Questions from the Webinar

This is a set of common questions from the audience during the September 29, 2011 Webinar entitled “Best Practices for Taxonomy and Managed Metadata in SharePoint 2010” presented by WAND, Inc. and Term Management LLC. We’ve divided the questions into groups for convenience.

This webinar was recorded and can be viewed again at


Please contact us if you would like any details about a question below or if you have any other information needs:

Mark Leher
WAND, Inc.
mleher@wandinc.com
(303) 623-1200 x267

Mike Doane
Term Management, LLC
Mike.doane@gmail.com
(206) 683-8857
Pre-Built Taxonomies

Q: Do you have any mining taxonomies built?
A: WAND has taxonomies covering mining equipment and supplies currently available. A comprehensive Mining Taxonomy will be released in mid-October 2011.

Q: Do the pre-built taxonomies come with definitions?
A: WAND Taxonomies do not have pre-populated definitions.

Q: Do you have samples of tags that HR departments use as a starting point?
A: The WAND General Business Taxonomy includes branches for Accounting and Finance, HR, IT, Legal, and Sales/Marketing. The HR section will provide a set of terms that can be used as starting point for your own HR term sets. The WAND General Business Taxonomy, available as a free download from http://datafacet.com/signup.aspx?feat=GBT_SP2010

Q: Is there a foundation taxonomy available for {insert your desired topic here}?
A: WAND’s library of taxonomies is growing on a regular basis. Please contact us directly if there is a taxonomy that you need but is not on the list. We regularly announce new taxonomies as they are developed and released, usually on Twitter @taxonomies and at our blog which is located at blog.wandinc.com. If we do not have a taxonomy available that you need, WAND and Term Management can assist in building custom taxonomies on a professional services basis for any topic and we would be happy to assist in building this taxonomy.

Q: Are the taxonomies available in multiple languages?
A: WAND has taxonomies available in multiple languages, including German, Spanish, French, Italian, Portuguese, Traditional Chinese, Simplified Chinese, Japanese, Korean, and Vietnamese. Limited parts of the taxonomies are translated into Swedish and Finnish. Taxonomies can be translated into any languages quickly (using human translators).
**Taxonomies in SharePoint**

**Q: Can SharePoint 2010 handle associative relationships between terms (i.e., related terms)?**

A: SharePoint 2010 supports broader terms, narrower terms, and synonyms (labels). It does not support more complex relationship types, such as related terms.

**Q: Can SharePoint 2010 automatically pull out references to concepts or entities (e.g., people, products) from the content/text?**

A: SharePoint 2010 does not have any entity extraction capabilities.

**Q: Is SharePoint able to export term sets out of the box?**

A: SharePoint 2010 does not have the ability to export term sets from the Term Store. For this reason, it is a best practice to get as much taxonomy work completed on the import spreadsheet before the term set is imported into the SharePoint 2010 term store, because developing taxonomies in a spreadsheet will be more efficient.

As a side note, with WAND’s DataFacet Automatic Tagging Engine, there is a capability to export term sets into SKOS format. Support for export into other formats will be made available in future releases.

**Q: Do you need to clean up existing term sets if replacing with an updated Excel file?**

A: It’s best to not import a Term Set if one with the same name is already in the Term Store. Best practice would be to include all the terms in that Term Set that already exists in the Term Store in an Excel spreadsheet, add new terms or edit terms in the spreadsheet, delete the Term Set in the Term Store and re-import it from the spreadsheet. That way, you won’t have copies or duplications of terms. Be careful, though, if you’ve already used existing terms elsewhere in your taxonomies and then delete them. It will break the “re-use” relationship and may cause problems in other Term Sets.

**Q: What is the largest size (# of rows) CSV that can be imported into the term store?**

A: There are a few maximum limits to take note of as it relates to the term store:

- Maximum number of levels of a term set: 7
- Maximum number of term sets in a term store: 1000
- Maximum number of terms in a term set: 30,000
- Maximum number of items (terms and term sets) in a term store: 1,000,000

You could not have a 1,000 term sets, each with 30,000 terms, in a term store because this would exceed the maximum limit of total items of 1,000,000.
Q: Are keywords and terms the same thing?

A: Keywords and terms can be used interchangeably, in general, but there are a few specific important vocabulary items to be aware of as it relates to the term store. Managed terms are terms which have been placed in a Term Set. Managed terms can be organized into a hierarchy. Managed keywords are terms which have been entered by end users and are not part of a Term Set. Managed keywords do not have a hierarchy. It is a best practice to review managed keywords on a regular basis as a source of terminology which end users may have used and may be good candidates for “promotion” into a managed term set, either as a new managed term or as a new label/synonym.

Q: Is there a way to see statistics or metrics on how much a term is being used?

A: There is no capability to do this out of the box in SharePoint 2010.

Q: When you say Label - do you mean Term?

A: Label is what SharePoint 2010 calls a synonym. So, labels are other words for the Managed Term. For example, “HR” could be a good label for “Human Resources”

Q: Can taxonomies be created to make data dependent? For example, can you create cascading drop downs?

A: You cannot create cascading dropdowns in default SharePoint 2010, but you can use the hierarchy structure of Term Sets to see how terms relate to one another (e.g. you can see the Parent Term of terms, if available).

Q: What’s the best way to use managed metadata to assign and apply retention policies in SharePoint?

