

Analytics and Awareness:

A New Era of Insight in Early Case Assessment

Lana Schell, J.D.

Channel Partner Manager, Content Analyst, LLC.

INTRODUCTION

In its recent survey of 2,053 CIOs from 36 industries in 41 countries, Gartner found that the top technology priority for 2013 is analytics and business intelligence.¹ In fact, 55 percent highlighted that digital technologies impacting big data and analytics would be the most disruptive.²

From a social media standpoint alone, Facebook reported 665 million daily active users in its first quarterly filing of 2013³ and has revealed that its system processes 2.5 billion pieces of content and over 500 terabytes of data each day.⁴ And, Twitter reports that it has over 200 million active users who send 400 million tweets per day.⁵

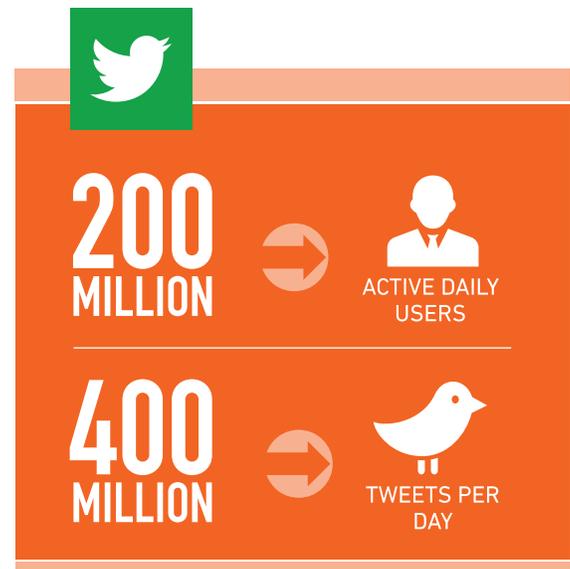
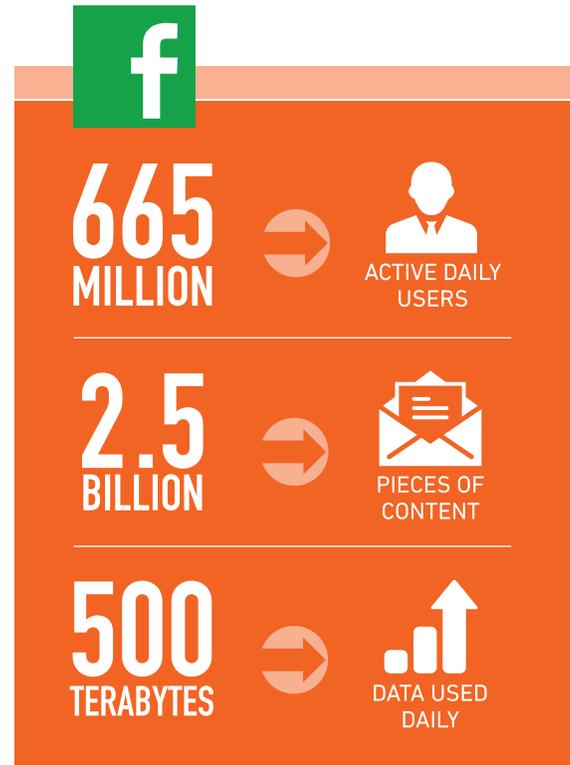
In this era of unprecedented data expansion, particularly in the litigation context, it is often more important to address data early in the process than to reduce the amount of data for assessment.⁶ By applying concept-aware analytics at the outset, a legal team can focus on insights and evaluation without concern for consolidation of the information.

Users can also practically organize and cluster larger pools of information into useful subsets without eliminating it during the entry stages of the litigation lifecycle. For instance, a trial attorney might be able to leverage advanced technology to find prior art in a patent matter by assessing the entire universe of material instead of wasting time eliminating irrelevant material since the computer will achieve that goal. Or, counsel might be able to reveal hidden meanings in coded conversations between employees, which presents evidence of malfeasance.

This effort often saves costs and enables practitioners to make better judgments with a larger pool of information during the nascency of a matter because they are seeking a comprehensive understanding. This is typically easier and more effective to acquire than inadequate answers based on limited data.

The process also permits legal teams to efficiently reapply earlier decisions and document prior designations to streamline their evaluation of a particular matter. By organizing through example-based categorization to quickly separate material used in a previous case, lawyers and executives can develop a more efficient strategy.

Of course, in the current collaborative environment, one must leverage the expertise of an individual familiar with the records and partner with a technology vendor that can maximize the talent of the legal team.⁷ It is essential to use the concept-aware analytics engine for a purpose with the human capital necessary to understand its operations.



