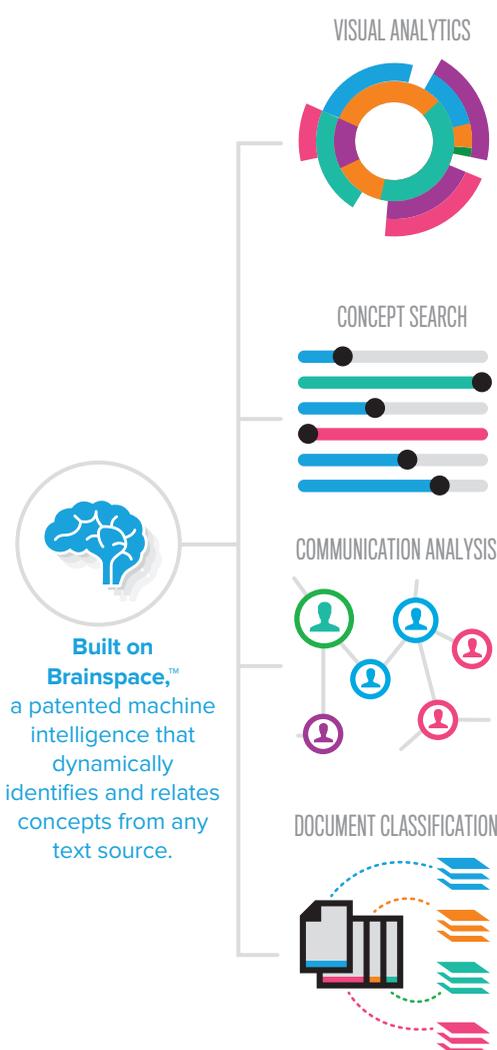# Brainspace Discovery™ 5

The industry's most advanced platform for E-Discovery and Text Analytics.

Brainspace Discovery™ 5 is the most comprehensive solution for analyzing unstructured data. From intelligent data preparation to exploration and collaboration, business professionals, analysts, and domain experts can use Discovery 5 to make decisions, evaluate and mitigate risks, and uncover insights and opportunities.

VISUAL ANALYTICS

CONCEPT SEARCH

COMMUNICATION ANALYSIS

Built on Brainspace,™ a patented machine intelligence that dynamically identifies and relates concepts from any text source.

DOCUMENT CLASSIFICATION

A BOLD STEP FORWARD

Rather than creating just another semantic search solution, Brainspace chose to rethink the method from the ground up. This new approach had to rise above the challenges and limitations that have hindered the other semantic search solutions on the market today — Brainspace semantic search had to be different. Applied first in e-Discovery, Brainspace Discovery™ had to learn from all of the documents loaded from a case — every single one. It had to scale to provide semantic search on extremely large and often changing data volumes as are encountered in early case assessment. Most importantly, Brainspace Discovery™ had to answer our clients' requests for complete transparency with control of each query — it could not be a black box. Having met and exceeded the exacting standards in e-Discovery, Brainspace is now an industry-leader offering a fully transparent, highly scalable, federated semantic search platform capable of handling the world's largest unstructured datasets.

THE PERFECT COMBINATION OF MAN AND MACHINE

Let's start with the machine. Brainspace's patented machine intelligence dynamically identifies and relates concepts from any text source and does so at a scale unmatched by any technology in the industry.

Now the important part — User Experience. What makes Brainspace Discovery™ so unique is the reimagining of "search" as a conversation between the machine intelligence and the human user. To make this possible Brainspace has developed several innovative, interactive visualizations.

Brainspace™

brainspace.com/discovery

SCALABLE POWER WITH ULTIMATE CONTROL

TRANSPARENT
No black boxes here. Each query is enhanced by the machine intelligence and shown to the user for their complete understanding.

FULLY FEDERATED
We can make your existing database smarter. We can also formulate semantic queries to run across virtually any index including internal document populations, intranets, extranets, enterprise content management systems, portals, email archives, and patent data.

COMPLETE USER CONTROL
Not only is our semantic search transparent, but users are given the ability to require, ignore, increase or decrease the importance of all query words in a unique visual query interface.

HIGHLY SCALABLE SEMANTIC TECHNOLOGY
Unlike other semantic technologies that must sample smaller sets of documents for learning, our largest semantic index currently includes learning from hundreds of millions of documents

INTEGRATED CONNECTIONS

INGEST DATA FROM COMMON PLATFORMS
Seamlessly transfer data back and forth through full integrations to systems such as Relativity® and Nuix.

