

Create Order Out of Electronic Chaos

Automatic concept organization and visual e-mail analytics significantly accelerate discovery and review of large document collections.

By Ramana Venkata, Stratify, Inc. President and CEO

You're facing a complex discovery process with boxes of documents and thousands of e-mail messages: a mountain of communications that pour in from various custodians and locations. You need a tool that can help you condense this data into the most relevant documents and delegate tasks to multiple reviewers. Where do you begin?

AUTOMATIC CONCEPT FOLDERS BRING ORDER TO CHAOS

Organizing documents by subject matter, or concepts, provides an intuitive overview of large, complex collections and can enhance your grasp of the matter from its inception. New discovery solutions that are deploying this approach can automatically recognize meaningful concepts that describe documents in the collection, organize the concepts into a hierarchy based on their interrelationships, and then sort documents into relevant concept folders (figure 1). The concept organization can evolve with the case (as new documents are processed) so reviewers can uncover new developments quickly. Here's a look at how the process works.

First, in response to a discovery request data is preserved and collected from various storage devices, computer systems and applications. Documents, records, and email messages are restored from these devices and loaded (usually by a discovery vendor) into the discovery system in their native file formats with embedded metadata intact.

Second, as documents are loaded into the system it automatically invokes the concept folder function that statistically analyzes every document based on the words it contains. The concept folder function uses a sophisticated algorithm that examines each word in every item of data (including e-mail messages and their attachments) in relation to each word in all the other documents, records, messages and other materials. Common words (which are usually contained in a standard stop word list but can be customized for specific matters) are usually discounted in the process. Each item is analyzed separately and in conjunction with every other document to determine its similarity to the others. Documents with a high degree of similarity

Figure 1: Concept folders organize a matter's documents in a concept hierarchy



(based on the characteristics of the documents in the collection) are grouped together into concept folders.

Concept folders based on document similarity differ fundamentally from term-based concepts that can occur within a document. Rather than indicate simply which key words or phrases occur within a document, concept folders provide a richer description since they categorize each document as a whole.

Third, based on the weight and density of the document groupings, highly similar groups are combined whereas others may be divided into more granular groups. This process results in a matter-specific hierarchical folder structure that is based on its unique document set.

Finally, since documents are sorted into concept folders based on their statistical word patterns, which are the set of statistically significant relationships between words in a document described above, a document may appear in more than one folder, situating the document in all appropriate conceptual contexts. You can then review and assign filters to the concept folders with generic criteria from a menu that includes custodian, batch number, tag, date/time, etc. to examine the distribution of specific sets of documents across the concept hierarchy.

CONCEPT FOLDER BENEFITS

Because the discovery software automatically creates the concept folders based on its analysis of the content of the documents, and then uses the folders to organize documents about a subject together, there is no need for you to manually create the concept hierarchy. However most tools allow you to create your own personal forms of organization and sort documents accordingly. By making the concept hierarchy of organized documents available via an on-line review application, this automated approach has a number of obvious benefits:

Ease of use

Concept folders immediately provide a list of subjects in the document collection. The hierarchical organization of concept folders supports easy browsing and navigation to help you locate relevant documents. And because documents often include information about more than one major subject, they will appear in multiple concept folders—as you would expect. This eliminates misfiled documents and provides the ability to find the same document from multiple perspectives.

Meaningful prioritization

Lawyers no longer need to dole out review assignments blindly by the banker's box. Instead, you might want to review documents about sales forecasts first and then turn to documents about employee compensation. You might also decide to "back burner" or ignore altogether spam e-mail messages or the documents about the client's social events. Subject-wise organization supports rapid prioritization, since documents about important subjects can be identified and reviewed first, while less important or irrelevant documents can be set aside.

Assignment by expertise

Lawyers and litigation support managers can thoughtfully direct documents to reviewers who are knowledgeable about particular subjects. For example, documents about product technical specifications could be assigned to reviewers with a technical background, while documents about sales forecasts could be assigned to reviewers with a business background. Documents in different languages, including Western European languages as well as double-byte languages such as Japanese, Chinese, Korean or Arabic (which are termed "double-byte" owing to the extra memory that is required to store their characters in the computer) are organized into concept folders within separate hierarchies for reviewers and lawyers with appropriate language skills.

