Pragmatics of RDF/OWL

Guus Schreiber
Free University Amsterdam
Co-chair W3C Web Ontology Working Group 2002-2004
Co-chair W3C Semantic Web Best Practices & Deployment Working Group

Overview

- Why ontologies?
- OWL and its relation to RDF
- Ongoing work in W3C Semantic Web Best Practices & Deployment Working Group
- Sample applications

2

Disclaimer

- This presentation describes work of many different people, including many participating in respective W3C Working Groups as well as others

3

The notion of ontology

- Explicit specification of a shared conceptualization that holds in a particular context
- The Semantic Web needs sets of shared concepts.
- "I wrote my own ontology" is a contradiction in terms.

4

Ontology types

- Domain-specific ontologies
  - Medicine: UMLS, SNOMED, Galen
  - Art history: AAT, ULAN
  - STEP application protocols
- Generic ontologies
  - Top-level categories
  - Units and dimensions

5

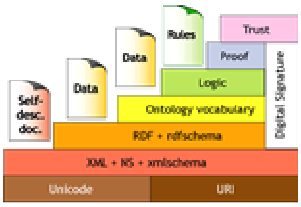
Ontologies and data models

- Main difference with data models is not the content, but the purpose (generalizes over applications)
- You cannot see the difference by just looking at the syntax!
- A conceptual model written in a ontology language is not necessarily an ontology!

6

W3C Web Ontology Working Group

- Chartered to develop the Ontology Vocabulary for the Semantic Web.
- Nov 2001 - Feb 2004



7

Bluffer's guide to RDF

- **Object ->Attribute-> Value** triples

- objects are **web-resources**
- Value can be an Object or a Literal

8

Use Cases for OWL

- Web portal
- Multi-media collections
- Corporate Website
- Documentation engineering & design
- Agents & Services
- Ubiquitous computing

9

Requirement: "classes as instances"

Ontology A

```
Class(Species)
Individual(Chimpanzee type(Species))
```

Ontology B

```
Class(Chimpanzee)
Individual(Joe  type(Chimpanzee)
```

10

Reaching consensus

- Issue 5.3 Semantic Layering

OWL is expected to be semantically compatible with RDF(S). Problems were foreseen with aligning a DL-style model theory with the RDF model theory, as the latter allows more or less unlimited metamodelling.

11

Consensus on Semantic Layering

- OWL Full ("Large OWL", "Great Horned OWL")
  - Free mixing of OWL and RDF = high expressivity
  - Non-standard formalization
  - Tractability not guaranteed
- OWL DL ("Fast OWL")
  - Maximize expressiveness while retaining tractability
  - Standard formalization
  - Same language constructs as OWL Full
  - Constraints on RDF/OWL vocabulary use
- Correspondence theorem links the two styles of semantics: entailments in OWL also hold in OWL Full.

12

A piece of OWL in RDF/XML syntax

```
<owl:Class rdf:ID="MozartDaPonteOpera">
  <owl:equivalentClass>
    <owl:Class>
      <owl:oneOf rdf:parseType="Collection">
        <Opera rdf:about="#NozzDiFigaro"/>
        <Opera rdf:about="#DonGiovanni"/>
        <Opera rdf:about="#CosiFanTutte"/>
      </owl:oneOf>
    </owl:Class>
  </owl:equivalentClass>
</owl:Class>
```

13

OWL presentation syntaxes

- UML Profile
  - ODM activity at OMG
- XML
  - More human-readable than the RDF/XML syntax
- Abstract syntax
  - See examples in this presentation
- N3 Turtle syntax

14

RDF language constructs

- classes and individuals
- subclasses
- properties
- subproperties
- domain/range of properties
- XML Schema datatypes

15

OWL language constructs

- equality, inequality
- inverse, transitive, symmetric, functional properties
- property constraints: cardinality, allValuesFrom, someValuesFrom
- conjunction, disjunction, negation of classes
- hasValue, enumerated type

16

Example: transitive property

```
TransitiveProperty(subRegionOf
    domain(GeographicalRegion)
    range(GeographicalRegion))
```

- If Tuscay is a subregion of Italy and Montepulciano is a subregion of Tuscany, then Montepulciano is a subregion of Italy.