IMPORT LOAD FILES
Using our simple import utility, you can load delimited text files such as DAT and CSV directly into Discovery.

The only analytics platform with a Brain.

BRAIN   BUILD

REDUCING NOISE

DUPE AND NEAR-DUPE DETECTION
Brainspace Discovery™ detects and marks all duplicates and near-duplicates during ingestion by comparing the body text of documents. This allows Brainspace to identify documents with the same or very similar content even if the native formats of the documents are different. With the duplicate and near-duplicate documents in your dataset clearly identified and grouped together, you can focus your analysis on unique documents without having to manually review and identify duplicative information. Typical dedupe processes rely on meta-fields, but the Brainspace methodology goes much further.

BOILERPLATE DETECTION
Brainspace Discovery™ automatically identifies boilerplate text in your documents and ignores it for clustering and learning. This ensures that semantically unrelated content does not distort the intelligence that Discovery gains from your documents. Boilerplate detection occurs automatically, however Brainspace also provides the ability to configure boilerplate detection settings, allowing you to specifically target text for inclusion or exclusion. This automated process significantly reduces the amount of operational time needed to get cases up and running.

Brainspace™

brainspace.com/discovery

BRAINSPACE ANALYSIS

CLUSTERING

We use a combination of hard clustering algorithms (where every document is assigned to a single cluster) and other propriety methods to produce a binary cluster tree. We later reshape this tree to a larger arity, also injecting synthetic clusters to create a cluster tree that better reflects the vocabulary found in the clusters.

PATENTED PHRASE DETECTION

Brainspace's patented Automatic Phrase Identification technology detects and creates phrase on the fly. Furthermore, it identifies word combinations with semantically similar context and creates them as phrases. Brainspace achieves all this automatically, without relying on lexicons, synonym lists, thesauri, or phrase lists. This allows for the detection of unique phrases not in lists, therefore increasing accuracy and relevance.

PATENTED MULTI-CONCEPT

Our patented Multi-Concept Detection technology recognizes and draws inferences from multiple concepts. In environments with multiple brains, each concept can access a different brain to draw relevant inferences. Unlike other solutions that depend on traditional Latent Semantic Analysis (LSA) that suffers from high recall and low precision, Brainspace's innovative approach that builds upon LSA with proprietary improvements significantly increases the accuracy without sacrificing recall.

FOREIGN LANGUAGE

Brainspace offers the ability to work with cases that span multiple languages. Brainspace is multi-lingual out of the box including CJK.

PEOPLE AND ORGANIZATIONAL GROUPING

Brainspace extract and reconciles all domains, senders, recipient and custodians

BRAINSPACE

While Latent Semantic Analysis (LSA) has been studied and used for decades, Brainspace has made significant strides beyond basic LSA to build multi-dimensional Brains of exceptional quality.

✓ Automatic identification and removal of repeated boilerplate text (e.g. email signatures).
✓ High-performance phrase recognition that far exceeds what is possible with n-gram phrases
✓ Identification and removal of near-duplicate documents
✓ Multi-Brain architecture enables larger scale analysis than a single-brain system.
✓ Multi-concept search can produce multiple brain queries.

REPORT GENERATION

EMAIL THREADING

Using Discovery's Email Threading feature, you can increase your review times by reviewing only the most inclusive email in a thread.

BRAINSPACE ID

Brainspace Discovery uniquely identifies each document and ranks them based on their textual content, effectively grouping semantically similar documents together. Using this powerful identification number, you can accelerate your review by smartly batching out documents based on their semantic relevance.

NEAR DUPLICATE CLUSTER ID

Brainspace automatically clusters together documents that are near duplicates of each other and assigns a unique identifier to each near duplicate cluster. This information can be exported out in a report or to another platform, allowing you to quickly identify documents that are near duplicates of each other even when you are not in Brainspace.

EXACT DUPLICATE CLUSTER ID

Brainspace automatically clusters together documents that are exact duplicates of each other and assigns a unique identifier to each exact duplicate cluster. This information can be exported out in a report or to another platform, allowing you to quickly identify documents that are exact duplicates of each other even when you are not in Brainspace.

