



Using e-Discovery Technology to Avoid Inadvertent Disclosure

By Shawn Huston

There have been many articles written on the topic of inadvertent disclosure of privileged information in litigation, but there are still mistakes being made in how technology is used (or not) in the identification of privileged documents in electronically stored information (ESI). For this reason, it might be helpful to describe some of the e-Discovery tools that can be leveraged to identify privileged communication within a company's ESI collection to reduce the risk of inadvertent disclosure. These same tools can be used to identify other groups of documents in the data set, but for this instance, we will discuss how they specifically relate to your firm's privilege review.

In many large-scale document reviews, the project management team will run predetermined terms and criteria, which likely relate to privileged items, against the collection of documents to pre-categorize them for a secondary review. During these types of review, the team of attorneys that perform the initial review, called a first pass review, will also identify potentially privileged documents and categorize them accordingly using tags in the document review tool.

While this is all well and good, when you have dozens (or hundreds!) of review staff, how can smaller teams manage to perform the same, or better, level of privilege review? Technology, of course!

Many of these options are available in a variety of document review tools. Even if you don't use the same ones that we do, your functionality is still readily available. Before we begin, please don't start with the mindset that this technology is too expensive for your matter. I've said it before and I'll continue repeating it, the cost of using e-Discovery technology is much less than it ever was dealing with all of those boxes of paper. Besides, in this day and age, you can't afford not to be using them if you plan to thrive in today's competitive marketplace.

Now that you have all bought into the value of technology and want to ensure you're leveraging the best methods to protect your client, let's dive in!

Keyword Searching

Contrary to popular reports, keyword search in e-Discovery is alive and well. Predictive Coding has failed to kill off keyword searching like many prognosticators were predicting (no pun intended), and instead, these tools are being used in tandem. For these purposes; however, we will be focusing on keyword searching on its own.

Keyword searching merely refers to performing searches on the text of a group of documents. The benefits increase when features such as stemming, fuzzy, proximity, and wildcard searching come into play. These features, which are common in e-Discovery technology, allow for the ability to quickly and easily find the documents that contain the terms you are looking for.



When performing privilege keyword searches, you may want to search for terms such as legal, attorney, privilege, confidential, counsel, etc. Keep in mind; however, that depending on the text contained in your documents, terms like these can result in a greater number of false hits. To remedy this, it is helpful to evaluate the results of your keywords to ensure you are finding the right balance between precision and recall. That is to say you want to try and find the sweet spot between locating the most pertinent documents while also finding the preponderance of the hits that exist. If you are unsure of how to do this in your application, it may be beneficial to engage an e-Discovery professional to provide assistance.

One major stumbling block to using keyword search to identify potentially privileged documents is the common use of footers in email communication. This becomes an issue due to the frequent use of terms which are associated with privileged information in the language used in these disclaimers. Below is an example of a confidentiality disclaimer which may be found in the footer of an email which was found on the www.policypatrol.com website (emphasis added):

This message is confidential. It may also be privileged or otherwise protected by work product immunity or other legal rules. If you have received it by mistake, please let us know by e-mail reply and delete it from your system; you may not copy this message or disclose its contents to anyone. Please send us by fax any message containing deadlines as incoming e-mails are not screened for response deadlines. The integrity and security of this message cannot be guaranteed on the Internet.

To lessen the number of false hits caused by language such as this, you can leverage proximity and Boolean searching to filter out the hits that would be returned based on this footer language. In addition, many e-Discovery search tools now allow for the document index to ignore blocks of text that commonly appear in your data set. While you need to know the exact disclaimer text used in your documents for this feature to be accurate, it can be advantageous when trying to eliminate the majority of false hits that may occur.

While the above may be useful when using e-Discovery tools, don't be fooled into thinking that search functionality in other software will achieve the same results. I've witnessed firsthand the failure of search functionality in productivity software like Adobe Acrobat and Microsoft Outlook when it comes to locating documents containing specific terms. I know many professionals rely on these tools every day, but it's not uncommon for tools such as this to completely miss the mark. I would suggest steering clear of their use when accuracy is critical.

Metadata Filtering

When we talk about privilege documents, most of them revolve around communication between attorneys and their clients, or between other attorneys. While some of this communication still isn't considered privileged, the ability to identify it through metadata serves to reduce the number of false hits typically associated with keyword searches. With email being the most common communication tool used in business we can benefit from, the various tools available to filter based on common metadata fields are found in this type of ESI.



The sender and recipient fields can easily be searched for by names and email domains to find all communication that included counsel. The CC (carbon copy) and BCC (blind carbon copy) fields can also be used in this effort. Along these same lines, the Author field with other types of electronic files can be used to the same effect.

With the advancement of e-Discovery tools over the last five to ten years, many search tools will show you the entire list of email addresses and domains associated with your data collection without needing the need to identify them ahead of time. The tools will allow you to then pick and choose the ones you would like to identify or filter. Many times this rapid insight into the data has allowed my clients to identify firms and attorneys that they had no idea were involved with the client, let alone related to the matter.

Concept Clustering (Conceptual Search)

Identifying and grouping your ESI by concept can be a great way to locate documents that have similar themes that may not otherwise be obvious without a full linear review. This is especially true early in the case when attempting to create a “bucket” (subset of documents) of potentially privileged material for later review.

Conceptual search typically works by evaluating the terms used in your documents and then grouping them together based on frequency and usage as compared to other documents in your set. While not right for every instance, the usefulness against large and diverse datasets can be immeasurable in the added efficiency and accuracy of identification.

Email Threading

The ability to view an entire email conversation in context is something we are all used to when using email tools like Gmail, Yahoo and Outlook. However, when trying to increase the consistency of your document review, viewing the entire chain of communication as one chain couldn't be more valuable.

Email threading is best when the appropriate metadata fields are available within the electronic data. This will help ensure an accurate representation of the line of communication. You can then bulk tag the entirety of the communication and avoid mistakes.

Production Time

Prior to delivering any document productions, it can be helpful to apply these same search features against the subset of documents you have deemed appropriate to produce. Search for those terms you know to be related to privileged communication. Identify the emails that include the addresses of counsel and double check them to be sure they legitimately belong in the production. Run some comparisons on the tags to be sure nothing with a privilege tag inadvertently made it into the set. Creating a production quality control checklist encompassing each of these steps can help ensure you have a repeatable and defensible process moving forward.



Technology can make the process of identifying privilege documents within your ESI easier and more precise. While all of the above tools will help in allowing you to identify privilege documents, it still comes down to accurate decision making by human users to get it right. Take advantage of the e-Discovery tools and experts at your disposal and you may just avoid those horror stories you read about in the legal publications.

This article discusses issues of general interest and does not give any specific legal or business advice pertaining to any specific circumstances. Before acting upon any of its information, you should obtain appropriate advice from a lawyer or other qualified professional.

4 of 4