INFO/CS 4302
Web Information Systems

FT 2012
Week 9 : Linked Data Technologies (SKOS)

- Bernhard Haslhofer -
Plan for today...

• Recap

• Discussion Homework 7

• SKOS

• Questions, Housekeeping, ...
URI

- Name and identify things (resources)
- Dereferencable HTTP URIs

http://dbpedia.org/resource/The_Shining_(film)
http://data.linkedmdb.org/resource/film/2014
http://rdf.freebase.com/ns/m.04fjzv
RDF

- A data model for representing data on the Web
- Several statements (triples) form a graph

```plaintext
http://dbpedia.org/resource/The_Shining_(film)
  rdfs:label 闪灵 (电影)
  rdfs:label The Shining (film)
  rdf:type http://xmlns.com/foaf/0.1/Person
  dbpedia-owl:birthDate 1937-04-22
  foaf:name Jack Nicholson

http://dbpedia.org/resource/The_Shining_(film)
  rdf:type http://dbpedia.org/ontology/Film
  dbpprop:starring http://xmlns.com/foaf/0.1/Person
  http://dbpedia.org/ontology/Film
```
RDF/XML, N3, Turtle, etc.

- **Data formats** for RDF resource representations
- Used to transfer RDF data between apps
RDFS

• A language for describing the syntax and semantics of vocabularies in a machine-understandable way

http://dbpedia.org/ontology/Film

http://dbpedia.org/ontology/Work

rdfs:subClassOf
OWL

- A more expressive (formal) language for defining the syntax and semantics of vocabularies
- Solves RDFS shortcomings but introduces quite some complexity
SKOS

• A language for describing controlled vocabularies (taxonomies, thesauri, classification schemes)
SPARQL

• A query language and protocol for accessing RDF data on the Web

```
SELECT DISTINCT ?x
WHERE {
}
```
RDF APIs

- Java
  - Jena Semantic Web Framework (http://jena.apache.org/)
  - Sesame RDF API (http://www.openrdf.org)
- PHP
  - ARC (http://arc.semsol.org/)
- Ruby
  - RDF.rb: Linked Data for Ruby (http://rdf.rubyforge.org/)
- Python
  - RDFLib (https://github.com/RDFLib)
- C
  - Redland RDF Libraries (http://librdf.org/)
Linked Data debugging

using cURL:

curl -iH "Accept: application/rdf+xml"
http://dbpedia.org/resource/The_Shining_(film)

curl -LH "Accept: application/rdf+xml"
http://dbpedia.org/resource/The_Shining_(film)

curl -iH "Accept: text/n3"
http://dbpedia.org/resource/The_Shining_(film)
Linked Data debugging

using raptor (http://librdf.org/raptor/):

rapper -o rdfxml
http://dbpedia.org/resource/The_Shining_(film)

rapper http://dbpedia.org/resource/The_Shining_(film)
> ~/Desktop/the_shining.nt
DISCUSSION HW7
HW7: Basic RDF Construction and Programming (Due: 11/01 11:59pm)

The purpose of this homework is to give you experience with using the Resource Description Framework (RDF) and existing RDF libraries (APIs). You will also extend the RESTful Movie Service from the previous assignment and bring it one step further to a Linked Data service.

You can download and use the provided code skeleton for this homework.

Task 1: RDF APIs

Model the movie data set from the previous assignments as RDF statements. In as many cases as possible use elements (classes and properties) from existing vocabularies. Hint: although most facts can be expressed in these vocabularies, there are a few that can't, in which case you should use a single new namespace of your choosing.

Write a program (e.g., convert.py) that takes the three movie data CSV files as input, converts them into an in-memory RDF graph, and serialize that RDF graph into at least two different RDF formats (e.g., e.g., RDF/XML, Turtle) on completion. Use an existing RDF library (e.g., RDFlib) for handling RDF models in your program.

Make sure you provide instructions on how to run your program README.TXT.

