

OLP eJournal

Summer 2012



†S The Best Kept Secrets to Using Keyword Search Technologies
Tough Negotiating for a Tough Market
Lawyers Managing Themselves?



contents

- 4** Letter from the Editor
- 6** The Press Releases: eLessons Learned
- 7** Mini-Briefs: What Is Loss Aversion
by Ralph Williams III, Esq.
- 8** Who's New? What's What?
- 10** Is eDiscovery Really the Problem? by
Linda Sharp, Esq., MBA
- 16** Lawyers Managing Themselves? Like Driving Without
a License? by Markus Hartung, Esq.
- 20** A Look Inside Hosted eDiscovery Pricing
by Megan Miller



26 Bridging the Gap in EDiscovery – The Emergence of Conceptual Semantic Search by Jeff Parkhurst

37 New OLP Chapters Now Forming

38 Zubulake’s Untold Story by Laura Zubulake

42 *Cover Story:*
Best Kept Secrets for Using Keyword Search Technologies
by Philip Sykes and Richard Finkelman

52 Tough Negotiating for Tough Times by Chere Estrin

56 A Certain Level of Trust by Charles Gillis

letter from the editor

Hello Readers.....

As someone who came to the legal field in the late '80's, I can tell you that technology then was in its early infancy. A computer was considered a threat to most lawyers, primarily because many felt it diminished their status (too secretarial) and decreased their billables. (It was, after all, supposed to save time.) That rationale didn't exactly move the field too far ahead. Let's thank the technology god, whoever she is, that those days are over.

As Editor-in-Chief of OLP's eJournal, it's become a difficult task to choose between the many stories which will run in the magazine. Perhaps that's because interest and passion has significantly emerged since the late '80's and technology has become a special interest of great importance. With each issue, I have both the agony and the pleasure of selecting between many excellent articles that readers will find informative, entertaining and far-sighted.

In this job, I've found that writers and readers alike have strong views. None are more evident than Linda Sharp's *Is eDiscovery Really the Problem?* or Jeff Parkhurst's *Bridging the Gap in EDiscovery – The Emergence of Conceptual Semantic Search*.

Megan Miller makes her views known in *A Look Inside Hosted eDiscovery Pricing* and Phil Sykes and Rich Finkelman let us know that despite our best efforts, we still don't know it all when they penned, *Best Kept Secrets for Using Keyword Search Technologies*.

Even if we step outside of technology for just a moment, Markus Hartung, chiming in from Germany, gives us his opinion on whether lawyers can manage themselves without specific and proper training. Charles Gillis, down in the great state of Texas, opines on trust, given the recent Dewey debacle. It seems everyone has an opinion.

I marvel at people who are so sure-footed on complex and important matters. My job, by definition, involves making choices on what (in the best interests of OLP members) matters and what is just going to have to wait - for now, anyway. One thing is certain: the late '80's this isn't and we're definitely never short on interesting material.

Happy Reading!

Chere Estrin
Managing Administrator and Editor-in-Chief
chere.estrin@theolp.org





OLP Announces the All New
Litigation Support
Certification Exam

Coming Late Fall 2012



eLessons Learned

www.eLLblog.com

Contact: Frank Gonnello Jr., Executive Editor

Frank@eLLblog.com

August 10, 2012

FOR IMMEDIATE RELEASE

eLessons Learned Appoints Borchert and Kiernan as Co-Editors in Chief

Cyberspace (August 10, 2012) – *eLessons Learned* – *Where Law, Technology, and Human Error Collide*, the premier eDiscovery blog written primarily by law students, announces its newest additions to the eLL Executive Team: Chris Borchert (2L at UConn Law) and Catherine Kiernan (2L at Rutgers Law-Camden).

eLessons Learned appoints Chris and Catherine to serve as Co-Editors in Chief for the coming 2012-2013 academic year, while Editor in Chief Frank Gonnello Jr. (J.D./M.B.A. 2012) will transition into a strategic role on eLL's Executive Team and serve as the blog's *Executive Editor*.

Chris is a second-year student at the University of Connecticut School of Law, where he is pursuing a certificate in Intellectual Property in addition to his J.D. Chris earned a B.A. in Political Communication from The George Washington University's School of Media and Public Affairs and wrote his thesis on the role of new media in modern American political discourse. Chris worked at Norris, McLaughlin & Marcus, P.A., where he focused primarily on matters relating to eDiscovery, cybersecurity, and data privacy.

Catherine graduated magna cum laude from Rutgers University, New Brunswick in 2010 with a B.A. in political science and expects to receive her J.D. from Rutgers School of Law – Camden in 2014. This year she will participate in the competitive Hunter Moot Court. This past summer she interned with a magistrate judge for the U.S. District Court, District of New Jersey where she contributed to a bench trial opinion, among other notable experiences. Catherine also has past experience working for a New Jersey state trial court judge, the United Nations, and as a research assistant for a prominent professor at Rutgers University's prestigious Eagleton Institute of Politics.

The new eLL editorial board is eager to continue *eLessons Learned's* tradition of providing subscribers with useful, timely information about the latest developments on how technology impacts the law. "One of our goals is to make eLLblog a destination site for the legal and technology communities to turn to for up-to-date, relevant information about technology's impact on the law," said Chris. Catherine added that "We will also work towards expanding the number of law student blog contributors and continue to improve eLL's features in a way that accommodates the needs of the increasingly fast-paced lives of our busy readers who need up-to-date information that is easy to digest and a click away."



Mini-Briefs

What is Loss Aversion?

Loss aversion describes the psychological phenomenon where people place a higher value on avoiding loss than on securing gain. This is why they buy certificates of deposit rather than invest in the stock market. The flip side of loss aversion is that once a person incurs a loss, they often engage in risk seeking behavior to recover the loss, such as putting money in a get rich quick scheme. Here's how loss aversion works, mediating a litigated case.

Plaintiffs have suffered loss and will take a disproportionately high risk in the form of large demands in an attempt to recoup the loss. However, once engaged in negotiations, when hard offers, perceived as fair or "in the ball park" are on the table, the risk continuum shifts to retaining the gain achieved and avoiding the loss of a settlement.

Defendants and insurance companies incur loss with every dollar offered; hoping to avoid further loss, they offer little. When negotiation proceeds to "the ball park", holding on to the gain, i.e. avoiding loss of a settlement, shifts the risk continuum to closing the negotiation.

The loss aversion process explains why parties start far apart, move to the "ball park" and then settle the case. It also illustrates that patience is a negotiating virtue and that rushing to the "bottom line" rarely works.

Ralph O. Williams III
ADR Services
rwilliams@adrservices.org



Who's New? What's What?

Welcome New Board of Governor Members

OLP takes great pride in announcing three new members to our prestigious Board of Governors. Please join us in welcoming:

Fernando M. Pinguelo, a U.S.-based trial lawyer, devotes his practice to complex lawsuits with an emphasis on business disputes, cyber security, media and employment matters.



A former prosecutor, Fernando is experienced in all facets of litigation (trial, mediation, arbitration, and appellate) in both federal and state courts. A leader in the emerging areas of cyber law and electronic discovery, Fernando works closely with clients to develop strategies for managing business and legal issues relating to electronic data. He recognizes the financial and public relations fallout that often results from high-profile data breaches; and, when a rapidly evolving crisis or emergency hits, Fernando mobilizes his inter-disciplinary team of professionals and implements a response strategy to mitigate damages. Fernando's active involvement with Meritas, a worldwide affiliation of independent law firms across the U.S. and in 70 other countries, enables him to provide clients with services outside his jurisdictional boundaries.

Recently, Fernando provided commentary on proposed U.S. federal legislation concerning cyber security, and received a Fulbright Specialist appointment for his work in electronic discovery. Fernando serves on federal and state court committees where he contributes to court rule amendments. He also lectured



Dave Tiller is the CEO of Phoenix-based Studeo Legal. His extensive 25+ years in litigation support and eDiscovery has given him an expertise in all facets of complex litigation in law firms and in-house legal departments. Previously, Dave held executive positions with Intel, Bowne and IPRO. Dave is involved with Hybrid Auto-coding, offshore and domestic coding, and leading Senior Project Managers and Review Attorneys to assist litigation teams. He has become an expert in Social Media monitoring of juries and hyperlinked eBrief services.

Dave is a co-founding member of OLP's Phoenix chapter, a former member of OLP's Advisory Council and President of the Phoenix Rotary Club



George I. Rudoy is a veteran of Legal Technology & Services industry. His consultancy (Integrated Legal Technology, LLC) provides a broad spectrum of advice on a variety of legal technology and services functions, including international and US eDiscovery, information governance, compliance and litigation readiness for the corporations and integration and consolidation of legal technology and support services for the law firms. Prior to forming his consultancy, George held senior management positions at a number of AmLaw 100 law firms, including Shearman & Sterling LLP and Cravath, Swaine & Moore LLP.

Mr. Rudoy has extensive expertise in development and support of information governance, compliance and litigation readiness models for companies of all sizes. In addition, George is assisting his corporate clients with document preservation and collection efforts worldwide, including serving as 30(b)(6) witness based on his advice, institutional knowledge and relevant information to adequately respond to the questions and categories identified in notices.

Mr. Rudoy graduated from New York University (magna cum laude). He is a seasoned veteran in the international legal support arena and a sought after speaker at legal conferences worldwide. George is a founding member of the eDiscovery Training Academy and Advanced eDiscovery Institute at Georgetown Law Center where he is teaching on a variety of topics. He periodically writes on a variety of legal tech topics for the main industry publications.

Welcome New Advisory Council Members



Marcus Ledergerber is an attorney and technologist with 15+ years of litigation experience at three top Am Law 100 firms. Currently, he handles electronic discovery projects and client service for Morgan, Lewis & Bockius LLP in New York, NY. Mr. Ledergerber completed digital forensics training from Guidance Software and he is an EnCase Certified eDiscovery Practitioner (EnCEP).

As an active member of the bar in New York and New Jersey, his knowledge of best practices for defensible handling of electronically stored information (ESI) spans the gamut of the Electronic Discovery Reference Model (EDRM), includes optimized methods to identify, preserve, collect, process, review, and produce ESI. Marcus is proficient in electronic discovery operations handled exclusively in-house. He has helped to achieve growth through a strategy that maximally leverages firm-wide collaboration, process improvement, and client service.

Linda Alele is Vice President of Client Relations of eStet and member of the company's Defensibility Consulting Practice where she offers strategic advice to clients facing the challenges of discovery and regulatory issues.



Prior to e-Stet, Linda was the Vice President at IMEF International Petroleum, a privately held oil and gas trading and investment company, where she managed the company's operations and consulted on national and regional regulatory issues, tribal concerns, e-discovery and data management strategies.

Alele was the firm-wide Manager of Practice Support at O'Melveny & Myers LLP, where she was a founding member of the Electronic Discovery Task Force. She spearheaded multi-million dollar e-discovery projects for high profile clients and consulted on e-discovery best practices, practical cost-cutting measures throughout the e-discovery lifecycle, and leveraged technologies across multiple law offices.



Is eDiscovery Really the Problem?

Or is it the **foundation** of our data management strategy or **lack** thereof!

By Linda Sharp, Esq., MBA

Did you know that the Leaning Tower of Pisa began to lean during the construction process? The architect and engineer had unknowingly built the tower on sand.

Yet, the builder continued adding floors, making one side taller than the other, in an attempt to remedy the problem. He didn't realize that his attempts were futile, as with each floor, the taller side was adding more weight, thus causing the tower to lean even further. He had spent too much money and had too much at stake to turn back. Perhaps we have the same problem in the world of eDiscovery.

Let's roll up our sleeves and understand the real problem.

Recently, you have probably read many articles that claim to provide insight into the benefits of "predictive coding", "automated document classification" or other technological assisted review, depending on the service provider that is sponsoring the article. In reality, the factors that are driving the costs of eDiscovery, is the manner in which information is managed and stored.

Let's dive in! There are largely two types of data stores:

1) Structured Data - which consists largely of databases that store information for HR, insurance policies, health care information, accounting data and many other data types. The bottom line is that these are largely stores of databases. You can efficiently run a search and derive the information that you seek.

These data sets contain only legitimate business information. That copy of your medical records, the paragraphs that once derived created your policy of insurance, or that information that is needed to populate and export your quarterly report, etc. The information is maintained in a fielded environment that allows for ease of searching. The search will render a very accurate and timely result.

2) Unstructured Data – this set of data has largely non-fielded information, has varying data types from social media, Blackberry devices, Bloomberg information, emails, Word, PPT, Excel, and the list goes on. It has no consistent structure from one application to the next. Thus, the name, "unstructured data". It is this group of data that provides most of the information that is now contained in what the industry is calling "Big Data".

Since structured data can be easily obtained by query of the database, this article is designed to target the obstacles of collecting information from unstructured data. This is where the issues for ESI largely arise. We're talking largely about those emails and working files that you are trying to get your hands on in some kind of a meaningful way, so that the information can be reviewed and ultimately pro-

duced.

Too much data

We all have read about the staggering amounts of information that are being created every day and how those numbers are growing exponentially. We need to understand why there is such a problem in getting to this information.

Let's take a look at email since it is the most commonly sought after group of information, as well as is the largest volume of data sought. I suggest that you take a look at your own mailbox. How much information is contained in your mailbox that has absolutely nothing to do with the business of the organization? Historically, you might expect to see in the area of over 50% to be non-corporate communications.

Additionally, corporations send out business communications regarding the day to day operations of the organization, those reminders of your annual enrollment for insurance, the holiday party, etc.

Further, most individuals do not work on the same project for their entire career at any given organization. Thus, of the 50% which represents corporate communications, what percentage is actually at issue in your matter? Depending on the individual, it could be as little as 1-5% of their total data volume. Starting to see the problem?

