Machine-Generated MetaData in SharePoint

How BA Insight AutoClassifier Integrates with the SharePoint Managed Metadata Service

BA Insight
June 2014

Table of Contents

Abstract

Metadata is essential for findability, yet the vast majority of content within organizations has poor metadata. Creating metadata is challenging, and doing it by machine is the only practical way to get consistent, quality metadata on a broad basis. SharePoint provides a facility for referencing and managing metadata called the Managed Metadata Service (MMS). The BA Insight AutoClassifier leverages this to provide a rules-based engine fully integrated with SharePoint. This paper discusses the need for Metadata and demonstrates how to use the AutoClassifier and MMS to create metadata and improve findability.

Findability and the Value of Metadata

Finding Information is Hard

In a Findability survey conducted by FindWise in 2013, 63% of respondents reported that locating information within their organization is hard or very hard. FindWise also reported that this percentage is a 10% increase since 2012. In other words, workers at every level are finding that the process of locating the information that they require to perform their jobs is an escalating challenge.

Metadata is Essential for Findability

The exponential growth of information requires new methods for managing content and search. If documents are not tagged with the appropriate metadata, then queries return every document that contains the search term, with no refinement of search results.

To meet today's information challenges, metadata must be used to enrich the end user search experience. With good metadata on content, search results are more relevant as priority is given to documents in which the query matches the metadata (over documents where the query matches only the text within the document). Also, faceted search (also called refinement or faceted navigation) can be used to quickly drill down into search results to get to the right information.

Microsoft SharePoint Incorporates Metadata

Microsoft introduced the Managed Metadata Service (MMS) as part of SharePoint 2010 and enhanced it in SharePoint 2013. MMS includes a term store that holds term sets (also known as taxonomies or ontologies). SharePoint OOB also includes verbatim extractors (string matchers) to populate managed properties in search and lets users manually tag content.

How the BA Insight AutoClassifier Improves SharePoint

The BA Insight AutoClassifier, which is the core of the BA Insight Content Classification engine, leverages MMS to precisely define the metadata tags and automate the process of classifying information to improve the user experience. Taxonomies that are stored in the term store are used to define the tags used as metadata on documents. The AutoClassifier automatically tags these documents and also assigns content types based on these taxonomies. The AutoClassifier rules that specify how documents are tagged are stored in the term store as well, unifying the management of terms and rules, making administration much simpler.

The Challenge: Creating and Managing Metadata

In order to meet today's information challenges, metadata must be:

- fully customizable in order to meet organizational requirements
- precise in order to eliminate false positives
- standardized and comprehensive across all subsets of information
- automatically and immediately applied with manual update capabilities
- utilized to enrich the end user search experience

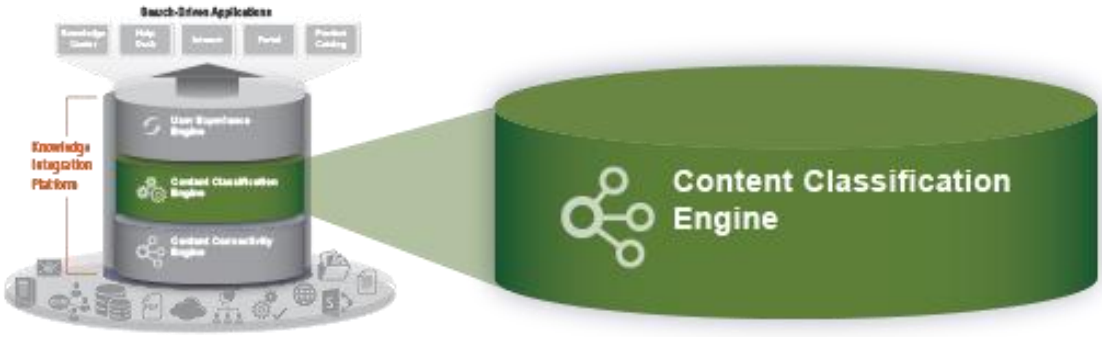
While it is well known that metadata is the key to effective search, in enterprises good metadata is rare. On the web, pages are tagged and curated manually by people who have a strong incentive to do so and a large SEO industry behind them. In the enterprise, it is impossible to tag everything by hand. Requiring authors and content contributors to tag everything usually backfires as it is very time consuming and people often avoid tagging or circumvent systems set up to force them to tag. For example, one organization found that over 40% of their repository was tagged with the first term on the dropdown list. People simply picked the quickest choice, not the correct choice. Inconsistent tagging and incomplete or simply wrong metadata results in poor search – and also degrades compliance.

