



LSE v. SEARCH

Arete Legal's cutting-edge technology can help you solve the problem of managing large volumes of potentially relevant content for e-discovery and investigations.

TABLE OF CONTENTS

Subject	Page
I. Introduction	3
II. Business Challenge.....	3
III. Solution Description.....	4
IV. Solution Benefits	5
V. Technical Specification	5
VI. Summary.....	6
 Addenda	
A. More Information	8
B. Sources	8
C. About Areté Legal.....	8

I. Introduction

Discovery of electronic data is a front-burner issue for litigants and the courts alike. Since the 1970 amendments to the Federal Rules of Civil Procedure, the “growth in electronically stored information and in the variety of systems for creating and storing such information has been dramatic.”¹ For over a decade the problem of how to handle huge volumes of potentially relevant electronically stored information (ESI) has exploded.

To address this explosion, a specialized group of Ph.D. linguists and attorneys within Arété Legal perform research and direct the implementation of products using advanced linguistic techniques. They have produced a proprietary methodology and software system based on computational and corpus-based linguistic science, and they apply this approach, with highly automated computer agents, to the electronic discovery and litigation process. We call this technology Linguistic Systems Engineering (LSE). It is based on a scientific approach to, and study of, language as it has developed over the past century.

The focus of LSE research and development is on the three primary obstacles faced by our clients (1) that the appropriate documents are collected; (2) that the appropriate documents are eliminated; and (3) that efficient legal review processes are affected by directing the reviewer to passages of interest.

Our investigational tool LinguaLexSM is built upon the foundation of LSE technology. This system allows investigators to use their precious human resources in the most effective ways. LinguaLexSM leverages advanced technology to gain deep insight faster, cheaper, more thoroughly and more accurately than traditional methods. LSE is also utilized in our ESI-Review system to identify documents by type for redirection to specialized reviewers. Additionally our ESI-Map system of document collection and preservation makes use of LingualexSM for data reduction.

The effect of LSE implementation is to eliminate the most striking failures of key word searching: the over inclusion of irrelevant documents and exclusion of relevant documents.

II. Business Challenge

Common methods do not effectively solve problems encountered in the discovery process. Keyword searching is being used to tackle the huge volume of discoverable ESI. Courts have accepted this method, but have not universally endorsed it as the preferred method and have been open to any new or better “process” that could accomplish the same task.² In fact, the courts have acknowledged the inherent inefficiencies in keyword searching, including the “unavoidable trial-and-error” nature of determining lists of terms and phrases.³ In a recent 2007 decision, the court did not require the use of a keyword list, but rather a “process” to identify relevant documents and avoid the burden of sorting and producing “truckloads” of irrelevant data.⁴ The absence of a scientifically-based

process reduces keyword searching to educated guesswork. Its application to discovery results has a tendency to include irrelevant documents and to exclude relevant ones.

High volume precludes human evaluation of each document because of time and cost constraints. A lack of language knowledge renders keyword-based searches ineffective because the client is unable to produce a reliable search term list to locate the desired topics of interest.

The Federal Rules of Civil Procedure, amended on December 1, 2006, addressed the growing problem of electronic discovery in part but did not address the more difficult issues related to the specifics of identifying relevant material.

III. Solution Description

Although keyword searching is the most common form of e-data recovery, interpretive cases and recent commentary suggest that it is not the only acceptable method.⁵ Thus far, the courts have neither approved nor established a uniformly acceptable method for identifying relevant electronic data for production. For example, the Sedona Principles allow for other methods by emphasizing consistency, “regardless of the method chosen.” As the Sedona Conference noted:

Automated collection involves using computerized processes to collect data meeting certain criteria, such as search terms, file and message dates, or folder locations. Automated collection can be integrated with an overall electronic data archiving or retention system, or it can be implemented using agents specifically designed to retrieve information on a case-by-case basis. *Regardless of the method chosen, consistency across the production can help ensure that responsive documents have been produced as appropriate* [emphasis added].⁶

Until now, no one has offered a superior, scientifically based alternative to keyword searching. Arété Legal’s Linguistics System Engineering (LSE) is such a process: it effectively reduces the discovery burden by improving the accuracy in determining relevance.

