29/08, to Paine Webber Asset Management for the Enton Corp. Omnibus Paine Webber Asset Mgr-54889-EN by or before Friday, June 1, 2011.

Please call Sherrin Gibson at 713-670-0103 should you have any questions.

Sincerely,


CC: Elaine Covert
 Pam Butler
 Joanne Covert
 Sherrin Gibson

5-25-01
 175491 Shares

Example 1: Stroz Targeted Data Extraction: Information including the document type, title, authors, and recipients of the document is extracted from the text of the document at right and is provided in an easily searchable form in the panel at left.

The second relevant advanced technology utilized in the Privilege Filter is Stroz Friedberg’s logical rules-based search methodology, Stroz Elevate. This technology automates the review process by applying “rules” that embody the logic that would be used by reviewers to manually classify documents into subjective issues, rather than simply relying on traditional Boolean search strings. In general, the search rules are created by Stroz’s linguistic logicians working in cooperation with a person, usually a lawyer, who is knowledgeable about the goals of the particular filtering project.

The utilization of logical rules, rather than simple Boolean search strings, is a powerful method for efficiently classifying documents by concepts, themes, and actions, including those attributes associated with privileged documents. Such rules can be crafted to take into account relational concepts and can be customized to distinguish the use of words in different contexts. For example, a rule can be constructed such that the term “regards” within the substantive text of a document is distinguished from the use of that word in a valediction at the end of a letter such as in “Kind regards” or “Warm regards.” Thus, for example, a rule can be constructed such that the placement of the term within the document and the surrounding language will be used to identify only those documents where the term is not being used in the valediction.

Other rules can be constructed to hone in on certain types of documents or correspondence from specific authors or recipients, based on the structure and particular traits of the document. For example, we have constructed our privilege filter rules to distinguish between non-public and public documents, such as newsletters, case decisions, and articles, to evaluate the context within which the label “Attorney-Client Privilege” appears. Or, a user may choose, for example, to review only those documents containing the term “shares” that are emails sent from John Doe to fewer than three recipients, and/or to exclude spreadsheets or other routine documents discussing “shares.”

Thus, logical rules provide for more nuanced and focused searches, and can reduce, in the first instance, the need to review large volumes of irrelevant documents. In addition, the tool provides an audit trail clearly showing which portions of a document were responsive to a particular rule or rules. By reviewing the results of applying a particular rule or set of rules, those rules can be revised or refined for additional accuracy.

2. Identification and Classification of Potentially Privileged Documents

Stroz Friedberg uses both Stroz Extract, our targeted data extraction technology, and Stroz Elevate, our rules-based search methodology, to identify and classify documents potentially protected by the attorney-client privilege and work product doctrine. Based on user-supplied search terms, as discussed more fully below, the Privilege Filter extracts key reference information from the document and then uses both that information and our internally-developed rule set to analyze and classify the identified documents into various categories based on different attributes associated with privileged documents. The rule set used in our Privilege Filter has been tested in real-case scenarios and will continue to be tested for each project to determine whether the logical rules need to be adjusted or supplemented to account for all potential bases for identifying privileged documents.

The Privilege Filter can be run against a document set either before an initial relevance filter is run or afterwards on the smaller subset of responsive data. Because of the sensitivities inherent in producing privileged documents, the Filter is purposely designed to be over-inclusive in identifying potentially privileged documents; that is, all documents responsive to the broad rules-based search criteria are segregated for closer scrutiny. As designed, the filter is not intended to replace manual review; rather, it is intended to supplement and streamline the review process by identifying potentially privileged documents in the first instance, classifying them by various attributes, and highlighting those sections of the documents that indicate possible privilege. The classifications correlate, to an extent, with the likelihood that a particular group will be privileged. In addition, subsets within each of the classifications exist to further

measure the relative likelihood that the documents will be privileged. Indeed, those subcategories include explicit designations as either “strong” or “weak” based on the context in which a particular term or attribute is found.