Custodial identification

Another issue arises in determining whether you have the correct custodians to begin with. As a general rule, the majority of matters brought are filed many years after the event actually occurred. Frequently, the designated custodians have changed positions or are no longer with the organization.

It is common for outside counsel to coordinate with individuals that may have had nothing to do with the underlying events. Appropriate individuals. Additional difficulties arise as custodians are attempting to remember events that occurred years earlier, to the extent that they can. Hmmm! This sounds like a recipe for disaster.

Garbage in, Garbage out!

The traditional method for handling ESI is: 1) ID the custodians the best you can, 2) do a full data collection or a selected data pull of their electronic information, 3) maybe do some sampling or ECA from those custodians' data, 4) possibly grab additional information based on the findings from the sample or ECA process (thus repeat of steps 2-4), 5) review data (as an iterative process), 6) export the production set. Sound familiar? No wonder we have a problem with the costs associated with ESI.

How do you know if you have the right custodians? Did you get the right data to start with? What key words should you use? This cumbersome process is compounded as attorneys attempt to identify key words, prior to having the ability to start looking at the data. I can't think of an instance where the saying, "Garbage In, Garbage Out," holds truer than in the world of ESI. Yet, clients, lawyers and judges struggle with understanding why ESI is so expensive and are trying to find the right solution to fix the problem.

Let's throw technology at it!

Today's blogs, legal periodicals, articles and of most recent, even cases are written around the benefits of various types of technology automated document review. Why? Because clients are tired of paying the costs associated with document review. I can't say that I blame them. We all know that the statistics show that the costs of review can be as high as 10x the costs associated with processing and hosting combined.

Thus, something has to be done. Throwing more technology at a flawed process isn't going to fix the problem. If you're trying to build a block wall, you'll need certain things, but most importantly, a viable foundation. If the foundation that you are starting with, much like the Leaning Tower of Pisa, has issues, it doesn't matter how you stack the blocks, there is going to be a problem.

Unfortunately, in the world of ESI, these blemishes may be like the Wizard in the "Wizard of Oz". You may not know that you have a problem until you pull back the curtain. This could be at the time that Plain

Plaintiffs challenge your methodology; you print off your privilege logs or Plaintiffs come across custodians which you have not identified, collected or reviewed their data.

What is the problem?

To fix the problem with the costs of ESI, you have to understand the underlying problem. Much like the builders of The Tower of Pisa, we find ourselves continuing along, utilizing the same process, hoping that this strategy will resolve the problem. The reality is it isn't going to go away until we resolve the underlying issues. For corporations, it is the manner in which information is maintained.



these blemishes may be like the Wizard in the “Wizard of Oz”. You may not know that you have a problem until you pull back the curtain. This could be at the time that Plaintiffs challenge your methodology; you print off your privilege logs or Plaintiffs come across custodians which you have not identified, collected or reviewed their data.

What is the problem?

To fix the problem with the costs of ESI, you have to understand the underlying problem. Much like the builders of The Tower of Pisa, we find ourselves continuing along, utilizing the same process, hoping that this strategy will resolve the problem. The reality is it isn’t going to go away until we resolve the underlying issues. For corporations, it is the manner in which information is maintained.

Paper days. In the “old days”, information was placed in labeled folders, which ultimately went into red welds, etc. This information, once the employee decided it was no longer useful, was shipped to storage with an assigned retention period. As the period arrived, absent a legal hold, the information was summarily destroyed. Keep in mind that the folders only contained information that was relevant to the subject of the file.

Today. Let’s contrast the paper days with today. In the “paper days”, our business records didn’t contain those “honey pick up milk” communications, family photographs, etc. as we do today in the electronic world. We complain about the costs associated with ESI, its review and production, however, the largest volume of data either is a non-business record or has absolutely nothing to do with the matter at hand. Yet, we can’t seem to figure out why the costs of document review are so high. Thus, the new buzz, technological assisted review.

This new buzz attempts to resolve the problem of reducing the costs associated with document review. The reality, we need to reduce or eliminate the volume of information that does not have a useful business purpose to begin with. Additionally, we need to reduce the number of incidents of the information so that we know what we have, where it came from, and which custodians have ownership. Lastly, we need to reduce the number of times that this information resides external to the organization.

These simple steps, if implemented, would significantly reduce the volume of information that would have to be potentially collected, processed and “reviewed”.

Under the traditional model, we compound the problem of managing corporate information by “collecting” data on a case-by-case basis, sending it to various law firms who, in turn, send it to outside vendors. Unfortunately, how many times is the same information, or portions thereof, sent to different providers for different matters?

There may be cases where you are compelled to provide assurance that certain information is no longer available. Can you comfortably make such a representation? Was an image of a hard drive or copies of backup devices provided to a service provider or counsel? Does the information exist outside of your environment and, thus, outside of your records retention policies and practices? What processes and protocols do you have in place to ensure that providers that you are working with are not “storing” your information after the conclusion of the matter? What about their back up devices?

Keep in mind that data may be responsive to another matter should it raise its ugly head in the future. Since the information is available, through one of your prior agents, you may be subject to produce that information, or at a minimum, they may be subject to a third party subpoena.

Bottom line, who has your data?

How do we resolve the problem?

The largest problem that we face today is the sheer volume of non-business information or business records that no longer have a useful business purpose that are being stored and maintained in our IT infrastructures to then be collected, processed and reviewed when litigation arises. We need a solution that allows you to maintain your viable corporate information, yet, eliminate data that has no useful corporate purpose.

This solution needs to ensure that corporate records retention policies and legal requirements for electronic data mirror that of paper documents. Then, when a matter does arise, you aren’t spend-

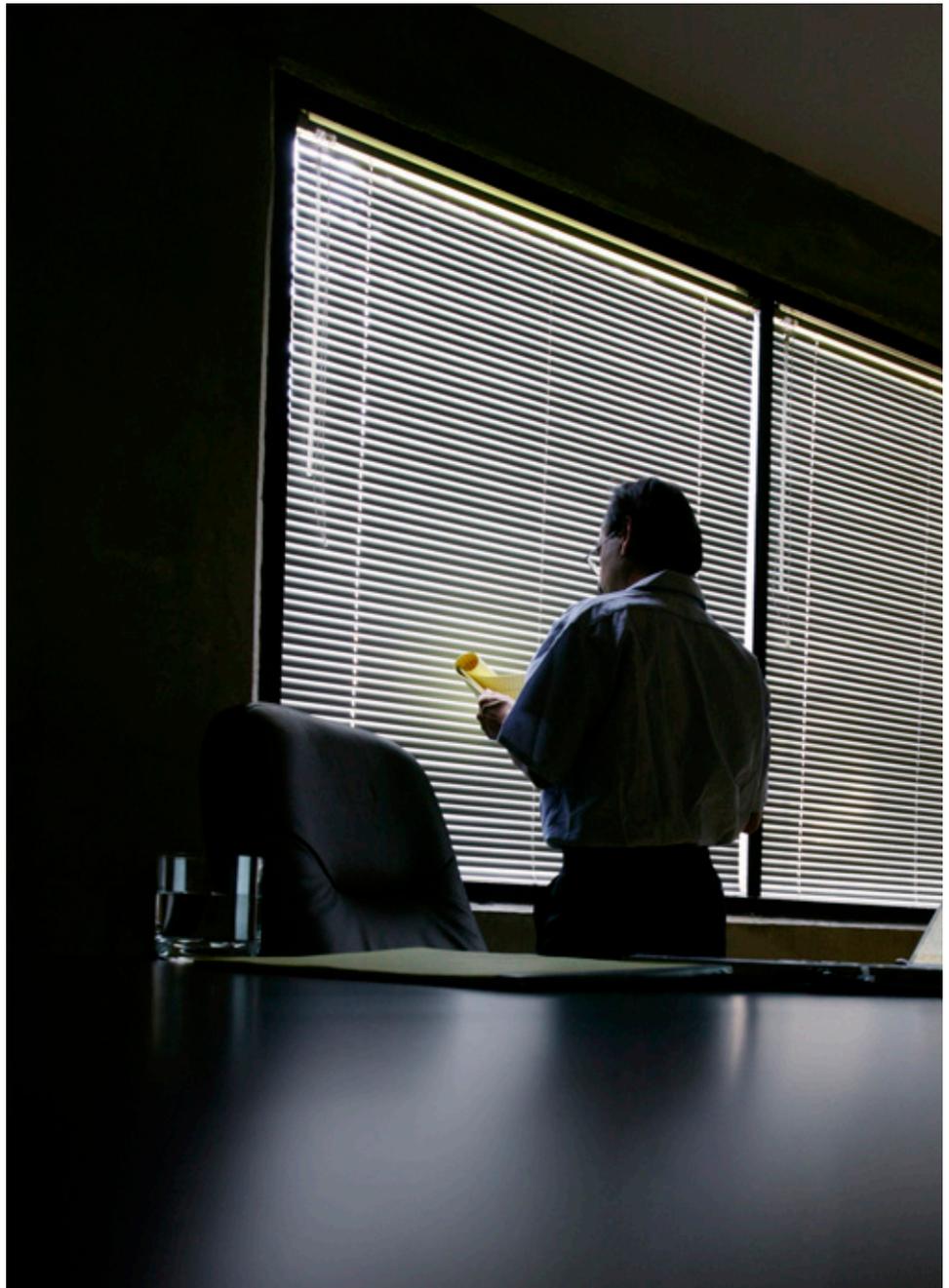
ing precious resources grabbing, processing and reviewing outdated or non-business records. Imagine how much more successful (and less expensive) a technology automated document review process could be if it was only manipulating viable data. What a novel idea...

Enterprise Strategy

What if compliance, records, IT and legal could actually collaborate on a holistic approach to resolve the issues surrounding enterprise data management? This strategy would provide the organization with a total information governance solution. It would reduce the costs of the IT infrastructure, eliminate information in keeping with records and legal policies, as well as provide a streamlined process for handling information.

This technology is available today and has been for a number of years. Unfortunately, many companies may be resistant to change, aren't looking at data from a holistic approach, or they treat data and business needs as silos, rather than resolving the problem. Companies are slowly recognizing that there truly is a better solution.

Let's fix the problem...



About the Author

Linda Sharp, Esq., MBA is the Associate General Counsel for ZLTechnologies. She is a member of OLP's Advisory Council, a member of the Los Angeles Chapter of Women in e-Discovery and a well-known author and speaker.

Lawyers Manag
- like driving a car w



Driving Themselves?
without a license?



Lawyers Managing Themselves?

It is hard to understand how a group of highly-regarded lawyers, known for their technical excellence and expertise at one of the finest law firms in the world, manages to ruin a century-old law firm in a rather short period of time. In its golden days, Dewey and LeBoeuf was home to more than 1200 lawyers, including some 300 partners, with more than 20 offices around the world. Now the firm is bankrupt.

Dewey's defenders, no doubt ex-partners amongst them, may argue that it was a perfect storm. Articles about Dewey's management failures don't make happy reading, and the sheer sum of these failures is ludicrous. But, given that U.S.-firms tend to move in flocks, one does wonder why this perfect storm hit only Dewey while the rest survives – or is it just a question of time and can we expect other law firms to follow suit?

Dewey may be regarded as an extreme example for management or leadership failure, but it is certainly not the only example of a poorly-managed law firm. It appears that lawyers lack certain core competencies to run law firms. Are they to blame, though?

They are excellent legal technicians and they are good at what they learn in law schools or universities, but what do they learn there? Certainly nothing about how to lead and manage an entity consisting of highly-qualified lawyers; some argue that they don't even learn "lawyering" there. Lawyers; some argue that they don't even learn "lawyering" there.

Teaching management is the reign of business schools, so why not think about recruiting business school graduates to run law firms?

There are many experienced managers out there who have proven that they are able to manage big companies far more complex than law firms. And yet, business school graduates only find a home in law firms within the lower ranks, not in a position to tell lawyers – especially not partners! – what to do.

There is a simple reason for this, in the words of Norwegian professor Bente Lowendahl, Grande Dame of professional service firms: "Professional service firms are different to such an extent that a direct application of traditional strategic management assumptions and tools is at best misleading and at worst disastrous."

It seems that a combination of skills is required to run a law firm. It is not enough to be a good lawyer. What is needed, in addition, is managerial skill in a number of key areas for the future, including:

“It appears lawyers lack core competencies to run law firms. Are they to blame?”

1. Understanding the fundamental changes taking place and grasping how to apply that understanding to your firm and practice.

It may sound easy, but training lawyers to apply non-legal knowledge and change standard procedure is hard work. Lawyers still present expertise and experience in a pitch, rather than really trying to understand a client's need.

2. Managing not only capabilities, time sheets and - most importantly – people, but also complex legal projects.

With the increase of global legal assignments, managing deliverables from outside and inside a firm takes planning and often a full-fledged project management skills.

3. Knowing the business and the environment your clients operates in.

Many global businesses have become increasingly complex, where the line between competitors and partners is blurry, like in the technology sector. Financial services developed sophisticated products that are even difficult to understand for industry professionals, but lawyers of all kinds are expected have a general understanding about the basic principles, even when they have just graduated.

Unfortunately, there are few law schools that combine legal education with management classes to understand the entrepreneurial part of the legal industry.

All in all, the competencies and skills needed to run a law firm extend far beyond technical legal skills. The earlier in life such management skills are obtained and the stronger they are integrated with the technical skills in daily law firm life, the more optimistic a law firm should be about riding out future storms like the one that sank Dewey and LeBoeuf.



About the Author

Markus Hartung is a lawyer and mediator. In 1999, he was elected as Managing Partner of Oppenhoff & Rädler, the first full-time Managing Partner in Germany.

During his tenure, he oversaw the merger with Linklaters and served as the Managing Partner of Linklaters in Germany from 2001-07 and as member of the Global Executive Committee of Linklaters. He is a member of the German Bar Association (DAV) and Chair of the Committee on Professional Regulation.

