The Bell Curve
& Document Indexing/Imaging

This white paper applies the "bell curve" from statistics class to the document collections produced during discovery. Out of every thousand cases, how many involve 1,000,000+ documents? 100,000+? 100+? And what does this distribution suggest regarding document indexing and the possibility of using Adobe® Acrobat PDF files for document images?

Introduction

Remember the "bell curve" from statistics class? The bell curve, so named because of its shape, illustrates the frequency distribution of many phenomena, for example, height. Measure a thousand people. For every person over 7', you'll have a mob between 5'6" and 5'10".

Let's apply the bell curve to the document collections produced during discovery. Out of every thousand cases, how many involve 1,000,000+ documents? 100,000+? 100+? And what does this distribution suggest regarding document indexing and image handling strategies?

LexisNexis®

Giant Cases— Special Tools Required

We're all familiar with cases in which millions of documents are produced during discovery. But we've also seen individuals over 7' tall. Both are outliers occurring infrequently. Out of every thousand cases, maybe two or three have 1,000,000 or more documents.

Cases with document collections of over 100,000 are also relatively rare. Do even a hundred cases out of every thousand involve this many docs? Widespread use of email has definitely increased the volume of documents produced in many cases, but it hasn't turned every case into a document monster.

Dealing with 1,000,000+ documents or even 100,000+ justifies a substantial investment in scanning and coding. This type of case also demands sophisticated software products such as Concordance®, iCONECT®, IPRO®, Litigator's Notebook®, or Summation® to assist with document indexing, image handling and more.

Our CaseMap® case analysis tool includes document indexing functionality analogous to that found in the applications mentioned above. However, while CaseMap is the place to organize facts and issues for every size case, it isn't intended to be the document indexing solution for mammoth cases. CaseMap isn't philosophically or physically designed to house an index of 100,000+ or 1,000,000+ documents. (Please note that most products used in mass document cases, e.g., all of the products listed above, feature tight integration with CaseMap. This integration makes it easy to marry your CaseMap fact analysis to the source documents stored in another application.)

So that's the story for the giant cases lurking out in one tail of the bell curve. But what about the cases that fall under the rest of the curve? How many documents do these cases involve? What's an appropriate image handling and document indexing solution for them?

Normal Cases— Perfect for Adobe® PDF Files

Cases with very small document collections fall at the other end of the spectrum. For every 1,000,000 document case, there's a case that involves, literally, a handful of documents. For every case with 100,000 documents, there's a case with 100. But these matters with only a single redwell of docs are as atypical as those with massive quantities of documents.

Which brings us to the approximately 70% of all cases that fall under the center area of the bell curve. Our experience suggests these cases have between 1,000 and 25,000 documents—a small number relative to a gargantuan million doc case, but still a heap of paper. These matters have far more documents than any trial team can memorize the details about or can organize effectively without indexing and imaging.

The good news is that the ubiquitous Adobe PDF format provides an excellent method for handling document images for this size case. It's likely you're already adept at working with PDF files, perhaps in connection with court-filing requirements. It's also a good bet that expert witnesses and even clients are familiar with using PDF files and have either a full Adobe Acrobat® license or the free Adobe Reader, making it a cinch to share case documents.

The PDF format has become the de facto standard for electronic versions of paper documents. Why? Because a single PDF file can contain the images of all pages of the paper document as well as the associated document text, which can be captured by optical character recognition (OCR) software.

Prior to the introduction of the PDF format, the standard technique for creating electronic versions of paper documents was to generate a series of single page TIFF images and a separate OCR text file. Thus, scanning a 15-page document would yield a total of 16 separate electronic files—15 TIFFs and a text file. When scanning first became available, the Many Electronic Files = 1 Paper Document approach was as good as it got and undoubtedly beat having nothing at all. However, with the advent of the PDF, which meant that 1 Electronic File = 1 Paper Document, it wasn't long before this new format ruled the roost.

The argument in favor of using PDFs to handle document imaging on small to mid-sized cases gained strength following Adobe's release of Acrobat 7. It has become stronger still with the introduction of Acrobat 8. These new versions offer numerous enhancements that are particularly useful to trial teams. For example, Acrobat 7 offers improved document mark-up functionality. It also provides cross-PDF text searching. This feature permits you to search a folder containing any number of PDFs for those documents that include a particular word or phrase.