SHIFTING DATA UPSTREAM IN ECA IS A KEY LITIGATION TREND

While traditional Boolean searching or keyword-based document tagging are effective once a prospective reviewer is familiar with a data set, the volume of information in current matters makes this a very random and inefficient process. To reposition this effort to an earlier stage, savvy teams are supplementing keyword search strategies with advanced text analytics to harness the power of linguistics, mathematics, and probability, among other elements. Through this form of concept-based search, advanced software can help lawyers maximize their time analyzing critical findings, rather than worry about finding essential material to analyze.

Start By Creating a Roadmap to Organize Your Data

Although lawyers once based case strategy on the known facts and the prevailing law, they now focus on how quickly they can properly assess the universe of relevant data. Speed is a paramount concern because of the accelerated pace of modern litigation.

In the initial stages, teams must organize documents first and review second in an effort to provide correct information to the appropriate experts.⁸ This organizational shift is transforming the entire discovery process. In fact, courts are prompting parties to take a more holistic view of their litigation closer to commencement of the case, rather than as they approach trial.⁹ It is this broader view that is helping to offer an enhanced projection of the data requirements, anticipate potential information challenges, and generate a more accurate estimation of potential resource needs.

Categorizing, Clustering, and Tagging Are Critical Best Practices

Proven techniques for classifying data are helping legal teams formulate their plans faster, more dynamically, and in greater detail than ever before. Categorizing unfamiliar data, for instance, helps practitioners understand the general landscape and broadly identify the key issues associated with a matter.¹⁰ It prevents a blind evaluation of information, while revealing the prospective costs given the total number of documents, their location(s), language(s), and other variables.¹¹ This practice is no longer a novel approach; it is part of a strategy that many expect to be standard in every engagement.

By sorting documents that are conceptually related, reviewers can build efficiency by simply grouping material containing similar terms and word usage. In doing so, they can begin breaking down an unmanageable volume of information into a series of separate, but unified components, teams of reviewers can assess. The tactic enables disparate groups, including outsourced talent, to combine their abilities to form a potent armada of legal analysts.

Focusing on this type of concept clustering and analysis can provide insight during the early case assessment phase and support the execution of other common searching techniques. Further, by carefully scrutinizing the words and phrases contained in a particular data set, legal teams can quickly craft a simple story in a sea of complex facts.

From near de-duplication technology that compares and groups similar records, e.g., weekly status updates that use identical formats, but feature variations in key statistics, to e-mail threading techniques, which group messages that share subjects, forwarding details, or attachments, there are a variety of practices that now accelerate review without sacrificing a team's ability to uncover critical information. And, by engaging in this process at the outset, non-responsive material is less likely to populate later search results, which could eliminate a substantial portion of the litigation cost.

Apply Critical Details to Empower Preliminary Assessments

Applying analytics in this early stage helps counsel plan for the initial meet and confer conference, as well as identify key search terms. In fact, there is an expectation that each individual attending the meeting will have a specific understanding of how the discovery process is likely to proceed. There is little tolerance for uncertainty from either adversaries or the courts, which have become much more familiar with e-discovery over the past few years.¹²

In addition, litigators have a responsibility to understand even the most granular details associated with a particular production of information. For instance, they can gain tremendous advantages by quickly studying obvious e-mail relationships, including frequency of the communication, senders and recipients, and domain names to both eliminate irrelevant material (e.g., shipment receipts from Amazon.com) and highlight unusual patterns (e.g., daily correspondence between a senior executive under investigation for stealing trade secrets and a direct competitor). This type of analysis might focus a meet and confer discussion on excluding certain records during the initial exchange, as well as limiting or expanding the potential custodian pool.

It is the ability to identify these hidden details that make analytics so powerful and so important in the current era of high-stakes disputes. While the contents of a particular group of files once provided some type of revelation on the veracity of any given allegations, it is now the patterns in those records or the gaps in their sequencing that typically presents the most useful clues.

THERE ARE SIGNIFICANT ADVANTAGES TO ADAPTING EDA TO COMPREHENSIVE ECA

Although there are many factors that impact the success of an early case assessment protocol, filtering the universe of data is frequently one of the highest priorities. In order to do so, analysts must leverage language usage to capture key conceptual details faster.¹³

They must also recognize that context, not always content, is king. It is, therefore, essential for legal teams to determine where a particular document fits within the broader data environment. By comparing latent semantic analysis (analyzing connections between sets of documents and their terms by producing a set of related concepts), probabilistic analysis (gauging the probability that certain words and phrases will appear in a given set of documents), and lexical techniques (identifying patterns in language and tone), one can select a method that is ideal for a given matter. While each method has certain advantages, they all serve as valuable weapons in the early stages of an investigation or litigation.