Consistent review

Concept organization promotes consistent tagging. As you or your reviewers review documents, messages and other materials that are part of the document collection, you need to identify them as "responsive", "confidential", "hot" etc. Because documents are organized into concept folders concept organization you can review identical, nearly identical, and related documents together and tag them consistently. By contrast, lawyers using traditional approaches are forced to shift gears and attempt to understand documents on a wide range of topics. Even more importantly, reviewing the same, similar, and related documents separately creates significant risk that reviewers will make different judgment calls about privilege, confidentiality, and responsiveness. Reviewing similar documents grouped together in concept folders greatly diminishes the risk of producing non-responsive or privileged documents inadvertently.

HOW CONCEPT FOLDERS AND SEARCH COMPLEMENT EACH OTHER

Traditional Boolean search delivers results in a flat list based only on the query terms entered. Often many of the results from a keyword query are not relevant to the matter under investigation. With concept folder functionality, you can use search to further narrow the document universe based on specific metadata including date range, sender, custodian, and more, rather than using search to try and discover the detailed subject matter of the document universe.

But the combination of concept organization AND search can create order out of electronic chaos like no other paradigm. For example, the results of a traditional search can be automatically grouped so that results that match different meanings of the search terms are bucketed into

different concept folders. You can focus on the most important documents based on the concept folder organization of the search results. Lawyers and reviewers can also navigate back and forth from the search buckets to the full concept folders and easily access closely related documents that weren't exact search hits.

THE CHALLENGE OF E-MAIL REVIEW

Lawyers confront a problem of a different order when they are required to analyze and review large e-mail collections. During a recent NPR Marketplace interview, David Brown, Investment Protection Director in New York State Attorney General Eliot Spitzer's office said: "Lawyers used to say the great engine of truth was cross-examination...but today, it's e-mail. That's where people now put their most unguarded, and often most damaging thoughts down." Not only do most people treat e-mail more informally, but the sheer number of messages and correspondents creates a virtual investigatory goldmine. It is no wonder that e-mail analytics has become the focus of investigations by state and federal authorities. These large e-mail collections pose both practical and technical challenges for counsel who need to assess risk and potential liability for their clients, even as they strive to comply with demanding discovery requests.

E-mail review presents a number of different challenges compared to traditional, document-based discovery. E-mail is characterized by the notion of threading (usually associated with the subject line of the message), and individual messages can include documents as attachments or embedded within the message (depending on the e-mail system). It's important to maintain these connections in the discovery environment, and to ensure that lawyers can examine messages and documents in the context of e-mail threads and relationships to correctly understand their meaning. In addition, as we all are painfully aware, e-mail is characterized both by an extraordinarily high volume of messages and the fact that discussions often occur in parallel threads involving different custodians, all of who are potentially relevant to an issue or matter. These factors easily overwhelm traditional paper or search-based discovery approaches. Lawyers need a new, more powerful approach to help them identify key relationships, communication patterns, and critical documents from large e-mail collections.

THE VISUAL APPROACH TO E-MAIL ANALYSIS

The industry is mobilizing to take advantage of new visualization technology to overcome these challenges. For instance, Stratify, a leading discovery vendor, has developed an innovative, user-friendly visual solution designed to help attorneys analyze large, complex collections of e-mail messages. Many lawyers feel comfortable using flow-diagrams and graphs to represent their matters. Stratify adapted these graphs to represent e-mail pathways and relationships, calling them Email Maps (figure 2, next page).

Tagging and Assigning Documents

Review and categorization of documents is central to the discovery process.

Initially, lawyers must examine documents to decide their relevance and responsiveness. During this process, attorneys may also mark documents as confidential, privileged or with other notes depending on the case. In an electronic review system, lawyers apply a "tag" to documents to mark them as "responsive", "privileged", "hot", "confidential", etc. The tags can then be used to create production sets.

During review, lawyers can mark documents that should be used in depositions, or that are relevant to particular interrogatories or are specific to an expert witness. They may also organize documents using categories specific to their case (in parallel with the automatic concept organization that the discovery solution provides). When documents were in paper form, attorneys may have copied them and sorted them into folders. With an electronic review system, lawyers can create custom case management "work folders" electronically and "assign" documents to them.