Normal Cases—Perfect for Document Indexing in CaseMap

On any case where you use Acrobat PDF files to handle document images, please consider employing our CaseMap case analysis software to manage document indexing and to organize the PDF collection.

Every CaseMap file contains a fact spreadsheet, a cast of characters spreadsheet and an issue spreadsheet that are designed for use in even the largest of cases. In addition, each CaseMap file includes a document spreadsheet that provides an easy-to-implement, yet powerful document indexing solution appropriate for matters with small to mid-sized document collections.

The document spreadsheet in each CaseMap file contains a series of predefined columns for building a detailed doc index. These columns include Date, Bates-Begin, Bates-End, Description, Type, Authors,

"The PDF format has become the de facto standard for electronic versions of paper documents."

Recipients, Copied To, Mentioned In and Linked Issues. There's also a Linked File column, which can be used to connect the document index spreadsheet to PDF files. Something missing? Select from other predefined columns or easily create any number of new columns from scratch.

The display of the document spreadsheet can be customized as you desire. Choose which columns are included in the spreadsheet and the order in which they appear. Determine whether spreadsheet rows are listed by document date, by Bates number or by some other criterion.

The spreadsheet can also be easily filtered based on values in any columns. For example, it takes two mouse clicks to filter the spreadsheet down from all documents to just those written by a particular witness, or related to a particular case issue, or flagged as being critical. Compound filters can be run as well. For example, filter the document spreadsheet down to just those documents drafted by a specific author and sent to a specific recipient.

Begin Bates	End Bates	Full Name	Date	Type +	Key	Author(s)	Recipient(s)	Linked Issues
P1234	P1256	Hawkins Employment Agreement	06/??/1998		☑	William Lang	Philip Hawkins, Frank Vanaro	Wrongful Termination
P1257	P1265	Hawkins Performance Review	08/??/1999	Performance Review	☑	William Lang	Philip Hawkins, William Lang	Transfer.
P1266	P1266	Hawkins Letter of 9/23/99	Thu 09/23/1999	Letter	☑	Philip Hawkins	???	Hawkins Deserved Termination
P1267	P1278	Lang Reduction In Force Memo	Wed 11/10/1999	Internal memo	☐	William Lang	Anstar Biotech Industries Staff	Wrongful Termination, Age, Related Acts.
P1279	P1279	Lang Memo to Regan	Fri 11/19/1999	Internal memo	☑	William Lang	George Regan	Demotion
P1280	P1283	Sheridan Memo	Mon 01/10/2000 12:13 p.m. ET	E-mail	☐	Susan Sheridan	Frank Vanaro, George Regan	Pattern & Practice
P1284	P1284	Anonymous Complaint Letter	Thu 01/20/2000	Letter	☐	???	Frank Vanaro	Hawkins Deserved Termination
P1315	P1340	Workforce Analysis Report - Draft	09/??/1999		☐	Henkle & Lee	???	Wrongful Termination, Age, Related Acts.
P1341	P1341	Vanaro Alert Email	Fri 01/21/2000 9:31 a.m. ET	E-mail	☐	Frank Vanaro	Philip Hawkins, George Regan	Demotion
P1342	P1342	Regan Reply Email	Fri 01/21/2000 11:15 a.m. ET	E-mail	☑	George Regan	Frank Vanaro	Hawkins Deserved Termination
P1343	P1349	Sheridan Performance Review	Mon 01/10/2000	Performance Review	☐	Frank Vanaro	Susan Sheridan	Pattern & Practice

The CaseMap document spreadsheet makes it easy to organize and explore knowledge about case documents.

Dozens of reports can be generated from the document spreadsheet. Printing is "what you see is what you get." When you add a column to the spreadsheet, hide a column, change the sort order, run a filter, etc., a report based on this revised spreadsheet view is just one mouse click away. Each CaseMap report automatically includes a polished and customizable cover page.

Dozens of reports can be generated from the document spreadsheet.

