



Best Practices: eDiscovery Search

Improve Speed and Accuracy of Search Using the Latest Techniques

September 14, 2016

**eDiscovery
Webinar Series**

Erin Derby
eDiscovery Specialist
Lexbe

eDiscovery Webinar Series

About our Webinars

- Webinars take place monthly and cover a variety of relevant eDiscovery topics
- If you have technical issues or questions, please email webinars@lexbe.com
- Lexbe webinars are available for viewing (video), and as PDF or an MP3 podcast. This Webinar and a complete listing of other onDemand webinars is part of the: [Lexbe eDiscovery Webinar Series](#)
- For notices of future live and on-Demand webinars as part of this series, email us at webinars@lexbe.com; [Follow us on LinkedIn](#)

Next Month's Topic: In-house eDiscovery

**eDiscovery
Webinar Series**

Best Practices: Best Practices: eDiscovery Search | September 14, 2016

lexbeSM

eDiscovery Webinar Series

Erin Derby bio

- eDiscovery Specialist at Lexbe, specializing in working with clients handling eDiscovery in complex litigation, requiring a high level of precision and expertise.
- Certified by the Association of Certified E-Discovery Specialists (ACEDS).
- Erin is also a Litigation Paralegal with 10 years experience with both plaintiff and defense law firms in a variety of practice areas.



Erin Derby
800-401-7809

(512) 956-5594

ederby@lexbe.com



**eDiscovery
Webinar Series**

Best Practices: eDiscovery Search

Agenda

- Understanding How eDiscovery Search Indexes Work
- Multi-Index Approach to Prevent Missing Critical Case Data
- Pros and Cons of Keyword Search
- Constructing Quality Search Queries
- Search Expanders and Search Limiters
- Testing Keyword Search Results
- Practice Tips

Best Practices: eDiscovery Search

Introduction

- By understanding the latest tools, indexing techniques, and features associated with advanced eDiscovery search, you will conduct a highly effective review while still delivering your production on time.
- Applying the information from today's webinar will aid in finding the needle in the ESI haystack with increased speed, efficiency and accuracy for your current and upcoming cases.

Best Practices: eDiscovery Search

Search Indexing Types

- **Imaged Based:** Search results are generated only from imaged/OCRed documents, which capture image-based text.
- **Native Extraction:** text that is excluded from print and/or hidden/embedded text is extracted and converts the Native files into PDF and runs the documents through OCR.
- Lexbe, offers a “Multi Index,” which is a combination of these two. Please check with your vendor to see if your platform provides this multi index approach.
- Lexbe also offers an additional index to search across translated documents, using a multi-lingual search index, as well as a foreign language concept search.

Best Practices: eDiscovery Search

Multi-Index Index

- The Lexbe search engine indexes both text extracted from Native files (email, attachments and spreadsheets), a paginated file converted from Native files into PDF or TIFF and OCR'd, and any translated versions.
- Most comprehensive approach and minimizes potential of lost data.

Benefits of Multi-Index Approach

Index Method	Captures Embedded Text	Captures Text Excluded From Print	Captures Hidden Text
Imaged/OCR	Yes	No	No
Native Extraction	No	Yes	Yes
Lexbe Multi-Index	Yes	Yes	Yes

Best Practices: eDiscovery Search

Multi-Index Index

Native text extraction picks up changed text, but imaged OCR does not

Page 1/12

Chewco 2000 Pro Forma Balance Statement Sheet

CRITICAL ENRON EVIDENCE

	9/30/2000	9/30/2000	12/31/1999
ASSETS (Pro Forma)			
Current Assets			
Cash and Cash Equivalents	\$ 629,461	\$ 85,688	\$ -
Accounts Receivable, net	15,289	15,289	2,880
Stock Subscription Receivable	-	-	100,000
Inventory	48,429	48,429	-
Prepaid Expenses	6,417	6,417	2,302
Total Current Assets	699,596	155,823	105,181
Fixed Assets, net	10,163	10,163	235
Other Assets	11,531	11,531	7,698
Intangible Assets, net	17,059	17,059	-
Total Assets	738,350	194,577	113,115
LIABILITIES AND STOCKHOLDERS' EQUITY			
Current Liabilities			
Accounts Payable	50,998	50,998	1,615
Accrued Liabilities	4,624	4,624	215
Due to Affiliate	16,854	16,854	10,045
Deferred Compensation	8,836	8,836	-
Total Current Liabilities	81,312	81,312	11,876
Stockholders' Equity			
Common Stock, \$.01 par value	-	-	-
Preferred Stock \$.01 par value	576	244	200
Paid-in Capital	-	-	-
Accumulated Deficit	768,800	225,359	101,039
	(112,338)	(112,338)	-
Total Stockholders' Equity	657,038	113,265	101,239
Total Liab. & SH Equity	\$ 738,350	\$ 194,577	\$ 113,115