Although Stroz’s Intelligent Privilege Filter can be used to make document-by-document privilege review far more efficient and reliable, its design is strikingly well-suited as a methodology for classifying, reviewing, and producing potentially privileged documents as described in the Facciola-Redgrave Framework. The following description of its practical application makes this point abundantly clear.

i. User-Created Search Term List

At the beginning of every new project, the user is asked to supply a list of specific categories of key terms to be imported into the Privilege Filter. In particular, the user provides three categories of information: (1) the identities of inside counsel relevant to the matter (or a broader group of inside counsel to protect against loss of privileged information relating to other matters); (2) the identities of outside counsel relevant to the matter; and (3) the identities of non-lawyers whose inclusion on protected materials would be expected and would not otherwise result in waiver (hereinafter the “Client” list). Each of these categories of key people can be identified by proper name, law firm, or organizational name, email address, and/or domain.

With the exception of the “Client” list, we have found that lawyers regularly use keyword lists comprised of attorneys’ names and email addresses to perform privilege searches. By including an optional “Client” list, the filter enables users to more easily target correspondence between attorneys and key business people. It also aids reviewers in determining whether previously unidentified authors or recipients are engaged in sensitive privileged discussions, as well as whether the involvement of extraneous authors or recipients arguably results in waiver of privilege as to a particular document.

ii. Categories of Potentially Privileged Documents

The key term list provided by the user is imported into our internally developed “privilege” rule sets and the data is processed and filtered as described earlier. The responsive data is automatically classified and foldered into the following four major categories:

1. Named Author **and** Named Recipient (where at least one of the Named parties is a Counsel)
2. Counsel is an Author **or** a Recipient
3. Counsel is mentioned either implicitly or explicitly in text

4. Contains Attorney Client Privilege / Work Product Label

Documents falling into particular categories are more likely to be privileged than documents in other categories. For example, we believe the “Named Author and Named Recipient” identifier is a strong indicator of privilege. Based on the results of applying the Privilege Filter to a real case data set that had been manually reviewed for privilege, as discussed more fully below, we found close correlation in results between manual review designations and the documents identified by the Filter as falling into this first category. Thus, for example, under the Facciola-Redford Framework, the parties could agree to deem certain subcategories of documents falling into category 1 as presumptively privileged for purposes of production, or they could subject such documents to limited sampling for quality control purposes.

Moreover, within these four major categories, the Privilege Filter also automatically classifies the documents into subcategories for even more refined decisions as to relative strength of privilege indicators. For example, within the first category the Filter identifies whether each of the Named Author and Named Recipients are Inside Counsel, Outside Counsel, or Clients. Furthermore, there is a particular category comprised only of documents between Inside Counsel and Outside Counsel. This subcategory of documents is likely to contain highly privileged correspondence, compared to some of the other subcategories of documents.

The filter also specifically identifies and highlights within the text of the document all “extraneous” authors or recipients in each document within categories 1 and 2; that is, the tool will provide information about all authors and/or recipients who are copied on a document but who do not appear on any of the user-supplied key term lists. Segregating such documents provides a means for a reviewer to quickly determine whether the inclusion of extraneous people may result in waiver of the privilege. Including information about extraneous people not only helps to prioritize documents within subcategories based on the likeliness of their being privileged, it also provides a means for users to determine whether the inclusion of specific extraneous authors or recipients on a particular document is noteworthy to the case.

The subcategories under the third and fourth major categories, above, are based on subtle refinements by utilizing the power of rules-based text analytics. For example, we differentiate between explicit mentions of a named counsel in the text of the document and implicit, or generic, references to an attorney or law department. An example of a document identified by the filter as containing an implicit mention of counsel within the text of a document follows:⁴

⁴ For demonstration purposes, we applied the Stroz Discovery Intelligent Privilege Filter to the publicly available data associated with the Enron litigation. Inside Counsel, Outside Counsel, and Client Lists were developed simply for the purposes of testing the filter, and not based on full listings of the relevant people involved in the case.