In addition to printing reports on plain old paper, they can be output as PDF files using the built-in PDF creation capability that doesn't require a full Acrobat license. In two mouse clicks, you'll have a PDF of any document index report. CaseMap also offers "Send to Word" and "Send to WordPerfect" functionality that creates an editable table of spreadsheet information directly in a word processing document. It takes just three mouse clicks to accomplish this task. Finally, the CaseMap ReportBook feature makes it easy to include one or more document-based reports in a compilation of reports from any of the five CaseMap case analysis spreadsheets.
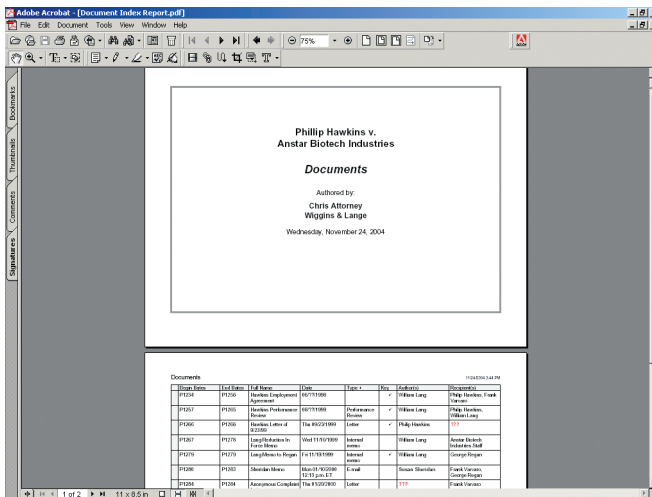
CaseMap makes it easy to create dozens of styles of document reports. Reports can be printed or, as shown here, output as PDFs using the built-in PDF capability.

Another benefit of using CaseMap to create document indexes for cases with small to mid-sized document populations is the availability of many ease-of-use features that aren't present in the industrial-strength software products required for the 1,000,000+ document case.

For example, CaseMap offers live spell checking and autocorrect just like Microsoft® Word and WordPerfect®. We're all hooked on live spell checking's red squiggly line and right-click list of spelling suggestions. This and many other ease-of-use features in CaseMap save energy and result in a higher quality document-index work product.

The Tight Integration Between Acrobat & CaseMap

Acrobat can certainly be used to manage document images without employing CaseMap to organize knowledge about these documents. And vice versa. However, the tight integration between these applications provides a strong incentive for using them in tandem.

The document descriptions in a CaseMap spreadsheet are easily linked to PDF files containing document images. Once PDFs are linked to a CaseMap file, they can be opened in one click anytime you review the document index.
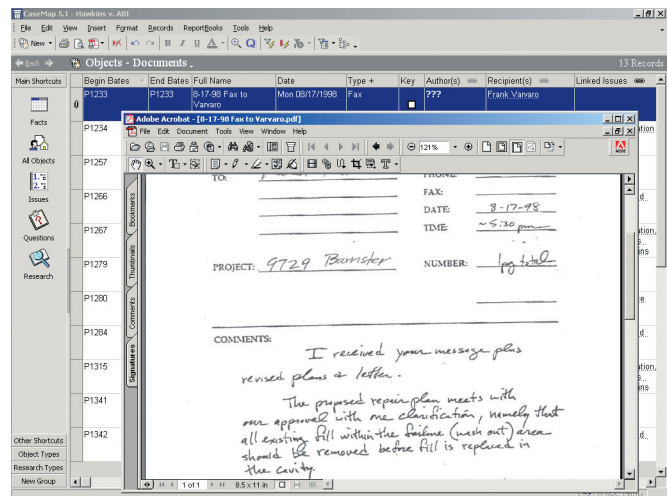
Each spreadsheet row can be linked to the PDF containing the image and full text of the document. Clicking the paperclip icon at the left edge of the selected row opens the PDF.

While links between CaseMap and PDF files can be made manually, CaseMap offers a feature that automates the process—the Adobe PDF Bulk Importer. Point this utility at any folder containing PDFs of case documents. CaseMap analyzes the contents of the folder, creates a new row in the document spreadsheet describing each PDF and links each row back to the PDF file so it can later be displayed with one click. Say you have a folder containing 500 PDFs. It would take less than two minutes for CaseMap to process these files and add 500 rows to the document index, each linked to the appropriate PDF.

