Session 1: Business Taxonomy and Metadata Design
Welcome

John Challis  
CEO and CTO  
johnc@conceptsearching.com

Don Miller  
VP Product Development  
donm@conceptsearching.com

Jill Hannemann  
Principal Consultant  
jhannemann@ppc.com
About Concept Searching

- Company founded in 2002
  - Product launched in 2003
  - Focus on management of structured and unstructured information

- Technology
  - Delivered as a web service
  - Automatic concept identification, content tagging, auto-classification, taxonomy management
  - Only statistical vendor that can extract conceptual metadata

- 2009 and 2010 ‘100 Companies that Matter in KM’ (KM World Magazine) and Trend Setting product of 2009 and 2010

- Authority to Operate Enterprise wide USAF and Enterprise wide NETCON US Army

- Locations: US, UK, & South Africa

- Client base: Fortune 500/1000 organizations

- Managed Partner under Microsoft global ISV Program - “go to partner” for Microsoft for auto-classification and taxonomy management

- Microsoft Enterprise Search ISV, FAST Partner

- Product Suite: conceptSearch, conceptTaxonomyManager, conceptClassifier, conceptClassifier for SharePoint, contentTypeUpdater for SharePoint
About PPC

**Energy/Environment**
Green strategies for government and industry:
- Air quality and climate change
- Greenhouse gas reduction
- Carbon management
- Environmental risk mitigation
- Environmental impacts of transport
- Information and data management

**Infrastructure**
Design, build, and manage innovative IT infrastructure that provides reliability and performance for public and private sectors

**1,000-person multi-disciplinary team of scientific and technical experts**
- Scientific subject matter experts
- Systems engineers and architects
- Policy and regulatory specialists
- Project management professionals
- Certified Information technology experts
- Security professionals

**Solutions Development and Strategy**
Technology solutions for improved data and information management, streamlined business processes, and accomplishment of government and commercial missions

**Information Management**
In-depth strategy and strong implementation expertise for our top commercial and government customers

---

Session 1: Business Taxonomy and Metadata Design
Session 1: Business Taxonomy and Metadata Design

Jill Hannemann
Principal Consultant
jhannemann@ppc.com
Taxonomy and Metadata

1. Definitions
2. Benefits
3. Design
1. Definitions
Taxonomy and Metadata

Metadata
1. Data about data

Taxonomy
1. The classification of organisms in an ordered system that indicates natural relationships
2. The science, laws, or principles of classification; systematics
3. Division into ordered groups, categories, or hierarchies
Primary tools to provide structure to unstructured information

Depending on system design and use, may be front-end or back-end functionality

Taxonomy (categorization) is often actualized by applying metadata to documents

Enable Findability
Taxonomy and Metadata

Metadata Values (As Taxonomy)

**Audience**
- Internal
  - Executives
  - Managers
- External
  - Suppliers
  - Customers
  - Partners

**Topics**
- Employee Services
  - Compensation
  - Retirement
  - Insurance
  - Further Education
- Support Services
  - Infrastructure
  - Supplies
- Products and Services
- Finance and Budget

Metadata Values (As Taxonomy)
Business Taxonomies

- Tend to be less rigid and constrained
- Influenced by usability concerns
  - Minimize number of “clicks”
- Often content-driven
  - Ensure balanced content distribution
- Allow flexibility, redundancy
  - Items may be organized into multiple categories
  - May support multiple taxonomies for disparate audiences
- May use one or more different categorization approaches
Controlled Vocabulary

- A pre-selected list of words or phrases used to tag information
- Promotes easy retrieval through web browsing or searching
- Can be used in taxonomies
- Leverage controlled vocabularies to improve tagging consistency across lists and libraries in SharePoint

Product Taxonomy
Sports & Recreation
Home Entertainment
Personal Use
Yard & Garden
Home Maintenance
Home Furnishings & Fixtures
  - Bathroom Fixtures
  - Beds, Mattresses, Pillows
  - Carpets, Rugs
  - Chairs, Sofas, Sofa Beds
  - Desks, Cabinets, Shelves
  - Ladders, Stools
  - Holiday, Party Supplies
  - Electric Fixtures
  - Lamps
  - Heating, Cooling, Ventilation
  - Household Container
2. Benefits
Front-end Visibility

