

Automatic Classification

Its Role and Benefits

At Your Service...Today...Tomorrow
We Appreciate The Privilege Of Serving You!



In This Issue: Automatic Classification

Automatic Classification Role

By Eleonora Babayants

An ability to find content in a content management system is crucial. One of main goals of having a CMS is to make content easy to find, so you can take an action, make a business decision, do research and development work, etc.

The main challenge to findability is to anticipate how users might look for content. That is where categorization comes into play. The quality of the categorization of each piece of content makes or breaks its findability. Good tagging will last the lifetime of the content.

Success Factor

The success factor for good content tagging is the automatic categorization techniques and processes. It applies the rules more consistently than people do. It does it faster. It frees people from having to do the task, and therefore has fewer costs.

You can experience these benefits by using concept-based categorization driven by an analytics engine integrated into a CMS. An automatic classification system analyzes example documents you provide to calculate concepts that can be used to categorize other documents.

Identifying hundreds of keywords per term, they are able to distinguish relevance that escapes keyword and other traditional taxonomy approaches. They are even highly likely to make connections that a person would miss.

Conceptual Analysis

Taxonomy systems require someone to enter the newly developed or discovered terms. In conceptual analytics, it is simply a matter of providing additional example documents that automatically add to the system's conceptual understanding.

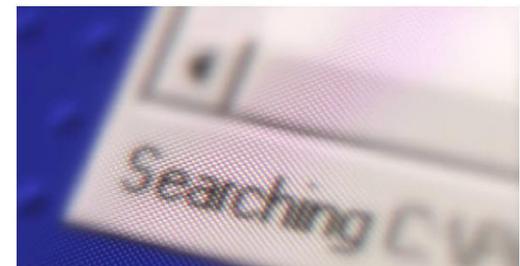
Good categorization and tagging systems are essential to good content management and to controlling expense and exposure.

Effective and thorough tagging can identify obsolete content to reduce costs and exposure related to legal or regulatory requirements.

Efficiently removing obsolete content means less storage, bandwidth, and archive volume.

Superior categorization through conceptual analysis also affects operational efficiency by enabling fast, accurate, and complete content gathering. A significant benefit for any enterprise is that it allows more time for actual work by reducing the time it takes to find necessary information.

Concept-based analysis makes content findable and actionable, regardless of language, by automatically categorizing it based on understanding developed from example documents you provide. Both internally and externally, the company becomes more competitive with one of its most important assets which is unstructured information.



Automatic Classification Role

The success factor for good content tagging is the automatic categorization.

Automatic Classification Benefits

Specific benefits are: consistency, larger context, persistency, cost effective.



Industry News

- Splice Machine Partners with Lucidworks
- Unified Enterprise Information from Colligo
- Cognitive Scale Launched Cognitive Cloud Platform
- Expert System Launched Cogito API

Automatic Classification Benefits

by Eleonora Babayants

Organizations receive countless amounts of paper documents every day. These documents can be mail, invoices, faxes, or email. Even after organizations scan these paper documents, it is still difficult to manage and organize them.

To overcome the inefficiencies associated with paper and captured documents, companies should implement an intelligent classification system to organize captured documents.

With today's document processing technology, organizations do not need to rely on manual classification or processing of documents.

Organizations that overcome manual sorting and classification in favor of an automated document classification and processing system can realize a significant reduction in manual entry costs, and improve the speed and turnaround time for document processing.

Recent research has shown that two-thirds of organizations cannot access their information assets or find vital enterprise documents because of poor information classification or tagging.

The survey suggests that much of the problem may be due to manual tagging of documents with metadata, which can be inconsistent and includes errors, if it has been done at all.

Technology Solutions

There are few solutions for automated document classification and recognition.

Some of them are: SmartLogic's Semaphore, OpenText, Interwoven Metatagger, Documentum, CVISION Trapeze, and others. These solutions enable organizations to organize, access, and control their enterprise information.

They are cost effective and eliminate inconsistency, mistakes, and the huge manpower costs associated with manual classification.

Putting an effective and consistent automatic content classification system in place that ensures quick and easy retrieval of the right documents means better access to corporate knowledge, improved risk management and compliance, superior customer relationship management, enhanced findability for key audiences and an improved ability to monetize information.

Specific Benefits

Specific benefits of automatic content classification are:

More consistency. It produces the same unbiased results over and over. Might not always be 100% accurate or relevant, but if

something goes wrong, it is at least is easy to understand why.

Larger context. Enforces classification from the whole organizations perspective, not the individuals. For example, a person interested in sports might tag an article which mentions a specific player, but forget/not consider a team and a country topic.

Persistent. A person can only handle a certain number of incoming documents per day, whilst an automatic classification works round the clock.

Cost effective. Possible to handle thousands of documents much faster than a person.

