Knowledge Organization and Semantic Search

Thomas Baker
Dublin Core Metadata Initiative

Joint NKOS/CENDI Workshop on Semantic Search
Department of Transportation Media Center
Washington DC, 6 December2012

Concept Search

- Search for: *Ford films*
- Get back: documents also mentioning just *movies*
- Search engine knows how *film* relates to *movie*:
  - By inferring from statistical co-occurrence or…
  - **By using a taxonomy**

Seth Grimes, Information Week, January 21, 2010, http://is.gd/UxreBr

Ontology-Based Search

- Search for: *What does a dog chase?*
- Get back: results about *cars*, *cats*, *tails*… as they relate to *dogs*
- Search engine knows how *dogs* relate to *cars, cats*, and *tails*…
  - **By using an ontology**

Seth Grimes, Information Week, January 21, 2010, http://is.gd/UxreBr

Semantic Web Search

- The Semantic Web seeks to capture data relationships and make the resulting Web of Data queryable.
- "This lofty and worthy goal is years from practical usability…"

Seth Grimes, Information Week, January 21, 2010, http://is.gd/UxreBr

This Talk

- **Concept-Based** *vs* **Ontology-Based** Search
- **SKOS Concept Schemes** *vs* **OWL Ontologies**

2001: a Semantic Web odyssey

Tim Berners-Lee, Scientific American (2001)

- Your Semantic-Web-enabled browser finds a good doctor
- …nearby, with good ratings,
- …makes an appointment by itself,
- …using “ontologies” …in the future…

2004: OWL

- Web Ontology Language (OWL), starting 2001
  - For describing a “cartoon universe”
  - An “explicit specification of a conceptualization”…
- Ontologies
  - Use URIs to denote classes and properties of *entities* (which refer to real-world objects)
  - Can specify complex logical relationships between classes of entities
  - Are used to describe “individuals” – instances of classes
  - Support logical inference (“entailments”)…

Example of OWL entailment

- If we specify in our ontology that:
  - *Dog* is a sub-class of the class *Mammal*
  - The class *Male* is disjoint with the class *Female*
- …we might then state in our data that:
  - "Oscar" is an instance of the class *Dog*
  - "Oscar" is an instance of the class *Male*
- …from which we can infer that:
  - "Oscar" is an instance of the class *Mammal*
  - "Oscar" cannot be an instance of the class *Female*

2006: Linked Data

- Tim Berners-Lee: “Publish your data on Web using URIs and Semantic Web technology”
- Less about fancy ontologies than “raw data”

Photo by Joshua Wanyama, http://www.flickr.com/photos/wa-j/3256903146/

2009: SKOS

- Simple Knowledge Organization System
  - Goal: make KOSs usable in Linked Data space
  - Like OWL, denote conceptual entities with URIs
  - Started 1997, W3C draft 2005, finalized 2009

Photo by CACC North Library, http://www.flickr.com/photos/ccacnorthlib/3554628032/

Circa 1750: KOS

- KOS = Knowledge Organization Systems (*see NKOS…*)
- Knowledge in a form useful for resource discovery
  - Sets of terms with definitions (**glossaries**)
  - …in hierarchies (basic **classifications** and **taxonomies**)
  - …with more complex semantic relations (**thesauri**, **subject heading lists**)
- Typically
  - based either on *terms* (words) or *concepts* (underlying abstractions)
  - lack formal axioms for complex inference

SKOS Design

- Translate KOSs as **SKOS Concept Schemes**
  - Concepts, “labeled” with (natural-language) terms
  - Specify semantic relations between concepts
- Design goals
  - Capture structures common to many types of KOS
  - Easily convert KOSs into Linked-Data-ready form
  - Principle of “Minimal Semantic Commitment”
  - Avoid encouraging false semantic precision

Deliberately weak semantics

- *Avoids* supporting many kinds of inference
  - *Car* is broader than *Wheel*
  - *Machine* is broader than *Car*
  - *Machine* is not "broader than" *Wheel*
    - …though one could choose to interpret "broader than" relations as being "transitive"
- Turning KOS into OWL Ontology is hard work
  - When does *broader* mean *class subsumption*, *class instantiation*, or *part-whole relations*?