- Traditionally, taxonomies are used as values to populate one or many metadata fields
  - Back-end visibility
  - Used for integration and classification
- Business Taxonomies are typically front end “folder” structures
  - Front-end accessibility
  - Used for navigation as well as integration and classification
Usability and Simplicity

• What you lose with a Business Taxonomy
  – Absolute granularity
  – Ultimate classification

• What you gain
  – Usability
  – Simplicity
Consistency

- Metadata facets/term sets in SharePoint:
  - Can be used across several lists and libraries
  - Search can be driven by facet terms
  - Content can be tagged with commonly used terms
How Taxonomy Impacts SharePoint

- Information Architecture
- Core Navigation
  - Site Structure
    - Sites, pages, lists, and libraries
  - Tree view and breadcrumb trail
- Managed Metadata Service
  - Hierarchical term sets
- Supporting Metadata Uses
  - Content Types
  - Site Directory (by Metadata)
  - Columns
  - Content Targeting

Energy Working Group

Energy Working Group » Calendar » Calendar »
Use the Calendar list to keep informed of upcoming meetings, deadlines, and other important events.
Faceted Browsing and Searching

Find computers by category, weight, screen size, etc.

Session 1: Business Taxonomy and Metadata Design
How Metadata Impacts Search

- Implementation of Faceted Search capabilities takes metadata fields and enables users to browse for information by a specific metadata value.

Faceted Search

By Document Type
- Brochures (3)
- Forms (3)
- Policies (5)
- Presentations (10)
- More...

By KM Topic
- Content Management (4)
- Governance (15)
- Search (2)
- Usability (5)
- More...
3. Design
**Taxonomy Building**

**Manual/Top-down**
- Work with librarians and functionally- or subject-based individuals or Focus Groups
- Identify overall ontology and major categories of information
- Subdivide categories as necessary to build taxonomy
- Individual-driven; may entrench obsolete or arbitrary categories

**Automated/Bottom-up**
- Identify overall ontology and major content collections
- Analyze content collections using automated textual analysis tools
- Reveal major and minor topics of information; build taxonomy based on the relationship of these topics
- Content-driven; may reveal new associations of information

---

Concept Searching | Project Performance Corporation:

**Expert Webinar Series**

Session 1: Business Taxonomy and Metadata Design
## Categorization Schemas

<table>
<thead>
<tr>
<th>Method</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Facet-based</strong></td>
<td>Information categorized into multiple taxonomies or “stackonomies” based on unique but pervasive characteristics including topic, function, etc.</td>
<td>Wines by region</td>
</tr>
<tr>
<td></td>
<td></td>
<td>France &gt; Alsace</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wines by type</td>
</tr>
<tr>
<td></td>
<td></td>
<td>White &gt; Chardonnay</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wines by price</td>
</tr>
<tr>
<td><strong>Subject-oriented</strong></td>
<td>Information categorized by subject or topic.</td>
<td>Water pollution, soil pollution, air pollution, etc.</td>
</tr>
<tr>
<td></td>
<td>➢ Instantive - each child category is an instance of the parent category</td>
<td></td>
</tr>
<tr>
<td></td>
<td>➢ Partitive - each child category is a part of the parent category</td>
<td></td>
</tr>
<tr>
<td><strong>Functional</strong></td>
<td>Information categorized by the process to which it relates</td>
<td>Employment, staffing, training</td>
</tr>
<tr>
<td><strong>Organizational</strong></td>
<td>Information categorized by corporate departments or business entities</td>
<td>Human Resources, Marketing, Accounting, Research...</td>
</tr>
<tr>
<td><strong>Document Type</strong></td>
<td>Information categorized by the type of document</td>
<td>Presentations, expense reports, press releases, etc.</td>
</tr>
</tbody>
</table>
• Users can easily search documents or list items based on taxonomy and metadata fields

**Document Type**
- Budget
- Marketing Material
- Policy
- Presentation
- Publication
- Regulation
- Report
  - Quarterly
  - Annual
- Template

**Metadata Field = KM Topics**
- Strategic Planning and Analysis
- Governance
- Taxonomy and Metadata
- Content Management
- Knowledge Continuity
- Portal Strategy and Development
- Search
- Usability

**Filtered Content**
- Presentation about Governance
- Quarterly Report about Usability
- Template for Portal Strategy
Risks and Challenges

