Webinar 1: Designing Information Architecture for SharePoint: Techniques for Tailoring SharePoint for your Organization
Welcome

Tom Reamy
Chief Knowledge Architect
KAPS Group
tom@kapsgroup.com

Jill Hannemann
Director of Knowledge Management
Project Performance Corporation
jhannemann@ppc.com

Donald Miller
Vice President Sales – Commercial Accounts
Concept Searching, Inc.
donm@conceptsearching.com
Agenda

• Introductions
• Overview of Concept Searching
• Introducing The “Smart Content Framework™ Building Blocks”
• PPC - Information Architecture for SharePoint
• KAPS Group - Text Analytics and Taxonomies – Bridging the Taxonomy Gap
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


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

- Locations: US, UK and 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
• Enterprise Content
• Information Governance
• Introducing The “Smart Content Framework™ Building Blocks”
  • Metadata
  • Insight
  • Governance
  • Policy
  • Privacy
  • Enterprise 2.0
What is Enterprise Content?

Spectrum of enterprise content
- Transactional content
- Business content
- Persuasive content

Types of content:
- Transactional content:
  - Scanned images
  - Computer reports
  - Forms
  - Fax
  - Corporate records
  - Document correspondences

- Business content:
  - Office documents
  - Web content
  - Corporate records
  - Messages

- Persuasive content:
  - Web content
  - Product catalogs
  - Rich media
  - Collateral
  - Blogs/wikis

Examples:
- Transactional content:
  - Invoices
  - Loans
  - Claims
  - Tax returns

- Business content:
  - Sales proposals
  - Technical documentation
  - Employee training
  - Business plans
  - Contracts

- Persuasive content:
  - Customer self-service
  - eCommerce
  - Multichannel marketing
  - Corporate communications
  - Channel extranet

Technologies:
- Transactional content:
  - Imaging
  - E-forms
  - Cold/ERM
  - Document output management
  - Workflow/BPM

- Business content:
  - Document management
  - Records management
  - Web content management
  - Team collaboration
  - Workflow
  - Portal

- Persuasive content:
  - WCM
  - Personalization
  - Campaign management
  - DAM
  - Web analytics
  - Document output management
  - Portal
  - COLD/ERM

Source: Forrester Research, Inc.
What is Information Governance?

- Managing the information lifecycle of structured and unstructured information to improve business performance and address:
  - Regulatory requirements
  - Organizational risk
  - Privacy/Security
  - Compliance

Why do you care?

- Without effective governance, most technology focused metadata projects will fail *(Forrester Research)*
- The 2009 digital universe was estimated to be 800,000 petabytes and total volume growing to 1.2 million petabytes by 2010
  - Unstructured data and metadata are increasing at an average annual growth rate of 62%
  - Corporations will be responsible for the security, privacy, reliability, and compliance of 85% of that information *(IDC 2010 Digital Universe Study)*
Building Blocks

Smart Content Framework™ Building Blocks

- Metadata
- Insight
- Governance
- Policy
- Privacy
- Enterprise 2.0
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
- Systems Engineering and Technical Assistance (SETA)
- Capability Maturity Model Integration (CMMI)
- Earned Value Management
- Configuration Management
- Technical and Advisory Support
- Independent Verification & Validation (IV&V)

Information Management
- Master Data Management and Data Governance
- Business Intelligence
- Adaptive Data Warehousing
- Enterprise Architecture
- Infrastructure Systems Engineering
- Knowledge Management
- Portal Solutions
- Enterprise Content Management
- IT Optimization/Virtualization

1,200-person multi-disciplinary team of scientific & technical experts
- Scientific subject matter experts
- Systems engineers and architects
- Policy and regulatory specialists
- Project management professionals
- Certified Information technology experts
- Security professionals

Enterprise Solutions
- Program and Project Management
- Earned Value Management
- Performance Measurement
- Program Assurance and Evaluation
- Business Process Improvement
- Security Policy and Compliance
- Communications/Outreach and Facilitation
Agenda

• Information Architecture for SharePoint
  – The What
  – The Why

• Defining Business Objectives

• Intersection of Taxonomy and IA
  – Site Map
  – Taxonomy usage in SharePoint
  – Classification Schemas
The What

- Definition: Information Architecture (IA)

“The art and science of organizing and labeling websites, intranets, online communities and software to support usability and findability.”