>
From: michelle.cash@enron.com
Sent: Monday, August 27, 2001 10:18 AM
To: teobaldo.camejo@enron.com; roberto.deleon@enron.com
Cc: coralina.rivera@enron.com
Subject: PR Internet Issue

I have had some research done on the issue of what was found on the computers in the Ecoelectrica offices. It looks like the best course of action is to destroy all of the identified materials. Mere possession of these materials is a violation of law. Our outside counsel called the FBI, which said that if the company does not know who put the materials out there, it does not have a duty to report it.

We also received advice that the Puerto Rico office of the FBI might be able to assist in narrowing down the possible users to determine who actually was involved. Our outside counsel suggested that we might want to call the local FBI office and talk with a duty agent. The number is 787-754-6000. Given the situation of the possible sale, however, we may choose not to take that route.

In any event, if we don't report it, all illegal materials must be destroyed and removed from our systems as soon as possible.

Michelle

Example 2: Implicit Reference to Counsel

Not only is this document a striking example of the type of subtle reference to attorney communications that will be identified by the Privilege Filter, it also highlights the strengths of using rules-based searches to detect relevant text where typical Boolean searches would fail to do so. Moreover, these “implicit” references to counsel are further distinguished as “strong” or “weak” indicators of privilege based on the context in which they appear. That is, we distinguish those references that have a high likelihood of relating to an attorney-client communication from those that have a lower likelihood. Thus, the example document above would be specifically marked as having a “strong” likelihood of being privileged because of the logical and textual context within which the phrase “counsel suggested” appears.

Similarly, category 4, “Contains Attorney Client Privilege / Work Product Label” contains subcategories of documents labeled as having a “strong” or “weak” likelihood of privilege based on where such labels appear, the surrounding text, and the general nature and type of document. One such example of a document identified as having a “strong” likelihood of being privileged based on its labeling is as follows:

>
From: michelle.cash@enron.com
Sent: Thursday, August 10, 2000 8:00 PM
To: hoyt.thomas@enron.com
Cc: melissa.laing@enron.com
Subject: Re: Privileged and Confidential communication to my attorney > Re:
Henry Bath LA

< Privileged and Confidential >
< This email contains legal advice >

Hoyt and Melissa,

I am assuming that we are discussing contract personnel who provide some sort of services for MG/RW. It looks like there may be some key personnel we want to employ, but that there are many whom we don't want to hire. I am not really familiar with the specific facts here.

The question of a joint employer arises when a company treats its contract

Example 3: Contains Attorney-Client Privilege / Work Product Label

Thus, the tool provides users with a preliminary system for ranking various types of documents for purposes of determining which categories of documents can simply be sampled for accuracy, and which categories should undergo a more intensive document-by-document review process. It is precisely this kind of rational, testable categorization of documents that is envisioned in the Facciola-Redgrave Framework.

iii. Preliminary Testing Results

Since its introduction to the market in the fall of 2009, Stroz Friedberg has applied the Intelligent Review Filter to discovery data sets that have already been manually reviewed for privilege. Preliminary results on active cases are extremely promising, and we look forward to further testing and refinement of the filter.

In an initial sample set of 22,032 documents, manual reviewers using traditional review approaches determined that 13,564 of the documents were either privileged or had some basis for privilege requiring a secondary review; manual reviewers determined that the other 8,468 documents had no basis for privilege.⁵

⁵ It is important to note that during the manual review, attorneys in that case followed the following convention: upon determining that one document in a family of documents, i.e., an email and its attachments, was privileged, then every document in that family was automatically

Applying Stroz Friedberg's Intelligent Privilege Filter to the same set of data, the Privilege Filter automatically identified 95% of the documents that the manual reviewers had determined were either privileged or had to be sent for secondary review for a final determination. Of the 713 documents deemed privileged pursuant to manual review, but not identified by the Privilege Filter as being privileged, it appears that a large number of those documents included: general business documents containing confidential information or strategies that were neither marked as privileged nor otherwise identifiable from their face as containing attorney advice; general legal forms and contracts without labels; or correspondence with lawyers who were not identified as being relevant lawyers in the keyword list.