In this environment, automated tagging is the only practical solution. Manually generated or process-generated metadata can be supplemented (or corrected) by machine-generated metadata very cost-effectively. Machines may not be perfect at tagging based on human language, but they are much more consistent and often more accurate than humans.

Using the AutoClassifier

Overview of the AutoClassifier

The BA Insight AutoClassifier provides automatic tagging and classification of content based upon a taxonomy, using a robust rules engine.

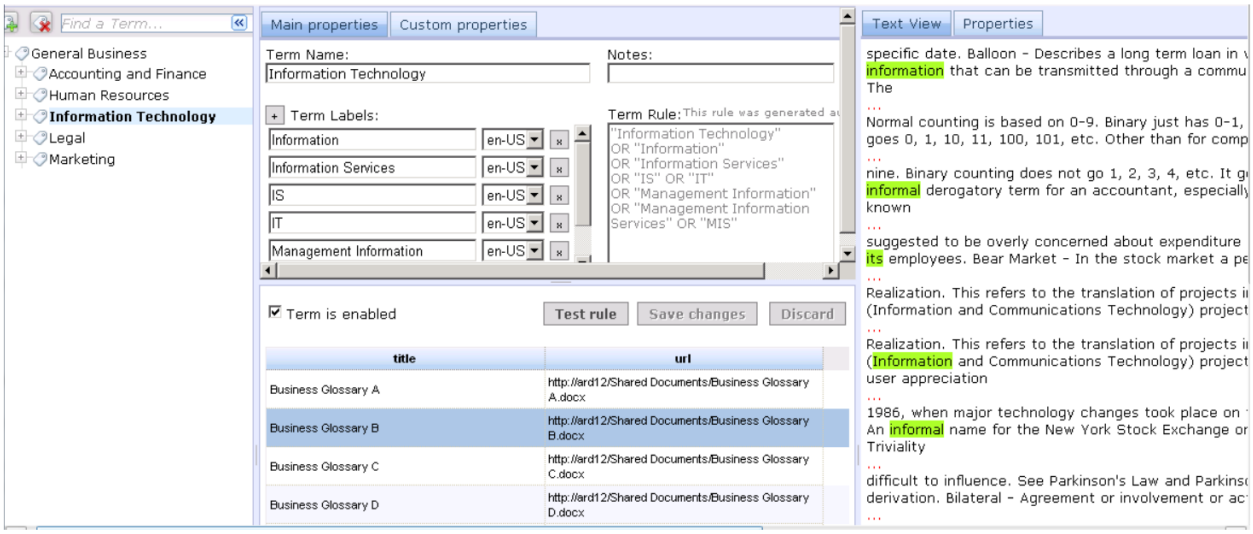


Instead of manually tagging documents, BA Insight AutoClassifier can be used to automatically classify crawled documents according to a defined taxonomy. AutoClassifier automatically adds the defined taxonomy node names, or terms, to the MMS term store and tags the matching documents after these files are crawled.

Managing Taxonomies and Rules

The AutoClassifier includes a taxonomy manager to develop both taxonomies and rules. These are stored in the term store and are usable throughout SharePoint since they are fully integrated with MMS. Each taxonomy node contains associated terms that are stored in the term store. A tag, or taxonomy node name, is applied to a crawled document if the document is a match based on the associated rule. Rules provide control over the application of metadata. They establish the conditions under which metadata should be applied to a document and may involve multiple criteria. For example, a rule might specify that a matching document must include the terms *President* and *Bush* and exclude the word *plant* in order to be classified with the metadata tag Presidents.

Rules are generated automatically by default, and can then be specified within the taxonomy manager. Taxonomies can also be developed in the same environment.