Advanced and Accelerated Culling Streamlines the Process

The advantages of culling documents are heightened by the Big Data revolution that is consuming every industry. Legal teams no longer have the option of analyzing massive amounts of data in the ordinary course. They must apply accepted techniques for separating and prioritizing their information.

Given the ease with which litigants can now find, organize, and securely distribute their information, there are fewer burdens associated with culling material in advance. This ease of access only heightens the benefits of doing so by ensuring a more complete records search, enhancing

defensibility of the process, and supporting the integrity of the review. By eliminating these variables, reviewers can streamline their activities and focus on the answers, rather than waste valuable time or resources responding to often-irrelevant questions.

Threading and Grouping Offer Tremendous Economic Benefits

Like culling, further dividing information into comprehensive components offers reviewers a deeper understanding of the macro issues involved in any given matter. This knowledge can be very useful in developing a search strategy with an adversary or explaining the likelihood of success or failure to a client. The ability to do so is a distinguishing characteristic in a hyper-competitive marketplace.

When the American Lawyer asked law firm leaders to identify the most common change in client behavior, 75% highlighted that more of them are requesting discounts.¹⁴ 49% noted that they are then asking for deeper discounts.¹⁵ This focus on cost reduction is pressuring law firms to continue to improve performance with no expectation of generating additional revenue. It is a new normal of legal practice that simply mandates greater efficiency, more concise analysis, and expedited responsiveness.

Incorporating techniques that link documents with similar themes together and focus on e-mail threads can help legal teams develop better search protocols, but also showcase their innovative approach to clients, colleagues, and the court. At a minimum, each individual representing a party at a meet and confer conference must be familiar with these issues and comfortable addressing them in the proper context.

■ ■ IT IS THE ABILITY TO IDENTIFY THESE HIDDEN DETAILS THAT MAKE ANALYTICS SO POWERFUL AND SO IMPORTANT IN THE CURRENT ERA OF HIGH-STAKES DISPUTES. ■ ■

MARKET SHIFTS AND THE NEED FOR LITIGATION READINESS ARE PROMPTING ECA IMPROVEMENTS

With case law¹⁶ and market sentiment swaying in favor of meaning-based search, including predictive coding, there is broader anticipation of its use.¹⁷ In addition, the likelihood of needing multiple tools to address a more complex process is mandating the integration of different platforms.

As the technology grows more sophisticated and widespread, linguistic anomalies are no longer tolerated. In addition, the quality of the results are fueling higher levels of trust in machine learning. Organizations that harness that trust and use it to instill confidence in judges and their adversaries are generally successful in navigating bottlenecks in an often-confused conversation.

Value and Relevance in Machine Learning are the New Criteria for Case Evaluation

In order to maximize the value of a machine-based auto-categorization solution to heighten the level of analysis at the outset of a given matter, legal teams must offer minimal but direct feedback to ensure the best results. Once the tool learns the intent and purpose for coding a document in a particular way, it will be capable of perfecting that process alone.

Start by selecting specific portions of text to define each category, rather than an entire sample document, most of which may be irrelevant. Ensure that the examples are consistent and clear. The most precise solutions are reliant on human interaction at the outset, including reconfiguring the initial categorization as the document set changes or grows.

■ ■ IT IS A NEW NORMAL OF LEGAL PRACTICE THAT SIMPLY MANDATES GREATER EFFICIENCY, MORE CONCISE ANALYSIS, AND EXPEDITED RESPONSIVENESS. ■ ■



Clients focus on cost reduction

Corporate Categorization Yields Greater Consistency

Proactive litigants focused on trial readiness tend to be better prepared and develop a greater understanding of their ideal strategy than their peers in the current reactionary environment. A corporation, however, often has to suffer the pain of a lawsuit before it shifts priorities. Although cost tends to be the greatest disincentive to implementing improvements, defeat is an even stronger motivator for doing so.

Teams that leverage analytics are able to identify and categorize records in a manner that is consistent with the global movement toward information governance, which will make production easier and quicker in the later stages of the process. It is that future level of preparation to which many organizations aspire, but routinely struggle to reach. By applying these best practices and readily available technology, they can achieve both success and a litigation-ready culture.