SKOS Concepts vs OWL Classes

- OWL Class *Butterfly*
  - A collection, or set, of individual butterflies ("extensional" semantics)
  - Represents things in the world (cartoon reality)
- SKOS Concept *Butterfly*
  - An individual "idea or notion" of butterflies
  - Represents a thing in the mind
  - Relationship to reality is, by default, unspecified

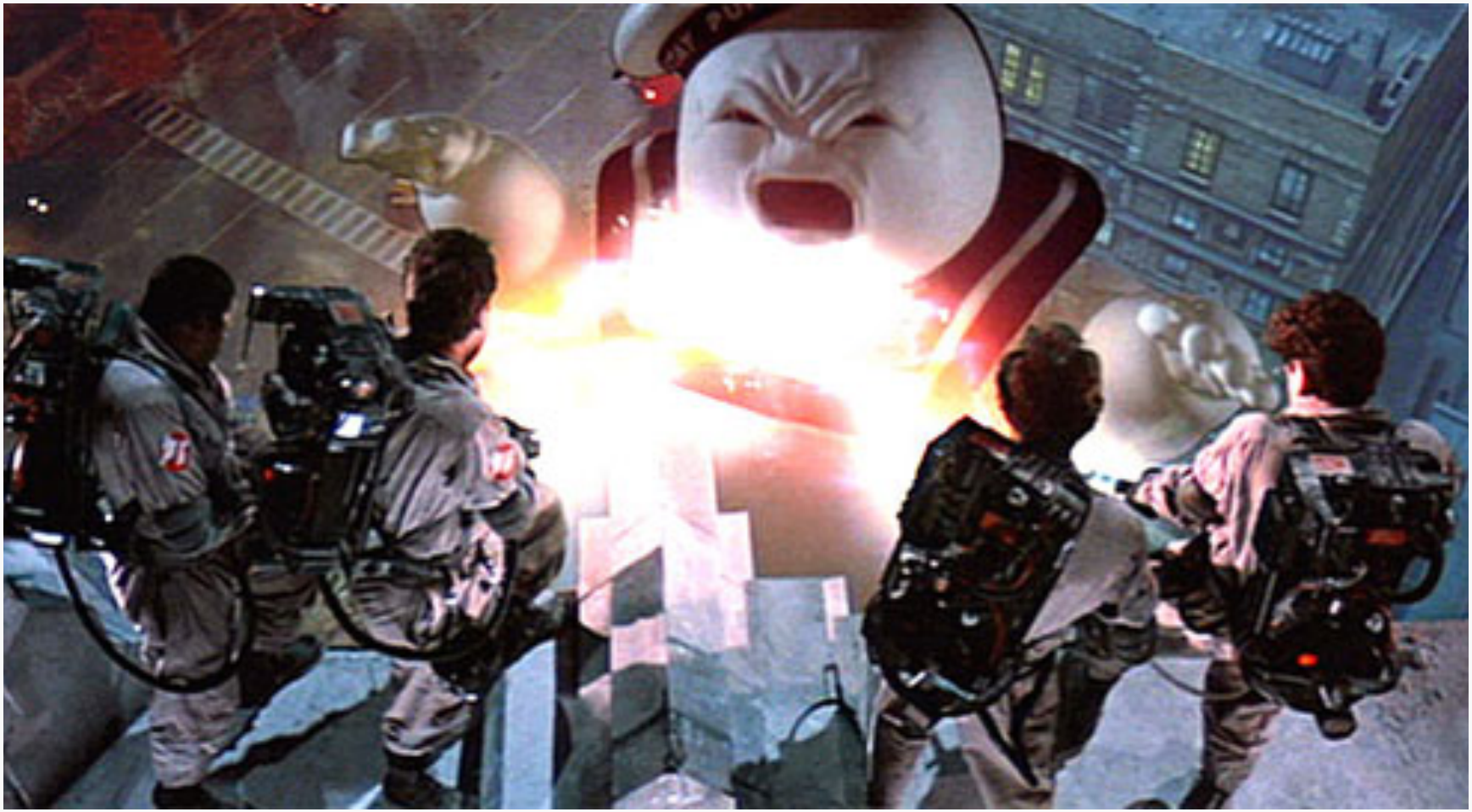
What's *not* a SKOS Concept?

- Almost anything can *also* be a SKOS Concept
  - Fine print: Except for SKOS Concept Schemes, Collections, and Labels.
- A SKOS Concept explicitly related to an OWL Class:
  - http://viaf.org/viaf/85312226/#skos:Concept
    - Tim Berners-Lee, the SKOS concept in the OCLC Virtual International Authority File (VIAF)
  - …has a foaf:focus relationship to:
  - http://viaf.org/viaf/85312226/
    - Tim Berners-Lee, an instance of the OWL class Person

