The Value of SKOS for Business Vocabulary Management

Semantic Technology 2012
Agenda

- Named entity managed vocabularies
- Business value of managed vocabularies
- Why SKOS
- Semantic metadata
People

Arroyo, Gloria
Atkinson, Rowan
Baldwin, Alec
Bale, Christian
Banderas, Antonio
Baron Cohen, Sacha
Burrymore, Drew
Beck, Glenn
Beckham, David
Bernanke, Ben
Berry, Halle
King Abdullah
Biden, Joe
Bin Laden, Osama
Blair, Tony
Blanchett, Cate
Bono
Branson, Richard
Brown, Dan
Brownback, Sam
Bryant, Kobe
Buffett, Warren
Burton, Tim

Gingrich, Newt
Giuliani, Rudolph
Goldberg, Whoopi
Gore, Al
Grace, Nancy
Grant, Hugh
Griffin, Kathy
Grisham, John
Hagel, Chuck
Hanh, Thich Nhat
Hanks, Tom
Hannity, Sean
Harper, Stephen
Hastert, Dennis
Havel, Vaclav
Hawking, Stephen
Hayek, Salma
Hilton, Paris
Hoffman, Dustin
Holmes, Katie
Houston, Whitney
Howard, John
Hudson, Kate

Powell, Colin
Prince
Prince Charles
Prince William
Putin, Vladimir
Qaddafi, Muammar
Queen Elizabeth II
Queen Latifah
Rather, Dan
Redford, Robert
Reid, Tara
Reynolds, Burt
Rice, Condoleezza
Richardson, Bill
Rivera, Geraldo
Robbins, Tim
Roberts, Julia
Robertson, Pat
Rock, Chris
Rove, Karl
Rowling, J.K.
Sandler, Adam
Santorum, Rick

* courtesy of mondostars.com
What are managed vocabularies

- Names of people, organizations, products, events, locations, etc.
  - Alternate labels
    - Synonyms
    - Abbreviations
    - Acronyms
    - etc.
  - Additional information
    - Unique identifiers
    - Coverage dates
    - Descriptions
    - etc.

- A set of concepts, optionally including statements about semantic relationships between those concepts.
Business value of managed vocabularies: Product catalogs

- Increase conversion and lift
- Speed to market
- Editorial efficiencies
- Product data consistency
- Merchandising efficiencies
- Faceted navigation/product filtering
- Ingest new product categories
- Ingest whole new product catalogs

- Related products recommendations
- Gather and publish customer reviews
- Monitor and respond to customer sentiment
- Alert customers about new products
- Multi-platform publishing
- Data governance
- Linked open data
- SEO
Conversion rate for product findability
- $80M web sales net income
- 10% conversion rate increase

Lift in order size from satisfaction
- $80M web sales net income
- 20% lift in sales
Product catalog case study

http://www.bestbuy.com/

- Reflect new and changing consumer electronic technologies and products.
- Flexible to SKU growth in general merchandise categories.
- Conformance with taxonomy design best practices for better user experience.

- Online strategy recognized as a leader among consumer electronics websites.
Business value of managed vocabularies: Web publishing

- Monetize content
- Aggregate content
- Reuse existing content
- Editorial efficiencies
- Create new information products
- Ingest new content feeds
- Improve search results
- Faceted navigation/content filtering of search results

- Current awareness services
- Content targeting
- Recommend related content based
- Gather and publish user generated content
- Federated search
- Multi-platform publishing
- Content governance
- Linked open data
- SEO
SAMHSA Taxonomy promotes the agency behavioral health program.

The taxonomy powers the SAMHSA Store as illustrated in a YouTube video, and also provides:

- The framework for agency key performance indicators.
- Increases the availability and visibility of SAMHSA information.
- Offers tools for analysis, visualization and mash ups with other sources.
Oracle taxonomy is based on brand name products marketed by product line, technology, application, and industry solution:

The taxonomy powers the Oracle events website, provisioning:
- Filters based on search results.
- Mashup with Google maps.
- RSS feed to subscribe to criteria of this result set.
- Register, view details, or email info.
The taxonomy also powers the Oracle newsroom website, provisioning:

- Filters based on search results.
- RSS feed to subscribe to criteria of this result set.
- Social networking services including:
  - Email item
  - Bookmark and share
Oracle top-level taxonomy architecture

- Person
- Organization
- Location
- Content Type
- Audience
- Products
  - Product Line
  - Technology
  - Application
  - Industry Solution

"Is a" groups of Products
Q: How do you share a taxonomy across (and outside of) the enterprise?
A: With standards

- **ISO 2788:1986** Guidelines for the Establishment and Development of Monolingual Thesauri
- **ISO 5964:1985** Guidelines for the Establishment and Development of Multilingual Thesauri
- **ISO 25964** (combines 2788 and 5964) Thesauri and Interoperability with other Vocabularies
- **Zthes** specifications for thesaurus representation, access and navigation
- **W3C SKOS** Simple Knowledge Organization System
Why SKOS? According to Alistair Miles …