BODY TEXT

Page 1/12

'Chewco 2000 Pro Forma Sheet'

	9/30/2000	9/30/2000	12/31/1999
ASSETS (Pro Forma)			
Current Assets			
Cash and Cash Equivalents	\$ 629,461	\$ 85,688	\$ -
Accounts Receivable, net	15,289	15,289	2,880
Stock Subscription Receivable	-	-	100,000
Inventory	48,429	48,429	-
Prepaid Expenses	6,417	6,417	2,302
Total Current Assets	699,596	155,823	105,181
Fixed Assets, net	10,163	10,163	235
Other Assets	11,531	11,531	7,698
Intangible Assets, net	17,059	17,059	-
Total Assets	738,350	194,577	113,115
LIABILITIES AND STOCKHOLDERS' EQUITY			
Current Liabilities			
Accounts Payable	50,998	50,998	1,615
Accrued Liabilities	4,624	4,624	215
Due to Affiliate	16,854	16,854	10,045
Deferred Compensation	8,836	8,836	-
Total Current Liabilities	81,312	81,312	11,876
Stockholders' Equity			
Common Stock, \$.01 par value	-	-	-
Preferred Stock \$.01 par value	576	244	200
Paid-in Capital	-	-	-
Accumulated Deficit	768,800	225,359	101,039
	(112,338)	(112,338)	-
Total Stockholders' Equity	657,038	113,265	101,239
Total Liab. & SH Equity	\$ 738,350	\$ 194,577	\$ 113,115

'Body Text'

How Doc Appears Natively

OCR Based Index



Best Practices: eDiscovery Search

Use of Keyword Search in Discovery

- Early Stage Culling - Reduce amount of ESI for review by using keywords, date ranges, custodians to cull document collections.
- Keyword-Based Responsive and Privilege Review - Construct search queries to return documents that are likely to be responsive and/or confidential.
- ID Documents for Depo Prep - Find and assign key documents related to specific case participants to prepare for depositions.
- ID of Key Docs for Trial - Find and mark key case documents. Code documents that will be needed for trial.

Best Practices: eDiscovery Search

Pros of Keyword Searching

- Fast - Keyword search is very fast compared with other document search methodologies.
- Inexpensive - Good results are obtained at little cost compared with manual review or other computer-assisted methodologies.
- Quality - Search can deliver high quality results, particularly if keyword terms are carefully developed and tested.
- Avoids Manual Review Errors/Inconsistencies - Search results are computer generated and avoid known human review errors that can result from fatigue and inadequate training or lack of focus.

Best Practices: eDiscovery Search

Cons of Keyword Searching

- Search Can be Over or Under-Inclusive - Search terms can bring back too many junk results or miss good results. These are known as “false positives” and “false negatives.”
- Difficulty of Creating Good Search Terms - Constructing good search terms takes design time, testing, iterations and analysis.
- Non-Searchable Text - Search results are only as good as the underlying searchable text. ESI collections and review tools can miss text that a human reviewer might catch for a variety of reasons.
- Some file types cannot be indexed - There is little consistency in what files can be indexed across databases.

Best Practices: eDiscovery Search

Alternatives to Using Search

eDiscovery searching is not a remedy for all cases and other techniques are used when appropriate.

- Predictive Coding - For very large cases, predictive coding, aka technology assisted review, may be used to identify documents for production or for other purposes. Predictive coding requires training of a seed set of documents that are marked responsive or non-responsive, and a computer algorithm applies learning from the seed set to the remainder of the review set.
- Manual Linear Review of All Documents - For smaller document sets, or even larger sets with a large review budget, attorneys or other reviewers may manually review all potentially responsive documents.