Task 2: Enhance movie service to support RDF

Extend your Movie Service from the previous assignment so that it also supports at least two RDF serialization formats (e.g., RDF/XML, Turtle) for use cases UC3 (Retrieve a specific actor) and UC4 (Retrieve a specific movie). You don't need to implement resource creation / update functions based on RDF.

You can either use an RDF API or template mechanisms such as those provided by Tornado.

Test your implementation using cURL and/or raptor and provide example commands in README.TXT.
SKOS
Simple Knowledge Organization System (SKOS)

- A language for describing knowledge organization systems (taxonomies, thesauri, classification schemes)
data.nytimes.com

For the last 150 years, The New York Times has maintained one of the most authoritative news vocabularies ever developed. In 2009, we began to publish this vocabulary as linked open data.

The Data

As of 13 January 2010, The New York Times has published approximately 10,000 subject headings as linked open data under a CC BY license. We provide both RDF documents and a human-friendly HTML versions. The table below gives a breakdown of the various tag types and mapping strategies on data.nytimes.com.

<table>
<thead>
<tr>
<th>Type</th>
<th>Manually Mapped Tags</th>
<th>Automatically Mapped Tags</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>People</td>
<td>4,978</td>
<td>0</td>
<td>4,978</td>
</tr>
<tr>
<td>Organizations</td>
<td>1,489</td>
<td>1,592</td>
<td>3,081</td>
</tr>
<tr>
<td>Locations</td>
<td>1,910</td>
<td>0</td>
<td>1,910</td>
</tr>
<tr>
<td>Descriptors</td>
<td>498</td>
<td>0</td>
<td>498</td>
</tr>
</tbody>
</table>

|       |                     |                           | 10,467  |

**Browse individual data records:**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

**SKOS Files**

Download all of the data records as SKOS Files.

- People
Authorities and Vocabularies

About

The Library of Congress Authorities and Vocabularies service enables both humans and machines to programatically access authority data at the Library of Congress via URIs.

Read more >
The AGROVOC thesaurus contains more than 40,000 concepts in up to 22 languages covering topics related to food, nutrition, agriculture, fisheries, forestry, environment and other related domains.

A global community of editors consisting of librarians, terminologists, information managers and software developers, maintain AGROVOC using VocBench, an open-source multilingual, web-based vocabulary editor and workflow management tool that allows simultaneous, distributed editing. AGROVOC is expressed in Simple Knowledge Organization System (SKOS) and published as Linked Data.

- Explore search terms or browse hierarchies
- Download or access via web services
- Suggest new terms not present yet

LATEST NEWS

"Design and Implementation of the ZH/EN Bilingual Retrieval System Based on the CAT/AGROVOC Mapping"
13 Sep 2012

AIMS at Agricultural Ontology Service (AOS) 2012
05 Sep 2012

Dutch FSA to complete project "Apollon" for detection of Emerging Risks in Food Chains
06 Aug 2012

AGROVOC 2012 edition released
17 Jul 2012

International Conference on Information and Knowledge Dynamics, Bangalore
15 Jul 2012
Knowledge Organization Systems (KOS)
Folksonomy

• **Not controlled** by some central authority
• Usually no explicit definition of meanings
• Used in
  – collaborative tagging systems
  – social media
Folksonomy Examples

All time most popular tags

animals architecture art asia australia autumn baby band barcelona beach berlin bike bird birthday black blackandwhite blue bw california canada canon car cat chicago china christmas church city clouds club color concert dance day de dog england europe fall family fashion festival film florida flower flowers food football france friends fun garden geotagged germany girl graffiti green halloween hawaii holiday house india iphone island itaie italy japan kids la lake landscape light live london love macro me mexico model mountain museum music nature new newyork newyorkcity night nikon nyc ocean old paris park party people photo photography photos portrait raw red river rock san sanfrancisco scotland sea seattle show sky snow spain spring square street summer sun sunset taiwan texas thailand tokyo toronto travel tree trees trip uk urban usa vacation vintage washington water wedding white winter women yellow zoo