A LOGIC-BASED APPROACH TO AN ILLOGICAL PROCESS CAN BE BENEFICIAL

While litigation is inherently uncertain, the application of mathematical reasoning provides a sense of predictability and a universal set of standards. The results may differ, but the common premise that applying logic to a process of this type will yield better results is often undisputed.

Advanced Analytics are Accelerating ECA

Despite the variety of options available, latent semantic indexing is one of the most advanced analytics strategies available, due to the sophisticated computations used to assess key linguistic associations and the similarity to human reasoning itself. Supplementing professional judgment with technological advancements in this fashion is naturally accelerating the early case assessment process.

While legal teams have always been proficient in conveying assessments based on a limited review of cursory information, they are now empowered with the ability to apply a broad view of a vast pool of information to draw conclusions, rather than make inferences. Advanced ECA tools permit them to identify details about document types and topics in an effort to form a holistic picture of the ultimate task. When they combine these preliminary characteristics with a key set of responsive documents to match across the document spectrum, they can make recommendations faster, with more confidence, and greater accuracy.

Mathematical Algorithms Simplify the Complexity

While those studying search once focused their attention on linguistics and the rules of language usage to find patterns within any given body of text, they now realize that it is more of a mathematical equation than a reading comprehension dilemma. Current analytics engines commonly rely on statistics and probability or some type of algebraic analysis, to identify concepts within a particular record.

Ironically, although this appears to be a complex approach, it actually simplifies the experience for the end user. Math-based advanced analytics tools enable more reliable and flexible conceptual search or classification in less time, with minimal effort. It offers initial conclusions and continues to build on that foundation to make more advanced

determinations as the matter progresses. In addition, the tool is often more successful when there is a larger pool of data on which to base its findings.

Ranking Data Can Result in Better Case Strategy

Efficiency, effectiveness, and precision are ultimately the goal of any technology-assisted review. Increasing the quantity of results is less important than the quality of those findings. Legal teams are less concerned with eliminating what is not important, than finding the most important material faster.

By ranking relevant records based on early analytics, dynamic counsel can develop strategies at an unprecedented pace, and with an exceptional level of accuracy. That ability will enhance their competitive positioning and support critical performance claims.

AN INTEGRATED SUITE OF TOOLS PROVIDES A STRONGER SOLUTION

The analysis of data for early case assessment purposes is both an art and a science. In many ways, it is a mathematical equation as well. Given the multi-faceted nature of this modern approach to making key determinations early in the lifecycle of a case, it is essential that legal teams leverage a diverse set of tools to supplement their judgment and experience.

They must do so, however, within a continuous workflow of integrated operations, rather than a disparate conglomeration of technology. That integration provides a consistent user experience, which fosters efficiency, but also enables an analytics engine to maximize a reviewer's

“ ALTHOUGH COST TENDS TO BE THE GREATEST DISINCENTIVE TO IMPLEMENTING IMPROVEMENTS, DEFEAT IS AN EVEN STRONGER MOTIVATOR FOR DOING SO. ”

analysis options. From simplifying maintenance and enhancing improvements, to lowering costs and raising productivity, it provides a critical foundation on which to build a long-term strategy.¹⁸

In fact, despite the exploding volumes of data that are suffocating reviewers, technology can help them evaluate their options more holistically than ever before. With the proper tools, the amount of data will become less important than the ability to categorize and interpret key components of the entire set.

As the legal community progresses in its effort to adapt to the changing nature of data, and with the modified expectations associated with offering comprehensive conclusions more rapidly than its members are accustomed, technology will become more common and trustworthy. As it becomes a defensible mainstream utility, lawyers will use analytics earlier, more extensively, and without hesitation.

EFFICIENCY, EFFECTIVENESS,
AND PRECISION ARE
ULTIMATELY THE GOAL OF ANY
TECHNOLOGY-ASSISTED REVIEW.

About Lana Schell

Lana Schell has spent more than a decade managing large, complex electronic evidence projects for AmLaw250 law firms and Fortune 500 companies both in-house at law firms as well as with service providers and software companies. Currently, as the Manager of Channel Partnerships at Content Analyst Company, Lana works with other software and service providers delivering to their corporate and law firm clients expertise in coordinating the litigation lifecycle with use of cutting-edge advanced analytics software.

Prior to joining Content Analyst, Lana served as a Director of Client Advisory Services at an international legal services company, routinely advising corporations and their counsel on litigation readiness and electronic discovery in support of all phases of litigation. Lana's focus on strategic consulting to address the cost and risk associated with litigation has provided her with a diverse understanding of data systems and the laws that regulate them. She also held senior consulting positions providing e-discovery project management to clients using a proactive process-oriented approach to ensure the successful delivery of large-scale evidence productions. She first worked for two AMLAW250 firms in Litigation Support roles after receiving her J.D.