- Information Architecture Institute
The Why

• Creating a plan for SharePoint
  – Site Map
  – Library and list definition
    • Document/file management plan
  – Information management plan
    • Managed metadata
      – Term sets
      – Content types
  – User Access
  – Governance
    • Policies
What will SharePoint do for you?

• Defining the business objectives
  – The business requirements for SharePoint 2010
  – End user expectations
  – Organizational goals

• Examples:
  – Replace file shares
  – Collaboration project sites
  – Department intranets
  – Global intranet
Intersection of Taxonomy and IA

- Classification to structure and respond to business requirements
- Depending on system design and use, taxonomy can be front-end or back-end classification
- Taxonomy (categorization) is often actualized by applying metadata to documents
- Enable Findability and Usability
Taxonomy Use for SharePoint

- Search
- Navigation
- Faceted browse
- Personalization or content targeting
How Taxonomy Impacts SharePoint

- 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
Defining a Site Map

Web Application

Site Collection A

Site

Sub-Site

Doc Library

Doc Library

Site

Sub-Site

List

Site Collection B

Site

Sub-Site

Doc Library

List

List
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
Why Term Set Columns?

- Libraries, Lists
- Records Management
- Search refiners/filters
- Faceted browse
- Content types
# Categorization Schemas

<table>
<thead>
<tr>
<th>Method</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facet-based</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>Subject-oriented</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>Functional</td>
<td>Information categorized by the process to which it relates</td>
<td>Employment, staffing, training</td>
</tr>
<tr>
<td>Organizational</td>
<td>Information categorized by corporate departments or business entities</td>
<td>Human Resources, Marketing, Accounting, Research…</td>
</tr>
<tr>
<td>Document Type</td>
<td>Information categorized by the type of document</td>
<td>Presentations, expense reports, press releases, etc.</td>
</tr>
</tbody>
</table>
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
About KAPS Group

• Knowledge Architecture Professional Services
• Virtual Company: Network of consultants – 8-10
• Partners – Concept Searching, SAS, SAP, etc.
• Consulting, Strategy, Knowledge architecture audit
• Services:
  – Taxonomy/Text Analytics development, consulting, customization
  – Technology Consulting – Search, CMS, Portals, etc.
  – Evaluation of Enterprise Search, Text Analytics
  – Metadata standards and implementation
  – Knowledge Management: Collaboration, Expertise, e-learning
  – Applied Theory – Faceted taxonomies, complexity theory, natural categories
Text Analytics and Taxonomies
Bridging the Taxonomy Gap
Agenda

• Introduction – Semantic Context, Taxonomy Gap
• Elements of Text Analytics
  – Categorization, Extraction
• Text Analytics Software
  – Adding Applied Intelligence to SharePoint
• Text Analytics and Taxonomies
  – Integration of the Two and Implications
• Development and Applications
  – SharePoint as Platform
• Semantic Infrastructure – Approach and Benefits
• Conclusions
• Thesauri, Controlled Vocabulary, Glossaries, Product Catalogs
  – Resources to build on
• SharePoint – Managed Metadata Services
  – Term stores – corporate taxonomies
  – Enterprise Keywords (Folksonomy)
• Metadata standards – Dublin Core - Mostly syntactic not semantic
  – Semantic – keywords – very poor performance, no structure
• Facets – classes of metadata
  – Standard - People, Organization, Document type-purpose
  – Requires huge amounts of metadata
• Multiple Types of Taxonomy
  – Browse – classification schemes
  – Formal – Is-Child-Of, Is-Part-Of
  – Small business taxonomies – combination of industry-wide and company-specific concepts
  – Large Subject taxonomies - MeSH

