conceptSearching

8 Reasons to Use conceptClassifier for SharePoint 2010

Using Concept Searching's **Smart Content Framework™** conceptClassifier for SharePoint delivers automatic conceptual metadata generation, auto-classification and taxonomy management, fully integrated with SharePoint 2010 and running natively with the Term Store.

1. To Auto-classify Content

The Issue: SharePoint 2010 provides the facility to create taxonomies via the Term Store, but no mechanism exists to automatically classify documents against these hierarchical metadata structures.

What This Means: Microsoft has provided the mechanism to create taxonomies and integrated these metadata models into the search engines, but without an auto-classification capability these facilities have no value, because manual classification of documents will lead to incomplete and inconsistent tagging of documents.

The Solution: conceptClassifier can automatically classify content in SharePoint 2010 against any number of taxonomies. Content stored in SharePoint is processed in real time and the automatically generated metadata is added in real-time.

The metadata is stored in SharePoint properties, and authorized users can manually adjust the tags if necessary. This eliminates manual metadata tagging by end users and provides comprehensive and consistent tagging of documents.

2. Classify Using Concepts Not Just Keywords

The Issue: Accurate document classification needs to use multi-word concepts in matching algorithms because single words in isolation are inherently ambiguous and leads to inaccurate tagging.

What This Means: To deliver quantifiable business benefits an accurate document classification solution must be able to identify multi-word concepts, not just keywords, when matching the documents to the nodes in the taxonomies.

The Solution: conceptClassifier uses a unique advanced statistical concept identification technology - compound term processing - to classify unstructured documents against taxonomy structures.

This mechanism provides highly accurate automatic document classification because the ambiguity in single words is eliminated.

3. Classify Using Powerful Rules

The Issue: Accurate document classification is about more than simply finding word patterns in a document.

What This Means: You need a variety of classification rules to perform accurate classification for many different types of documents and taxonomy structures.

The Solution: In addition to unique automatic concept extraction, conceptClassifier also provides advanced classification rules including:

- Standard
- Case-Sensitive
- Metadata
- Phonetic
- Regular Expressions
- ClassID
- Language

4. Advanced Taxonomy Management

The Issue: Taxonomy Management must be easy to use, accurate and efficient.

What This Means: Administrators must be able to easily create the rules for auto-classification, tune the classification results, and test the accuracy of classification.

The Solution: conceptClassifier's taxonomy manager component provides many advanced features required by business driven taxonomy deployments:

- Integrated environment combines classification rules development with the actual documents to be classified
- Automated clue suggestion based on document content analysis
- Instant feedback of rule refinement based on re-classification in real time

5. Secure Taxonomy Management

The Issue: Taxonomy Management must be secure, provide an audit trail, and highly scalable.

What This Means: It must be possible to restrict access to the taxonomy management tools audit changes made to the classification rules, and also allow multiple administrators to work on the same structures simultaneously.

conceptSearching

The Solution: concept**Classifier's** taxonomy manager component provides a secure and scalable platform:

- User management based on Windows or non-Windows identities
- Full audit trail, with rollback facility
- Locking facilities to enable concurrent taxonomy development by many users

6. Integration with the Refinement Panel
The issue: Microsoft has enhanced the SharePoint 2010 search engines with easy-to-use faceted search facilities via the Refinement Panel, but without accurate and consistent classification these facilities do not provide improved search results to end users.

What This Means: SharePoint 2010 users will realize no value if the search refinement facilities fail to deliver meaningful search results because the documents have not been accurately tagged.

The Solution: concept**Classifier** stores the metadata that it generates in SharePoint Properties that are designed to work with the new search refinement facilities. When all documents have been accurately and consistently classified then the refinement facilities will be able to improve the search experience.

7. Tag and Secure Content Assets
The Issue: SharePoint 2010 cannot provide a comprehensive solution for the identification of Personally Identifiable Information (PII), Personal Healthcare Information (PHI), For Official use Only (FOUO) or other types of organizationally defined sensitive information, thereby making them candidates for security exposures and available to unauthorized users during the search process.

What This Means: Regardless of industry, sensitive information exists in many forms within an organization. The most serious is PII and PHI, but may also include content about new product development, Board of Directors reports, mergers and acquisitions, or any confidential information that the organization would like to protect from unauthorized viewing or use.

The Solution: concept**Classifier** is used to detect sensitive information or potential data security breaches automatically, in real time.

While other technologies can detect patterns, concept**Classifier** is the only solution that combines pattern matching with associated vocabulary. Suspect documents can be immediately tagged and locked down if necessary.

Using Concept Searching's concept**ContentTypeUpdater** the Content Type can be automatically changed when the identified documents contain the sensitive information, so that workflows and rights management can be initiated automatically to process and protect the suspect documents.

8. Enable More Effective Records Management
The Issue: SharePoint 2010 does not have the functionality to automatically tag documents with retention codes for automated Record Management purposes.

What This Means: In SharePoint 2010 the records management capabilities have been largely improved. However, to declare records the end user must still select the correct retention code that has an associated Content Type. Due to the inconsistency of human tagging, records cannot be declared or may have the wrong content type assigned.

The Solution: concept**Classifier** is used by many organizations to identify and declare documents of records in SharePoint and across file shares, then tag these with appropriate retention codes as well as vocabulary.

This enables content that is created or ingested to be automatically tagged with both vocabulary and retention codes. Where a record refers to multiple retention policies it can also invoke a content type change automatically and route the records management application for disposition.

Smart Content Framework™
Concept Searching's **Smart Content Framework**™ is a toolset that provides the enterprise framework to mitigate risk, automate processes, manage information, protect privacy, and address compliance issues. The flexible and innovative framework and unique technology solutions deliver rapid and quantifiable results.

Microsoft Partner
Gold Independent Software Vendor (ISV)
Silver Content Management
Silver Search
Silver Portals and Collaboration