To remedy these problems, LSE automated methods allow effective forensic examination of large document collections. This is accomplished without the necessity for manual examination of each document, and without extensive *a priori* knowledge of the matter or insight into the language used to discuss the case. By examining document collections at various levels, from word-level syntactic and semantic (meaning) analysis, to document-level evaluations of pragmatics (context), to collection-level establishment of norms, LSE methods provide for the discovery of focused content and criteria for use in overcoming the three primary obstacles. This is done through a unique, proprietary combination of proven methods from the fields of Corpus Linguistics and Statistics.

LSE can be applied to case preparation, discovery, and investigations for plaintiff firms, in-house legal departments, defense counsel, or government agencies. In support of

document collection and review, LSE is a powerful tool for eliminating spurious documents and data from the collection, for ensuring the correct material is considered, and for identifying within documents relevant, focused passages of interest. Use of a superior method of finding information in a large collection of documents enables a client to investigate rapidly and thoroughly, to prepare its own case, and to oppose that of its opponent more effectively.

IV. Solution Benefits

Unlike simple keyword searching, LSE provides visibility into the actual words of the document collection, as well as the context of the words, affording greater insight into the relevance of the data reviewed. Accounting for synonyms, misspellings and context, in addition to exclusionary criteria, LSE allows the documents themselves—and the words, ideas, and meanings expressed within them—to dictate relevance. The result is a more accurate and comprehensive recovery of relevant electronic files. The linguistic approach achieves everything that keyword searching achieves, and more.

Moreover, even though it has not yet received the widespread attention of traditional keyword search, the LSE method has been applied successfully in litigation settings. To date, Arété Legal has applied its proprietary linguistic science method to large-scale electronic discovery programs for Fortune 500 companies in a variety of civil and criminal litigation. It is also being applied in pending litigation for Arété Legal’s clients.

V. Technical Specification

Fundamental Differences between Search Engines and LSE Evaluations

The proprietary LSE evaluation engine has several fundamental differences compared to traditional search engines, whether on the Internet or otherwise. In many ways, this compares apples to oranges: although there are apparent similarities, they perform radically different functions. The following is a brief comparison to highlight the power of the LSE evaluation engine, as opposed to the traditional search engine.

The first difference involves how LSE and search engines handle the documents that they are searching. Often, search engines only search the first part of a document. They assume that each document is a coherent, homogeneous whole. If that is true, the bottom of a document probably will not add enough information to make it worth the effort to index the entire text. If documents talk about multiple topics or have widely varying style—as email chains and business documents often do, for instance—this assumption does not hold. Because of this, LSE analyzes the entirety of every document, every word.

Search engines also remove the ends of the words in a document as they index it. For example, *taxes*, *taxed*, *taxing*, and maybe *taxation* would all become *tax*. This is useful for catching different forms of a word, but it also loses information. In LSE analyses, the endings of a word often contain important information about how the word is being used and what it means in context. A search engine also removes words from documents to be searched. Typically, most search engines remove words that are common, but which do not represent content, such as *the*, *she*, or *with*. This dramatically decreases the amount

of information search engines must store for each document. However, like word endings, LSE analysis has found these function words to be critical to determining the content and type of communication in a document. Therefore, LSE maintains information about both word endings and common words in our analysis. In fact, LSE does not utilize an index in any way, that is, words are not indexed, and they are exhaustively analyzed.

Another area where LSE differs completely from search engines is in both the number of language markers supported and the complexity of the analysis logic allowed. Generally, search engines limit each search query in both size and complexity. Although they can handle simple Boolean (*and*, *or*, and *not*), context, and query grouping they limit how many of these can be combined together. LSE's use of extensive scientifically created and validated language marker sets and complex rule-based content analyzers, both of which can be combined and grouped in an infinite number of ways, allows much more powerful analysis capability, one that is impossible to perform with a search engine. Search engines identify documents based on keywords, but LSE exhaustively analyzes text using language marker sets and inclusion/exclusion rule based logic.

The final area where LSE differs from search engines is in how the result set is determined and presented. This difference is particularly pronounced for Internet search engines, which index too many documents to search in real time. Instead, they have to present the searches found from a large subset of the total pages they index and present the results from that. The number of documents reported is actually an extrapolation.