In total, the Privilege Filter identified 18,825 documents of the 22,032 documents as being potentially privileged; specifically, the Filter identified an additional 3,974 distinct documents as potentially privileged that manual reviewers had determined had no basis for privilege. Interestingly, of those documents, over 3,000 of them were identified as potentially privileged by the Filter because of some explicit or implicit mention of counsel in the actual content of the document. Thus, the Privilege Filter is identifying precisely those documents that are often hardest to detect through traditional key word searches as containing potentially privileged information. The sample data also suggests that there is a very high correlation between those documents identified as potentially privileged through manual review, and those identified as potentially privileged by the Privilege Filter because they contain attributes classifying them as falling into category 1, Named Author and Recipient, and category 4, Contains Attorney-Client Privilege/Work Product Label.

Although these initial results suggest the Privilege Filter is performing impressively in identifying potential privileged documents, we will continue to analyze the efficacy of the filter and make refinements as necessary as we continue to apply it in other cases.

3. Reporting Requirements: Indexes and Privilege Logs

Although this paper is intended to focus primarily on the conceptual parallels between the Stroz Friedberg Intelligent Privilege Filter and the Facciola-Redgrave Framework, it is worth noting that the Privilege Filter also automatically generates many of the reports envisioned in the Framework. All of the documents imported into Stroz Friedberg's review platform, Stroz Review, are assigned a unique identifying number for tracking. Documents categorized in accordance with the Privilege Filter's classification systems can be foldered and subfoldered in accordance with users' customized specifications.

marked as privileged. For purposes of comparing apples-to-apples, we followed the same convention using the Privilege Filter.

In addition, users can specify folders and subfolders for purposes of tagging documents during review. Stroz Review permits customized tagging, including bulk, email thread, and document family tagging, and escalation to secondary review. The tagging function also enables the reviewer to quickly concur with, override, or modify the Filter's automated designations.

The review tool also highlights the text sections of each document that are responsive to the Privilege Filter, creating a full audit trail for review. Indeed, Examples 2 and 3, above, are screen shots from our review platform depicting the renderings of documents identified by the Privilege Filter as potentially privileged. Similarly, the names of all key players, labels, and any other attributes contributing to a document being designated as potentially privileged will be highlighted; moreover, those attributes are searchable within the document set. Such classification and highlighting further contribute to efficient, cost-effective review.

Finally, all information pertaining to the document, including attributes contributing to its being identified as potentially privileged, bibliographic information extracted through Stroz Extract, tagging information, and user comments can be exported directly into spreadsheets that can be used as a privilege log, or a practical index during the meet & confer processes as described in the Framework. All such information can be excluded, included, or modified in accordance with user specifications.

Conclusion

The legal, business, and technical landscape in which electronic discovery is performed is undergoing dramatic and fundamental changes because of the large volume of information routinely stored by all types of enterprises. Concerned with costs, reliability, and resources, all parties involved in this process are seeking legal and technical solutions that radically reform the way in which litigation is conducted. Sensing these seismic shifts, thought-leaders from a variety of fields and perspectives are coming together to discuss ways in which we can coordinate legal and technological advancements in search and review methodologies to develop comprehensive, compatible approaches.

One striking example of how the legal and technical developments have intersected, albeit on distinct parallel tracks, is the almost simultaneous publication of the Facciola-Redgrave Framework for conducting privilege reviews and the introduction by Stroz Friedberg of the industry-first Intelligent Privilege Filter. Indeed, the Privilege Filter provides the technical means necessary for practitioners and courts to promptly and confidently adopt and apply the methodologies described by the authors of the Framework.