What are tags?
You can give your photos and videos a “tag”, which is like a keyword or category label. Tags help you find photos and videos which have something in common. You can assign up to 75 tags to each photo or video.
Controlled Vocabulary

• A simple list of terms, definitions and naming conventions
• Some process “controls” adding and removing term definitions to ensure consistency
• Terms are often defined in relationship to each other
• Glossaries, Taxonomies, and Thesauri are example types of controlled vocabularies
Taxonomy

- Classification of entities in a hierarchy
- A taxonomy node stands for a real-world entity type
Introduction

In February 1999, the statistical agencies of Canada, Mexico, and the United States launched a joint multi-phase initiative to develop a comprehensive demand-oriented product classification, known as the North American Product Classification System (NAPCS). Work to date has focused on the products produced by service industries in 12 NAICS sectors 48-49 through 81. With that work nearing completion, this web page provides an overview of and progress report on the NAPCS initiative and presents the final versions of the product lists developed so far for the service industries included in those 12 sectors.

Overview and Progress Report

Final NAPCS Product Lists

- Description of product lists
- View PDF files of individual product lists
- Download sequential Excel file of all product lists
Thesaurus

• Associates the meaning of a term with the meaning of another term
• hypernym, hyponym, synonym, antonym
Noun

- **S:** (n) *cat*, **true cat** (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
- **S:** (n) *guy*, *cat*, *hombre*, *bozo* (an informal term for a youth or man) "a nice guy"; "the guy's only doing it for some doll"
- **S:** (n) *cat* (a spiteful woman gossip) "what a cat she is!"
- **S:** (n) *kak, khat, qat, quax*, *cat*, *Arabian tea*, *African tea* (the leaves of the shrub Catha edulis which are chewed like tobacco or used to make tea; has the effect of a euphoric stimulant) "in Yemen kat is used daily by 85% of adults"
- **S:** (n) *cat-o'-nine-tails*, *cat* (a whip with nine knotted cords) "British sailors feared the cat"
- **S:** (n) *Caterpillar, cat* (a large tracked vehicle that is propelled by two endless metal belts; frequently used for moving earth in construction and farm work)
- **S:** (n) *big cat*, *cat* (any of several large cats typically able to roar and living in the wild)
- **S:** (n) *computerized tomography*, *computed tomography*, *CT*, *computerized axial tomography*, *computed axial tomography*, **CAT** (a method of examining body organs by scanning them with X rays and using a computer to construct a series of cross-sectional scans along a single axis)

Verb

- **S:** (v) *cat* (beat with a cat-o'-nine-tails)
- **S:** (v) *vomit*, *vomit up*, *purge*, *cast*, *sick*, *cat*, *be sick*, *disgorge*, *regorge*, *retch*, *puke*, *barf*, *spew*, *spue*, *chuck*, *upchuck*, *honk*, *regurgitate*, *throw up* (eject the contents of the stomach through the mouth) "After drinking too much, the students vomited"; "He purged continuously"; "The patient regurgitated the food we gave him last night"
# National Library of Medicine - Medical Subject Headings

## 2010 MeSH

### MeSH Descriptor Data

Return to Entry Page

Standard View. Go to Concept View; Go to Expanded Concept View

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Heart</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Number</td>
<td>A07.541</td>
</tr>
</tbody>
</table>