CLUSTER RELATED SET ID

Brainspace automatically clusters together multiple exact duplicate and/or near duplicate clusters that are lexically similar to each other. Each parent cluster gets assigned a unique identifier, which can be exported out in a report or to another platform. This allows you to quickly identify documents that are in the same parent cluster even when you are reviewing the data set outside of Brainspace.
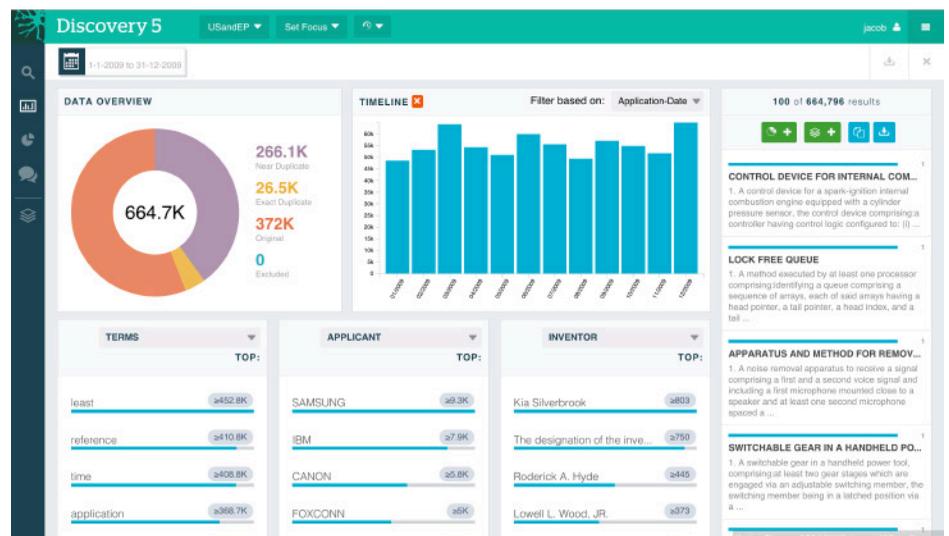
VISUAL ANALYTICS

THE POWER OF A 360° VIEW

Brainspace enables a truly unique Visual Analysis experience by Linking Multiple Views of Data in context including: Transparent Concept Search, timeline, clustering, communication analysis, and any structured data elements as well.

DASHBOARD

From a zero-state with no human decision-making, the Dashboard presents the essence of your documents. Every element on the dashboard is interactive, so with just a few clicks you can greatly narrow the field of investigation. You'll notice that as searches are executed, documents on the far right match, and the dashboard itself changes to reflect the narrowed population. The Dashboard makes it easy to find meaning in even the largest datasets.
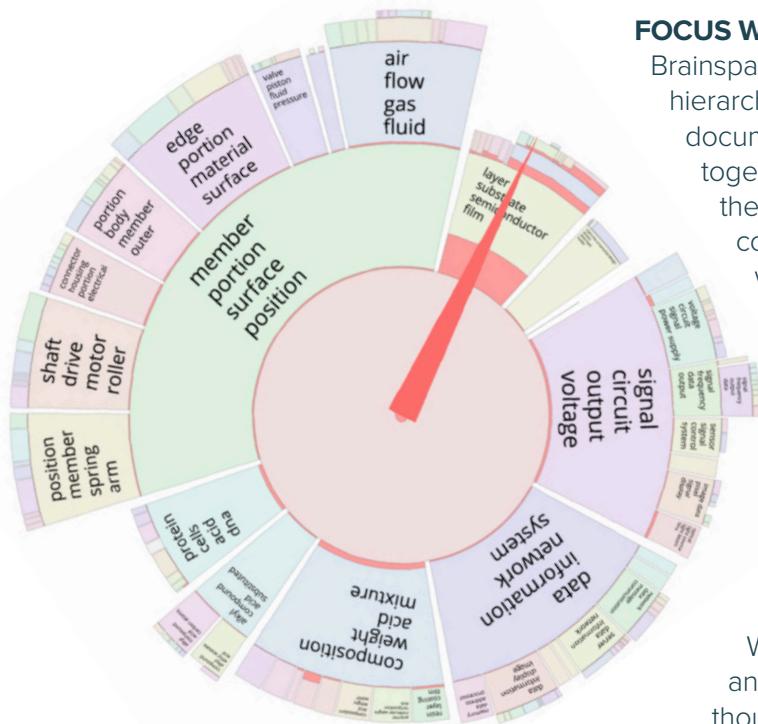

Brainspace™

DUPLICATE VIEW

A visual representation of the total volume of data including duplicates, near duplicates, and original documents. This view dynamically adapts to clicks, searches, and filters.