• Structure for Subject Metadata
  – An answer to information overload, search, findability, etc.
  – Consistent nomenclature, common language
  – Application platform – adding meaning

• Structure for facets
  – Organization, functional, subject, document type

• SharePoint Front end – browse, sort, filter
Introduction – Taxonomy Gap
Mind the Gap – Documents and Taxonomy

- How do you bridge the gap – taxonomy to documents?
- Tagging documents with taxonomy nodes is tough
  - And expensive – central or distributed
- Library staff – experts in categorization not subject matter
  - Too limited, narrow bottleneck
  - Often don’t understand business processes and business uses
- Authors – Experts in the subject matter, terrible at categorization
  - Intra and Inter inconsistency, “intertwingleness”
  - Choosing tags from taxonomy – complex task
  - Folksonomy – almost as complex, wildly inconsistent
  - Resistance – not their job, cognitively difficult = non-compliance
- Text Analytics is the answer(s)!
Introduction to Text Analytics

Text Analytics Features

- **Noun Phrase Extraction**
  - Catalogs with variants, rule based dynamic
  - Multiple types, custom classes – entities, concepts, events
  - Feeds facets

- **Additional - Summarization, Fact Extraction, Sentiment Analysis**

- **Auto-categorization**
  - Training sets – Bayesian, Vector space
  - Terms – literal strings, stemming, dictionary of related terms
  - Boolean – Full search syntax – AND, OR, NOT, DIST#, SENTENCE

- **This is the most difficult to develop, fundamental**

- **Combine with Extraction**
  - If any of list of entities and other words
  - Build dynamic rules with categorization capabilities - disambiguation
From Taxonomy to Text Analytics Software

- Software is more important in Text Analytics
  - No Spreadsheets for semantics
  - SharePoint is still missing this essential capability
- Varieties of Software
  - Vocabulary and Taxonomy Management
  - Taxonomy and Text Analytics Platform
  - Embedded - Content Management, Search
  - Specialty – Sentiment Analysis
- Integration with SharePoint
  - Issue of Level of integration and effort
- No standards for Text Analytics
  - Everything is custom job, context sensitive
Evaluating Text Analytics Software – Process

• Start with Self Knowledge
  – Why and What of software, not social media bandwagon
• Eliminate the unfit
  – Filter One- Ask Experts - reputation, research – Gartner, etc.
    • Market strength of vendor, platforms, etc.
    • Feature scorecard – minimum, must have, filter to top 3
  – Filter Two – Technology Filter – match to your overall scope and capabilities – Filter not a focus
  – Filter Three – In-Depth Demo – 3-6 vendors
• Deep POC (2) – advanced, integration, semantics
• Interdisciplinary Team – IT, Business, Library
• Focus on working relationship with vendor, POC for 1?
• SharePoint environment – Concept Searching
Text Analytics and Taxonomy
Complimentary Information Platform

• Taxonomy provides a consistent and common vocabulary
  – Enterprise resource – integrated not centralized
• Text Analytics provides a consistent tagging
  – Human indexing is subject to inter and intra individual variation
• Taxonomy provides the basic structure for categorization
  – And candidates terms
• Text Analytics provides the power to apply the taxonomy
  – And metadata of all kinds
• Text Analytics and Taxonomy Together – Platform
  – Consistent in every dimension
  – Powerful and economic
• Standard Taxonomies = starter categorization rules
  – Example – Mesh – bottom 5 layers are terms
• Categorization taxonomy structure
  – Tradeoff of depth and complexity of rules
  – Easier to maintain taxonomy, but need to refine rules
• Analysis of taxonomy – suitable for categorization
  – Structure – not too flat, not too large
  – Orthogonal categories
• Smaller modular taxonomies
• Different kinds of taxonomies
  – Sentiment – products and features
    • Taxonomy of Sentiment, Emotion - Expertise – process
Text Analytics and Taxonomy: Applications
Content Management – SharePoint