### Annotation

heart as pump, not tissue (MYOCARDIUM); qualif permitted: /anat /drug eff /embryol (FETAL HEART also exists) /growth /innerv (HEART CONDUCTION SYSTEM also exists) /microbiol /parasitol /physiol (see also ATRIAL FUNCTION & VENTRICULAR FUNCTION & their specifics; see also MYOCARDIAL CONTRACTION & its specifics) /physiopathol /rad eff /radiogr (do not confuse with ANGIOCARDIOGRAPHY) /radionuclide /viro; /blood supply = CORONARY VESSELS but consider also CORONARY CIRCULATION; /drug eff: consider also MYOCARDIAL DEPRESSANTS see ANTI-ARRHYTHMIA AGENTS & CARDIOTONIC AGENTS; /transpl = HEART TRANSPLANTATION: do not coord with TRANSPLANTATION, HOMOLOGOUS unless particularly discussed; /ultrasonogr = ECHOCARDIOGRAPHY or ECHOCARDIOGRAPHY, DOPPLER; inotropism: INOTROPISM, CARDIAC see MYOCARDIAL CONTRACTION & INOTROPIC AGENTS, POSITIVE CARDIAC are available but see MYOCARDIAL CONTRACTION note; chronotropism: see HEART RATE note; DIASTOLE & SYSTOLE are also available; heart-lung prep: index HEART /physiol (probably NIM); Manual 21.48+; carditis = MYOCARDITIS; mural thrombosis: index HEART DISEASE (IM) + THROMBOSIS (IM)

### Scope Note

The hollow, muscular organ that maintains the circulation of the blood.
Why do we need “knowledge organization” in the age of full-text search engines?
What is SKOS?

• A model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, taxonomies, folksonomies, and other similar types of controlled vocabularies

• Allows concepts to be composed and published as Linked Data on the Web

• Hides the complexity of OWL - easy to use
SKOS Concepts are...

- ... identified by URIs
- ... labeled with 1..* natural language strings
- ... documented with various types of notes
- ... semantically linked to each other
- ... aggregated into concept schemes
Parody films

From Library of Congress Subject Headings

**Parody films**

This heading is used as a topical heading for works about films that comically imitate another work or group of works of a more serious nature. When used as a topical heading it is subdivided by the appropriate geographic, topical, and/or form subdivisions.

General works about the use of parody in motion pictures are entered under Parody in motion pictures.

**URI**

<http://id.loc.gov/authorities/sh2001000475#concept>

**Type**

Topical Term

**Alternate Labels**

- Film genre parodies
- Film parodies
- Genre parodies (Motion pictures)
- Genre parody films
- Motion picture parodies
Concepts are

- the **units of thought**: ideas, meanings, categories of objects, etc.
- **abstract entities** which are independent of the terms used to label them

```xml
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix lcsh: <http://id.loc.gov/authorities/> .
```
**skos:(pref|alt|hidden)Label**

- Labels refer to concepts’ natural language(s)
  - `skos:prefLabel`: the preferred lexical label
  - `skos:altLabel`: alternative lexical labels (e.g., synonyms)
  - `skos:hiddenLabel`: labels useful for indexing (e.g., typos)

![Diagram showing relationships between labels](image.png)
SKOS Semantic Relationships

• The meaning of a concept is also defined by its links to other concepts
  – skos:broader: hierarchical link to a more general concept
  – skos:narrower: hierarchical link to a more specific concept
  – skos:related: associative (non-hierarchical) link
SKOS Documentary Notes

- Add further **human-readable** documentation
  - `skos:scopeNote`: info about intended meaning
  - `skos:definition`: complete explanation of meaning
  - `skos:example`: example concept use

```
  "Comedy films"@en
```

```
“This heading is used as a genre/form heading for films that comically imitate another work or group of works of a more serious nature.”@en
```

```
lcsh:sh2007025038#concept
```

```
skos:scopeNote
```

```
skos:prefLabel
```

```
"Comedy films"@en
```

```
This heading is used as a genre/form heading for films that comically imitate another work or group of works of a more serious nature.”@en
```
skos:ConceptScheme

• Allow the organization of skos:Concepts in some Knowledge Organization Scheme (KOS)
Another Real-World Examples

• NY Times topics: http://topics.nytimes.com/

• Dereference and analyze “Jack Nicholson” @ New York Times
  – http://data.nytimes.com
    N5761411277431266513
QUESTIONS & HOUSEKEEPING