Document Classification with Visual Similarity Clustering


Visual-Similarity Clustering automatically processes and groups documents together for document boundary detection and document type classification, regardless of source and format — seamlessly processing native electronic files and scanned documents.

BeyondRecognition™ Visual Similarity Technology enables: (1) use of graphical rules to process, extract and validate words; (2) creation of clusters of 99% visually similar documents and their classification into document types for extraction; and (3) use of graphical and textual rules for information appearing in particular zones on a page or in a document for extraction and validation.

Features and Capabilities

1. **Processing of words**: spelling and context corrections can be done once on many documents to speed later review. Word validation allows for three stages of automatic and human assisted text correction. Accuracy and Validation are persistent.

2. **Similar document typing** allows for automatic document clustering and classification based on graphical rules. Accuracy and Validation are persistent. Clusters presented as groups of documents for comparative review and decision as to status:

   ![Automatic document clustering based on graphical rules](image)
3. **Data Extraction & Validation** allows for automatic fielded coding based on both graphical and textual rules. Accuracy and Validation are persistent. Search on documents using text and graphical rules to find key words or phrases at a particular place on a page such as in a form, i.e. find and return all documents with the word “Jones” at this place on the page or within 1 inch (up, down, side-to-side) from this place on the page.

**BeyondRecognition™ by the Numbers**

- 560% faster, 75% lower cost and better accuracy than manual
- Scales to Big Data and enables very effective management of records and documents that are part of enormous datasets

**Flexible Deployment Options**

BeyondRecognition™ is offered as a hosted service or as licensed software for deployment behind the firewall. Licensed software and recommended hardware configurations are available with guaranteed capacity of 3 million, 6 million and 12 million pages per day. Hosted service pricing is available on a per document and/or estimated cost savings from existing processing regime. A cost effective storage system (as compared to Documentum™ and others) is available that is tuned to handle the output of the Appliance.

**About Xpriori, LLC**

Xpriori is a privately held company headquartered in Denver and Colorado Springs, CO. The company has developed proprietary technologies, products and services for the eDiscovery, compliance and review markets since 2006 as well as for XML-based data management. Xpriori offers the BeyondRecognition™ Visual Similarity and Clustering technologies and other leading eDiscovery solutions. Learn more about Xpriori at [www.xpriori.com](http://www.xpriori.com).

**Contact Xpriori at 720-439-9937 for more information, arrange a demonstration or discuss your needs and requirements.**