A: You could set up a series of Term Sets that contain tags for the processes in retention policies and apply those tags in a custom workflow when creating and managing policies. This sounds like something that would have to be customized in the policy retention process, but is something that most likely can be done.

Q: when is the best time to start addressing governance in a taxonomy project?

A: It’s best to address governance right from the start of a taxonomy project, but it certainly can be addressed anywhere along the course of the project. In fact, sometimes it’s better to have some of the taxonomy work done before you start to set up governance. That way, people on the governance committee already have something to work with rather than having to start from scratch.
Q: Is there any detailed documentation or whitepaper about the Term Store Management available?


Q: Do you have any guidance on when to use one managed metadata service versus multiple managed metadata services?

A: SharePoint 2010 can support multiple Managed Metadata Services (MMS)/Term Stores. The same site collection can subscribe to more than one MMS. One approach is to keep a default MMS which contains centralized corporate terminology that is common across the organization and can be easily maintained at a high level. Then, domain specific MMS can be set up which would be managed and populated by specific subject experts in a given area. This would be terminology that would not be easily managed at a centralized level.

Q: A managed term or keyword is available and tagged to documents by end users. Then, you delete the word from the Term Store. What happens to the documents that were applied with that tag?

A: The tag will remain on the document, but it will be identified as an “orphan” term that is no longer managed by the Term Set.

Q: Is there any way to reuse parts of the taxonomy term store, but allow departments to customize beyond a certain sublevel in the hierarchy?

A: Branches of term sets can be re-used across other term sets using the “Re-use” option in the term store taxonomy manager. This will copy a branch of taxonomy over to a new term set. If you make changes to any terms in this re-used set, such as renaming it, changing the description, adding a label, or adding a translation, this change will populate to any place the term was re-used. However, if you add a narrower term to a re-used term, it will not be automatically added to other re-used instances of the term. This allows you to re-use higher levels of the taxonomies and then customize them locally.

Q: What are the folksonomy terms?

A: The Folksonomy capabilities of the SharePoint 2010 term store are really expressed through the use of Managed Keywords. Managed Keywords are terms which users have keyed into tagging fields in SharePoint 2010 versus selecting from a Managed Term.
Automatic Tagging Questions

Q: Does Automated Tagging cause additional load on the server?

A: The DataFacet Automatic Tagging Engine can be installed on existing SharePoint 2010 hardware and day to day tagging should not cause any noticeable degradation on the performance of the server. However, if tagging a large document set, we recommend scheduling this for off-peak hours.

Q: Can the managed metadata auto-classify uploaded documents to content types?

DataFacet cannot be used to automatically set content types. DataFacet can only populate managed metadata terms. Content types cannot be changed based upon managed metadata values natively in SharePoint 2010.

Q: Is the DataFacet web part similar to the refiners feature?

A: The DataFacet search web part is very similar to the out of the box search refiner web part. The major difference is that the out of box refiners web part presents the managed metadata tags in a flat list while the DataFacet web part provides a hierarchical view of the taxonomy so that users can more easily navigate through and find the term which they wish to refine by in context of the complete taxonomy.

Q: You mention Auto tagging as preferable to manual tagging but we mention that a key reason for tagging is that the content of the document does not contain those keywords --- does this auto tag use content of the document to derive the tags?

A: An important value for tagging documents is that the metadata will be indexed by the search engine and so end users will be able to do a keyword search for the managed metadata and find tagged documents even if the tagged document does not contain the keyword in its text. Automatic tagging applies tags to documents based upon the text of a document, so it does rely on keywords that are already inside of the document and thus would have shown up regardless. There are a couple of key items to note, however. First, rules that drive the automatic tagging are Boolean style rules that can encompass a variety of keywords or concepts. For example, if an organization wants to tag its human resources documents with the managed metadata value “HR”, but many of these documents might not have the keyword HR or Human Resources inside of it. In this case, a rule can be created that looks for other keywords that may indicate a document is an HR document. In another case, a search for a keyword may be too broad. For example the word “Contract” may appear in a lot of documents which are not actually contracts, A keyword search will bring back all of these documents with many irrelevant results. By creating a more focused rule for the taxonomy term “Contracts”, for example including rules for words you may often see in contracts such as “parties”, “recitals”, “governing law”, and others, then you can make sure that only contract documents are tagged as “Contracts” providing a more relevant and effective way to find these documents.

Finally, an end user may be searching for keywords that are not managed metadata values. In this case,
it is valuable to have keywords available as managed metadata values because they will be refinement options which the searcher can use to filter his keyword search results to a more manageable set.

Q: Does your product support FAST Search in addition to Enterprise Search?
A: Yes, DataFacet has an integration with both SharePoint enterprise search and FAST search, among a few other engines.

Q: Is DataFacet an add-on to SharePoint?
A: Yes, DataFacet is an add-on to SharePoint that is installed as a managed service application. SharePoint 2010 does not provide any native rules based tagging capabilities and mostly relies on manual tagging as well as some tag inheriting through mechanisms like Document Sets.

Q: How are rules created for automatic tagging purposes?
Rules can be created using standard Boolean operators such as And, Or, And Not, parenthesis, ranges, wild cards, and more. Please contact us if you’d like a copy of the DataFacet Query Language, which outlines all of the operators available to create rules.

A rule is created for each of the terms in the taxonomy and then each rule is applied to a document at the time it is checked in. Documents are then tagged based upon those rules which are valid for the given document, meaning the rule was “True”.