As a lawyer, he focuses on conflict management, regulatory issues and professional indemnity issues. In addition, he advises law firms in strategy and management questions and coaches partners in management functions.

At the Bucerius Center for the Legal Profession in Hamburg, Germany, Markus Hartung is responsible for the conception of educational and continuing education programs for legal professionals. His expertise in the framework of the CLP lies in market development and trends, management and strategic leadership, as well as corporate governance of law firms. Markus Hartung holds regular public lectures on these topics.



A Look Inside Hosted e-Discovery Pricing

Secrets your service provider
may not want you to know.

By Megan Miller, Gallivan Gallivan & O'Melia

Lawyers and litigation support teams researching alternatives for e-discovery have a daunting task, complicated by the fact that information on which they will have to base an important purchase decision is often fuzzy at the time the e-discovery solution is being selected:

- The location, scope and size of the data set may not be known. Data volume is difficult to estimate.
- The technologies used in processing, analyzing and producing digital documents are confusing, and vary from one provider to the next.
- Pricing models are often convoluted, or based on technical aspects of the project that won't be known until later in the game - think GB, document counts, page counts....
- "per GB" pricing is very prevalent, and yet probably the most volatile parameter in the process.

It's reasonable and prudent to evaluate several potential providers, request quotes, and compare the cost of services to make an informed choice.

Here's the rub: some service providers charge an hourly fee to collect data or image a hard drive; some charge a flat fee. Some charge a processing fee based on the number of Gigabytes (GB) of data. Some might waive or reduce processing fees, but make it up with higher hosting fees or another line item.

So how can you effectively assess multiple proposals for one e-discovery project and determine which will be most cost-effective? What questions should you ask service providers to be sure you understand what's included in each line item?

This paper will provide some insight into the various elements that drive the cost of e-discovery processing, hosting and review. To create our comparison we gathered several e-discovery bids; 3 representative examples are provided in the table at the end of this paper.

We studied the bids, 'translated' them line by line, in order to compare them in an apples to apples view, side by side. The pricing information is real – all of it collected from real service providers in major metropolitan areas across the US, in the last 90 days. However, it is also perishable.

The market is constantly changing, and pricing changes on a regular basis. While the bids are from real service providers, we have changed their names in our example.

GENERAL TIPS

Size Matters (in a per GB world)

All of the service providers we reviewed charge on a per GB basis for at least some of their services. A smaller data set (say, 20 GB or less) becomes very expensive in these models.

Many service providers are unwilling to incur the costs of project setup and management on a small matter. Volume-based ("per GB") fees won't generate enough revenue to be profitable.

We translated the bids line-by-line to compare them in an apples-to-apples view, side-by-side.

The pricing information is real.

If your ESI corpus will be 20 GB or less, but you still want the services of a hosting center and some project management, ask the provider if there are minimum fees, or if they will provide a fixed-fee bid.

If you have some skill and resources to manage software in-house for smaller matters, ask whether the provider has a software tool you can run on your own server or PC. Digital WarRoom Pro™ installs quickly on your PC, where you can then process, analyze, review and produce documents at a price point under \$1,000.

Collected Data May Well Grow in Size

The approach used to collect custodian data will have an important impact on the GB volume of data flowing into the processing step. A forensic image of an entire hard drive will typically have a higher volume of content, perhaps 100 to 200 GB, but a good portion of that will be program and system files that are not reviewable documents.

A procedure called de-NIST'ing removes these files as part of the processing phase. More selective collection, via selective copying, or use of ECA search and collection tools, may reduce the volume of data the e-discovery provider receives.

The processing phase of the EDRM is probably the least well understood. In addition to de-NISTing, processing involves the use of specialized programs to open container files (mailboxes full

of messages, folders full of Word documents, .zip files full of a variety of documents), identify and remove duplicates, and index the document contents to create a database that is key-word searchable.

The important tip to know about processing: in a normal collection of email and documents, for example, the GB volume of documents after processing may be 10-30% larger than the original collection. The growth occurs during 'expansion', when a content of condensed folders is extracted into individual documents, messages and attachments.

So it's important to ascertain whether the fees your provider charges on a per GB basis will apply to the 100 GB collected, or the ~125 GB after expansion and indexing.

Monthly Hosting Fees: Volume-based

Most hosting providers charge a per GB/monthly fee for hosting the documents, some charge a flat monthly rate. Be sure your provider is clear about which GB volume represents the hosting fee basis: is it the collected volume, the volume post-processing, or the post-filtered volume for review.

Monthly Hosting: User Access

A 'user access fee' or license is commonly charged; it typically includes account logins and credentials for authorized users. Have the provider clarify whether the user access fee is charged monthly per named user, or per account.

Your hosting needs can vary month by month over the course of a matter. If, for example, the case is delayed in the courts, you may incur hosting fees for months of downtime when attorneys are not accessing or reviewing documents. If you anticipate this could happen in your case, ask about the availability of a 'standby' or 'on hold' rate for reviewer access licenses.

A reduced rate during a lull in activity can realize significant savings, particularly on large reviews.

Efficiency Tips to Contain Costs

- Start early! Decisions made under the pressure of a litigation hold and pending filing may not serve you well. Dedicate a small team to the evaluation process, and allow them the time to research and make recommendations.

- Request a hands-on trial of the software, if you have not used it before. If your trial can include a test on your own data (you'll have greater familiarity), you'll know how intuitive the software is. By far, an easy-to-use application for processing and review gives you greater control and will save custom consulting costs (you'll need less help, and incur fewer hours of consulting assistance).

- Take advantage of flat-fee processing rates if available, and then manage the ESI collection process to result in fewer and larger, rather than many smaller, collected data sets. Combining several custodians' data on one drive will achieve greater efficiency and lower cost when you are paying a flat fee for processing.

- Don't put client data at unnecessary risk. Each time data is moved from one system to another, or is saved in a new format, the risk of data corruption or loss increases. A system that can process, support review, and produce responsive documents while keeping the entire collection intact in one application will reduce time delays involved in importing/exporting between various systems, and will reduce risk considerably.

- Investigate Repository storage. Do you expect follow-on litigation? Maybe a civil case on the heels of a government investigation? Often the same documents will be responsive in multiple matters. Ask the service provider if they have the ability to build a repository of documents and work product that may be re-purposed for future matters. Don't collect and process the same documents multiple times.

A 'user access fee' or license is commonly charged.

native documents into images prior to starting review. Imaging incurs avoidable up-front time and cost.

Know that ultimately, only 5-30% of the original documents will prove to be responsive and require production. At that point, (if you agreed to provide TIFF), you will only have to pay for imaging on the smaller production volume of documents.

- Look for a provider who is also the creator and owner of the software tools. Many companies are 3rd party re-seller or service providers, who add a 'middle-man' layer of cost to the project. You can realize savings, and sometimes great benefits by working directly with the company that created and owns the review platform:

- o The principal who owns the technology is not saddled with all the licensing fees and royalty costs that a 3rd party service provider is required to pay; with a lower cost basis, they can price more competitively or have room to make price concessions.

- o The team who created the software knows it better than anyone, often can provide better technical support;

- o That same team may also be more responsive on bug reports or new feature requests – you have a direct line of communication to the source, rather than through a 3rd party service provider.

A final caveat: Don't Sacrifice Quality for Cost

There are many steps and technologies at work in e-discovery. It goes without saying that your company's data, or your client's data must be carefully guarded, and the e-discovery process should be conducted by professionals who have proven skills and experience to handle the challenges that inevitably arise.

Seek to control cost, but don't sacrifice quality. The service provider you rely on should be ready to provide credentials held by project managers and technical staff, and those should include:

- A healthy mix of technical talent and legal training; degrees in information sciences, data management, database administration, systems architecture, machine learning
- Field experience in a law firm, in litigation support

or e-discovery management

- Industry certifications among the following:

- o CeDP (Certified eDiscovery Professional, Organization of Legal Professionals)

- o CISSP (Certified Information Security Specialist)

- o ISCFE (International Society of Computer Forensics Examiners)

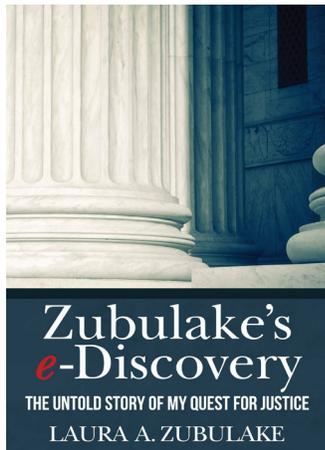
Now you are privy to some of the secrets that e-discovery service providers seldom divulge. Use this information to demand fair pricing and high quality results. You will be a hero for your firm or corporation, and the industry will be better for it.

Interested in Learning More? This topic was addressed in a live webinar, "A Look Inside Hosted e-Discovery", by Bill Gallivan in May 2012. Readers are invited to download the presentation and FAQ documents on the Digital WarRoom website.



The Thumbed Page

Books by Members of OLP's Board of Governors



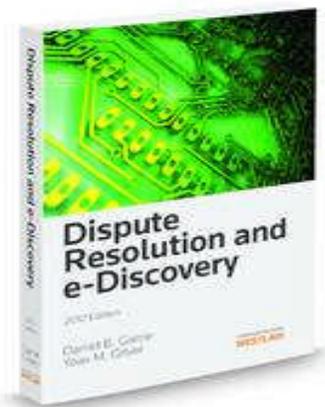
Zubulake's e-Discovery: The Untold Story of My Quest for Justice by Laura A. Zubulake

Here is Zubulake's own story. Her case ("Zubulake") was one of, if not the largest, jury award in the US in an employment discrimination suit. Establishing the foundation for standards in the evolution of discovery, Zubulake resulted in the issuance of landmark legal opinions. Known as the Zubulake opinions, they established precedents in the area of electronic discovery ("e-discovery"). Considered the first definitive case concerning a range of e-discovery issues, it influenced the 2006 amendments to the FRCP written from the plaintiff's perspective. It is a book about perseverance, vindication, accountability, and justice.

To purchase this book:

www.LauraZubulake.com

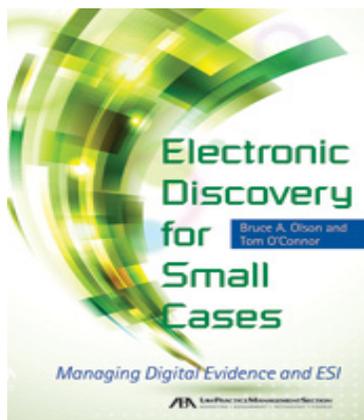
Price: \$29.95



Dispute Resolution and eDiscovery By Daniel Garrie and Yoav Griver

This book is a one-stop guide for counsel, arbitrators, mediators, special masters, and others that seek to develop a better grasp of e-discovery and its application in various alternative dispute resolution forums. Each chapter author is a recognized authority in the topic of the particular chapter, and particular emphasis has been placed on offering the reader practice tips and practical knowledge useful in successfully resolving actual disputes in diverse areas of practice.

To purchase this book: store.westlaw.com Price: \$149.00 USD



Electronic Discovery for Small Cases: Managing Digital Evidence and ESI by Tom O'Connor and Bruce Olson

For firms with limited technology budgets, or cases with small amounts of electronically stored information (ESI), e-discovery can be challenging. This book offers effective, budget-friendly solutions for collecting, viewing, and analyzing electronic evidence that will benefit any litigator.

To purchase this book: American Bar Assoc.

<http://bit.ly/MOXLcD> Price: \$79.95 (Regular)

Bridging the Gap in EDiscovery – The Emergence of Conceptual Semantic Search

By Jeff Parkhurst



Much has been written lately about the volume of data that must be sifted through in today's litigation. For decades the "gold standard" for sifting through document populations has been keyword searching. Of late, there has been a movement to understand the limitations of keyword searching and to replace it with more advanced and cost effective methods of data gathering and parsing.

Recently focus has been on the positive impact that Technology Assisted Review and specifically, Predictive Coding (as a more refined type of TAR), can have on our industry. These technologies are not without controversy. The theory is easy: Predictive Coding involves the use of software to help identify potentially relevant documents thus reducing the volume of documents that need to be examined.

The Courts are finally beginning to weigh in on technologies that augment other human document review and keyword searching to find responsive information. Corporate data volumes are estimated to be growing at 40% a year. As expected, ESI volumes are also growing exponentially, yet lawyers must remain in compliance with the Federal Rules of Civil Procedure which govern process "... to secure the just, speedy, and inexpensive determination of every action and proceeding".

The only way to achieve this is to look outside the box and identify technologies that can bring greater proportionality and cost savings to the discovery process. Predictive Coding is the most often discussed alternative today, but it may also be somewhat outdated in its current iterations. There are virtually limitless variations of training and algorithms used in predictive coding and vendors frequently tout their approach as being superior to other variations. Arriving at a "standard" model of predictive coding that produces consistent results has been difficult.

This article is about data analytics. Data analytics have been around for decades. Virtually all corporations and consulting organizations have analytics technology. Certainly it is being used to sample and review the efficacy of functions within organizations.

However, most of the data being analyzed is not currently text-based. Text or semantic data analytics have been around in one form or another for well over a decade. It has been largely gone unimplemented because the need was not there.

However, in a world that creates 1.3 zettabytes of new text annually, we now face a crisis of "not knowing what we know." With the rise of keyword indexing decades ago in corporate America, our law firms and our EDiscovery populations have deployed keyword searching to solve our knowledge problems.

Therefore, when I read Judge Shira Scheindlin's recent opinion that is described below, I saw the opportunity open up the discussion to consider a better technology. I have chosen to focus on conceptual semantic search technology as a more comprehensive search technology that extends beyond keyword searching and predictive coding tools.