• Lack of Understanding
  – The primary concepts of taxonomy and metadata are misunderstood

• Complexity
  – Too deep and too wide; too much jargon
  – Too detailed

• Lack of Compliance
  – Many users will not provide effective metadata

• Resistance to Change
  – Users fear losing control of their content

• Delay and Avoidance
  – Content is difficult to find; too many documents and folders
Best Practices

• Define your Use Case
  – Understand how and why you will be using taxonomy and metadata

• Keep your Audience in Mind
  – Recognize that users may think about and look for information in different ways

• Define Governance
  – Roles, responsibilities, policies, and procedures

• Control Depth and Breadth
  – A “flat” taxonomy ensures that users can find information quickly
  – A focused taxonomy ensures that users can easily “digest” the scope of information

• Make a Long-Term Investment
  – Taxonomy development is an iterative and on-going effort
Session 1: Business Taxonomy and Metadata Design

John Challis  
CEO and CTO  
johnc@conceptsearching.com

Don Miller  
VP Product Development  
donm@conceptsearching.com
What’s Missing in 2010?

Driving business value with metadata

1. Enterprise Managed Metadata Service and the Term Store
2. Term Store Applications
3. Manual vs. Automated approach to applying metadata
4. Does a native integration into Term Store for SharePoint 2010 really matter?
5. Demo
The Term Store

• The purpose of the Term Store is to store and manage and hold all of your corporate taxonomies (i.e. Term Sets)
• Term Sets provide guidance to your valuable information and semantics
• However, Term Sets do more than that—they can be used to drive records management with purposeful classification (manual tagging) that is aligned with federal and corporate mandates
Introducing **EMM**, the **Term Store**, and **Term Store Management Definitions**

The Term Store is integrated to the Enterprise Managed Metadata (EMM) Service. Multiple Term Sets can be used in conjunction with a Content Type. The EMM allows you to leverage Content Types via the Content Type Hub.
Enterprise Managed Metadata

- Enterprise Managed Metadata Services drive the following native 2010 applications:
  - Term Store (Taxonomies – Tagging Structure)
  - Manual Classification (Corporate tagging)
  - Content Type Publishing (Providing structure for tags and content targeting)
  - Folksonomies (Personal organization tagging)
  - Social Tagging (Leveraging the tagging of the masses)
  - Records Management (Aligning tagging with corporate/federal policies)

- The problem is that every application requires an end user to apply a manual metatag (PHYSICAL COST $) to drive value, which is not necessarily always going to be accurate.
A Manual Metadata Approach Will Fail 95%+ of the Time

<table>
<thead>
<tr>
<th>Issue</th>
<th>Organizational Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inconsistent</td>
<td>Less than 50% of content is correctly indexed, meta-tagged or efficiently searchable rendering it unusable to the organization <em>(IDC)</em></td>
</tr>
<tr>
<td>Subjective</td>
<td>Highly trained Information Specialists will agree on meta tags between 33% - 50% of the time <em>(C. Cleverdon)</em></td>
</tr>
<tr>
<td>Cumbersome - Expensive</td>
<td>Average cost of manually tagging one item runs from $4 - $7 per document and does not factor in the accuracy of the meta tags nor the repercussions from mis-tagged content <em>(Hoovers)</em></td>
</tr>
<tr>
<td>Malicious Compliance</td>
<td>End users select first value in list <em>(Perspectives on Metadata, Sarah Courier)</em></td>
</tr>
<tr>
<td>No perceived value for end user</td>
<td>What’s in it for me? End user creates document, does not see value for organization nor risks associated with litigation and non conformance to policies</td>
</tr>
<tr>
<td>What have you seen</td>
<td>Metadata will continue to be a problem due to inconsistent human behavior</td>
</tr>
</tbody>
</table>

The answer to consistent metadata is an automated approach that can extract the meaning from content eliminating manual metadata generation, yet still providing the ability to manage knowledge assets in alignment with the unique corporate knowledge infrastructure.
Should I Write My Own Auto Tagging Tool?