• Mind the Gap – Manual, Automatic, Hybrid
• All require human effort – issue of where and how effective
• Manual - human effort is tagging (difficult, inconsistent)
• Automatic and Hybrid - human effort is prior to tagging
  – Build on expertise – librarians on categorization, SME’s on subject terms
• Hybrid Model
  – Publish Document -> Text Analytics analysis -> suggestions for categorization, entities, metadata - -> present to author
  – Cognitive task is simple -> react to a suggestion instead of select from head or a complex taxonomy
  – Feedback – if author overrides -> suggestion for new category
  – Facets – Requires a lot of Metadata - Entity Extraction feeds facets
• Hybrid – Automatic is really a spectrum – depends on context
Text Analytics and Taxonomy: Applications
Enterprise/SharePoint Search

- Simple Subject Taxonomy structure
  - Easy to develop and maintain
- Combined with categorization capabilities
  - Added power and intelligence
- Combined with people tagging, refining tags
- Combined with Faceted Metadata
  - Dynamic selection of simple categories
  - Allow multiple user perspectives
    - Can’t predict all the ways people think
    - Monkey, Banana, Panda
- Combined with ontologies and semantic data
  - Multiple applications – Text mining to Search
  - Combine search and browse
3. Roles and Responsibilities

• Sample roles matrix:

News Clusters

president: thompson
john mccain: lieberman
republican convention:
blue dog members:
dow jones:
fox news:
murdoch:
news international:
h syndication

Companies

Xcel Energy Inc
General Motors Corp
Fannie Mae
Countrywide Financial Corp
C Span
Freddie Mac
Anheuser-Busch Companies
InBev
Societe Generale
NBC Universal

---

Obama says McCain is trying to run from Republican Party's bad economic record
Associated Press Newswires, 3 September 2008, 9:51 PM GMT, 647 words, (English)

Obama met with Fox news executives
The Anniston Star, 3 September 2008, 816 words, (English)

Why Obama Can't Close The Sale
The Wall Street Journal, 3 September 2008, 896 words, (English)

Obama Met With Fox News Executives
The Washington Post, 3 September 2008, 878 words, (English)

With GOP struggling, Obama content to keep it local and low-key
Associated Press Newswires, 3 September 2008, 12:07 AM GMT, 792 words, (English)

Obama’s Motion To Suppress
Investor’s Business Daily, 3 September 2008, 674 words, (English)

NA POL US Elections Obama Palin Comparison
Associated Press Newswires, 3 September 2008, 10:44 PM GMT, 424 words, (English)
3. Roles and Responsibilities

- Common Roles and SharePoint Permissions:
  - **Site Administrator**
  - **SharePoint Owner**
    - Full Control
  - **SharePoint Member**
    - Contribute
  - **SharePoint Visitor**
    - Read
Taxonomy and Text Analytics

Multiple Search Based Applications

• Platform for Information Applications
  – SharePoint – Multiple Navigation Paths
  – Content Aggregation
  – Duplicate Documents – save millions!
  – Text Mining – BI, CI – sentiment analysis
  – Records Management, Security – PII Info, etc.
  – Social – Hybrid folksonomy / taxonomy / auto-metadata
  – Social – expertise, categorize tweets and blogs, reputation

• Use your Imagination!
Semantic Infrastructure: 4 Dimensions Foundation / Platform for Information

• Ideas – Content and Content Structure
  – Map of Content – Tribal language silos
  – Structure – articulate and integrate
  – SharePoint – Enterprise Metadata Repository (Block #1)

• People – Producers & Consumers
  – Communities, Users, Central Team

• Activities – Business processes and procedures
  – Semantics, information needs and behaviors
  – Information Governance Policy