TIMELINE VIEW

See the volume of trends throughout the entire timespan of a datset, identify anomalies such as peaks and troughs, and drill into specific months or days for deeper analysis.

FACET TABLES

Brainspace Discovery automatically builds tables using available facets within your dataset and displays key information such as top terms, top individuals and top domains.

FOCUS WHEEL

Brainspace automatically groups documents in a wheel of hierarchical clusters, each containing lexically connected documents. Each document is placed into only one cluster together with documents similar based on the meaning of their text. Each cluster is identified with the main topics that connect the underlying documents. The wheel provides you with a bird's-eye view of the data set and allows you to interactively navigate and zoom into it to identify topics of interest, which then can be expedited for review. In a similar fashion you can identify areas of little interest that can be deprioritized.

CREATE FOCUS WHEELS ON THE FLY

You can create a focus wheel on demand based on any group of documents resulting from a concept search, communication chain, or even another document cluster. With this powerful feature you can quickly visualize your data and identify other topics of interest that you may not have thought about.
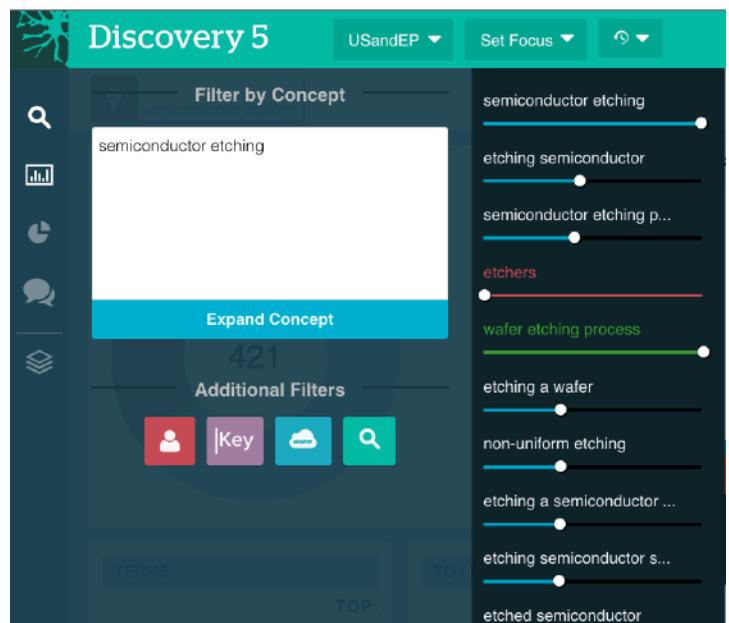
COMPASS NAVIGATION

After finding an exemplar document, Brainspace can easily navigate to the cluster of interest. Just click on the compass navigator and it will zoom to the document in the Focus Wheel. Zoom out to quickly identify the document's "neighborhood." This is extremely powerful for investigation and prioritizing review.

TRANSPARENT CONCEPT SEARCH

Using Brainspace Concept Search functionality you can use a word, phrase, or even a whole document to run searches against data sets in excess of hundreds of millions of documents to identify related terms. Brainspace achieves this by using patented technology that transformed traditional Latent Semantic Analysis (LSA) including Automatic Phrase Detection and Multi-Concept Detection.

Brainspace's developed a patented method for identifying multiple concepts in a corpus of text, extracting explicit and inferred terms from the Brain. Typically, semantic technologies average all concepts in a corpus - attempting to find a "semantic center." This is a significant deficiency and departure of how ideas

TRANSPARENT CONCEPT SEARCH (cont.)
are understood and communicated. Brainspace multi-concept queries enable a more human-like learning from any unstructured text, extracting multiple themes instead of an artificial average.

DEFENSIBLE CONCEPT SEARCH
Brainspace has the capabilty to take any concept search and turn it into a boolean expression. Unlike alternative "Black Box" concept search tools, this transparency takes the guess work out of concept expansion and delivers a versatile and defensible platform for attorneys.

ADJUSTABLE TERM WEIGHTS
Each query is expanded to a list of the most relevant inferred words and phrases (concepts). By using the slider under each term and phrase, the user can influence the intelligence, indicating importance, required terms or ignored terms, zeroing in on highly relevant results. Brainspace allows you to interactively manipulate the list of related terms and their weights using the search term 'sliders'. The result is a transparent and defensible query using an explicit list of terms that can return exactly the documents of interest.

