



VISUAL EVIDENCE / E-DISCOVERY LLC

## VeREVIEW 3.0 Document Analytics

**Visual Evidence** has partnered with **Content Analyst Company**, a leading developer of advanced analytics for searching and analyzing unstructured text – such as documents, emails, and publications. VeREVIEW includes **CAAT** from Content Analyst, a unique patented software that provides advanced Conceptual Search based on Latent Semantic Indexing. This mathematical technique enables **CAAT** to acquire all its search intelligence from the actual documents being indexed. Unlike other solutions, **CAAT** doesn't require external dictionaries or thesauri and can work across languages, even searching – and finding – relevant documents in other languages without prior translation.

The sheer volume of data now available to companies and government organizations challenges their ability to act quickly in a “sense and respond” environment. **CAAT** is a powerful example of a new class of technologies known as “Text Analytics” designed to transform large volumes of unstructured data into relevant actionable information. Text Analytics products automate most of the human activity traditionally associated with understanding, organizing, prioritizing and retrieving information from large sources of unstructured data.

While a variety of companies have developed applications with limited Text Analytics features, **CAAT** is fulfilling the promise of Advanced Text Analytics, giving organizations the ability to organize, access, and share information across multiple languages without the need for extensive human intervention.

---

### Capabilities

**Concept Search** – Using advanced mathematics to allow users to find similar, related, and relevant documents based purely on the concepts those documents are discussing – without using keywords, and without getting back “matching” yet irrelevant content.

**Categorization** - Allows users to define categories by means of examples. Based on the exemplars, Content Analyst technology automatically assigns incoming documents to categories.

**Instant Context (Contextual Explanation)** - Helps users understand unfamiliar terminology. The user clicks on an unfamiliar term, and Content Analyst technology highlights similar terms found in related text.

**Language Analytics** - Within any single language, Content Analyst technology can be applied to any topic, vocabulary or language that can be represented in the Unicode encoding system. In a cross-lingual mode, users can submit queries in English while searching documents in other languages.

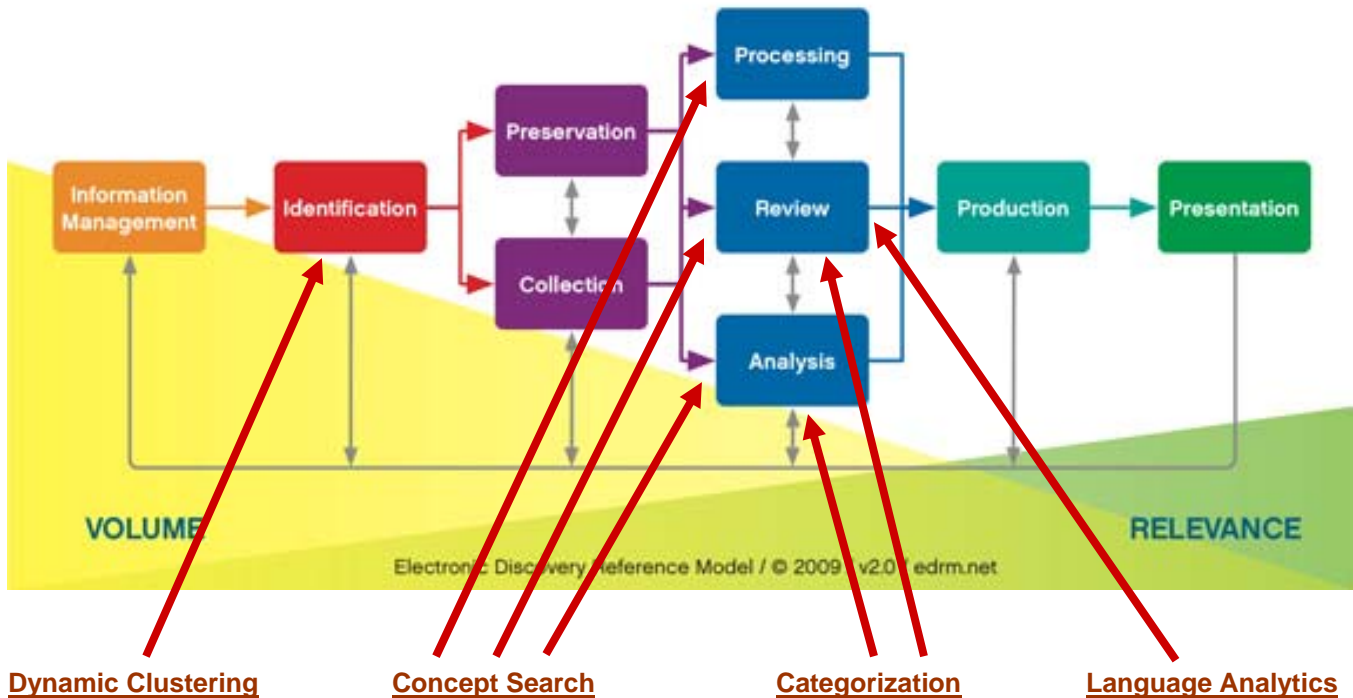
**Summarization** - Automatically identifies sentences in a document that best represent key concepts, and uses those sentences to give users a quick summary of the entire document.

**Dynamic Clustering** - Allows a user to point CAAT to a set of documents and then allow CAAT to dynamically group conceptually-similar documents together in a tree-type hierarchy and finally apply a descriptive title to each cluster of documents.

**Near Duplicate Document Detection** - Uses CAAT's Dynamic Clustering to identify and group documents that are duplicates or near-duplicate documents as well as identifying the extent of duplication.

## Document Analytics Demo

### Electronic Discovery Reference Model



---

**Dynamic Clustering** – We'll use a 500K multi-custodian collection of emails to show how clustering quickly separates work-related from non-responsive emails

**Concept Search** – We'll use depositions from the tobacco litigation suits to show how Concept Search finds relevant documents that are not easily identified via keyword

**Categorization** – We'll use the same emails and use categorization to find emails that relate to two specific and relevant accounting practices

**Language Analytics** – We'll perform cross-lingual searches against a set of United Nations documents written in five different languages