The PDF Bulk Importer is designed to handle the growth of your collection of document images over time. When additional discovery documents are scanned as PDFs, the Bulk Importer can be run again. It's smart enough to skip PDFs that have already been processed and to only add spreadsheet rows and links for new documents.
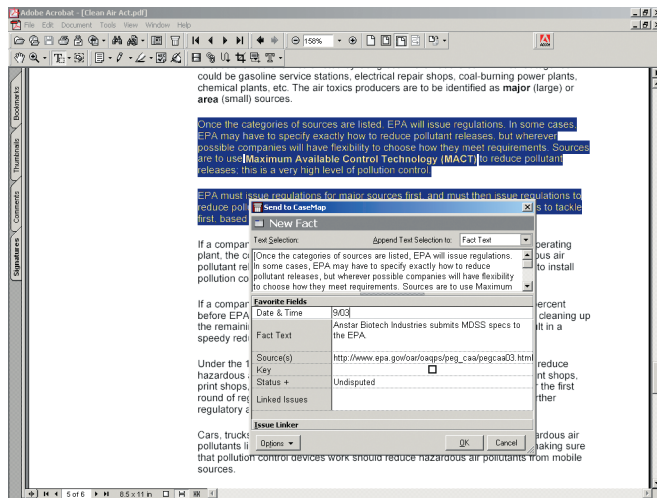

The "Send to CaseMap" Plug-in for Acrobat makes it easy to cull passages from PDF files and organize them as facts in CaseMap.

Another important example of Acrobat/CaseMap integration is the "Send to CaseMap" Plug-in that's available for Acrobat. The "Send to CaseMap" Plug-in makes it possible to cull critical knowledge from PDFs and instantly organize it in CaseMap.

Here's how the "Send to CaseMap" Plug-in works following installation: Open a PDF of a case document and start reading. When you spot an important passage, select it and click the Send to CaseMap button the Plug-in adds to Acrobat's toolbar. A new fact containing the text selection is created in CaseMap and linked back to the specific section of the PDF.

As you later review your fact chronology spreadsheet in CaseMap, a click on any fact sent from a PDF file reopens the PDF document and reselects the original excerpt in context. The "Send to CaseMap" Plug-in works with the full version of Acrobat and also with the free Adobe Reader.

When you use the Acrobat/CaseMap combination to handle document imaging and indexing, also consider our TextMap transcript summary utility for dealing with electronic transcripts. TextMap indexes all words in case transcripts so they are easily searched for any word or phrase. While reviewing a transcript, key passages can be selected and sent to CaseMap using the same process you've mastered by employing the "Send to CaseMap" Plug-in for Acrobat.
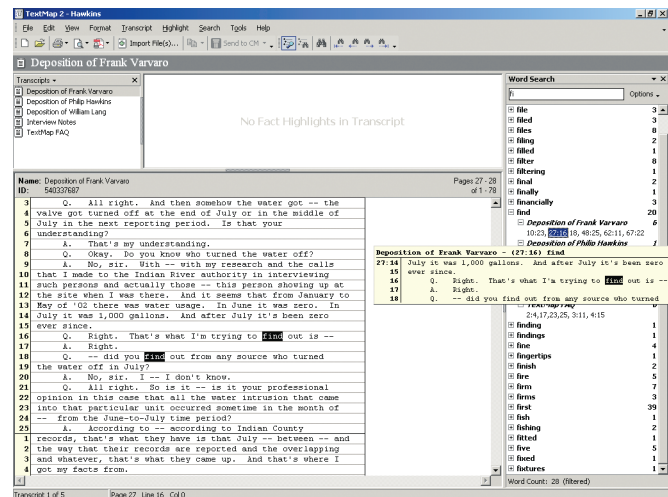
When you're using Acrobat and CaseMap to handle document indexing and imaging, consider using TextMap for transcript summaries.

Here's a final tip for any reader who's yet to experiment with document imaging: Using CaseMap and Acrobat together is a great way to get comfortable using electronic documents without jumping into the deep end of the pool. Don't scan every case document until you're sure it's worth the effort. Instead, identify the 100 or so most critical documents, have these scanned as PDFs and linked to your CaseMap file. You'll be able to evaluate the benefits of using electronic versions of case documents with a minimal investment of time and expense.

Not All PDFs are Created Equal

A word of warning before you fire up the office scanner or ship boxes of documents to a scanning service. Take care or you could end up with PDFs that contain images, but not text.

If there's no text in your PDFs, you won't be able to use Acrobat to search the collection of PDFs for those containing specific words or phrases. If there's no text in your PDFs, there's nothing to select when using the "Send to CaseMap" Plug-in for Acrobat.