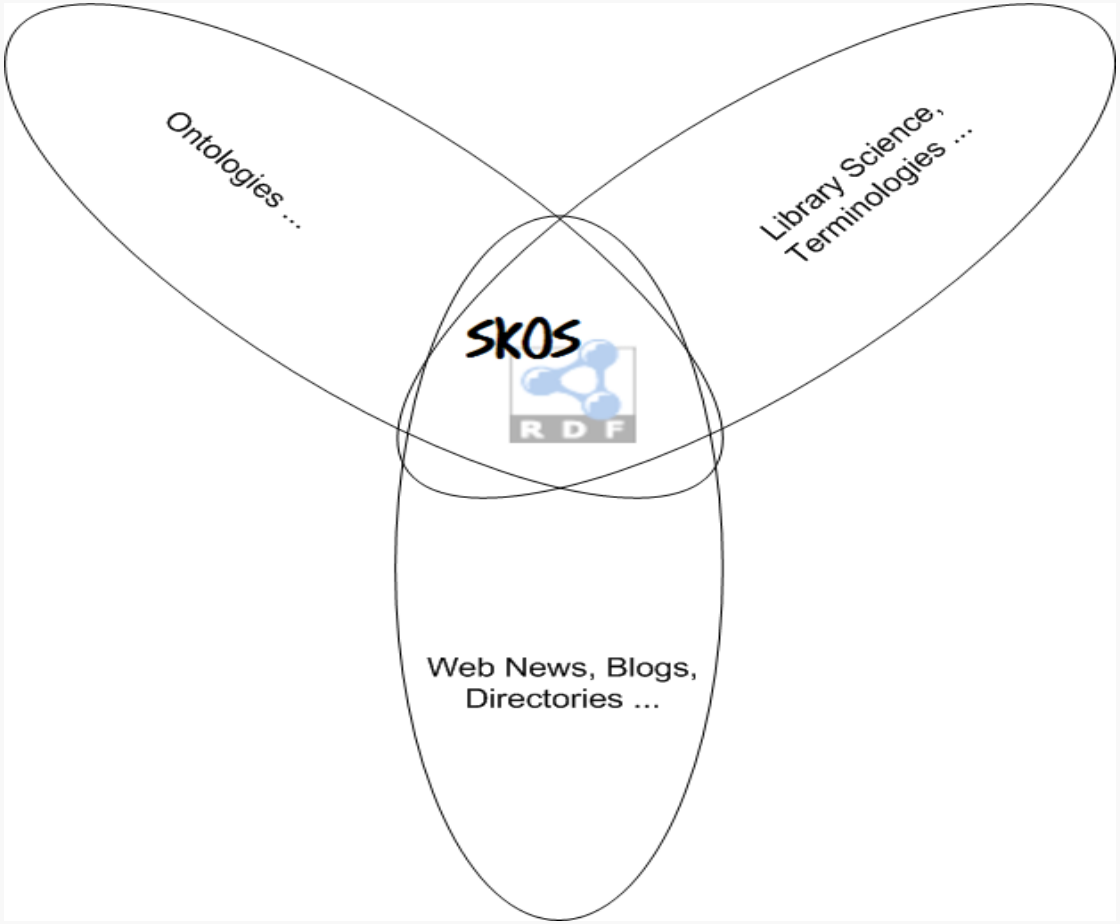
Can SKOS and OWL be mixed?

- What if OWL and SKOS data are mixed?
- The Ghostbuster Warning
  - “Never cross the streams” of the proton throwers
  - “Life as you know it would end”…
  - “It would be bad”...

Crossing the Streams



At the Confluence?



Source: Dan Brickley or Alistair Miles (not sure), circa 2005

Conclusion: Either/Or?

- OWL Ontologies
  - Formally precise “cartoon realities” that support inference
  - For Ontology-Based Search (see above)
- SKOS Concept Schemes
  - Conceptualizations by default informal and pragmatic
  - About retrieval and query expansion more than knowledge representation per se
  - For Concept Search
- Both provide URIs useful for semantic retrieval
- In practice, SKOS and OWL co-exist peacefully…

tom@tombaker.org

“Minimal Semantic Commitment”

An ontology should require the minimal ontological commitment sufficient to support the intended knowledge sharing activities.

An ontology should **make as few claims as possible about the world being modeled**, allowing the parties committed to the ontology **freedom to specialize and instantiate the ontology as needed**.

Thomas Gruber, 1995

SKOS in a Nutshell

- Using SKOS, concepts can be
  - **identified** using URIs,
  - **labeled** with lexical strings in one or more natural languages,
  - **assigned notations** (lexical codes),
  - **documented** with various types of note,
  - **linked to other concepts** and organized into informal hierarchies and association networks,
  - **aggregated** into concept schemes,
  - **grouped** into labeled and/or ordered colletions,
  - and **mapped** to concepts in other schemes.