When conceptual semantic search is used in conjunction with predictive coding methodology and document review software systems, it forms the foundation for a very strong platform in EDiscovery that should improve everyone's understanding of the search results and help identify the truly important documents. In effect, what is the best way to enhance existing systems and help alleviate some of the concerns that surround current keyword search term limitations and predictive coding complexities?

Keyword Searching and the Courts

U.S. Magistrate Judge Andrew J. Peck's opinions on keyword searching continue to be at the forefront of EDiscovery. Discussions go beyond his landmark *Da Silva Moore v. Publicis Groupe* opinion approving of the use of predictive tagging technology. Judge Peck recently further refined his opinion when he said:

'In too many cases, however, the way lawyers choose keywords is the equivalent of the child's game of 'Go Fish' ... keyword searches usually are not very effective.'

In July, U.S. District Court Judge Shira Scheindlin of the Southern District of New York quoted Judge Peck in her opinion regarding *National Day Laborer Organizing Network v. United States Immigration and Customs Enforcement Agency*, No. 10 Civ. 3488 (SAS), 2012 U.S. Dist. LEXIS 97863 (S.D.N.Y. July 13, 2012):

"Simple keyword searching is often not enough: 'Even in the simplest case requiring a search of on-line e-mail, there is no guarantee that using keywords will always prove sufficient.' There is increasingly strong evidence that '[k]eyword search[ing] is not nearly as effective at identifying relevant information as many lawyers would like to believe.'

The opinion offered further advice on "emerging best practices," encouraging collaboration between parties and technology-assisted review. Judge Scheindlin continued, citing the "shortcomings" of keyword searching:

"There is a "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." And beyond the use of keyword search, parties can (and frequently should) rely on latent semantic indexing, statistical probability models, and machine learning tools to find responsive documents.

EndNote 1 contains information on the Basics of Predictive Coding for background purposes

The Current State of Predictive Coding

Predictive coding sounds promising on its own merits; but still, the judiciary, lawyers and technologists have been unable to come together to completely accept this newer technology. Why?? I believe that there are as number of reasons:

1. Having been involved in the legal industry for 30 years, I have concluded that the legal system as a whole is not a leader or a fast adopter of new technology. "Precedent" is strongly woven into more than the just the judicial interpretation of a set of facts, it is part of the fabric of every stage of the practice of law. Change is difficult to implement.

2. Predictive coding is highly scientific and difficult for many lawyers to understand. It is based on math, with complicated algorithms and advanced statistics. As many of my lawyer friends are fond of saying, "I became a lawyer because I wasn't good at math".

3. It is proprietary in nature. Companies have spent millions creating software that they feel represents the best approach to the problem based on their experience and the interpretation of that experience by highly intelligent program designers. So no one wants to open up the "black box" and reveal exactly how their system operates, requiring us to take it on faith that the results are valid and accurate.

4. Each company uses algorithms in a slightly different manner to produce a set of documents that they are sure are better than the competition. This means that no two products will produce the same set of search result documents, which is troubling to attorneys and judges when discussing the completeness of document productions.

5. Predictive coding involves implementing a series of complex steps leading to the production of a final set of responsive documents. Some of the steps include identifying a small set of training documents (also known as seed documents), reviewing that subset using humans to train the machine, and then evaluating the system results, including complicated sampling and confidence ranking... and then repeating the process with further refinement. It can be time consuming and require attorneys to learn how to effectively use the software to obtain good results.

How Can We Improve on the Existing Model?

I believe that while legal professionals are focused on EDiscovery and how to improve document retrieval, we need to recognize that what we are quickly encountering is a problem with 'Big Data'. The number of documents which must be quickly analyzed in the discovery process has grown dramatically over the last few years. It goes far beyond the documents created or received by a few key custodians.

In addition to the documents that may be within the direct control of a corporation and its employees; it has grown to include vast quantities of email, social media, document repositories and other information that are exchanged on the Internet and often controlled by third parties.

It has truly become a world of Big Data that keyword search engines just can't handle effectively or accurately. As the data volume expands, it becomes harder for attorneys to know what documents to look for in the first place, in order to find seed documents. We don't know what we need to know in order to make the process work effectively.

Conceptual semantic searching offers some of the best approaches to early data analytics that can provide users with an exceptionally unique visual overview and approach to analyze unstructured data. It's very visual, fast and provides an abundance of information about your documents in the early phases of EDiscovery.

The goal of analyzing data is to simply identify information that is responsive to your request, quickly, effectively and efficiently. In fact, semantic searching should help educate you about what might be important about a topic, rather than the other way around.

What is Conceptual Semantic Search?

Semantic search goes far beyond keyword search. It seeks to improve search accuracy by understanding the intent of the searcher and the contextual meaning of terms as they appear in a searchable text. Semantic search processes consider multiple points of information simultaneously; including context of search, location, intent, word variation, synonyms,

generalized and specialized queries, concept matching and translation of natural language queries to provide a result set of documents. In summary, while full-text searching lets you query exact words in a document; semantic search lets you query the underlying meaning of the document.

What does this really mean? Using semantic search, there is no particular document that the user knows about that they are trying to uncover. Rather, the user is trying to locate documents that when examined, will give them the concepts that they are trying to find. The goal of semantic search is to deliver targeted information queried by concept rather than have a user sort through a list of documents loosely related by only the presence of a keyword. This is what has been missing from EDiscovery tools: the ability to quickly and effectively perform complex conceptual analytics on document populations and to help a company and its legal counsel determine what information they have. This conceptual analysis helps litigants determine what data they have, what it means and whether you should litigate or settle the case.

“No two products will produce the same set of search result documents.”



The advantage of semantic search originates in the engine's ability to match on the meaning of words regardless of what words are used in a user's query or in a source document. In short, semantic search engines are able to go beyond keyword matching and match on concepts.

Most of the semantic search engines can produce relevant search results that do not contain any of the original query words. In addition, many of the linguistic challenges that typically wreak havoc on keyword engines like polysemy (words with multiple meanings) and synonymy (multiple words with the same meaning) are handled intelligently and naturally by semantic engines.

While it is much more complicated than I have just written here (doctoral dissertations and lengthy white papers have been written on semantic search technology), the essence is that semantic search differs greatly from keyword search. It is a new way of identifying useful and relevant information.

How is Semantic Search Different?

Keyword searching involves linking keyword queries to taxonomies, lexicons and thesauri as a way to provide a form of structured query. This technique for recognizing relationships in data has been around for years with every new list creator claiming a new breakthrough in learning.

The reality is that these structured systems utilize manually created (but machine generated) lexicons in an attempt to understand meaning and improve search results. In fact, they are not constantly 'learning systems' but rather static, non-dynamic systems that require constant updating (re-indexing) in order to produce any type of accurate results. In today's world of Big Data, it is not possible to constantly update large federated index lists at the same rate that information changes and new documents are added to a dataset for analysis. So the resulting searches become less effective over time.

In summary, semantic search generally does not suffer from the same limitations of keyword indexing technologies. Perhaps most importantly, semantic search is dynamic, with the ability to continually update concepts as new information is introduced to the system.

In addition, semantic search improves your search results by adding concepts that relate to your original search query. It is not limited to keyword lists. You can then instantly review the concepts the system suggests, assign weighting measures marking the relative importance of the documents, eliminate concepts you don't want to examine at that time and add more concepts of your own for another round of analytics by pressing a button.

Shortly after loading text searchable documents into the system, you can begin to examine them to help you determine what you need to know about your case. Semantic searches actually produce the seed concepts and documents for you by grouping document concept clusters. You can then determine which documents are of most interest and prioritize a review of that material first, immediately examining cluster contents to determine their utility for your case.

Some Current Limitations of Semantic Search Products

Semantic technologies have been around for years. Most technologies deployed today have some limitations depending on the underlying algorithms. Semantic search engines offer a variety of potential benefits to search technology in applications; however there have been a number of early products that contain limitations on their utility. These limitations include things such as:

- A lack of transparency:** Most eDiscovery "Black Boxes", with the internal search generation

“Semantic search improve
concepts that relate to yo

approach hidden from the end users. Search results remain somewhat elusive to understand, describe and defend.

•**Defensibility:** It is very difficult to provide a defense for a set of search results when it is not clear exactly what that result set is based on. If the contents of the Black Box are not opened for examination, it is difficult to convince opposing counsel and the judge that you have provided all the information.

•**Lack of Control:** In addition to being black boxes, many semantic search engines do not provide the searcher with the easy ability to immediately enhance or interact with the search result or the search query.

•**Index Scaling:** Some semantic search engines can only search the documents they index and further do not allow other engines to search their indexes. As the volume of content continues to grow exponentially, it is difficult to re-index data sets.

•**Intelligence Scaling:** Semantic indexes are generally larger than keyword indexes, typically reside in RAM and cannot scale to learn from tens of millions of documents, making it difficult to handle large data sets.

These limitations and an understanding of how litigation proceeds has created the opportunity for a semantic search solution that can overcome one of the most difficult problems facing litigators: the ability to provide an easy to use system that augments the human ability to make legal determinations.

What does “BrainSpace” by Pure Discovery Bring to the Table?

I have been attending Legal Tech New York for well over 18 years. I scour the floor each year looking for something “new”, which is not just an updated version of an existing product or a slick repackaging of an old one. About every 4-5 years (My own Moore’s Law for EDiscovery), there seems to be a quantum leap by a forward thinking company that creates what I would

consider to be a watershed moment in the legal market. In 2012, that discovery came while I was attending a session on semantic review by a company called Pure Discovery and test drove a product called LegalSuite deploying their technology called BrainSpace.

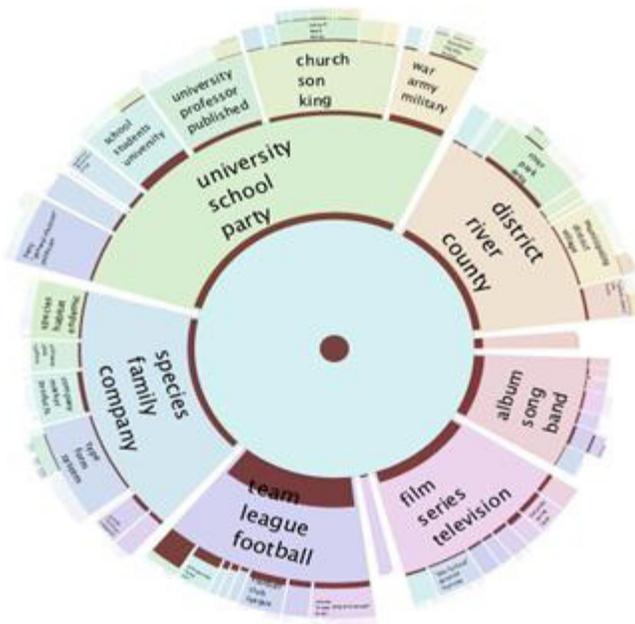
Whereas predictive coding technologies rely on linking together pieces of data within specific indexes using industry-specific ontologies, BrainSpace doesn’t try to index data at all. It certainly recognizes when documents contain similar words or concepts, but it also has a wider capability. BrainSpace actually learns about documents the same way that we as humans learn, by processing relationships. This learning includes its ability to determine concepts, interests, people and perspectives.

Since full-text searching engines entered the legal industry in the 1980’s, our analysis has focused on “what” was in a document, partly because that was a limitation of the technology. However, today we live in a new “social” world when networks are employed at virtually every client.

We are all wired to our co-workers in multiple ways. We interact with email, documents, data and social media and material is no longer attributed directly to us, rather many people in an organization touch it along the way. With data existing in a social world, so does evidence.

The answer is no longer simply what, but perhaps even more importantly, who. BrainSpace provides machine learning of your data, providing information on both what and who. BrainSpace, through the Pure Discovery Legal Suite (PDLs) acts as a device that exists between a user’s cognitive thoughts and the data sources where the data resides. It is “post-search” technology that presents you conceptual data in a way that is tunable, fast and easy-to-understand. BrainSpace reads your query, analyzes it against what it has .

es your results by adding
ur original search query.”



Some BrainSpace Considerations

1. The current version of BrainSpace requires that the documents be in a text readable format, it does not read or process native files or non-text documents.
2. BrainSpace requires you to point it at a data set, but it doesn't care how large or what type of data set it examines. Semantic queries can be formulated to run across virtually any document populations including intranets, extranets, enterprise content management systems, portals, and email archiving.
3. BrainSpace does not create an index. Rather, it creates containers of document "intelligence" with similar concepts which allow you to quickly examine them for relevance.
4. Within approximately an hour of "brain building" (loading) a dataset that includes 1 million documents, you could begin to examine the contents by clustered document container.

By quickly returning easy to understand visual maps of the contents of your data, you can quickly review, edit, tag and parse information as your knowledge of the contents increases

- Improved document recall

Focused Semantic search quickly returns more results that are related to your search query, whether or not you used the specific terms in your initial search.

- Greater precision

By combining Boolean and focused semantic search technology, your results will be highly relevant to the topics that are important to you.

- Increased transparency

The terms and concepts suggested by the semantic technology are returned to you for your review as the actual search is run. You know exactly what the system is searching on.

BrainSpace also allow you to tune or refine your conceptual search results and outputs a standard Boolean statement as to how the weighted search results were arrived at.