To produce PDFs of discovery documents that contain text in addition to images, the documents must be processed using Optical Character Recognition (OCR) software when scanned. Don't assume this work will be done automatically. Make sure the individuals performing document scanning have a clear picture of your expectations.
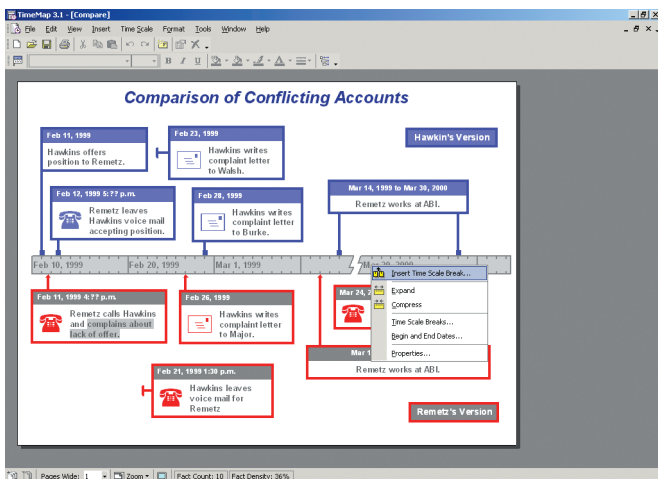
Send document index data from CaseMap to our TimeMap tool (shown above) to instantly create timeline graphs.

The "Send to CaseMap" Plug-in makes it possible to cull critical knowledge from PDFs and instantly organize it in CaseMap.

Conclusion

CaseMap's traditional long suit is fact and issue analysis. For the vast majority of cases, it can also be an ace document indexing solution. Similarly, while Acrobat is a poor image-handling/text-searching solution for the matter with a gazillion docs, it's a great one for the average matter. And, when used together on appropriate cases, the unique integration between Acrobat and CaseMap enhances the value of both tools.

If you have yet to try the Acrobat/CaseMap document management strategy, please review your current cases and find one or two where you can put it to the test. We would be glad to show you the Acrobat/CaseMap integration at work. Just write us for a quick phone tour at phonetour@casesoft.com

Don't have Acrobat?
Download the free Adobe Reader at www.adobe.com. Not a current CaseMap client? Obtain a full-featured trial version of CaseMap and the "Send to CaseMap" Plug-in for Acrobat at www.casesoft.com

About the Author

Greg Krehel is a co-founder of CaseSoft, now part of the LexisNexis® Litigation Services family.
If you have questions or suggestions about this article, please contact him at
greg.krehel@lexisnexis.com or 904.273.5000 x233.

Greg has written a series of other white papers you may find of interest,
e.g., "Chronology Best Practices" and "Creating and Using Issue Analysis Memos."
You can obtain PDF versions of these articles at no charge by visiting www.casesoft.com/articles.shtml.

About CaseSoft

The CaseSoft suite of software tools include:

- CaseMap®—our case analysis tool
- NoteMap®—our outlining tool
- TimeMap®—our timeline graphing tool
- DepPrep®—our witness preparation tool
- TextMap®—our transcript summary tool

Full-featured trial versions of all five products are available for 100% free, no strings attached download at
www.casesoft.com

Our tools are in use at 10s of 1000s of small and large law firms, government investigative and prosecutorial agencies, and private investigation and forensic accounting firms. Two examples of our many clients: The United States Attorney's Office has 15,000 CaseMap/TimeMap/TextMap license sets and the Securities and Exchange Commission has a CaseSoft Suite Enterprise License for 1,100 users.

We're totally devoted to providing every client with excellent support and training.

About LexisNexis® Total Litigator

LexisNexis Total Litigator is built around the way you work—the only litigation solution available with trusted content and tools organized around the tasks of the litigation process. From case assessment to court appearance, find, manage and use vital information seamlessly and gain a strategic edge at each step of the litigation process. CaseMap provides the workflow platform into which all of the key products that make up the LexisNexis® Total Litigator suite are linked and integrated.

CaseMap®   TimeMap®   TextMap®   NoteMap®   DepPrep®

Case Analysis Tool   *Timeline Graphing Tool*   *Transcript Summary Tool*   *Outlining Tool*   *Witness Preparation Tool*

Download full-featured trial versions: www.casesoft.com

Our software tools and support have won over 50 awards in the last 5 years.

TOTAL PRACTICE SOLUTIONS

Client Development   Research Solutions   Practice Management   Litigation Services

LexisNexis®