Best Practices: eDiscovery Search

Construct Quality Searches

- Start with Request for Production or Subpoena - Translate the demands of the RFP/SDT into a keyword search strategy.
- Interview Custodians - Ask key case participants/data custodians about their ESI. Use their insights and their terminology to find obscure key documents.
- Include Jargon - Seek out industry or company sub-culture specific terms and abbreviation/acronyms you may not be familiar with.
- Included Misspellings - Include misspelled versions of keywords or (use "fuzzy search" settings or Boolean limiters) in your search string to account for emails or other words with typos.

Best Practices: eDiscovery Search

Use Search Expanders

Search Expanders Enable Easy Expansion to Reduce False Negatives

- Fuzzy - Fuzzy searching will find a word even if it is slightly misspelled, or converted improperly by the OCR process. For example, a fuzzy search for "litigation" will find "liti3ation."
- Phonic - Phonic searching looks for a word that sounds like the word being searched and begins with the same letter.
 - a phonic search for "Invoice" will also find "Investors", "Infection" and "Investment."

Best Practices: eDiscovery Search

Use Search Expanders

Search Expanders Enable Easy Expansion to Reduce False Negatives

- Concept - Concept searching expands a search beyond specific keywords to include related concepts (synonyms and derivatives).
 - A search without concept search enabled for the term "injury" would return documents that included the term injury and derivatives, like injuring and injuries. However, a concept search on "injury" would additionally return documents that mention trauma, abrasion, wound, etc.
- Stemming - Stemming extends a search to cover grammatical variations of a word. A search for "fish" would also find "fishing."

Best Practices: eDiscovery Search

Use Search Expanders

Fuzzy Search Example

Fasto = Fastow

Show Searches For

- Current Case
- All Cases In Account

Options

All of the words

- Concept Search
- Fuzzy Search
- Stemming Search
- Phonic Search

Fasto

Select Docs: [All](#) [Page](#) [None](#) Page: [Prev](#) [Next](#)

Documents found: 698


	Title	Rank	Bates
<input type="checkbox"/>	ORIGINAL Hearing Enron ...here today will appear only briefly. Mr. Fastow , Kopper, Causey &	100	LEX 0000366-0000855
<input type="checkbox"/>	2001-10-23 01:08:02PM - Thank you and Stay Strong ...PM To: Lay KennethKenneth.Lay@ENRON.com; Fastow Andrew	40	
<input type="checkbox"/>	fastows-brady-letter-june-1-2004 ...s interest in the Nigerian Barge. Andrew Fastow Fastow was int	38	enron 0006540-000654

Best Practices: eDiscovery Search


Use Search Expanders


Phonic Search Example

Invoice = Investigation/Invest/Investor

Show Searches For 

Current Case
 All Cases In Account

Options 

All of the words 

Concept Search
 Fuzzy Search
 Stemming Search
 Phonic Search

Invoice

Select Docs: [All](#) [Page](#) [None](#) Page: [Prev](#) [Next](#)

Documents found: 3,482


	Title	Rank	Bates
<input type="checkbox"/>	ORIGINAL Hearing Enron ...HEARING BEFORE THE SUBCOMMITTEE ON OVERSIGHT AND INVESTIGATIONS OF	100	LEX 0000366-0000855
<input type="checkbox"/>	Memorandum re Investment ...Pickennng PRIVATE PLACEMENT MEMORANDUM CONFID~ IJM2 CO- INVESTMENT , L.	94	LEX 0000856-0000905
<input type="checkbox"/>	ConfMemo-1 ...all of the information that a prospective investor may desire. In all cases, interested p	22	enron 0000142-0000192

Best Practices: eDiscovery Search


Use Search Expanders


Concept Search Example

Trade = Swap

Show Searches For 

Current Case
 All Cases In Account

Options 

All of the words 

Concept Search
 Fuzzy Search
 Stemming Search
 Phonic Search

trade

Select Docs: [All](#) [Page](#) [None](#) Page: [Prev](#) [Next](#)