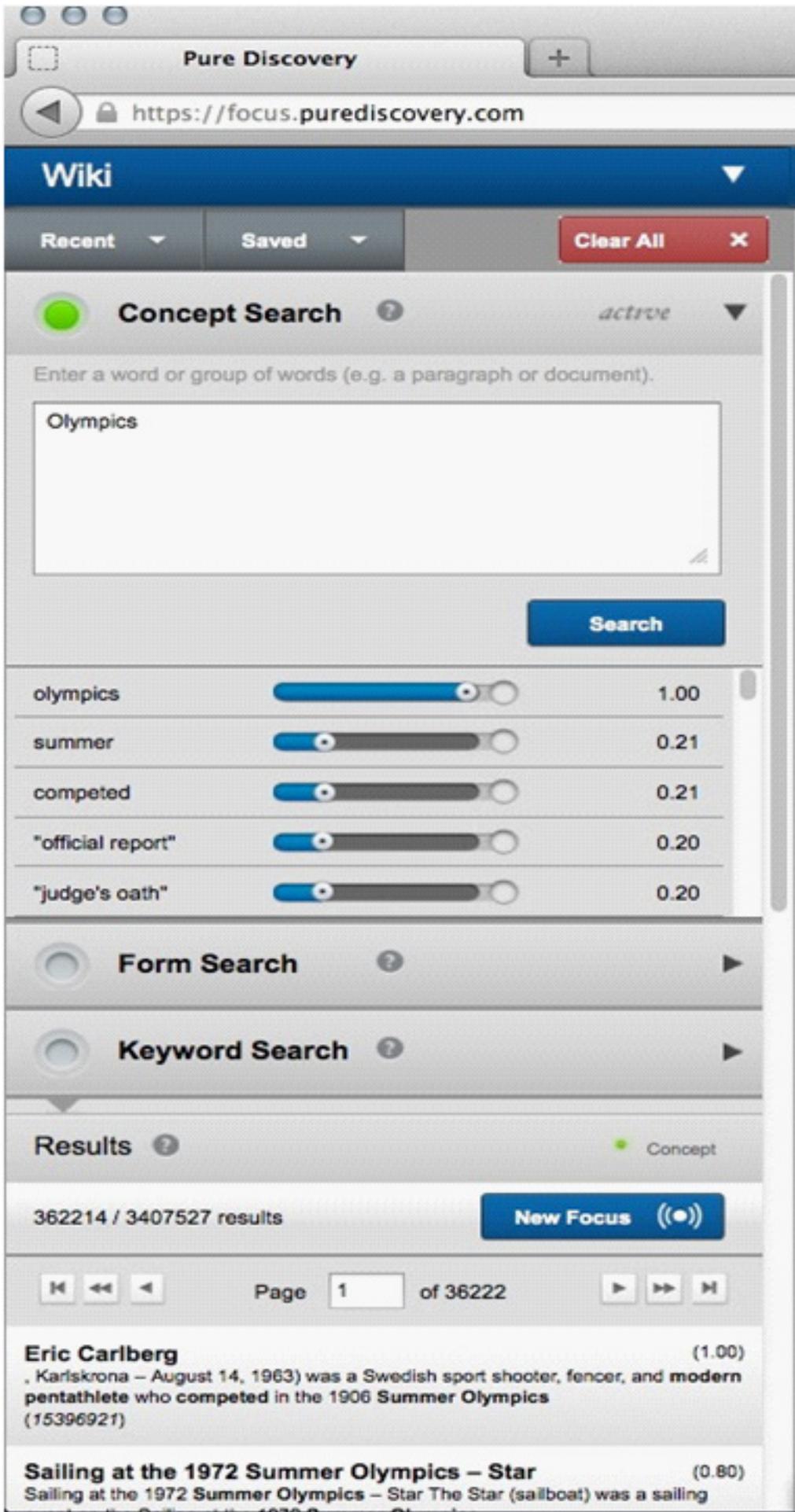
This overcomes the mystery blackbox results set that will plague many semantic and predictive coding technologies. The system presents the user with a weighted result set which then allows the user to adjust the weight of the various lesser included concepts to get the most complete results possible. This is done through the use of slider bars in the results box.

What are the Benefits of BrainSpace Semantic Search?

- Saves time

You enter what you do know about the case or topic you're researching in your search query. The semantic technology uncovers related concepts and terms, educating you about the documents in the database.

- Immediate control over the process



5. The BrainSpace query takes a plain language question that you are asking and converts to a conceptual Boolean search statement and then searches the data set. Results are displayed in concept clusters which allow you to easily identify documents that are responsive to questions you didn't even know to ask.

6. BrainSpace maintains the Boolean statements it created so that results can be repeated and described to the opposition or the court, revealing how the results were obtained in a defensible manner. It makes transparent to the user what search statement is being sent to the data set.

7. Functionality built into the product:

a. The Semantic Near Dupe Identification Engine detects and groups near duplicate documents, identifying redundant documents with only slight variances which reduce review time.

b. It also contains text-based deduping, which goes beyond hash value in identifying duplicate documents by comparing the text of documents, exclusive of metadata differences.

c. Concept clustering is displayed as a "Focus" wheel of relevant containers that can be continually parsed into concept subsets, all the time displaying the concept that brings the documents together. Users can quickly generate visual

maps of responsive documents, identify and tag key areas of hot documents and store and/or export these subsets for immediate review by a team.

d. BrainSpace assigns a PDID Document Tagger number that goes beyond Bates numbering. It is a semantic Bates number because it assigns documents with contextual similarity a similar PDID number. This allows the user to sort documents from related containers by using the PDID number.

e. Users are given the ability to add, delete, increase or decrease the importance of all query words in a unique visual query interface as they are examining results.

8. Document Containers of concept related documents that are quickly reviewed and determined to be relevant and likely require further analysis can be tagged and then exported into any document review platform or predictive coding system for a secondary analysis.

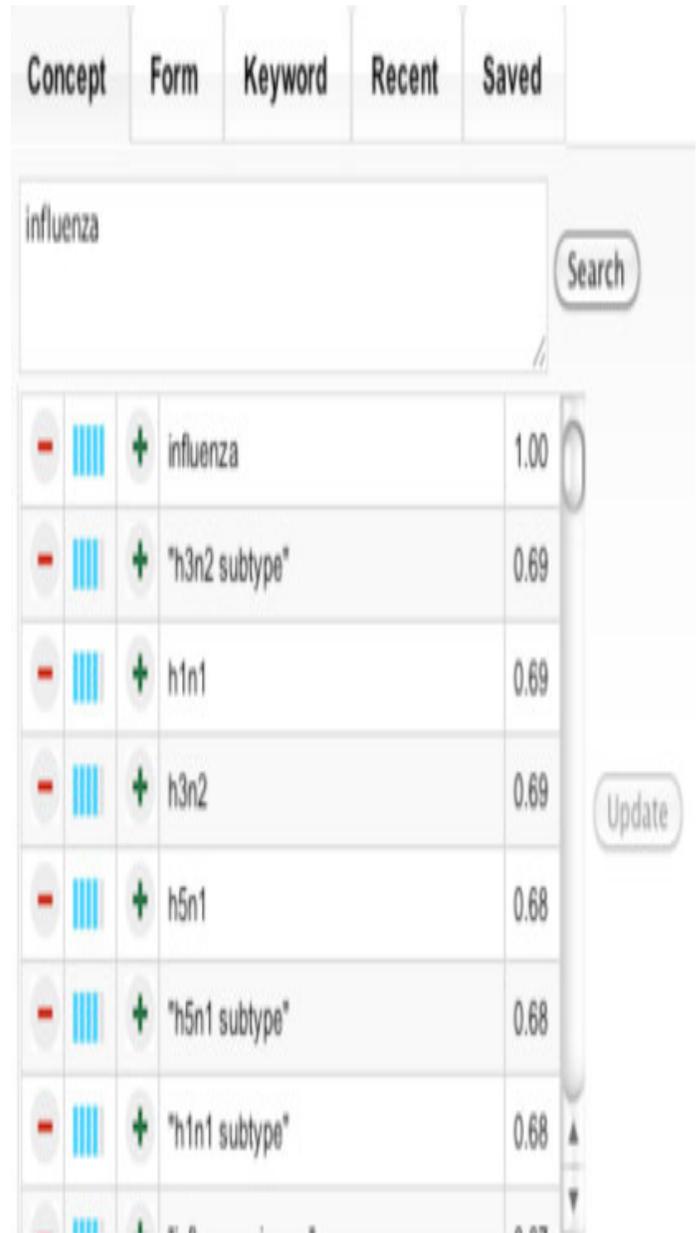
Why is this type of analysis different?

One of the most important features of the software is not really a feature, but rather forms the basis for a new approach to using semantic search. BrainSpace has been designed to be an interactive experience or conversation between the user of the system and the machine learning concept search.

The goal of the system is to encourage continued interaction with the data so that the search results continue to educate you about the information in the documents, quickly increasing your level of knowledge. It is less transactional and more interactive and hands on.

Creating a tool that is easy to use and understand means that attorneys can easily spend time with the data, quickly educating them as to concepts and documents that may be relevant to their ongoing discovery. The interaction creates a place to work and learn, rather than a long result set of documents that are similar to a keyword that was preselected.

This level of understanding is made possible by transforming your queries into a QueryCloud, which is a visual portrayal of the newly generated semantic query. It effectively places the user in the center of the transaction, encouraging interaction between the query and the data. Each user query is transformed into a list that shows the most relevant extracted and inferred words and phrases (which are



Conclusion – The Advantages of Semantic Search

The goal of litigators in handling the large volume of data in today's discovery is to provide the most cost-effective and comprehensive solution to analyzing the data that is potentially involved in EDiscovery and discover what is relevant and why. And the earlier this is done in the process, the better! When used in conjunction with other data management and review tools, semantic search can improve the state of EDiscovery. I have listed 4 key factors that indicate why and how semantic search can be used to improve your handling of Big Data in EDiscovery. It is time to take a long look at how this can impact EDiscovery:

1. Know your data – you have to be aware of what data you have, what it means and how it might impact your case as early in the process as possible. Your knowledge may result in your pursuit of a settlement rather than proceeding to trial based on what you learn. Including semantic searching in your plan dramatically reduces your learning curve by pointing you towards information that is likely relevant; more quickly and easily than other methods.

2. Semantic search improves your results – Semantic search queries take plain language questions that you are asking and convert them to a conceptual Boolean search statement which then examines the data set.

3. Explain your approach – you need to provide an explanation to the opposition and the judge about how you have achieved your search results and why the document population you are turning over is in fact responsive and relevant to the discovery request. This level of search transparency is at the heart of the semantic search product Pure Discovery which turns all the plain English search requests into a conceptual Boolean statement which can be clearly understood and replicated when necessary.

4. Be transparent and cooperative – Judges require parties to come to the meet and confer with definitive plans that have been worked out between the parties. They are looking for reasonable and well thought out approaches to discovery that are based on some degree of proportionality.

Results are displayed in concept clusters which allow you to easily identify documents that are responsive to questions you didn't even know to ask. Semantic searches are dynamic, with the ability to continually update results as new information is introduced to the system. The better knowledge you have, the better you are able to negotiate during the meet and confer to limit document production, understand your case and determine litigation strategies.

Using semantic search as part of your overall preliminary document strategy will help improve your knowledge about the document population and allow you to improve everyone's understanding of what and how documents have been selected. You will not be taken by surprise at the meet and confer since you will be in control of the information on behalf of your client.

±The Basics of Predictive Coding

Without going into a complete discussion about predictive coding, the essential element that is relevant to understand is that predictive coding is based on some type of document seeding in order for the machine to "learn" what kinds of things you are interested in finding. The legal team puts together several representative populations of documents dealing with key areas of interest and the machine begins to locate documents of a similar nature. Predictive Coding requires:

- Input from case experts: both substantive legal issue and software consultants
- Keyword analytics to first locate important documents and create seed sets for the machine to use as their matching sets.
- A defined workflow that includes strong statistical sampling analysis to help insure accurate results
- Iterative rounds of machine "learning" (augmented by software and case experts) to find other documents that are "like this" based on keywords and some concepts.

Predictive coding is not designed to replace human review of documents, it is meant to optimize the review and help reduce the volume of documents that must be examined during discovery. The output from predictive coding during discovery is to take all the documents the computers identify as "related" to an issue identified by the case experts and then rank them and tag them so that they can be reviewed by humans for relevance and responsiveness.

One of the advantages of this technology is that you are using human decisions to "teach" the computer to locate documents, increasing the accuracy and relevancy of search sets over time. Whether you call it predictive coding, computer-aided review, or technology-assisted review, it employs a combination of human beings and computer algorithms that are used to determine relevant doc

uments by creating "seed sets" -- and then using the seeds (controlled by algorithms) to have computers produce subsets of responsive documents.

About the Author:



**Jeffrey Parkhurst
Consultant,
Studeo Legal**

Jeff provides consulting and business development leadership to legal service providers regarding new business opportunities and increased service offerings. He delivers consulting to clients on EDiscovery procedures, processes and software alternatives to process discovery data. He writes a weekly blog, Support for Litigation on EDiscovery and litigation support issues highlighting discovery trends, consulting services and the impact of recent court rulings on the practice of law.

He is also responsible for all aspects of Trial Brief Pro, a new cost-effective hyperlinked electronic brief service from Studeo Legal that focuses on all levels of the judiciary. He spent 10 years of his professional career with litigation support firms RealLegal and CounselPress, developing software to facilitate the creation and delivery of digital deposition and trial transcripts and the creation of hyperlinked electronic briefs. He pioneered the efforts of law firms seeking to improve methods of electronic filing and provided extensive education to court personnel on the advantages of electronic technology in the legal market

“BrainSpace does not create an index. Rather, it creates containers of document “intelligence” with similar concepts which allow you to quickly examine them for relevance.”



New OLP Chapters Now Forming!

You don't have to wear any "Hello My Name is" tags that stick-to-your shirt, eat soggy, day-old Cobb salads or listen to talking heads drone on and on about yesterday's technology.

We promise.

What you will want to do is join the newest OLP chapters or help create one in your city. Why? Because OLP is forming innovative, exciting and informative chapters. Latest topics. Discussion groups. Peer-to-peer connections. Meetings you won't want to leave.

Contact OLP today and find out how you can advance your career by joining the organization that brings top quality continuing legal education and certification exams to the legal community at an affordable cost.

