

What You Need to Get Ahead Using eDiscovery Analytics

By Pete Feinberg



42

The challenge in eDiscovery has been to develop technology that reliably and accurately collects, processes, searches and presents information efficiently. With the latest wave of tools and workflows, that goal has been largely accomplished. But in just the last few years, the information landscape has profoundly changed in both data volume and complexity. Unfortunately, reliance on specialists to tame this data surge introduces risk and cost. But there is a path forward by leveraging analytics.

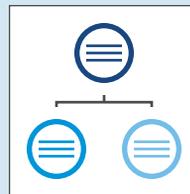
In this article, we'll explore just what we mean by analytics and show why you need analytics to get ahead in modern eDiscovery. You'll learn how next-generation eDiscovery platforms make the power of analytics readily available.

TYPES OF ANALYTICS

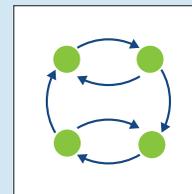
New terminology has surfaced as the techniques and tools of analytics have evolved. Predictive Coding, Technology Assisted Review (or TAR), Continuous Active Learning (or CAL) and Machine Learning are just a few of the naming permutations. But more fundamentally, there are some key details about how the technology works that can help us further understand how analytics plays within eDiscovery.

Syntactic Analytics

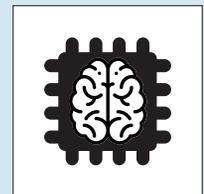
Syntactic analysis segments content into subgroups based upon defined structural elements within the content-specific text, domains, locations, and dates. These techniques are used to filter data and determine pieces of content that are related. This enables



Syntactic Analysis
Segment content based on specific text or metadata



Conceptual Analysis
Find and cluster content that is similar in meaning



Machine Learning
Define similarities based upon previous human interaction with data

email thread detection, gap identification, determining language and geographical groupings and detecting duplicate or near duplicate content.

Conceptual Analytics

Conceptual analysis uses the meaning of words and content — not just the letters in the text — to find documents with similar meaning to search terms or sample documents. Examples of conceptual analytics are keyword expansion, concept searching, concept clustering and categorization.

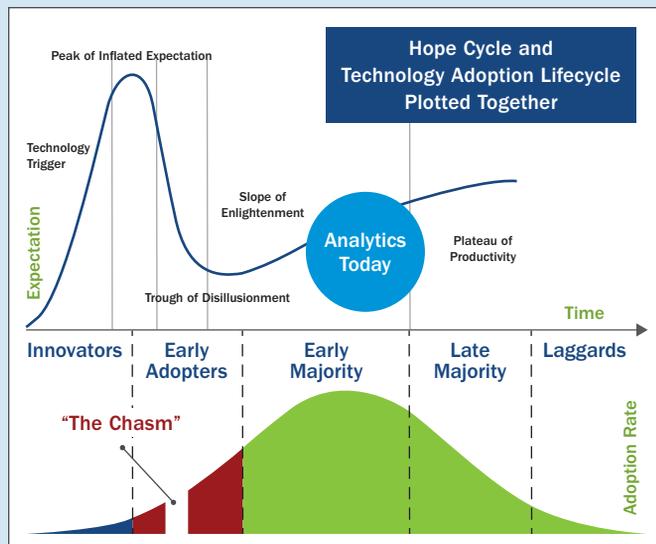
Machine Learning

A Machine Learning platform is trained to learn directly from the data and previous human interaction with that data. Techniques like Technology Assisted Review and Continuous Active Learning teach the system based on positive and negative feedback from human analysts and can reliably “predict” how

a human analyst would code for the rest of the collection. The systems are also now commonly used for analysis that requires a level of “judgement” to determine next steps, such as privilege review, compliance monitoring, contract review and uncovering hidden relationships within collections of documents.

ADOPTION OF ANALYTICS

The recent 2018 CTRL survey reported that “the overwhelming majority of respondents agree that data analytics ‘will be very important, will be considered indispensable, and [their] use will be widespread’ among the legal profession over the next 10 years.” And yet, in a study from FTI and Ringtail, the actual adoption of eDiscovery analytics hovers at around 50%.



Analytics has moved past the earliest stages of adoption but is still mostly in the hands of experts. Classic technology adoption theory tells us that increasing the usability of analytics will be key to broadening the user base for the technology.

PRACTICAL IMPACTS OF ANALYTICS IN eDISCOVERY

Some types of analytics reduce data volumes, some help uncover patterns and relationships and some group similar documents in an efficient manner. Let’s look at three important areas where analytics benefits eDiscovery.

Making Sense of It All — Organizing the Chaos

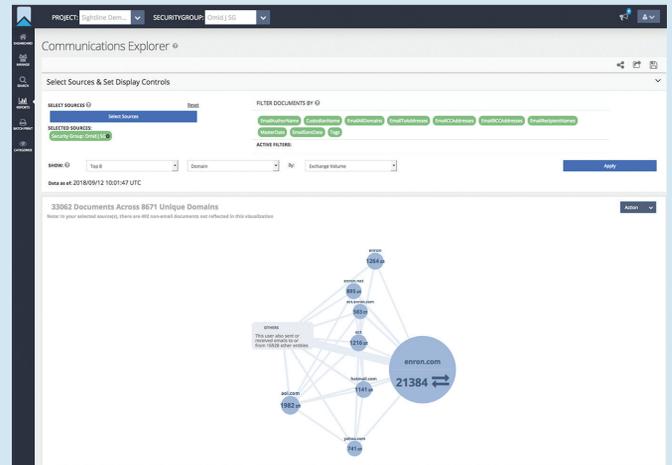
One of the overlooked benefits of analytics is logical organization of a project. Analytics tools excel at organizing documents by both textual and conceptual similarity. The ability to group documents into e-mail thread groups, near duplicate groups or clusters is incredibly valuable as work is planned and early estimates of effort are made.