As the taxonomy is developed, the AutoClassifier test bench can be used to:

- Test taxonomy rules
- See the matching documents
- View the highlighted matching terms within the text of the classified document
- Edit rules as required

The AutoClassifier automatically applies these rules to content in SharePoint or to content outside of SharePoint that is crawled for search. This results in consistent metadata on all content. When users search, their results are more relevant, and the refiners are fully populated resulting in better exploration. End users:

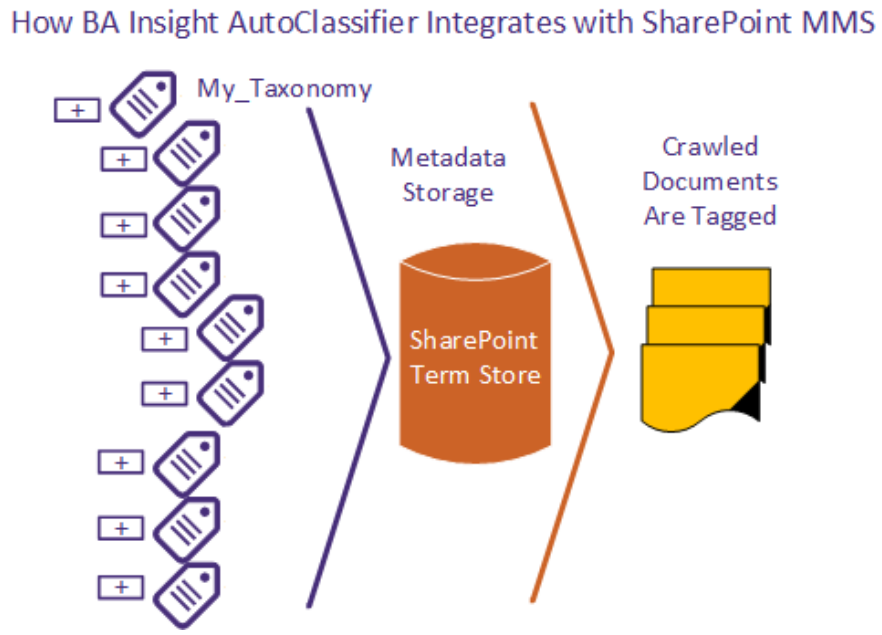
- Experience faster search
- Get a more accurate set of matching documents
- Enjoy a more intuitive search experience
- See pages that are generated based on the taxonomy hierarchy

Benefits Summary

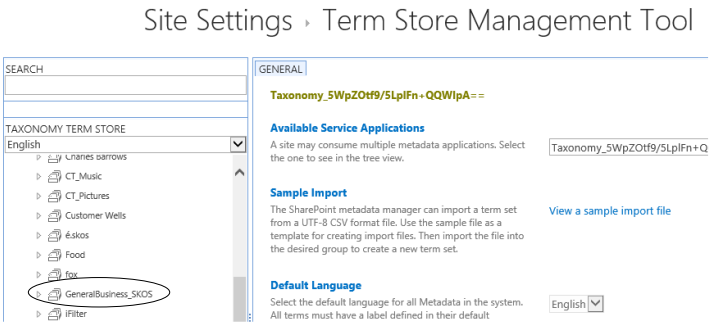
Metadata is the backbone of successful information architecture. Microsoft's Managed Metadata Service is a powerful facility that centrally controls taxonomies, enterprise keywords, and content types. With the addition of the BA Insight AutoClassifier, an organization can realize the full potential of an organized information architecture that enriches the end user search experience.