Phoenix Chapter - Founding members: Paul McIlroy, Dave Tiller, Brenda Studebaker, Cindy Perez. Kick-off Networking Event: Sept. 20th

Washington D.C. - Founding members: Jeff Fehrman, Christina Ayotis, Joe Kanka, Leigh Issacs. Kick-off Networking Event: October (dates TBA)

Los Angeles, Miami, Atlanta and New York starting soon!

www.theolp.org info@theolp.org 760-610-5462 Join today!



Zubulake's *e*-Discovery

THE UNTOLD STORY OF MY QUEST FOR JUSTICE

LAURA A. ZUBULAKE

During the years 2002 through 2005, I made a series of decisions that not only altered my life, but also inspired historic changes to the practice of law and influenced the way in which organizations manage information.

I became the plaintiff to a litigation that has become one of the most written-about federal cases in history. The voice of the plaintiff, particularly the plaintiff in a litigation with far-reaching consequences, is rarely heard. Zubulake's e-Discovery: The Untold Story of my Quest for Justice is my story.

In one of the most difficult professional decisions I would ever make, I filed employment-related claims against my former employer, a multi-billion-dollar Wall Street firm. My case ("Zubulake") came to be about much more than the allegations.

It was a classic David versus Goliath story: a multi-year dispute pitting an individual against a financial giant. Culminating at trial, the confrontation was unusual on Wall Street where management seldom airs internal matters in public. Resulting in a notable verdict, it was one of, if not the largest jury awards in the United States for a single plaintiff in an employment discrimination suit.

Establishing the foundation for standards and seminal in the evolution of discovery, Zubulake became renowned. Years I spent searching for electronic evidence that I was led to believe did not exist resulted in the issuance of landmark legal opinions. Known as the Zubulake opinions, they established precedents in the area of electronic discovery ("e-discovery"). Considered the first definitive case in the United States concerning a range of e-discovery issues, it influenced the 2006 amendments to the Federal Rules of Civil Procedure.

E-discovery has grown into a multi-billion-dollar business and is one of the fastest-growing legal specialties. E-discovery has changed the way in which an organization creates, manages, preserves, uses, and disposes of electronically stored information.

E-discovery has changed the way in which an organization creates, manages, preserves, uses, and disposes of electronically stored information. It concerns management's ability to achieve corporate goals and to control the exponentially-increasing costs and risks associated with the growth of digital information.

Zubulake's e-Discovery explores the factual underpinnings that caused the establishment of legal and professional precedents. It is an analysis of the strategies I considered and the decisions I made about what happened to me and the resulting judicial rulings that changed the discovery phase of litigation. Zubulake's e-Discovery is written from the plaintiff's perspective - my perspective

I am a business person, not an attorney. The version of events and opinions expressed are portrayed by me from facts and circumstances as I perceived them. The purpose of my book is to educate and inform readers and provide them with a general overview of the topics discussed. It is my hope and intention that by exploring my failures and successes, and understanding the risks I assumed, the reader may learn from my mistakes and benefit from my experiences.

To become familiar with e-discovery requires, at a minimum, a reading of the Zubulake opinions. To appreciate e-discovery requires knowledge of how those opinions came to be. Zubulake's e-Discovery: The Untold Story of my Quest for Justice explains what really happened and how I did what I did. It is a story about perseverance, vindication, accountability, and justice.

Publication Date: Jul 17 2012

ISBN/EAN13: 0985064005 / 9780985064006

LCCN: 2012905837

Page Count: 230

Binding Type: US Trade Paper

Trim Size: 6" x 9"

Language: English

Color: Black and White

Related Categories: Law / General

Where to order: www.LauraZubulake.com

OLP

eDiscovery Salary & Job Duties Survey

Sponsored by Nuix

Law firms pay thousands of dollars to get these answers.
Participate and get free results

Can you answer these questions?

Are you paid fairly?

Do you know where you are on the org chart?

What duties should be corresponding with your title?

Who are preferred vendors?

Are women paid equally?

Does in-house pay more?

Where are the best paying jobs?

What are common billing rates?

How much overtime do most people work?

To find the answers to these and other important questions, take the eDiscovery Salary & Job Duties Survey sponsored by Nuix and get free results!

FASTER SIMPLER SMARTER eDISCOVERY

Smart companies have discovered a better way to navigate through the maze of electronic data using the world's most **advanced eDiscovery and electronic investigation software**.

Powerful enough to process the largest cases and the most complex unstructured data sets, yet intuitive and easy to use.

With Nuix, you can undertake accurate early case assessment and first pass review to understand the key facts of the case in the shortest possible time.

FIND FACTS FAST

info@nuix.com

nuix.com/fasterediscovery



Simple. Powerful. Precise.





The Best Kept Secrets to Using Keyword Search Technologies

By Philip Sykes
and
Richard Finkelman

*Part 1 – Understanding the Search Engines –
A Comparison of dtSearch and Lucene*

Introduction

Keyword searching is, and will continue to be, an important fundamental component of the electronic discovery workspace.

Effectively and correctly used, keyword searching is an important tool for the initial culling of large datasets prior to loading into a document review system. Constructing high-quality searches is a useful skill over the entire span of the litigation process.

General Information about dtSearch and Lucene

The first step in building powerful searches is to understand how the indexing and search engines work. The different eDiscovery tools incorporate various indexing and searching solutions. Some offer more than one.

For example, Relativity uses Microsoft's SQL Server Full Text Search for the document metadata and text. SQL Server Full Text Search has limited functionality, so Relativity also incorporates dtSearch for regular Boolean/proximity searching and Content Analyst for concept searching. Viewpoint also uses dtSearch. Clearwell, Intella, and SHIFT use Lucene for indexing and searching. This article compares these two indexing/search engines, explaining the similarities and differences.

dtSearch is a widely used indexing/searching tool that provides both Boolean and proximity searching options. A single-user dtSearch Desktop license costs \$200.00 and is so useful that you should have it in your suite of software for testing keywords regardless of the eDiscovery applications that you use. dtSearch Desktop provides you with the capability to perform preliminary testing of key custodians' data. This is particularly important when you will be using a solution that uses other indexing/searching technology, since it gives you a control set to use for comparison with the results from the solution's searching tools.

Another reason why dtSearch is useful is that the syntax of the searches is similar to the syntax used by Concordance for full-text searching. If you have built your searches with dtSearch, it's easy to transform the searches to run in Concordance. In addition, many people who work on drafting proposed keyword terms build their lists with a dtSearch/Concordance-compatible structure.

Lucene is an open-source (free) indexing and searching tool that has been used extensively to implement internet search engines. A number of eDiscovery solutions, including Clearwell, Intella, and SHIFT, use this technology, although there are differences in the way the developers have customized Lucene in the various tools. The same search may perform differently when used with these solutions, even with the same set of documents.

Differences between dtSearch and Lucene

Indexing

dtSearch and Lucene index some characters differently. Lucene treats all punctuation and symbols as word breaks. dtSearch is somewhat different, as it assigns all characters to one of four Character Types (Figure 1):

"If you have built your searches with dtSearch, it's easy to transform the searches to run in Concordance."

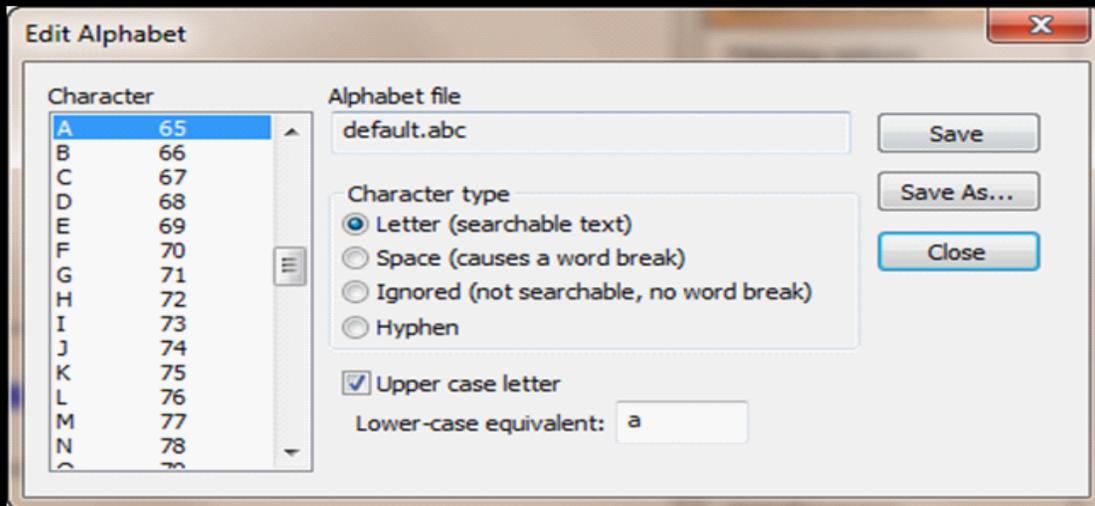


Figure 1: Character Types

dtSearch's treatment of characters is the same as Lucene except for two symbols—at least, if the default alphabet file hasn't been customized. The first difference is that dtSearch treats "_" as a letter, meaning it is indexed and there is no word break. So SmithTom would be indexed as a single word by dtSearch, whereas Lucene would treat it as two words: Smith and Tom.

The other difference is "%", which Lucene treats as a word break, while dtSearch sets it to Ignored and does not treat it as a word break. So Tom%Smith would be indexed as two separate words by Lucene (Tom and Smith), while dtSearch ignores the "%" and adds the word TomSmith to the index. (Note: the "%" is a reserved character in dtSearch that is used for fuzzy searching, so you may see it in search strings.)

The Hyphen Character Type, which by default only includes the hyphen character, is unique because when you build or update an index you have choices on how the Hyphen Character Type is treated (Figure 2).

"dtSearch is a widely used indexing/searching tool that provides both Boolean and proximity searching options. A single user license is \$200."

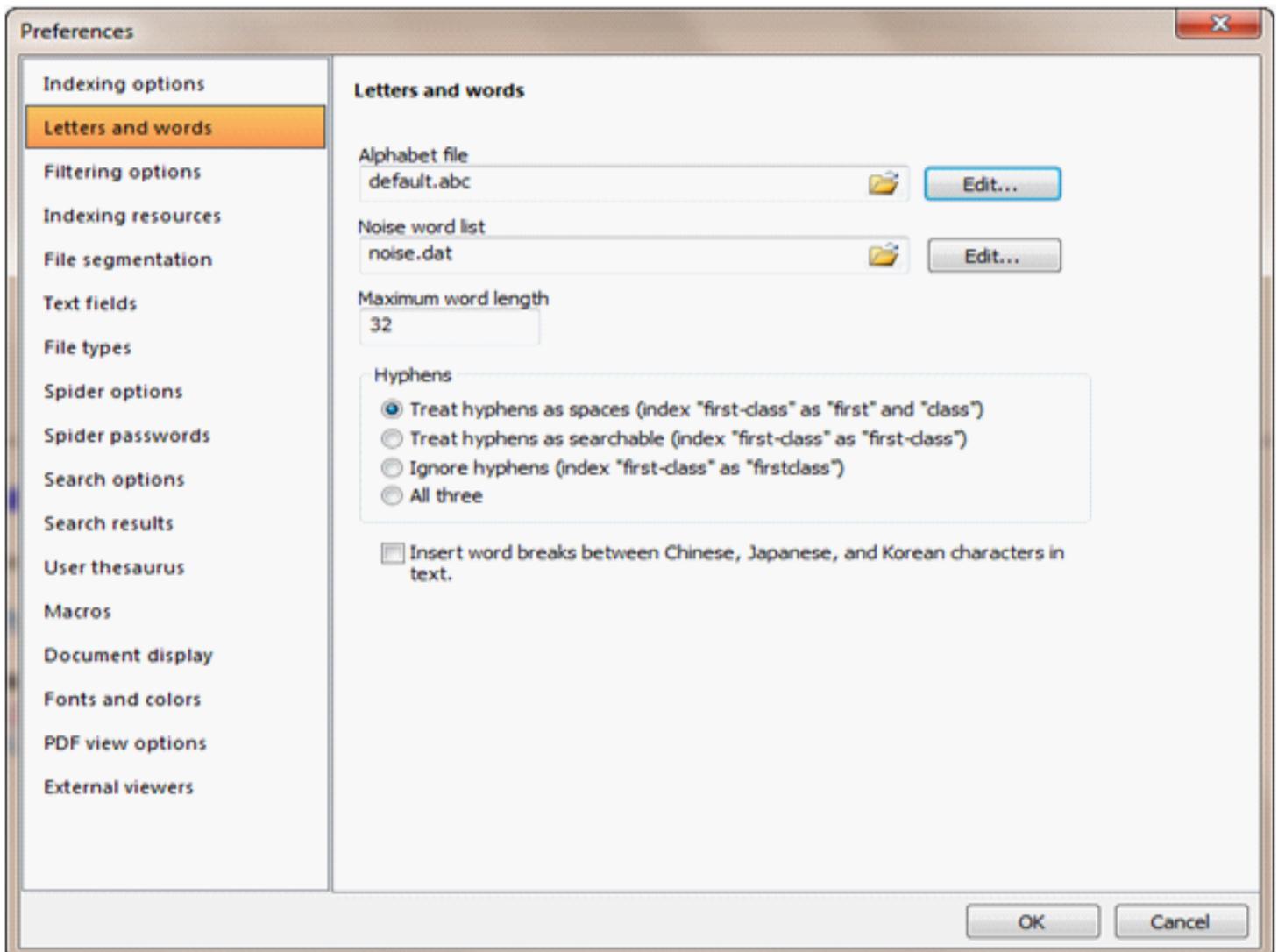


Figure 2

If you select "All three" for Hyphens, the indexing of company-wide would cause the word to be indexed all three ways:

- Spaces, so there are two separate searchable words: company and wide
- Searchable, so company-wide could be searched separately from company wide
- Ignored, so companywide would be searchable.

dtSearch uses a list of Noise (Stop) Words that contains words that are considered too common to provide any value. In addition to common words like all, for, it, this, and was, dtSearch includes the letters a and i as noise words. All other single letters are indexed. dtSearch Desktop allows customization of the noise words list, but embedded versions of dtSearch often do not. One thing to note with dtSearch is how the hit-highlighting appears when there are noise words in your search. If you search for the phrase the records, when the hits are displayed in context every occurrence of records will be highlighted, as well the word immediately preceding it, since the is a noise word. So all records and some records would be considered hits, and both pairs of words would be highlighted.