- **Ease of combination** with other standards
  - Vocabularies are used in great variety of contexts.
    - E.g., databases, faceted navigation, website browsing, linked open data, spellcheckers, etc.
  - Vocabularies are re-used in combination with other vocabularies.
    - E.g., [ISO3166 country codes](http://www.iso.org/iso/en/prods-services/iso3166ma/data-country-codes.htm) + [USAID regions](http://www.usaid.gov); [USPS zip codes](http://www.usps.com) + [US Congressional districts](http://www.house.gov); [USPS states](http://www.usps.com) + [EPA regions](http://www.epa.gov), etc.

- **Flexibility and extensibility** to cope with variations in structure and style
  - Variations between types of vocabularies
    - E.g., list vs. classification scheme
  - Variations within types of vocabularies
Why SKOS? (2)

- **Publish managed vocabularies** so they can readily be consumed by applications
  - Identify the concepts
    - What are the named entities?
  - Describe the relationships
    - Labels, definitions and other properties
  - Publish the data
    - Convert data structure to standard format
    - Put files on an http server (or load statements into an RDF server)

- **Ease of integration** with external applications
  - Use web services to use or link to a published concept, or to one or more entire vocabularies.
    - E.g., Google maps API, NY Times article search API, Linked open data

- **A W3C standard** like HTML, CSS, XML… and RDF, RDFS, and OWL
# Semantic relationships

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept</td>
<td>A unit of thought, an idea, meaning, or category of objects or events. A Concept is independent of the terms used to label it.</td>
</tr>
<tr>
<td>Preferred Label</td>
<td>A preferred lexical label for the resource such as a term used in a digital asset management system.</td>
</tr>
<tr>
<td>Alternate Label</td>
<td>An alternative label for the resource such as a synonym or quasi-synonym.</td>
</tr>
<tr>
<td>Broader Concept</td>
<td>Hierarchical link between two Concepts where one Concept is more general than the other.</td>
</tr>
<tr>
<td>Narrower Concept</td>
<td>Hierarchical link between two Concepts where one Concept is more specific than the other.</td>
</tr>
<tr>
<td>Related Concept</td>
<td>Link between two Concepts where the two are inherently &quot;related&quot;, but that one is not in any way more general than the other.</td>
</tr>
</tbody>
</table>
Some semantic relationships for IBM

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>lc:n79142877</td>
<td>skos:prefLabel</td>
<td>International Business Machines Corporation</td>
</tr>
<tr>
<td>lc:n79142877</td>
<td>skos:altLabel</td>
<td>IBM</td>
</tr>
<tr>
<td>lc:n79142877</td>
<td>skos:altLabel</td>
<td>I.B.M.</td>
</tr>
</tbody>
</table>
My company sells IBM’s XIV product

International Business Machines

A high-end disk storage server designed to provide high performance, scalability, and availability in disk storage.

Originally developed by Israeli company XIV, which was acquired by IBM in 2007.

XIV

Storage System

myco:p9324

myco:supplier

lc:79142877

skos:prefLabel

skos:definition

skos:altLabel

skos:prefLabel

skos:historyNote
Negotiations With Iran Over Nuclear Program May Resume

By STEVEN LEE MYERS and RICK GLADSTONE
Published: February 17, 2012

WASHINGTON — The United States and the European Union signaled on Friday that negotiations with Iran over its nuclear program could soon resume for the first time in almost a year, even as a telecommunications network vital to the global banking industry prepared to expel Iranian banks.

While senior American and European officials stopped short of declaring a diplomatic breakthrough, Iran dropped previously unacceptable preconditions for talks in a letter this week from its senior nuclear negotiator, Saeed Jalili, who declared his country’s “readiness for dialogue” at “the earliest possibility.”

After weeks of official bluster, ominous threats of military intervention and growing evidence of a covert arms shipment to Syria, the United States and its European allies appear ready to end a freeze that has stymied efforts to get Iranian leaders to agree to serious talks about their nuclear program.

Iran has repeatedly rejected the offer of introductory talks with the United States, and its foreign minister, Mohammad Javad Zarif, has repeated this stance in recent days.

But in a letter that Secretary of State John Kerry got from Mr. Jalili, Iran’s deputy foreign minister, the president’s chief nuclear negotiator, the letter provided what Mr. Kerry called “a major step forward,” and it could provide a basis for new direct talks.

The new overtures from Iran could also mitigate some of the pressure that the U.S. and European banks are facing as part of sanctions that have effectively shut off their access to the international economy.

The banks, which play a central role in international finance, are already in an already stressed system in which China and other nations are working to protect their own financial systems as part of a broader effort to resist the sanctions.