17

Example: necessary and sufficient conditions

- Having a red color is a neessary and sufficient feature to becalled a red wine.

```
equivalentClass(RedWine
    intersectionOf(
        Wine
        Restriction(wineColor
                    hasValue("red"))))
```

- "intersectionOf" represents a AND condition

18

Example: (in)equality of individuals

- No closed-world assumption in OWL!
- Used for ontology/vocabulary mapping

```
Ontology A
    Individual(Brunello, type(vin:Wine))
Ontology B:
    Individual(BrunelloDiMontalcino
        type(vin:Wine))
 sameAs(A:Brunello B:BrunelloDiMontalcino)
```

19

Summary: RDF/OWL family of languages

- OWL Full is a vocabulary extension of RDF.
- The RDF restrictions in OWL DL are there for good technical reasons
- Time will have to prove whether there is a place for OWL Lite or some other OWL subset.
- RDF/OWL: one can view it as an historical artefact that these are not grouped under the same acronym.

20

Is RDF/OWL just another datamodelling/KR language?

Key differences:
- All classes/properties/individuals have a URI as identifier
- RDF/XML exchange syntax enables interoperability

XML features
- UTF-8 character set
- Support for multilinguality
- Use of XML Schema datatypes: numeric, date, time, etc.

For the rest: RDF/OWL is state-of-the-art concept language

21

OWL tools and tests

- OWL ontology editor
  - popular is Protege-OWL.
- Repository of tests is part of recommendation:

  http://www.w3.org/2002/03owlt/
- See also test results of tools:

  http://www.w3.org/2003/08/owl-systems/test-results-out

22

Semantic Web Best Practices and Deployment Working Group

- Started 1 March 2004
- Co-chairs David Wood and Guus Schreiber
- 50 participants
- Objective: support for semantic-web application developer
  - Focus on “low hanging fruit”

23

Issues for publishing ontologies: good and bad ontologies?!

- Good ontologies are used
- Good ontologies represent some form of consensus in a community
- Good ontologies are maintained
- Good ontologies do not need to be complex
- Good ontologies may contain “mistakes”

24

Ontology engineering patterns

- Best practices for frequently occurring modeling problems
- WG documents outline alternatives with pros and cons
- Notes:
  - Classes as values
  - N-ary relations
  - Specification of value sets
  - Part-of

25

Representing value sets

- Intuitive representation of **color** value set:
  class/datatype **color**
  with instances/values "red", "white", etc.
- But suppose we want to talk about a subtype of "red", e.g. "vermillion red"
- Pattern:
  - Represent values as subclass hierarchy of value type
  - This preserves flexibility
  - Use anonymous instances as property values
    "This porcelain vase has as color some value of vermillion red"
- See note by Alan Rector
  http://www.w3.org/TR/swbp-specified-values/

26

Classes as values

- Common problem when using a hierarchy of concepts for indexing purposes
- Example: indexing books with concepts from the ACM computer-science subject hierarchy
- See draft technical note by Noy:
  - Four options with different merits
- See note by Natasha Noy:
  http://www.w3.org/TR/swbp-classes-as-values/

27

Numeric constraints and user-defined XML Schema datatypes

- Example: "an elevated diastolic blood pressure is a diastolic blood pressure of 90 Hg or more"
- Currently no simple way to represent this in OWL
- User-defined XML Schema datatypes could provide a solution
- Currently not possible for detailed technical reasons
- SWBPD task force is active in trying to solve this problem (Jeremy Carroll, HP)

28

Some task forces of W3C Best Practices WG

- Tutorial page
  http://www.w3.org/2001/sw/BestPractices/Tutorials
- RDF in XHTML
  http://lists.w3.org/Archives/Public/public-rdf-in-xhtml-tf/2004Oct
- Publication of vocabularies/ontologies
  - WordNet is first on the list
  - Units and measures is likely next target
- Applicatons weblog
  http://esw.w3.org/mt/esw/archives/cat_applications_and_demos.html

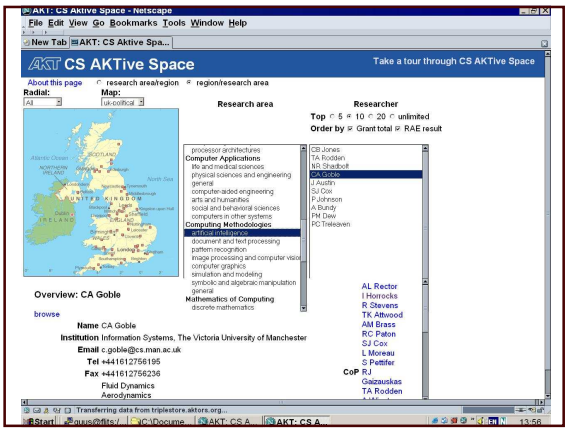
  Thesaurus/vocabulary management: see SKOS talk