• Technology
  – CMS, Search, portals, text analytics
  – Applications – BI, CI, Semantic Web, Text Mining
Knowledge Map - Understand what you have, what you are, what you want
  – The foundation of the foundation
Contextual interviews, content analysis, surveys, focus groups, ethnographic studies, Text Mining
Category modeling – Cognitive Science – how people think
Natural level categories mapped to communities, activities
  • Novice prefer higher levels
  • Balance of informative and distinctiveness
Living, breathing, evolving foundation is the goal
Semantic Infrastructure
Some Wrong Approaches

• Large Enterprise Taxonomy
  – Very difficult to develop - $100,000’s
  – Even more difficult to apply
  – Teams of Librarians or Authors/SME’s
  – Problems with maintenance
  – Cost rises in proportion with granularity
  – Difficulty of representing user perspective

• Web 2.0 / Wisdom of the Crowd
  – Social media requires a framework – doesn’t create one
  – Tyranny of the majority, madness of crowds
Semantic Infrastructure Approaches
Infrastructure vs. Projects

• Strategic Foundation vs. Short Term
  – Platform for Governance and Maintenance

• Integrated solution – SharePoint and Search and Applications
  – Enhance search and integrate with browse/sort/filter
  – Cost Effective and good quality keywords / categorization
  – More metadata – faceted navigation

• Semantics
  – Small comparative cost
  – Needed to get full value from SharePoint and other information technologies

• Foundation for Information Governance Model

• ROI – asking the wrong question
  – What is ROI for organizing your company?
Semantic Infrastructure Benefits

Why Semantic Infrastructure

- Unstructured content = 80% or more of all content
- Need to add (infra) structure to make it useful
- Information is about meaning, semantics
- Search is about semantics, not technology
- Can’t Google do it?
  - Link Algorithm – human act of meaning
  - Doesn’t work in enterprise
  - 1,000’s of editors adding meaning
- New technology makes it possible – Text Analytics
Semantic Infrastructure Benefits
General Time and Productivity

• **Time Savings – Too Big to Believe?**
  – Lost time searching - $12M a year per 1,000
  – Cost of recreating lost information - $4.5M per 1,000
  – Cost of not finding the right information – Years?
  – 10% improvement = $1.2M a year per 10,000

• **Making Metrics Human**
  – Number of addition FTE’s at no cost (enhanced productivity)
  – Savings passed on to clients
  – Spreadsheet of extra activities (ex. Training – working smarter)
  – Build a more integrated, smarter organization
Taxonomy and Text Analytics: Conclusions

• Text Analytics can fulfill the promise of taxonomy and metadata

• SharePoint
  – Hybrid model of tagging – Software and Human
  – Search – metadata driven
  – Faceted navigation and Search Based Applications
  – Integrated Enterprise, Enterprise 2.0 (3.0?)

• Future Directions - Advanced Applications
  – Embedded Applications, Semantic Web + Unstructured Content
  – Integration of Enterprise and External - Social Media
  – Expertise Analysis, Behavior Prediction (Predictive Analytics)
  – Voice of the Customer, Big Data
  – Turning unstructured content into data – new worlds
Welcome

KAPS Group
Knowledge Architecture Professional Services

Tom Reamy
Chief Knowledge Architect
KAPS Group
tom@kapsgroup.com

Jill Hannemann
Director of Knowledge Management
Project Performance Corporation
jhannemann@ppc.com

Donald Miller
Vice President Sales – Commercial Accounts
Concept Searching, Inc.
donm@conceptsearching.com
Webinar 2: Developing Enterprise Facets for SharePoint’s Managed Metadata Term Store

Date: Tuesday, March 27, 2012

Time: 11:30 AM - 12:30 PM EDT

Don't start from scratch - get advice on where to start with facets for SharePoint Term Store. This session will give strategy for developing your own facets, as well as common facets that many organizations use. We will examine the core principles behind designing a business taxonomy and metadata schema that will provide the foundation for SharePoint information architecture.

Guest Speaker: Raimund M. Wasner, Managing Director at Kollabria

To Register: http://www.conceptsearching.com/Web/home/events/webinars.aspx