**conceptClassifier for SharePoint 2010 provides an automated metadata approach for an immediate ROI and drives business value**

- Create enterprise automated metadata framework/model
  - Average return on investment minimum of 38% and runs as high as 600% *(IDC)*
- Apply consistent meaningful metadata to enterprise content
  - Incorrect meta tags costs an organization $2,500 per user per year – in addition potential costs for non-compliance *(IDC)*
- Guide users to relevant content with taxonomy navigation
  - Savings of $8,965 per year per user based on an $80K salary *(Chen & Dumais)*
  - 100% “Recall” of content, 35% Faster access to content “Precision”
- Use automatic conceptual metadata generation to improve Records Management
  - Eliminate inconsistent end user tagging at $4-$7 per record *(Hoovers)*
  - Improve compliance processes, eliminate potential privacy exposures
Proprietary or Fully Integrated Approach

Advantages:
- Configuration of the Taxonomy (Term Sets) using native SharePoint facilities
- Full Linkage between Taxonomies (Term Sets) and Managed Metadata Properties
- Full integration of “Managed Metadata Properties” vs. “Creating Custom Properties” into the Microsoft SharePoint Search and FAST Search for SharePoint search refinement panel
- Editing of Taxonomy (Term Sets) metadata using the native SharePoint editing facilities
- Reduced risk as Microsoft enhances the SharePoint 2010 platform advantage can be immediately taken of the new functionality

Disadvantages:
- Polyhierarchies unsupported by the Term Store
- Term sets with more than 30,000 terms unsupported by SharePoint 2010
- Term Stores with more than 1,000,000 terms unsupported by SharePoint 2010
3. Demo
Contact

John Challis
CEO and CTO
johnc@conceptsearching.com

Don Miller
VP Product Development
donm@conceptsearching.com

Jill Hannemann
Principal Consultant
jhannemann@ppc.com

Session 1: Business Taxonomy and Metadata Design
Upcoming Webinars in the Series

**Designing Information Architecture for SharePoint: Making Sense in a World of SharePoint Architecture:**
June 29th - 11:30am-12:30pm EST
Initial configuration of SharePoint 2010 information architecture can seem daunting navigating between collections, sites, lists, libraries, and web parts. This session will discuss how to implement a taxonomy and metadata schema to begin the functional planning of how users will interact with the various elements set up within a SharePoint environment. We also look at the out of the box functionality around the SharePoint 2010 Term Store and when to use a Third-Party tool such as Concept Searching’s conceptClassifier for SharePoint to drive value and deliver a superior end user experience.

**Leveraging Taxonomy Term Store for SharePoint: Defining a Multi-taxonomy Structure for Content Management:**
August 10th - 11:30am-12:30pm EST
This webinar will focus on providing strategy and best practices to designing a taxonomy and metadata schema to work with the Term Store for SharePoint 2010. There are a number of ways to manage content by leveraging this new service, whether the intent is for social collaboration or rigid categorization. PPC will review the different strategies you may take to leverage the Term Store effectively and in a manner that suits the business objectives and culture of your organization. We also review the difference between proprietary taxonomy solutions that replace the term store and those that integrate natively with the term store and why that is important to both end users and the Enterprise. Concept Searching will demonstrate their Term Store integrated Taxonomy and Auto-Classification solution that leverage PPC’s strategy and best practices.

**SharePoint Governance: Managing Content Sprawl:**
September 14th - 11:30am-12:30pm EST
Once deployed within your company, SharePoint’s popularity has the potential to become viral. This session will focus on how to apply a governance strategy against the SharePoint sites and objects, and how best to manage user expectations for leveraging SharePoint within your company. We also look at how using Concept Searching’s Concept Classifier for SharePoint you might automate much of the process designed to deliver a consistent user experience at retrieval time using taxonomy and automatic content tagging. Furthermore we explore using the tool to apply your Governance strategy to identify and lock down sensitive information such as PII from being published on uncontrolled portals.

**De-mystifying Content Types: Four Key Content Types to Leverage:**
October 19th - 11:30am-12:30pm EST
Content types are a powerful feature of SharePoint 2010 and are largely under-utilized. Learn more about content types, what they can do and how to implement them across your SharePoint environment. PPC will also share four key content types to implement that span multiple industries. We also review Concept Searching’s Content Type Updater, an automatic content tagging solution that can apply content types based upon vocabulary and metadata. The solution, fully integrated with SharePoint 2010 and the Term Store can then workflow specific types of content based upon policy and guidelines addressing such business issues as preservation and disposition, risk, and Governance.