Narrowing the Field — Making the Most of Your Time

It is generally relatively easy to assess any single piece of content.

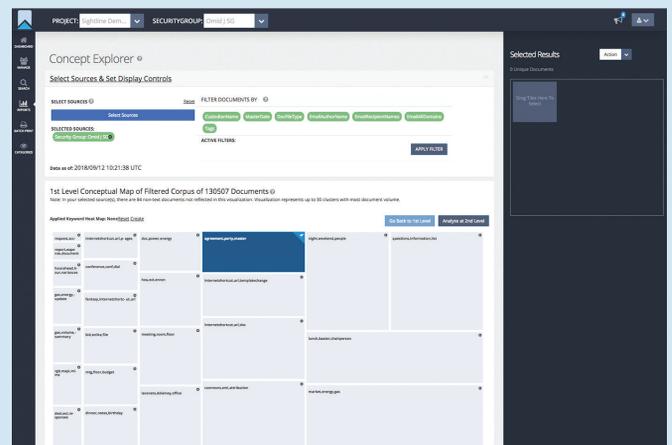
But examining duplicate content at an early stage is a drain on resources. It is best to start small, identify key individual documents as quickly as possible and then build context around those items later.

One helpful early approach is to limit data sets to only unique content. Achieving the smallest, most unique starting data set can be accomplished via a layering of analytics tools and metadata culling techniques. Three common forms of analysis are hash values, e-mail threading and textual near-duplicate identification.



Casting a Wide Net — Confidence in Completeness

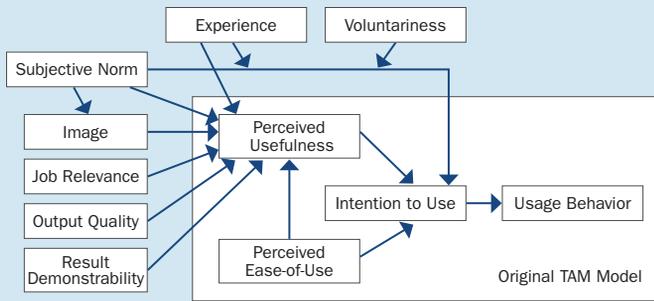
Moving efficiently through the starting set of data is only half of the battle. It is also important to establish confidence in the completeness of efforts. The identification of key documents should serve as a foundation for a more exhaustive effort that overcomes the deficiencies of standard keywords and the inconsistencies of human judgment.



There are a variety of analytics tools available that empower us to do just that. For example, conceptual analytics tools can be leveraged to suggest pockets of undiscovered information, identify inconsistent treatment of similar content and offer insights for improving our keyword search efforts.

FACTORS DRIVING SUCCESS IN ANALYTICS IMPLEMENTATIONS

It is no longer what systems do but rather how they do it that is defining eDiscovery platform success. This aligns with well-proven technology adoption models that combine perceived usefulness with perceived ease of use. Systems that deliver required functionality and lower the barrier to accessing and using that functionality are the new benchmarks in analytics.



Perceived Usefulness and Analytics

The factors that contribute to perceived usefulness are the functional elements that contribute to solving the user’s problem — that delivery quality results relevant to the user’s jobs. The technical requirements that define the functional needs have become table stakes.

Perceived Ease of Use and Analytics

Analytics must be both easy to access and easy to use. If users must hunt for the tools — if they are not included as a core part of the eDiscovery platform — then it is unlikely the users will become familiar with analytics and unlikely that they will make any effort to include analytics in their workflows.

ensure that analytics tools are readily available in the system at the point where they are most likely to be used. This enables users to easily and quickly organize content into related clusters making it easy to get a big-picture view of data landscapes and target areas for priority analysis.

LAST THOUGHTS

eDiscovery software is evolving beyond technical capabilities and is now being designed around users — how they work, how matters are structured, how technology can simplify processes — and lets practitioners focus on the law instead of the technology.

The entire eDiscovery process is becoming more user-friendly. It incorporates analytics at every stage to drive efficiency, cost savings and quality of results. This will change everything by moving eDiscovery from the exclusive domain of experts and enabling the concepts of modern eDiscovery to become mainstream.

44



PETE FEINBERG is Senior Vice President, Product and Marketing at Consilio. He is principally focused on bringing the voice of clients, the discovery market and operations into Consilio’s products and services line.
Pete.Feinberg@consilio.com

NEXT GENERATION eDISCOVERY AND ANALYTICS

Courts now embrace analytics and expect that some level of technology will be applied to eliminate completely irrelevant or unnecessary data and to make the review process itself more efficient and consistent. eDiscovery systems must meet these key functional requirements. But the next generation tools are also designed with users in mind — to be easy to implement, easy to use and provide transparency and predictability.

Modern eDiscovery platforms, such as Consilio’s Sightline,