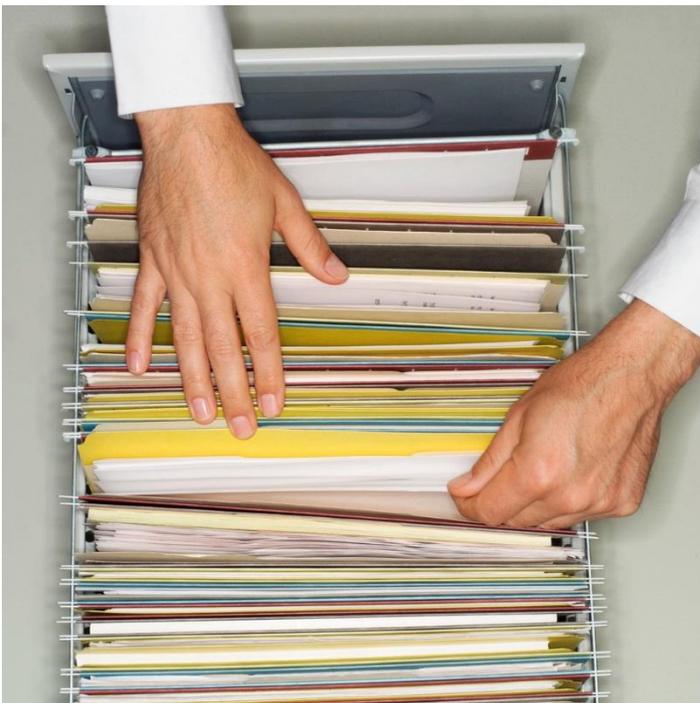
Automatic Classification Types

Automatic document classification can be divided into three types:

- supervised classification - some external mechanism (such as human feedback) provides information on the correct classification for documents;
- unsupervised classification (also known as document clustering) - the classification must be done entirely without reference to external information;
- semi-supervised classification - parts of the documents are labeled by the external mechanism.

Automate your content classification and consider using manual labor mainly for quality checking and approval of content.

To overcome the inefficiencies associated with paper documents, companies should implement an intelligent classification system



Industry News

Splice Machine Partners with Lucidworks

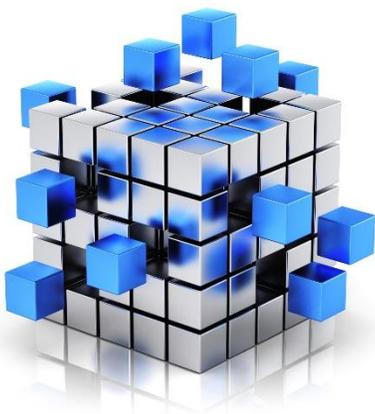
Splice Machine, provider of the only Hadoop relational database management system (RDBMS), has partnered with Lucidworks, the search platform for large companies and brands worldwide. The partnership will enable Splice Machine customers to access and analyze their unstructured data via Lucidworks Search, which delivers Apache/Solr for commercial-grade applications.

The companies say the deal will enable non-technical users of Splice Machine-powered applications to find data quickly, without being proficient in SQL. Application users will also have access to Lucidworks' enterprise-class features, such as advanced search and indexing.

Unified Enterprise Information from Colligo

SharePoint data synchronization provider Colligo has released a new collaboration platform that is said to securely unite corporate information across existing enterprise systems.

The Colligo Engage platform has been engineered to boost worker engagement and productivity with an organization's sanctioned systems through a family of applications that are built around everyday tools such as Outlook, Windows Explorer, Mac Finder, phones and tablets.



Cognitive Scale Launched Cognitive Cloud Platform

Cognitive Scale has announced Insights Fabric, which the company describes as a portable, open, standards-based cognitive cloud platform that protects and accelerates a company's existing investments in big data.

Setting the cognitive cloud platform apart, says Cognitive Scale, is its ability to securely source, process, integrate and extract insights from structured and unstructured data, often described as "dark data," which includes information found in databases, devices, blogs, reviews, e-mails, social media, images and other unstructured data sources. Dark data contains highly valuable and relevant information and makes up more than 80 percent of data, says the company.

Expert System Launched Cogito API

Expert System has released the Cogito API. The company claims the fully configured API series allows businesses to organize, link, find and understand unstructured information within corporate intelligence, CRM, human resources, data management, knowledge management and social media monitoring applications.

Using Cogito API, the topics, concepts, entities, relationships and sentiment expressed in any large collection of text can be analyzed and understood; the output is exported as XML, RDF or another format and made instantly ready to use in enterprise solutions ranging from customer care, sentiment analysis and advanced business intelligence.

About Galaxy Consulting



Galaxy Consulting was founded with the mission and vision of helping organizations to manage their valuable information assets. Many of our clients, both large and small, have dramatically improved efficiency and reduced unnecessary labor hours through efficient methods, processes, and solutions we created.

Galaxy Consulting believes in partnerships with our clients. We are committed to working with you and to helping you transform your business. We will increase efficiency and productivity, maintain regulatory and legal compliance, improve collaboration, enhance innovation, and reduce costs through effective information management!

Call us TODAY to schedule a free, no obligation consultation!

Contact Us

Office: 650-474-0955

Mobile: 650-716-3609

info@galaxyconsulting.net

www.galaxyconsulting.net



Galaxy Consulting provides services in business analysis and usability, content and knowledge management, records management, information architecture, enterprise search, taxonomy development and management, document control, and information governance.