29

AKTive Space

- AKT project (Shadbolt et al.), winner Semantic Web Challenge 2003)
- Integration of heterogeneous sources
  - Public archives of papers, researchers, projects
  - 430 Mb in total
- RDF/OWL used for syntactic interoperability
  - Storage/access issues
- Schema mapping is required
- Referential integrity is a problem
  - Use automatic techniques in combination with user approval

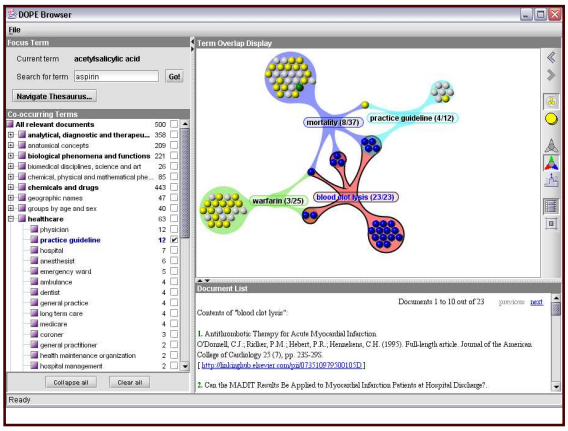
30



DOPE: thesaurus-based search of large document repositories

- Stuckerschmidt et al. (2003)
- EMTREE thesaurus (MeSH-based)
- Documents
  - 5M Medline abstracts
  - 500M of full-text articles
- Automatic document indexing
- RDF used for syntactic interoperability
  - RDF wrapper for SOAP-based access to documents
- Disambiguation of search terms
- Visualization of search results through semantic categories
  - Needed to prevent information overflow

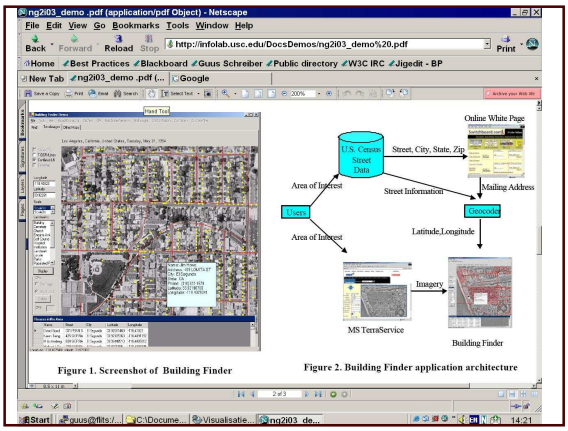
32



Building Finder: integrating image analysis and textual sources

- Knoblock et al. (USC/ISI)
- Multiple heterogeneous sources
  - Satellite images (Microsoft Terraservice)
  - Road map info (US)
  - Address information (white pages)
- Image analysis techniques to map satellite data to road map
- RDF used for syntactic interoperability

34



Application scenario: Paintings

Knowledge corpora
AAT ULAN
ICONCLASS
WordNet

Annotation Template
VRA 3.0
Scene descriptors

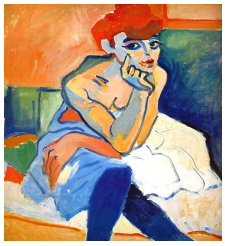
RDF Schema

Annotation & search tool → RDF image annotations

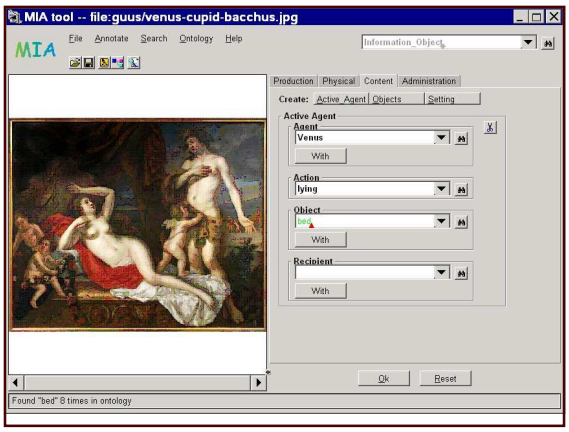
36

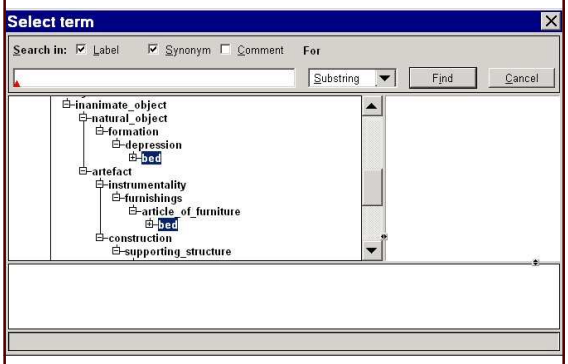
Semantic search in cultural heritage collections: use case