INTELLIGENTLY FILTER YOUR SEARCH RESULTS
You can quickly cull your search results by using a combination of filters including email addresses, keywords, and domains.

KEYWORD EXPANSION LIST
After each concept search Brainspace returns a list of keywords with an associated relevance weighting. This feature allows a user to expand Keyword lists and ensure they don't miss potential keywords

RELEVANCE RANKING
Discovery ranks your search results by conceptual relevancy based on your search criteria, allowing you to prioritize your analysis and review process.
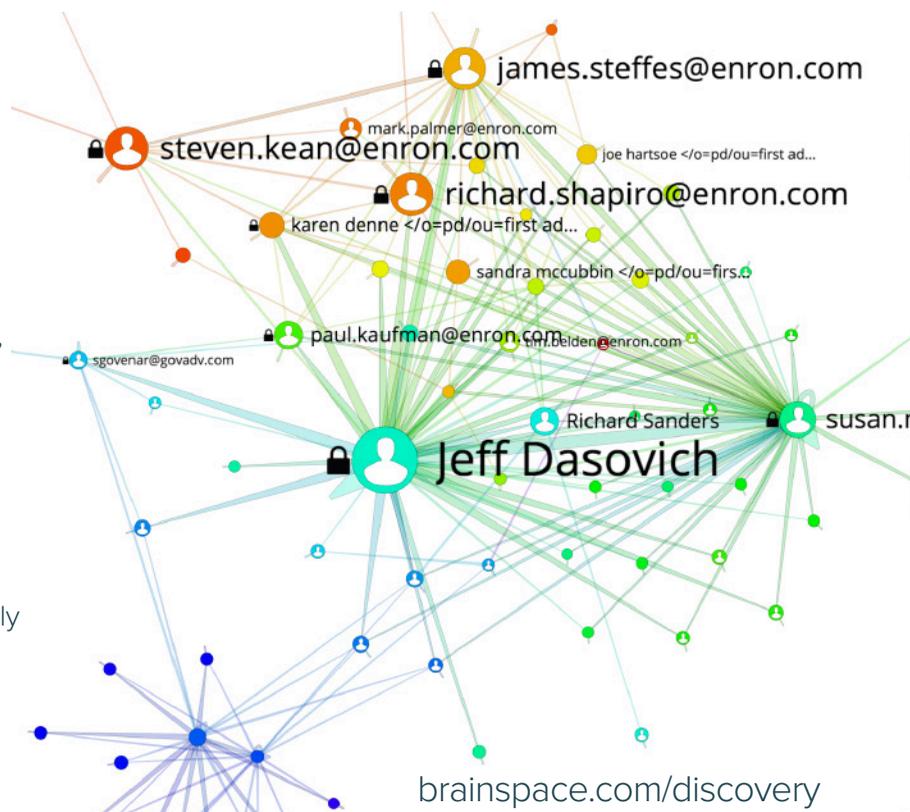
ADVANCED SEARCH
Using the Advanced Query Builder, automatically construct Lucene queries by combining any number of fields (metadata and those created by Brainspace) with Boolean operators. By using the query editor you can manually leverage powerful Lucene functions such as wildcard searches, fuzzy searches, proximity and range searches, term boosters and field groupings.

COMMUNICATION ANALYSIS
Communication Analysis interface provides you with an interactive map of all communication that took place in your data set. Using Communication Analysis, you can:

- View people-level or domain-level map of all communication traffic around your domain of focus.
- Immediately identify top sender, top recipients, and top terms.
- Analyze conversations between two or more individuals or domains including email traffic and topics discussed.
- Dynamically filter your view using a variety of attributes including To, CC, BCC, direction of traffic, and recipient count.
- Group to group communications- you can easily monitor communications between regulated and non-regulated divisions within your company and ensure they are in compliance.

brainspace.com/discovery

ACCLERATING REVIEW

BRAINSPACE ID
Brainspace Discovery uniquely identifies each document and ranks them based on their textual content, effectively grouping semantically similar documents together. Using this powerful identification number, you can accelerate your review by smartly batch out document based on their semantic relevance.

EMAIL THREADING
Using Discovery's Email Threading feature, you can increase your review times by reviewing only the most inclusive email in a thread.