In a video message to global bankers, Zhou Xiaochuan, governor of China’s central bank, has described the sanctions as a “blunder” and called on foreign banks to consider how they might help China evade them.

The letter from Mr. Jalili, which Mr. Kerry got by phone on Friday, is being received as a significant indication of Iran’s willingness to break the impasse.

The United States and the European Union issued a joint statement late on Friday offering a new framework for Iranian nuclear talks, saying they would meet with Iran in Washington on March 7, and in Brussels, Belgium, on March 30.

The European Union would be represented by Mr. Zarif, the senior Iranian negotiator.

In the letter, Iran’s foreign minister reiterated the country’s commitment to the 2005 treaty, known as the Non-Proliferation Treaty, but it was not clear whether Iran would agree to serious talks.

By the end of the week, Iran’s foreign minister, Mohammad Javad Zarif, had already started to tone down his previous objections to negotiations, and he said earlier this week that he was prepared to meet in Washington on March 7, in Brussels on March 30, and perhaps in a third location later on.

Mr. Zarif has formally rejected the offer of new talks, but he said in his letter that Iran was ready to meet under certain conditions.

Iranian officials say they remain convinced that the sanctions in place are essentially worthless and that the conditions set for resuming talks are too onerous to overcome.

The letter sent by Mr. Jalili and its implications will be closely watched by the Obama administration, which is eager to avoid a military confrontation with Iran, and by European governments, which are eager to avoid a confrontation but also face criticism from their allies over their reluctance to intensify the sanctions.

The new letter, however, is being received as a significant indication of Iran’s willingness to break the impasse.

The United States and the European Union issued a joint statement late on Friday offering a new framework for Iranian nuclear talks, saying they would meet with Iran in Washington on March 7, and in Brussels, Belgium, on March 30.

The European Union would be represented by Mr. Zarif, the senior Iranian negotiator.

By the end of the week, Iran’s foreign minister, Mohammad Javad Zarif, had already started to tone down his previous objections to negotiations, and he said earlier this week that he was prepared to meet in Washington on March 7, in Brussels on March 30, and perhaps in a third location later on.

Mr. Zarif has formally rejected the offer of new talks, but he said in his letter that Iran was ready to meet under certain conditions.

Iranian officials say they remain convinced that the sanctions in place are essentially worthless and that the conditions set for resuming talks are too onerous to overcome.

The letter sent by Mr. Jalili and its implications will be closely watched by the Obama administration, which is eager to avoid a military confrontation with Iran, and by European governments, which are eager to avoid a confrontation but also face criticism from their allies over their reluctance to intensify the sanctions.

The new letter, however, is being received as a significant indication of Iran’s willingness to break the impasse.
What is the semantic web

- Making content web-accessible in a format that can be read and used by automated tools, so that people and machines can find, share and integrate information more easily.

- Some current examples, especially if they use semantics as the basis for that integration.
  - **Dynamic Web Pages** – Content changes in response to different contexts or conditions.
  - **Personalization** – Tailoring to a user based on personal details or characteristics they provide.
  - **Mashups** – Combining data from more than one source into an integrated application.
  - **Linked Open Data** – Using managed vocabularies to identify links among named entities in near or distant applications.
SKOS and semantic web

- SKOS is an OWL ontology and uses the RDF data model
- This means that while SKOS tools are out there, you can use non-SKOS semantic web tools as well:
  - SPARQL query language
  - Data interchange tools
  - Triplestores
  - OWL and RDFS inferencing
Information collaboration: 
Managed vocabularies and business rules

- Vocabularies can do more than sell vacations, cars, and cruises.
- Vocabularies can help us decipher complex issues
  - Help citizens select health insurance policies.
  - Help parents find advice on handling with under-age drinking.
  - Help high school juniors find colleges with particular programs.
  - Help pharmacists find generic drugs to substitute for brand names.
  - Help nurses identify side-effects of medication or medical devices.
  - Help telephone sales reps correctly describe packaged products.
  - Help procurement professionals purchase computer equipment.
  - Help managers share better management practices.
Questions?

Joseph A Busch, Principal  Bob Ducharme, Solutions Architect
jbusch@taxonomystrategies.com  bducharme@topquadrant.com
twitter.com/joebusch  
415-377-7912
Abstract

Unlike earlier knowledge representation schemes, SKOS has really taken off because companies are getting committed to managing the names of entities (people, organizations, places, events and things). The names of products, especially well-branded names, have been managed by corporate communications for years. It’s relatively easy to “recognize” product names in product collateral like brochures or web pages and to mark those entities with URI’s and use those URI’s to drive customers to product information. Google maps API did the same for base geographic locations. The New York Times facets are making people, organizations, events and other named entities tagged and usable in similar ways.

This session will discuss how businesses and government agencies are doing clever things with managed vocabularies on the web, and describe the business value that explains why they are doing it—how they are making money, saving money and staying out of jail.