For Lucene, the use of noise word lists depends upon the tool being used. Some, like Intella, do not use noise words; others do. You will need to refer to the tool's documentation or execute test searches to determine whether or not all words were indexed.

Wildcards

Both dtSearch and Lucene support wildcard characters using "?" as a single-letter wildcard and "*" as a multi-character wildcard. However, Lucene does not support wildcards inside quotation marks. So the search ["sales agreement*"] is an acceptable search using dtSearch, but Lucene ignores the wildcard character and displays results for the search ["sales agreement"]. (Note: in all search syntax examples, the actual search syntax will be inside square brackets "[]".)

Boolean Searches

Both dtSearch and Lucene provide Boolean and proximity searching. When using dtSearch, always have the "Search for" option set to Boolean.

dtSearch

When using dtSearch with the "Search for" option set to Boolean, a series of words is treated as a phrase, and quotation marks are not needed. (Note: if "Search for" is set to "Any words," then the list will be treated as if the words were joined by "OR"; if it is set to "All words," the list will be treated as if the words were joined by "AND.")

Lucene

Lucene's behavior depends upon the customization of the searching by the tool's software development team. Some tools treat a list of terms as if each word is separated by OR, and others treat them as if they were separated by AND.

If you run the search [sales agreement] in dtSearch, it returns documents containing the exact phrase sales agreement. Lucene would return all documents containing either sales OR agreement, or containing sales AND agreement anywhere in the document, depending on the tool's configuration. The best search in Lucene would be ["sales agreement"], since the results would not depend on how lists of individual words are treated.

While Lucene supports a list of terms separated by spaces to search for words, it's recommended that the terms be separated by AND or OR to aid in clarity, since others may not be familiar with Lucene

Lucene's behavior depends upon the customization of the searching by the tool's software development team. Some tools treat a list of terms as if each word is separated by OR, and others treat them as if they were separated by AND.

If you run the search [sales agreement] in dtSearch, it returns documents containing the exact phrase sales agreement. Lucene would return all documents containing either sales OR agreement, or containing sales AND agreement anywhere in the document, depending on the tool's configuration. The best search in Lucene would be ["sales agreement"], since the results would not depend on how lists of individual words are treated.

While Lucene supports a list of terms separated by spaces to search for words, it's recommended that the terms be separated by AND or OR to aid in clarity, since others may not be familiar with the Lucene syntax, and so the query will perform the same in any tool based upon Lucene. Proximity Searching

Proximity searching with dtSearch and Lucene is very different, both in the syntax and how it works.

dtSearch Proximity Searching Syntax

Proximity searches in dtSearch can use either of two operators: W/ and PRE/. So the search [(records w/5 filed)] would return all documents where records has no more than four words between it and filed, regardless of the order in which records and filed appear. The search [(records PRE/5 filed)] would return all documents where records has no more than four words between it and filed and where

records appears before filed. dtSearch accepts phrases and Boolean operators within proximity searches; for example, [(sales agreement w/6 (jones OR smith))].

As an example, we will use two documents. The first, which is returned by this search, contains the text:

“Sales agreement that was executed by William Jones and Thomas Smith.”

The second, which is not returned by this search, contains the sentence:

“Agreement for sales to William Jones by Thomas Smith.”

Note that while the phrase sales agreement must exist in the document for it to be returned by the search, only the last word in the phrase (in this example, agreement) must be within six words of either jones or smith for the document to be returned by the search. In this case, the terms agreement and jones are highlighted. If we increase the word count from six to seven, then both sales and agreement are within seven words of jones, so sales, agreement, and jones are highlighted.

Lucene Proximity Searching Syntax

Lucene’s proximity search syntax is [“Term1 Term2”~N], where “N” is the “edit distance” between the terms. However, Lucene’s approach to proximity searching is different than that used by dtSearch. Where dtSearch counts words, Lucene counts the edit distance value needed to match the words in the text to the words in the search, in the same order they appear in the search.

Therefore, if the query is [“price stock”~4], it is possible that different results will be returned, counts words, Lucene counts the edit distance value needed to match the words in the text to the words in the search, in the same order they appear in the search. Therefore, if the query is [“price stock”~4], it is possible that different results will be returned by the search, only the last word in the phrase (in this example, agreement) must be within six words of either jones or smith for the document to be returned by the search. In this case, the terms agreement and jones are highlighted. If we increase the word count from six to seven, then both sales and agreement are within seven words of jones, so sales, agreement, and jones are highlighted.

Continuing with the example [“price stock”~4], if we have a document that contains the text “stock has a current price,” it takes four shifts to the left to get price in the first position, so the minimum edit distance value is -4; it takes one shift to the right to get stock in the second position, so the maximum edit distance is +1. Subtracting the minimum from the maximum ($1 - (-4) = 5$), so this document would not be returned by this search. It would be returned by [“stock price”~4], because stock is already in the first position, and only three shifts to the left are needed to get price into the second position. Again, subtracting the minimum from the maximum ($0 - (-3) = 3$), which is less than the specified edit distance of 4.

dtSearch and Lucene Proximity Search Comparison

For examples of how dtSearch and Lucene work differently, we will use four different text strings:

1. “price of the company’s stock”
2. “stock doesn’t have an increased price”
3. “price of the common stock”
4. “stock increase when the price”

Using dtSearch, the minimum word count to obtain results is four. The query [(price w/4 stock)] returns numbers 3 and 4. The query [(price w/5 stock)] returns numbers 1, 3, and 4. (Note: the reason why number 1 was picked up by "w/5" but not "w/4" is that the apostrophe in "company's" is Character Type "space" and thus creates a word break.) The query [(price w/6 stock)] returns all four of the strings. Reversing the term order in these queries has no impact on the results. If we switch to the query [(price PRE/6 stock)], only numbers 1 and 3 are returned. Conversely, if we search using the query [(stock PRE/6 price)], numbers 2 and 4 are returned.

Switching to Lucene, the minimum distance with results is three. The query ["price stock"~3] returns only number 1. Both numbers 1 and 3 are returned by ["price stock"~4]. The query ["price stock"~5] returns numbers 1, 3, and 4 (which contains the words in reverse order to the query). The query ["price stock"~7] returns all four strings. Reversing the order of the terms in the query changes the results. Using ["stock price"~3] returns number 4.

Table 1 summarizes the results:

	Query	price of the company's stock	stock doesn't have an increased price	price of the common stock	stock increase when the price
dtSearch Syntax	(price w/4 stock)	NO	NO	YES	YES
	(price w/5 stock)	YES	NO	YES	YES
	(price w/6 stock)	YES	YES	YES	YES
	(stock w/4 price)	NO	NO	YES	YES
	(stock w/5 price)	YES	NO	YES	YES
	(stock w/6 price)	YES	YES	YES	YES
	(price PRE/6 stock)	YES	NO	YES	NO
	(stock PRE/6 price)	NO	YES	NO	YES
Lucene Syntax	"price stock"~3	NO	NO	YES	NO
	"price stock"~4	YES	NO	YES	NO
	"price stock"~5	YES	NO	YES	YES
	"price stock"~6	YES	NO	YES	YES
	"price stock"~7	YES	YES	YES	YES
	"stock price"~3	NO	NO	NO	YES
	"stock price"~4	NO	NO	NO	YES
	"stock price"~5	NO	YES	YES	YES
	"stock price"~6	YES	YES	YES	YES

"Some tools treat a list of terms as if each word is separated by OR, and others treat them as if they were separated by AND."

Lucene also supports more than two terms within the quotes but, since the terms are inside quotes, wildcards are not permitted, and Lucene doesn't support the use of Boolean operators inside quotes. If we need to adapt the dtSearch Boolean operator example [(sales agreement w/6 (jones OR smith))] to Lucene, it would require two proximity searches in order to deal with the OR. The most precise search would be [("sales agreement" AND "agreement jones"~N) OR ("sales agreement" AND "agreement smith"~N)]. The value of N will need to be determined through testing, but a good place to start would be N=5, since it's the dtSearch distance minus one.

A less-precise search would be ["sales agreement jones"~N OR "sales agreement smith"~N]; it is less precise in that it doesn't require the exact string sales agreement, so the terms could be in any order and not adjacent as long as sales, agreement, and jones or smith were within an edit distance of N.

Table 2 shows the value of N required to return this document (they all actually return both documents—a smaller value of N will return the second document, which has sales and agreement, but not sales agreement) using the same two documents as above, with the six possible orders of the three terms:

Lucene Query 1	Lucene Query 2
"sales agreement smith"~8	"sales agreement jones"~5
"sales smith agreement"~10	"sales jones agreement"~7
"agreement sales smith"~9	"agreement sales jones"~6
"agreement smith sales"~11	"agreement jones sales"~8
"smith sales agreement"~11	"jones sales agreement"~8
"smith agreement sales"~12	"jones agreement sales"~9

The above example was designed to show the impact of the term order differences on the value of N. In the real world when using Lucene proximity searches, the best approaches are:

- If there are two words in the search, decide on the optimum number of words between the terms without regard to the term order, and run that search. For example, run the search "stock price"~4. After obtaining the results, run the search again with the terms reversed: "price stock"~4.

The results will normally be lower or higher than the first search. In order to check whether or not the value of N should be adjusted, perform "gap" queries for the search that returned the larger number of documents. For example, run the search [(("stock price"~5) AND NOT ((("stock price"~4) OR ("price stock"~4)))], which will show the documents found by the new value of N that were not found by the previous value of N in either term order. Once you are satisfied with the value of N, which we'll assume is 4, then just run the search as [("stock price"~4 OR ("price stock"~4)].

- If there are more than two words in the search, it becomes more difficult to test individual searches with the terms in different order (Note also that Lucene doesn't include the edit distances for each of the terms in calculating the total edit distance for the document; it only uses the one that is the maximum value and subtracts the minimum value from it.).

Therefore, the most reasonable approach would be to start with an acceptable value for N and note the results. Then increase the value of N by the number of terms, and note those results. Continue to run the search incrementing the value of N, and run "gap" queries to return the documents from the search with the largest N value that were not in the previous N value search until the "gap" queries are not bringing back documents that belong in the review set.

For example, if the search ["term1 term2 term3"~8] returns 500 documents and the search ["term1 term2 term3"~10] returns 600 documents, then run the "gap" query [{"term1 term2 term3"~10) AND NOT ("term1 term2 term3"~8)], and review the documents returned. If there are potential relevant documents, increase the values of N to [{"term1 term2 term3"~11) AND NOT ("term1 term2 term3"~10)], and review the documents returned.

Up Next

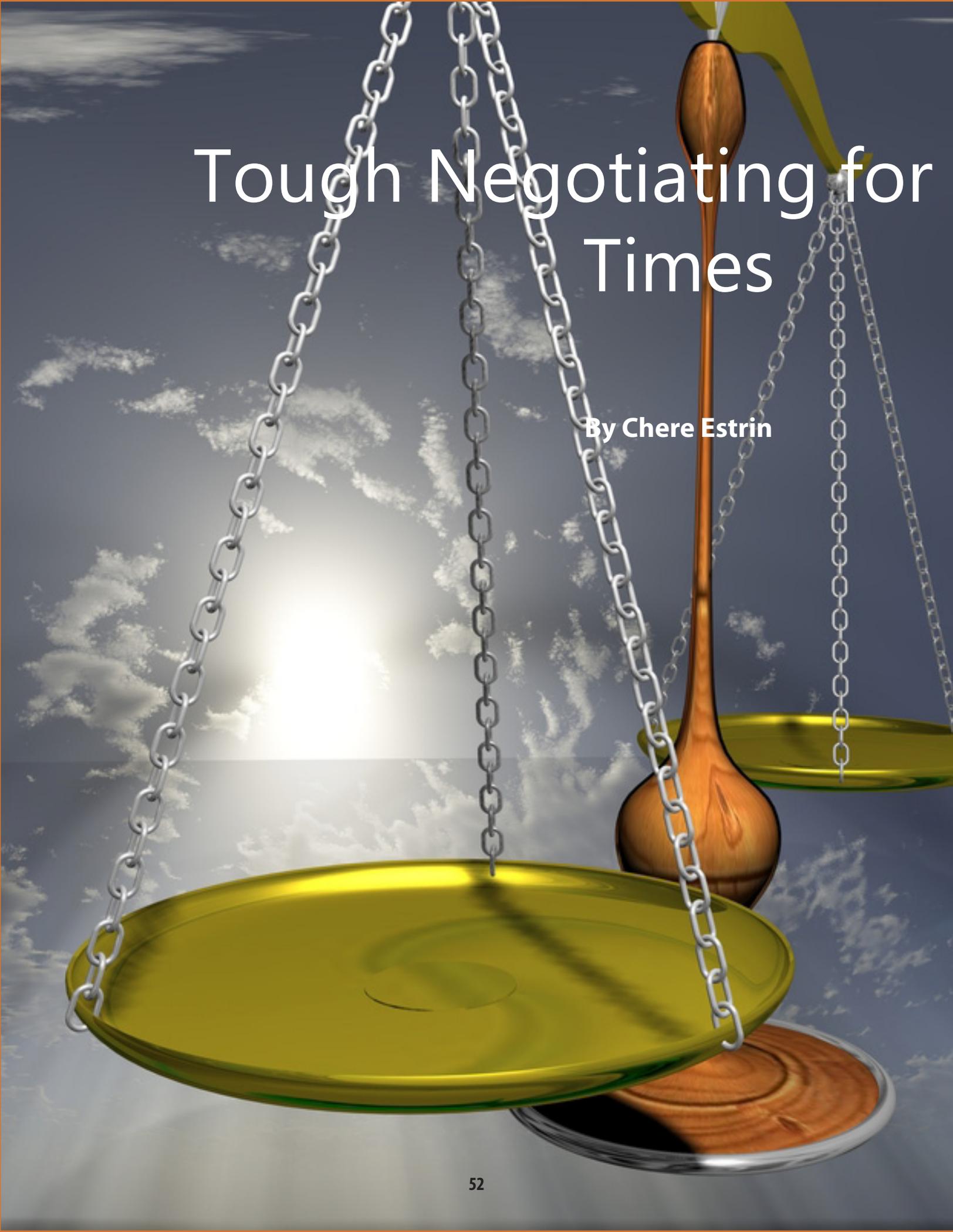
Part 2 will cover using structured search techniques to build your keyword searches.