How the AutoClassifier Integrates with MMS

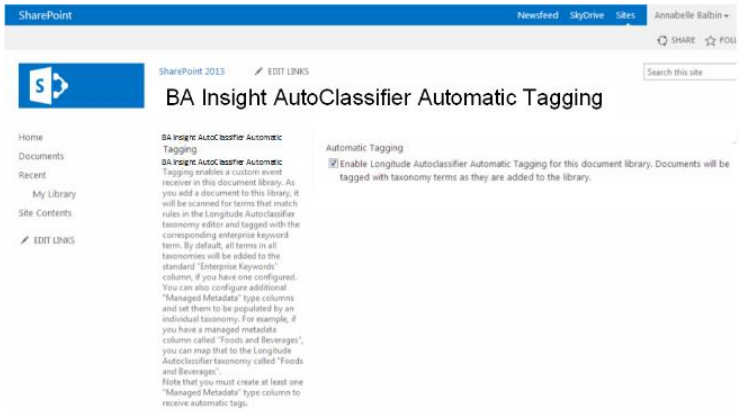
BA Insight AutoClassifier integrates seamlessly with SharePoint MMS to provide a customized set of metadata where the terms are stored in the SharePoint term store and automatically applied to crawled documents as they are indexed by AutoClassifier.



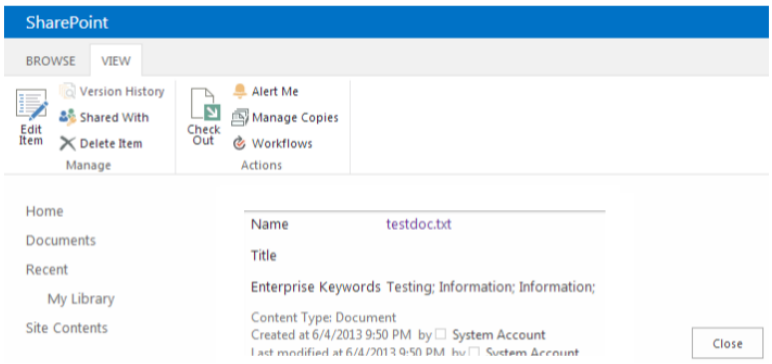
- The terms and content types that are defined in the AutoClassifier taxonomies are automatically stored in the SharePoint term store:



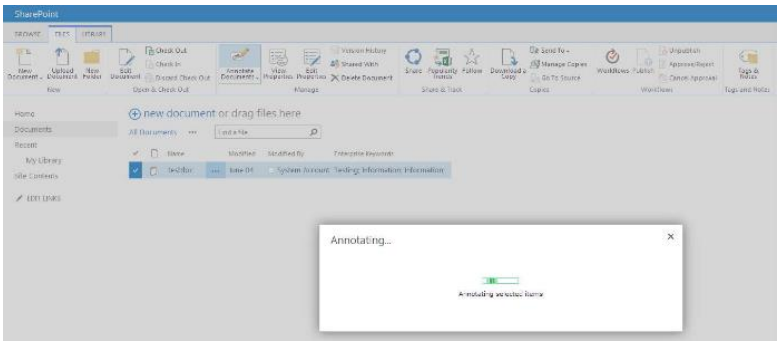
- The BA Insight AutoClassifier automatically tags crawled documents in SharePoint.



- The AutoClassifier document tagging within SharePoint can be verified.



- Documents can be automatically or manually tagged.



- The BA Insight AutoClassifier is an intuitive, easy-to-install application. A custom taxonomy can be developed, or the sample taxonomy shipped with AutoClassifier can be tweaked.

Summary

BA Insight AutoClassifier enables organizations to meet the challenges of customization, precision, and auto-classification while fully integrating with Microsoft SharePoint by leveraging MMS. It provides a rules-based engine, applying text analytics and semantic processing to add consistent, quality metadata to content surfaced within SharePoint.

About BA Insight

BA Insight provides software that enables organizations to transform SharePoint into a powerful unified information access platform for rapid implementation of search-driven applications at a fraction of the cost, time, and risk of alternatives. With our platform, our customers deliver a remarkable user experience, auto-classification, and connectivity to a wide variety of systems. It can function as a comprehensive solution or its components can be implemented in a phased approach to meet growing organizational needs.

We serve visionary organizations such as ADP, Australia Department of Defence, Bayer, Chevron, Deloitte, Ford Motor Company, Green Mountain Coffee, Pfizer, Rio Tinto, The Procter & Gamble Company, U.S. Army, and the U.S. Department of Veterans Affairs. Visit www.BAinsight.com for more information.