An accomplished presenter and trainer in the field of discovery and case management, she has presented numerous discovery management training sessions on process implementation for corporate clients and has also conducted training programs on several litigation support software platforms for end users.

Lana was a 2010 finalist for the Anita Borg Institute Women of Vision award and is the co-founder and executive director of Women in e-Discovery.

About Content Analyst Company

We provide powerful and proven Advanced Analytics that exponentially reduce the time needed to discern relevant information from unstructured data. CAAT, our dynamic suite of text analytics technologies, delivers significant value wherever knowledge workers need to extract insights from large amounts of unstructured data. Our capabilities are easily integrated into any software solution, and our support strategy for our partners is second to none.

For a live demonstration of the entire CAAT suite, visit www.contentanalyst.com, mail info@contentanalyst.com or call 1-888-349-9442.

References

- 1 Gartner Executive Program Survey of More Than 2,000 CIOs Shows Digital Technologies Are Top Priorities in 2013, Gartner, January 16, 2013 <https://www.gartner.com/newsroom/id/2304615>
- 2 Id.
- 3 Drew Olanoff, Facebook's Monthly Active Users Up 23% to 1.11B; Daily Users Up 26% To 665M; Mobile MAUs Up 54% To 751M, TechCrunch, May 1, 2013, <http://techcrunch.com/2013/05/01/facebook-sees-26-year-over-year-growth-in-daus-23-in-maus-mobile-54/>
- 4 Josh Constine, How Big Is Facebook's Data? 2.5 Billion Pieces Of Content And 500+ Terabytes Ingested Every Day, TechCrunch, August 22, 2012, <http://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/>
- 5 Hayley Tsukayama, Twitter turns 7: Users send over 400 million tweets per day, WashingtonPost.com, March 21, 2013, http://articles.washingtonpost.com/2013-03-21/business/37889387_1_tweets-jack-dorsey-twitter.
- 6 See Rachel K. Alexander, E-Discovery Practice, Theory, and Precedent: Finding the Right Pond, Lure, and Lines Without Going on A Fishing Expedition, 56 S.D. L. REV. 25, 26 (2011).
- 7 Jacob Tingen, Technologies-That-Must-Not-Be-Named: Understanding and Implementing Advanced Search Technologies in E-Discovery, 19 RICH. J.L. & TECH. 2, 76 (2012) (advocating a need for third-party discovery vendors when dealing with a large database of electronic information).
- 8 See, e.g., Jason R. Baron & Edward C. Wolfe, A Nutshell on Negotiating E-Discovery Search Protocols, 11 Sedona Conf. J. 229, 230-31 (2010).
- 9 See, EORHB, Inc. v. HOA Holdings, LLC, No. 7409-VCL (Del. Ch. Oct. 15, 2012) (Court issued order for parties to show cause why predictive coding should not be used and ordered them to consider using a single discovery provider.)
- 10 See, e.g., Tom Turner, John Tredennick, Smart Sampling in E-Discovery Reduce Document Review Costs Without Compromising Results, TENN. B.J., October 2011, at 16, 19 (explaining the "clustering" technique).
- 11 See, David Degnan, Accounting for the Costs of Electronic Discovery, 12 MINN. J.L. SCI. & TECH. 151, 158-82 (2011).
- 12 Chen v. Dougherty, 2009 WL 1938961 (W.D. Wash. July 7, 2009) (Court reduced attorney's fee for discovery-related work for failure to offer proper search terms, noting that her novice skills cannot command a higher rate.)
- 13 D. van Dijk, H. Henseler and M. de Rijke, Semantic Search in E-Discovery, Workshop on Setting Standards for Searching Electronically Stored Information in Discovery Proceedings (DESI IV Workshop), Pittsburgh, June 2011. <http://www.umiacs.umd.edu/~oard/sire11/papers/dijk.pdf>.
- 14 The American Lawyer, Survey of Law Firm Leaders (2012).
- 15 Id.
- 16 da Silva Moore v. Publicis Groupe, 2012 WL 607412 (S.D.N.Y. Feb. 24, 2012), approved and adopted, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012).
- 17 Jason R. Baron, Law in the Age of Exabytes: Some Further Thoughts on 'Information Inflation' and Current Issues in E-Discovery Search, XVII RICH. J.L.&TECH. 9 (2011), <http://jolt.richmond.edu/v17i3/article9.pdf>.
- 18 John Felahi, Why the Smart Money Checks the Analytics Engine, KM World, March 2013 at S8.