About the Authors

Philip Sykes is a Senior Managing Consultant with more than 20 years of experience working in the fields of litigation support and electronic discovery. His experience includes work on high-profile cases from HSR Second Requests from the FTC and DOJ to IP matters to complex securities cases. His expertise includes keyword analytics, data processing and analysis, data management, on-line review tools, and document productions. He regularly assists counsel and experts in understanding what information exists in databases and how it is relevant to their matters.

Richard Finkelman is a Director and Practice Group Leader of Berkeley Research Group's Electronic Discovery Practice. Mr. Finkelman brings more than 25 years of experience helping clients manage information in litigation, regulatory and business matters. His experience includes assisting clients with all aspects of litigation support in complex matters ranging from Securities Class Actions to Intellectual Property Disputes to high profile Regulatory matters.



A pair of golden scales of justice is the central focus, set against a dramatic sky with scattered white clouds and a bright sun on the left. The scales are tilted, with the right pan being lower than the left. The pans are a polished, reflective gold color. The background sky transitions from a pale blue near the horizon to a deeper blue at the top, with the sun creating a bright glow and lens flare effect.

Tough Negotiating for Times

By Chere Estrin

Tough

Want the right salary?
This is a harsh new market
calling for skillful handling
of the process.

If you've ever read the list for Top Ten Stressors, losing your job or looking for a new one is right up there with life altering events. Then, after finding that dream job, you hold your breath hoping the right amount of money is offered. That little stressor is guaranteed to give you hives, insomnia or worse, little crow's feet right around your perfectly smooth eye-area.

Relax. Chances are if your strategy is in place, you'll end up getting the dollars you want or at least, pretty close to it. You've got to know at this point in your career, that there are ground rules to the negotiating game. Some of those rules work and some, well, let's just say because nothing in life is a guarantee, they're good to know about anyway.

1. You will be asked, "What is the salary you are seeking?" Don't give a range. It gives employers permission to offer you the lowest amount. Many legal professionals will say, "I'm seeking between \$85 and \$95,000" and then are surprised when they receive the lower figure. It's much better to say, "I'm looking for around \$90,000", or, "I'm seeking the upper \$80's."

2. Never lie about your current salary. There are three things employers can check: dates of employment, reason for leaving and salary.

3. Get your hands around current survey data. If you are offered an amount below market, you might say, "According to recent salary survey information, this offer/raise is slightly below market. How can we work together to bring the up the base?"

4. Do not respond with close-ended statements such as:

- a. This offer is less than what I earn now.
- b. I would like the compensation part of the offer re-evaluated.
- c. Can you change the salary?
- d. The base salary is lower than what I expected.

There's no room for negotiation. You've responded negatively to probably a well-thought out offer/raise.

It is better to respond:

- a. I appreciate the pressure you must be under. I would like to find a way for us to resolve one area of concern. I'm hoping that the firm has some flexibility.
- b. Thank you for this insightful offer/raise. However, there is one remaining issue I hope we can resolve together.
- c. I have some concerns about the level of assignment compared to the corresponding pay. Can you clarify a few things for me?

5. You also might say:

- a. Given your need for someone with my background and given my interest in this firm, can you share with me how we might exercise some flexibility?
- b. How can we reshape this compensation package?

6. If you still can't get them to come up try:

- a. In what ways can we redefine this offer/raise?
- b. Other than compensation, can we think about changing the nature of this offer/raise?

7. You might be able to negotiate additional perks instead:

- a. A better or private office
- b. Additional vacation
- c. Credit for more experience with the firm
- d. Salary in lieu of health insurance (perhaps your spouse has you covered)
- e. Better or stronger title (You can leverage it for more money.)
- f. Continuing education

You are taking a risk when you start negotiating. Employers may say no or even, while rarely, withdraw the offer. However, if you accept an offer knowing that you are not happy, you'll probably be looking to leave in a very short time.

Decide what the salary range is that will be acceptable to you. Factor in current salary; market rate; location (some regions pay less); firm size; market conditions; practice specialty and level of assignment. Don't expect to go much higher than current market unless you have an exceptional skill you bring to the firm.

Walk in prepared for a collaborative discussion, not a fight. A decision to accept is based upon many factors. Salary is only one. Make an intelligent, informed decision. Above all, don't go to the session with an "all or nothing" attitude. You'll be defeated easily.

If you are leaving too much on the table, leave the door open to be reviewed in 3 to 6 months. Ask for a salary not a performance review. If you do get the firm to commit to a mid-year review, get



it in writing. Three months down the line, it will be difficult for the reviewer to recall exactly what was said, or it's even possible that the same person will not be in that position.

Portray yourself as a positive, confident, easy-going legal professional (even if you are not feeling quite so confident) and above all, don't issue an ultimatum. They aren't effective.

If you are working with a headhunter, for heaven's sake, make sure they are skilled in negotiations! Nothing is worse when the headhunter does not understand what you want. If you are negotiating on your own, here is a great opportunity to get involved in the process. Only you can take your career where you want it to be.

About the Author:

Chere Estrin is the Managing Administrator of OLP, an author of 10 books on legal careers. She is a national seminar speaker, recipient of the Los Angeles Women of Achievement award, former CEO of a national staffing organization and administrator in an major international law firm.

“In what way
can we reshape
this
compensation
package?”

A Certain Level of Trust

By Charles Gillis

In the predictable whirlwind of speculation that follows after a major international law firm implodes, many people register their opinions on what went wrong.

It's interesting for some of us to read the post mortem reviews where experts postulate on the poor decisions and missteps that led a troubled firm down the path to their ultimate ruination.



T R U S T

As the latest big firm poured out assets like a well hit piñata I, too, speculated on the root causes of their distress. While I cannot give fresh insight into the mistakes that most certainly happened in the latest spectacle, I can comment on a common problem that plagues law firms.

Although it sounds simplistic and naïve, there was a single moment when a firm like Dewey could have been saved. All failed businesses rest for a moment on a tipping point, teetering on the abyss and at that point decisions are made that will seal their fate. These decisions are based on a critical trust of all the surrounding elements and that's the problem. Trust is a resource that is often in short supply in the legal industry.

Certainly there are numerous factors to consider as we examine the wreckage of a failed firm, but hindsight does wonderful things to our sense of surety. Like any other business, law firms make mistakes. It's why they all carry professional liability insurance.

Despite what comedians might say, lawyers are in fact human, and humans are indeed fallible. We can still trust people who make an honest misstep.

However, it's not necessarily the mistake itself that will steer a firm off the cliff; it's the manner in which the firm reacts to the mistake. The weight of a bad decision is compounded by late reactions, over-correction and the unclear edicts from the leadership. Of course outright lies are like salt on the field – they kill current growth and make it hard to cultivate much in the future. Before you can fix anything you have to trust your surroundings and your colleagues.

The benefit of the doubt is not freely granted in our profession. It must be cultivated, even amongst colleagues. It is no great surprise that firms who engender trust amongst their teams create a more efficient, productive and ultimately more profitable work environment. Understanding that people need to be able to trust one another is the easy part of the equation. Actually getting to them to trust one another is entirely different.

A culture of trust is the result of ongoing shared experiences and a constant reinforcement of values. You cannot issue a proclamation to build trust. You cannot buy it. The inability to purchase trust confounds some firms. Firms who believed they were securing loyalty and commitment through lavish compensation, benefits and perks soon realized the shallowness of their connection once the economy caused the benefit pool to dry up.

Often the qualities that people value most are the non-financial benefits of their work environment. Satisfied employees feel a connection to something bigger especially if they have an opportunity to contribute to its success.

People who have faith in their employer help build traditions and memories. Engaged employees are not merely advocates for firm culture, they help create the firm culture. It is precisely this level of commitment that drives long-term trust and success in an organization.

**“The
benefit of the
doubt is not
freely granted
in our
profession.”**

So how do you earn trust? Clearly we must honor our commitments and be candid and truthful. This may not be enough. The desire for a trusting environment is not enough, you must pursue it. Sometimes we can make a connection with our colleagues through very simple means.

Recently at my firm, we strengthened our connection with our people and all that was required was a lunch, an hour of time and three bar stools. At our recent “Founder’s Lunch” three of the original named partners of the firm sat casually before a packed room and recounted the story of the Firm’s creation.

Without a script or teleprompters, they spoke of the early salad days of the firm, back when they had little but their wits. They spoke of the steep learning curve and the lessons learned early on. The war stories were humble and at times charmingly self-deprecating.

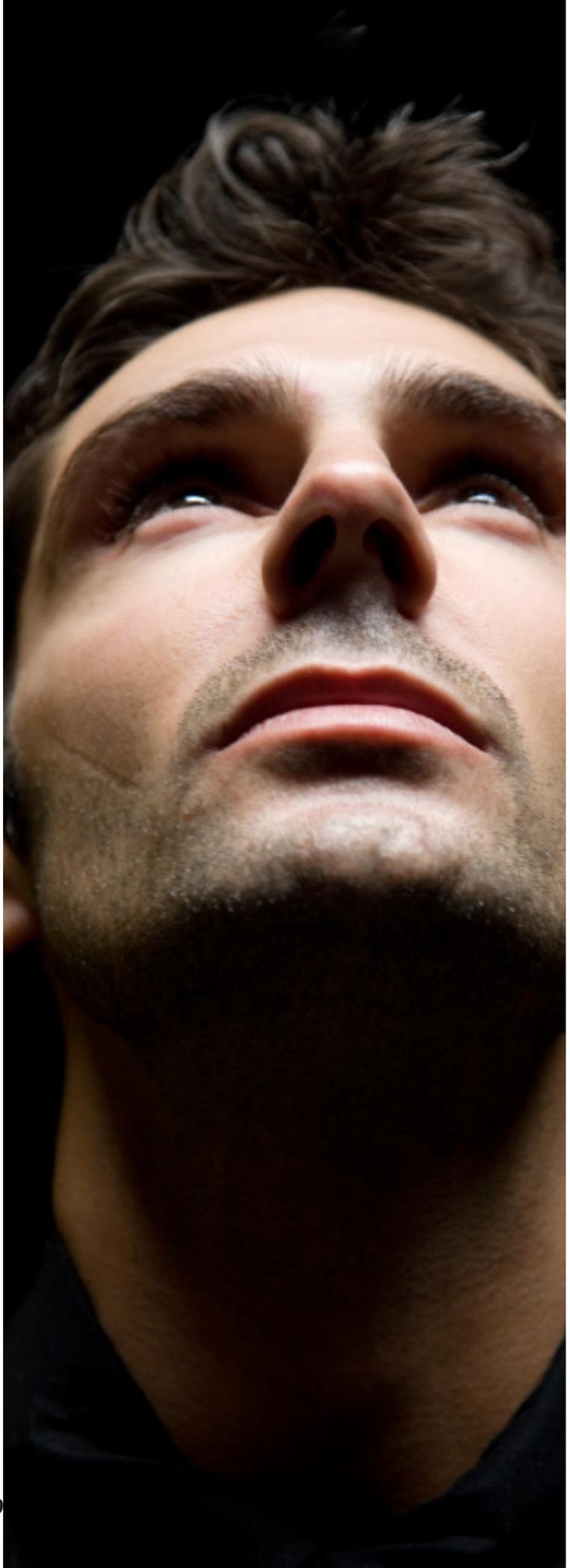
The investment of an hour of time brought people together in a shared experience. They were joined in the moment, but the stories made them feel as if they knew what it was like to be together 27 years ago when the firm was new and the future was uncharted. This outreach made everyone realize that they were part of the history. The decisions of the Founders helped create the present environment. It gave people faith in a shared destiny.

Having faith in the past made trusting the future a lot easier. As I recall the litany of failed firms, I can’t help but think how a simple event of this nature might have helped these other firms. When times get tough, trust may be the only thing left. I’m quite certain that makes all the difference.

About the Author:



Charles Gillis is the Executive Director of Texas law firm Munsch Hardt Kopf & Harr, P.C.





ORGANIZATION OF LEGAL PROFESSIONALS



Do you have the tools you need?

OLP is committed to gold-standard continuing legal education and eDiscovery certification exams. Whether your career is focused exclusively on litigation, eDiscovery or litigation support or you just want to know more about legal technology, OLP is your source for information and support.

**SHARPEN YOUR SKILLS AND STAY INFORMED BY
BECOMING A MEMBER TODAY**

www.theolp.org
info@theolp.org 760-610-5462



ORGANIZATION OF LEGAL PROFESSIONALS

Organization of Legal Professionals (OLP)
44-489 Town Center Way Ste. D436 Palm Desert, CA 92260
760-610-5462 info@theolp.org www.theolp.org