Documents found: 2,602

	Title	Rank	Bates
<input type="checkbox"/>	2000-09-20 08:33:00AM - trading day 9/19 ...CAL Swap 2000/09/01 2000 2000/09/19 9 Q3 Shoulder 0 0 0 ..	100	enron 0000496
<input type="checkbox"/>	west tot belden2 ...PWR COB COB West WSCC ST CAL Swap 2000/08/01 2000 200	99	
<input type="checkbox"/>	Enron email with attachments for upload	99	

Best Practices: eDiscovery Search

Use Search Expanders

Stemming Search Example

Trade = Trading

Show Searches For

Current Case
 All Cases In Account

Options

All of the words

Concept Search
 Fuzzy Search
 Stemming Search
 Phonic Search

Search Quick Links

trade

Select Docs: [All](#) [Page](#) [None](#) Page: [Prev](#) [Next](#)

Documents found: 3,051

	Title	Rank	Bates
<input type="checkbox"/>	Cash Analysis ...Exploration 0 23580 35004 35004 Aquila Capital & Trade , Ltd. C	100	enron 0004748-0005373
<input type="checkbox"/>	Market Data Costs 11-9 ...Data Costs by Cost Center NA-Currency Trading \$539,271.24 N	75	
<input type="checkbox"/>	Market Data Costs 11-9 ...Data Costs by Cost Center NA-Currency Trading \$539,271.24 N	75	enron 0001616-0002081
<input type="checkbox"/>	Cash Analysis		

Best Practices: eDiscovery Search

Using Search Expanders

Basic Boolean Operators:

- AND: returns results including both terms
- OR : looking for at least one of a list of terms
- NOT : exclude terms you don't want
- () : can be used to separate OR statements from the rest of the Boolean string
- PRE/*n* : First search term does not precede the second term by more than *n* words
- Wildcard Characters: '*' replaces a letter in your search term, '!' allows for stemming search within a Boolean query

Best Practices: eDiscovery Search

Using Search Expanders

Boolean Search Example

Lay W/5 Equity

Show Searches For

Current Case
 All Cases In Account

Options

Boolean ▾

Concept Search
 Fuzzy Search
 Stemming Search
 Phonic Search

Lay W/5 Equity

Select Docs: [All](#) [Page](#) [None](#) Page: [Prev](#) [Next](#)

Documents found: 2

	Title	Rank
<input type="checkbox"/>	ORIGINAL Hearing Enron ...smaller d stnbuton may indicate thai Bare lay's equity as disproportionate and perha	100
<input type="checkbox"/>	oct23allemployeemeeting1 ...a guarantee for the partnership 14 KEN LAY : We adjusted our equity . But we also toc	25

Best Practices: eDiscovery Search

Finding Personally Identifiable Information

Search for PII

- Social Security Numbers Search Options
 - Use Boolean option and wildcard '==' as a sequence of 09 digits:
 - 1- Wildcard "=== == =====" (spaces)
 - 2- Wildcard "===== " (w/ spaces)
 - 3- Wildcard "===-===-=====" (dashes)
 - 4- Search "SSN"
 - 5- Search "Social Security Number"

- The same logic can be used for credit card numbers, birthdates, phone numbers, or any other PII needing to be redacted from documents prior to production or Court filings.

Best Practices: eDiscovery Search

Finding Personally Identifiable Information

Search for PII

Phone Numbers "=== === ==="

Show Searches For

- Current Case
- All Cases In Account

Options

Boolean ▾

- Concept Search
- Fuzzy Search
- Stemming Search
- Phonic Search

Select Docs: [All](#) [Page](#) [None](#) Page: [Prev](#) [Next](#)

Documents found: 5,199

	Title	Rank	Bates
<input type="checkbox"/>	Exec Committee List ...CELL PHONE ASSISTANT'S PAGER Baxter, Cliff 713-853-7357 281-265-0813 713-	100	enron 0002305-0002306
<input type="checkbox"/>	Exec Committee List 091101withMgmtCmt ...CELL PHONE ASSISTANT'S PAGER Buy, Rick 713-853-4739 713-599-0794 713-5-	84	enron 0003338-0003339
<input type="checkbox"/>	Native-1f225b81-518c-4d23-9bd2-725d57bbddeb ...Interview Alatorre Carlos Associate EOL Commercial EB2639g 713-345-8705 carlos.a	75	

Best Practices: eDiscovery Search

Using Search Limiters

Search Limiters Reduce False Positives (Noise)

- Filter Out Unneeded File Types. Some file types are unlikely to lead to useful information and can be excluded.