Also, Internet search engines rank documents using a variety of methods. The primary one involves analyzing how many other documents link to each document in the result set. This works well for hypertext documents, but is largely meaningless for office documents or emails. Also, this is roughly a measurement of how “popular” a document is, and if the user wants to find documents that are not well known, it actually interferes with those results. Off-line search engines use some combination of looking at the proximity of search terms and occurrence to determine how they rank documents. Both of these assume that one method of ranking documents works for every query and every type of document.

In contrast, LSE does not make that assumption. Instead, it uses content analyzers with custom logic for every textual evaluation. Moreover, the analyzer logic has many more tools available than simply proximity or occurrence. It takes customized weighting, proportionality, and other factors into account.

Finally, LSE also provides multiple ways in which the results can be viewed. Typically, search engines will present a few words of context around the search keyword hits, along with a link to the entire document. LSE can show text lines around the content of interest, the chunk of the document that the content occurred in, or the entire document. This added flexibility makes the results much easier to review rapidly.

This brief comparison shows that LSE and search engines are very different in both their scope and approach. In fact, no overlap exists between a search engine and LSE evaluations.

VI. Summary

LSE technology is incorporated into a number of specialized legal process support applications being sold and supported by Arété Legal including ESI-Map (a system of document collection and preservation and data reduction), ESI-Review (a system for conducting and managing litigation document review and production), and LinguaLexSM (a system to support investigation, research, due diligence, and discovery).

Addenda

A. More Information

Thomas Y. Allman, *The Impact of the Proposed Federal E-Discovery Rules*, 12 RICH J.L. & TECH. 13 (2006), at <http://law.richmond.edu/jolt/v12i4/article13.pdf>.

Whitney Adams and Jeffrey Jacobs, *Ghost in the Machine: Legal Developments and Practical Advice in an Age of Electronic Discovery*, 22 No. 7 ACC DOCKET 48, 65-66 (July/August 2004) (emphasis added).

B. Citations

¹FED. R. CIV. P. 34, Advisory Committee Notes (2006).

²Medtronic Sofamor Danek, Inc. v. Michelson, No. 01-2372-MIV, 2003 WL 21468573 (W.D. Tenn. May 13, 2003); and Tulip Computers Int'l B.V. v. Dell Computer Corporation, No. CIV.A 00-981-RRM (D. Del. Apr. 30, 2002).

³Commonwealth v. Ellis, Nos. 97-192, et al., 1999 WL 815818, * 11 (Mass. Super. Aug. 27, 1999).

⁴Apsley v. Boeing Co., 2007 U.S. Dist. LEXIS 5144 (D.Kan. 2007).

⁵Medtronic Sofamor Danek, Inc. v. Michelson, No. 01-2372-MIV, 2003 WL 21468573 (W.D. Tenn. May 13, 2003). and Apsley v. Boeing Co., 2007 U.S. Dist. LEXIS 5144 (D. Kan. 2007).

⁶The Sedona Principles, Cmt. 11.c.

C. About Arété Legal

Areté Legal provides advanced, patented, fully integrated legal management systems. Our integrated strategic offerings and expertise in process engineering, enables our team to provide clients the keys to success combined with real efficiencies to manage complex business and legal matters.

From litigation hold and document collection through trial, we have the professional expertise to support and guide you through the most challenging litigation and business issues. Our significant expertise in electronic discovery and integrated case management drives development of the most advanced technologies to serve your needs.

Our expertise and our services combine to create an essential component for our clients' business and litigation success. Our patented methods enable your legal team to maintain coordination and control in a collaborative environment and to deal cohesively with massive records and collections in litigation or for other business needs. Our service offerings are extensible and flexible, providing our clients with both non-downloadable and downloadable software applications.

We have successfully supported some of the largest complex litigations in history, automating processes to save resources while ensuring best outcomes. Today we sustain hundreds of users working in collaboration around the globe, with solutions that are available only through Arété Legal.

For more information visit us at www.aretelegal.com or contact a sales representative at info@aretelegal.com.

Areté Legal
400 Lydia Street
Carnegie, PA 15106 USA
(412) 561-1900