"Tagging" documents and "assigning" them to work folders provide powerful data management capabilities. These characteristics can obviously be used in conjunction with other metadata (custodian, date range, etc.) to search the document collection for specific sets of documents. Finally, since they are directly accessible in the discovery interface lawyers can immediately examine documents with a particular tag or in a specific folder to facilitate review and analysis.

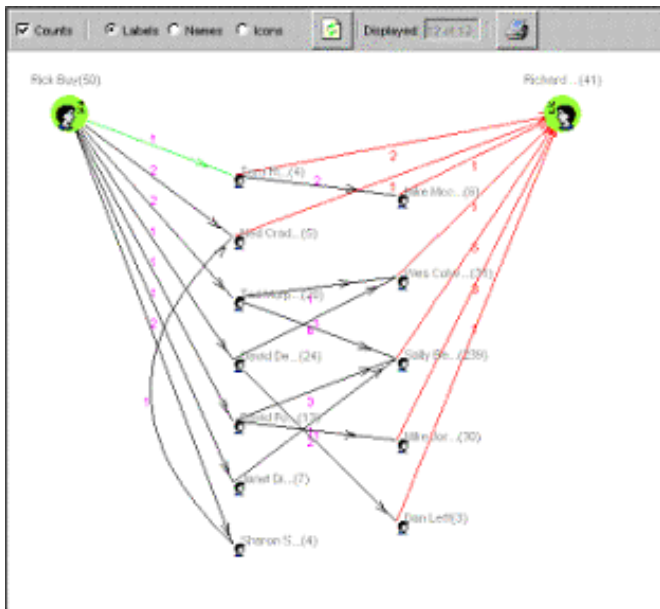


Figure 2: An Email Map from Stratify Visual Email Analytics™

An Email Map represents all the relationships that exist between a sender and a recipient within a selected set of messages and attachments (for example, from a search result set or concept folder). These messages also include those that involved intermediary correspondents, which are particularly difficult to identify, isolate, and review using traditional search technology. Email Maps portray the actual e-mail trail –including how messages were forwarded, copied, or reused (as part of a different thread).

WORKING WITH EMAIL MAPS

Visualizing e-mail relationships allows you to take advantage of the unique nature of e-mail communication and locate “smoking gun” messages. You can examine messages by clicking on a custodian’s icon (sender, receiver, or intermediary correspondent) or by clicking on pathways between correspondents to investigate message details between specific individuals. When you review messages in the visual environment, you have all the power of the review application and can take immediate action on a message or attachment, for instance by highlighting or tagging it, or by assigning it for secondary review. In addition, Email Maps allow you to focus on selective intermediary correspondents (particularly useful in high-density exchanges), to change the look and feel of the diagram, and to print a copy to visually document a set of key relationships.

You may also use this interactive environment to analyze the map by e-mail threads, for instance, by selecting one or more threads to see which custodians were involved in an exchange (and then directly view the pertinent messages

and attachments). You can also analyze the map by concept folder, or by a specific tag that was applied during the review process. This is particularly useful when addressing a large number of threads, and helps ensure you can analyze all the messages and pathways that are relevant to a specific subject. And you can analyze the relationships in the map by date range as well.

This new approach to e-mail analysis is especially valuable for depositions, when you need to quickly access all communication between the deponent and other key individuals to validate or disprove testimony.

CONCLUSION

Lawyers need powerful, razor sharp tools to bore through the growing mountains of electronic data. Concept folders and e-mail mapping technology is a direct response to the challenges of electronic discovery confronting attorneys today. Concept folder organization allows lawyers to understand the structure of the document universe, zoom in on relevant documents, and recognize irrelevant documents like spam, office memorandums, etc. Concept folders provide crucial information necessary to accelerate organizing, prioritizing, reviewing, and analyzing documents accurately and consistently in complex litigation. Email Maps are a user-friendly depiction of relationships and pathways between correspondents that accelerate the review process by allowing attorneys to locate critical, “smoking-gun” messages and uncover connections between key players involved in the case or matter under investigation.



Ramana Venkata is President and CEO of Stratify, Inc., a leader in electronic discovery that uses automatic concept folder technology and visual email analytics to facilitate review and improve analysis while controlling discovery costs.