- A person is interested in Fauve paintings
- There is a digital collection with images of paintings of Andre Derain
- The Derain images match the query, despite the fact that "Fauve does not appear in the annotation.



37





39

Parsing existing textual annotations into semantic annotations

VAN GOGH, Vincent

Olive Trees with the Alpilles in the Background

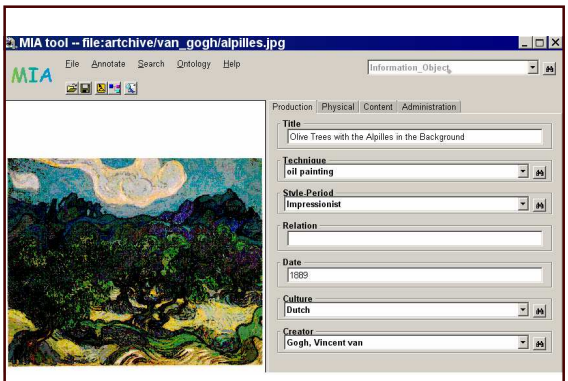
1889

Oil on canvas

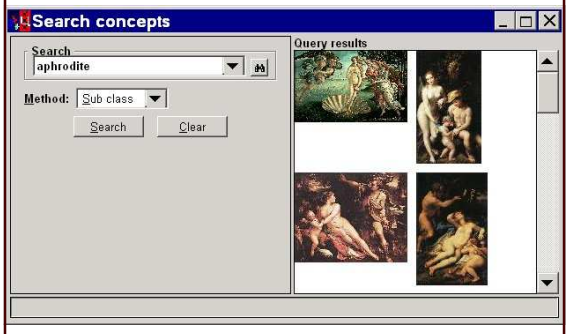
28 1/2 x 36 1/4 in. (72.5 x 92 cm)

Collection Mrs. John Hay Whitney

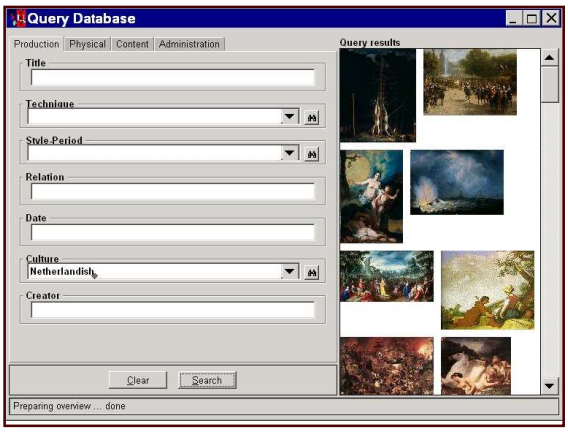
40



41



42



Issues in relation to vocabularies

- Public availability
- Concept mapping
  - Equivalence
    - Impressionism (AAT) = impressionism (WordNet)
  - Subclass relation
    - Artist (AAT) ⊂ Artist (WordNet)
  - Instance relation
    - Continent (WordNet) => Africa (TGN)
  - Domain-specific relations
    - Artist (ULAN) ⇔ Style (AAT)
- Vocabulary enrichment
  - Making implicit knowledge explicit
  - Adding additional knowledge

44

Will the Semantic Web succeed?

- One big plus: there is a growing need for semantic search of information
- Availability of large amounts of semantic content is essential
  - But: there is a lot of content already out there
- First applications are likely to be in area of large virtual collections
  - E.g., cultural heritage, medicine
- Web services will not work without ontologies

45

More information

- W3C Semantic Web page
  http:/www.w3.org/2001/sw/
- W3C Best Practices work
  http://www.w3.org/2001/sw/BestPractices/
- Semantic Web Challenge
  http://challenge.semanticweb.org
- European Network of Excellence Knowledge Web
  http://knowledgeweb.semanticweb.org

46