- Use Boolean Modifiers to Limit Overly Expansive Searches.
 - Boolean modifiers can reduce the number of documents returned from a query while increasing the relevance of those files.
 - Exclude certain words or combinations and specify word order.

Best Practices: eDiscovery Search

Test Keyword Searching Results

- Look at Results Returned. Searching without review and testing results in low quality results.
- Sample and Look for Ways to Limit Search. Create new queries that reduce false positives.
- More new keywords. Viewing search results prompt the discovery of additional keywords that could be used to expand or reduce search queries.
- Fuzzy and Concept Search. Find new keywords by searching and returning synonyms and near identical words.

Keyword searching becomes an iterative process.

Best Practices: eDiscovery Search

Ways to Improve Searchability

- Understand Data Collection. Understand composition of collection, file types, special issues. Not everything is ready to be indexed/searched. Some data will need processing.
- Understand Software Used for Processing & Search - Indexing tools create database entries for files by opening, scanning, and storing specific content in 'fields', often based on the data's location within the document, e.g. header data fields and body content fields.
 - If indexing tools can't open a file or see specific forms of data within the file, it will not be able to index it and the data will not be searchable.

Best Practices: eDiscovery Search

Practice Tips - Basic Searching

- Plan your document strategy in advance
- Abstract Requests for Production
- Cull document collections by using keywords
- Focus on keyword-based responsive and privilege review
- Code and index documents for trial, depositions
- Include misspellings in keyword searches

Best Practices: eDiscovery Search

Practice Tips

Use the following in your eDiscovery Searches:

- Concept and phonic search
- Fuzzy and stemming search
- Boolean search
- Keyword highlighting
- Indexing of extracted text and multi-indexing of natives/paginated file types
- Dual Index
- Search Expanders

Thank You

Thank You for Attending

We'll be making the following available to webinar attendees:

- A recorded streaming version
- MP3 podcast
- PDF

Please let us know if you have any questions or comments about this webinar or suggestions for future topics. This webinar is part of the [Lexbe eDiscovery Webinar Series](#). For notices of future live and on-Demand webinars as part of this series please email us at webinars@lexbe.com or [Follow us on LinkedIn](#).

Presenter:

Erin Derby
ederby@lexbe.com
(512) 956-5594

Moderator:

Christian Detrude
cdetrude@lexbe.com
(512) 213-6394

Webinar Questions:

webinars@lexbe.com

**eDiscovery
Webinar Series**

Best Practices eDiscovery Search | September 14, 2016



Thank You

What Attendees are Saying:

- "Excellent presentation! One of the best webinars I have attended!"
- "Time well spent."
- "Great in terms of content and presentation. Thanks!"
- "Excellent and informative piece!"

For notices of future live and on-Demand webinars as part of this series please email us at webinars@lexbe.com or [Follow us on LinkedIn](#).

Lexbe eDiscovery Platform

About Lexbe

Lexbe is an Austin, Texas- based eDiscovery software and services provider, specializing in serving small & medium-sized law firms and organizations.

- The Lexbe eDiscovery Platform, is our cloud-based processing, review and production tool. Designed for attorneys and legal staff to be DIY and easy to use with no users fees or case fees. Free standard loading with annual plans.
- Learn about our high-speed/high-capacity eDiscovery services, and expert professional services.

Request a personalized demo and expert consultation today!



"Cost-effective eDiscovery"



"Lexbe cost advantages, SaaS convenience and search capabilities appeal to many small firms."



"Secure, easy-to-use and a great review tool for consideration"



"Because of the Lexbe software, the entire playing field has been leveled for my firm."



"A powerful litigation document management service"



"Lexbe is the easiest eDiscovery software I have ever used."

Lexbe Sales
sales@lexbe.com
(800) 401-7809 x22

**eDiscovery
Webinar Series**