PREDICTIVE CODING
Brainspace Discovery's Predictive Coding feature uses our patented machine learning technology together with Logistic Regression to help you review less and decrease your associated costs. Unlike other black-box solutions, Brainspace gives you more control by allowing you to set your target recall at the beginning and allow you to adjust it by providing you feedback on depth for recall performance throughout the process. Furthermore, Brainspace's Active Learning methodology automatically identifies and selects not only those documents that the classifier is most uncertain about, but also uncertain documents that are diversified across the collection, and provide insight into concentrations of interesting material.

- Brainspace offers predictive accuracy that is at or close to the best across a range of published literature and internal experimental results. [Beyond naive Bayes (e.g. Autonomy) and decision trees.]
- Brainspace is quick to build its model and predict the relevance of target documents. (e.g. quicker than nearest neighbor approaches like Content Analyst.)
- Brainspace provides prioritized ranking over the target documents by decreasing predicted degree of responsiveness, not just binary partitioning.
- Brainspace provides reasonable predictions of the probability that a document is relevant and a degree of human interpretability; that allows, for instance, the system to highlight those parts of a document that have led to a prediction of responsiveness.
- Unique Active Learning Methodology maximizes the value of each round assigning reviewers the documents that have the best opportunity to train the classifier
- Brainspace Depth of Recall Measure shows how many documents must be reviewed to be done.

COLLECTIONS
Any search, cluster, or document set can be tagged and placed into a collection. Collections can be named on the fly or set up prior to a case. Like folders or tags, responsive documents within your data set can be easily categorized, organized and shared with a click of the mouse.

DOCUMENT VIEWS

HIT HIGHLIGHTING
All relevant concepts that exist within the body text of a document are highlighted. You can quickly navigate to the highlighted portions to gain a deeper insight into the context the terms were used within.

DOCUMENT TOPICS
Prominent topics that appear in a document are identified and displayed, thus allowing you to get a quick understanding of a document without having to read it in its entirety.

Importance: Normal
Priority: Normal
Sensitivity: None
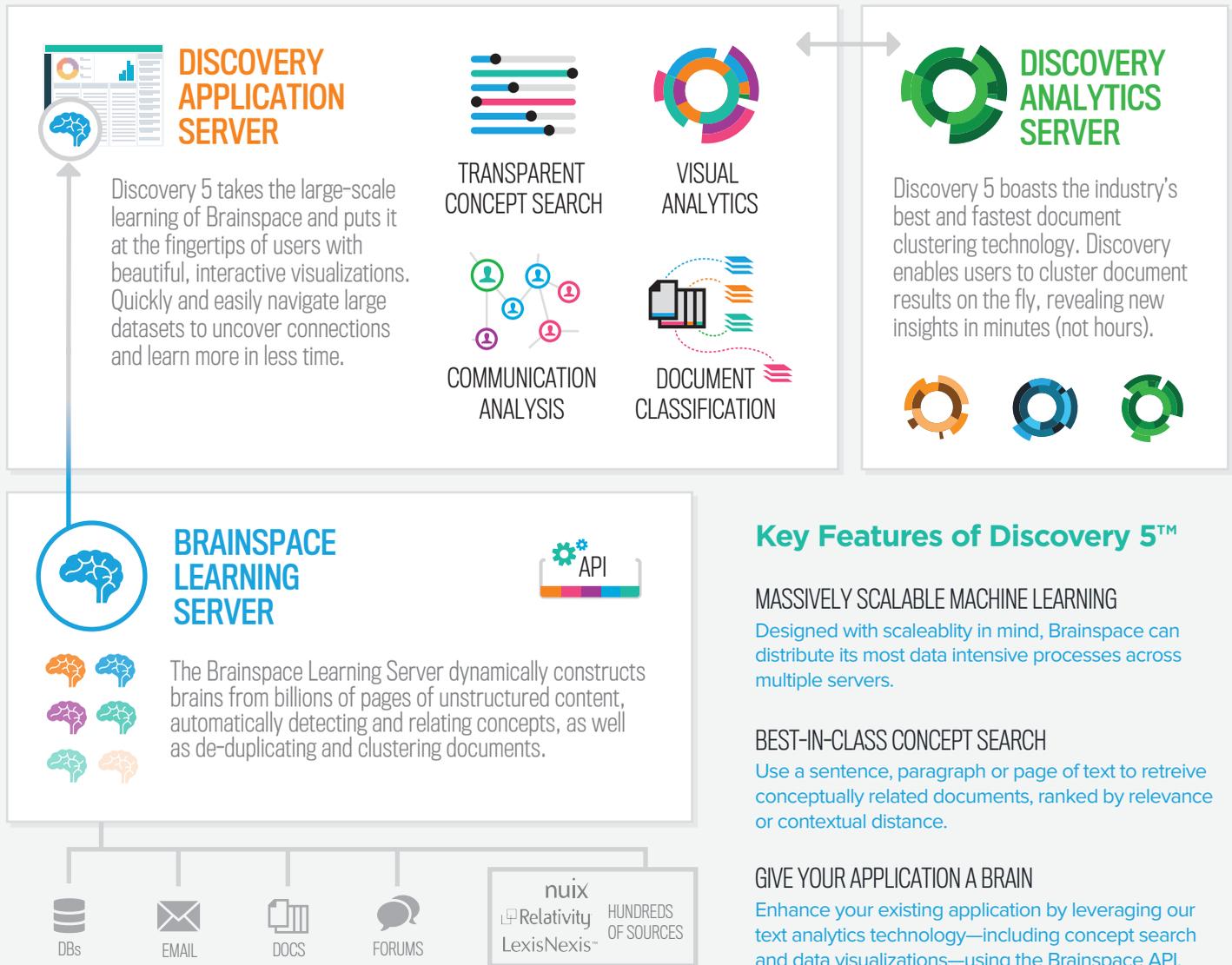
The attached swaps will be delivered by hard copy batch, together with the revised proposal sheets and another copy of the legal worksheets, to Julia Chin on Monday morning. Julia suggested due to the floor move on 27 that I wait until first thing Monday to have the batched execution versions delivered to her. Upon receipt, Julia will attach the valuation data that has been previously provided to her by Hope Vargas and coordinate execution. Michael Galvan and Anne Koehler are working together on final Active Power proposals sheets. The Active Power swaps will be drafted after the final proposal sheet has been approved by both Michael and Anne. As and when Proposal Sheets are provided on the privates we will commence drafting the private swaps.

DOCUMENT METADATA
All metadata that has been a part of the document since its creation as well as those that were created by Discovery are displayed, providing a complete picture of a document's history.

Brainspace™

The only analytics platform with a Brain.

Discovery 5 is powered by Brainspace™, the industry's most advanced, large-scale machine learning platform. Brainspace rapidly ingests millions of pages of unstructured text, dynamically learning without taxonomies or ontologies. This learning is surfaced through advanced, interactive visualizations, giving the full power of Brainspace to every user.

DISCOVERY APPLICATION SERVER

Discovery 5 takes the large-scale learning of Brainspace and puts it at the fingertips of users with beautiful, interactive visualizations. Quickly and easily navigate large datasets to uncover connections and learn more in less time.

TRANSPARENT CONCEPT SEARCH

VISUAL ANALYTICS

COMMUNICATION ANALYSIS

DOCUMENT CLASSIFICATION

DISCOVERY ANALYTICS SERVER

Discovery 5 boasts the industry's best and fastest document clustering technology. Discovery enables users to cluster document results on the fly, revealing new insights in minutes (not hours).

BRAINSPACE LEARNING SERVER

API

The Brainspace Learning Server dynamically constructs brains from billions of pages of unstructured content, automatically detecting and relating concepts, as well as de-duplicating and clustering documents.

DBs   EMAIL   DOCS   FORUMS

nuix
Relativity
LexisNexis™

HUNDREDS OF SOURCES

Key Features of Discovery 5™

MASSIVELY SCALABLE MACHINE LEARNING
Designed with scaleablity in mind, Brainspace can distribute its most data intensive processes across multiple servers.

BEST-IN-CLASS CONCEPT SEARCH
Use a sentence, paragraph or page of text to retreive conceptually related documents, ranked by relevance or contextual distance.

GIVE YOUR APPLICATION A BRAIN
Enhance your existing application by leveraging our text analytics technology—including concept search and data visualizations—using the Brainspace API.

Brainspace™

Brainspace Corporation is a pioneer and recognized leader in helping enterprise clients derive meaning, gain insights and identify human connections in unstructured data. Our unique solutions utilize our patented Brainspace platform and are leading the industry in text analytics, accelerating institutional learning and reinventing how organizations exchange knowledge and expertise. Our customers include the Fortune 500, leading consulting firms, legal service providers, and government agencies.