Publishing Vocabularies on the Web

Guus Schreiber
Antoine Isaac
Vrije Universiteit Amsterdam

Acknowledgements

- Alistair Miles, Dan Brickley, Mark van Assem, Jan Wielemaker, Bob Wielinga
- Participants of the W3C Semantic Web Best Practices and the Semantic Web Deployment Working Groups

Overview

- Issues in conversion to RDF/OWL
  - Example: Union List of Artist Names (ULAN)
  - Example: WordNet 2.0
- Work within the W3C Semantic Web Deployment Working Group
  - SKOS model for thesauri
  - Recipes for Web access to published vocabularies
  - RDFa: embedding RDF metadata in HTML

Thesauri / vocabularies

- Controlled vocabularies

  Thesauri, classification schemes, taxonomies, subject heading lists, authority lists…
- Large bodies of knowledge that represent consensus in particular domains
- Often lots of implicit semantics available
- Semantic Web Challenge showed that thesauri are important resources for SW applications
- Representation is typically relational database and/or XML

Example thesauri

- Domain-specific vocabularies
  - Medicine: UMLS, SNOMED, MESH, Galen
  - Art history: AAT, ULAN
  - Geography: TGN
  - Food: AgroVoc
  - Libraries: LCSH, DDC, UDC
- Generic vocabularies
  - Lexical vocabularies: WordNet, FrameNet
  - Currencies, country codes, …

ISO standard for representing thesauri

- Term
  - Preferred term (USE)
  - Non-preferred term (USED FOR)
- Hierarchical relation between terms
  - Broader/narrower term (BT/NT)
    - Generic
    - Partitive
- Association between terms (RT)

Typical conversion process

- Two steps
- Step 1: “As is” conversion
  - Keep original names/constructs
  - Make implicit semantics explicit (not trivial!)
  - Decisions on whether to keep all information
- Step 2: adding semantics
  - Separate file(s)
  - Interpretation of thesauri features, e.g. hyponym relation as **rdfs:subClassOf**
  - May require (lots of) additional research

Example thesaurus: ULAN

- 300,000 “Subject” records (artists and art institutions)
  - with biographical information (place/time birth/death)
  - and relations to other artists (student-of, …)
- Large XML file with all data
- Basic representation:
  - association links between subjects
  - preferred/non-preferred terms relations between subjects and terms

Research

Research Home ▸ Conducting Research ▸ Union List of Artist Names ▸ Full Record Display

Union List of Artist Names® Online
Full Record Display

New Search Previous Page Help

Click the icon to view the hierarchy.

ID: 500000351 **Record Type: Person**

Koninck, Philips de (Dutch painter and draftsman, 1619-1688)

Note: History and portrait painter who is today most well-known for his naturalistic panoramic bird's-eye view landscapes.

Birth and Death Places:

Born: Amsterdam (North Holland, Netherlands) (inhabited place)
Died: Amsterdam (North Holland, Netherlands) (inhabited place)

Related People or Corporate Bodies:

related to (familial) .... Koninck, Salomon
..... (Dutch painter, printmaker, and draftsman, 1609-1656) [500027532]
sibling of .... Koninck, Jacob, the elder
..... (Dutch painter and engraver, ca. 1616-1708) [500024292]
student of .... Rembrandt van Rijn
..... (Dutch painter, draftsman, and printmaker, 1606-1669) [500011051]

XML fragment of ULAN: links

```
<Associative_Relationships>
  <Associative_Relationship>
    <Historic_Flag>NA</Historic_Flag>
    <Relationship_Type>
      1102/student of
    </Relationship_Type>
    <Related_Subject_ID>
      <VP_Subject_ID>500011051</VP_Subject_ID>
    </Related_Subject_ID>
  </Associative_Relationship>
</Associative_Relationship>
```

Conversion issues

- XML and RDF/OWL are inherently different
  - XML = thesaurus document structure
  - RDF = thesaurus document content
- Redundant/meaningless information in XML file

  <Associative_Relationships>

  <Historic_Flag>NA</Historic_Flag>
- How to represent “student of”?
  - Subproperty of **Associative_Relationship** is probably preferred
  - Needs to be derived from the data; not part of schema

XML fragment of ULAN: terms

```
<Non-Preferred_Term>
  <Term_Text>Koning, Philips Aertsz. de</Term_Text>
  <Term_ID>1500207734</Term_ID>
  <Display_Order>34</Display_Order>
  <Vernacular>Vernacular</Vernacular>
</Non-Preferred_Term>
```

Conversion issues

- Do we include all information in the conversion?
  - Display order
- Should each term have a URI?
- Making language explicit
  - “vernacular” means the string is written in the original language
  - Multi-linguality is an important issue for thesauri

W3C Working Draft

RDF/OWL Representation of WordNet

W3C Working Draft 19 June 2006

This version:

http://www.w3.org/TR/2006/WD-wordnet-rdf-20060619/

Latest version:

http://www.w3.org/TR/wordnet-rdf/

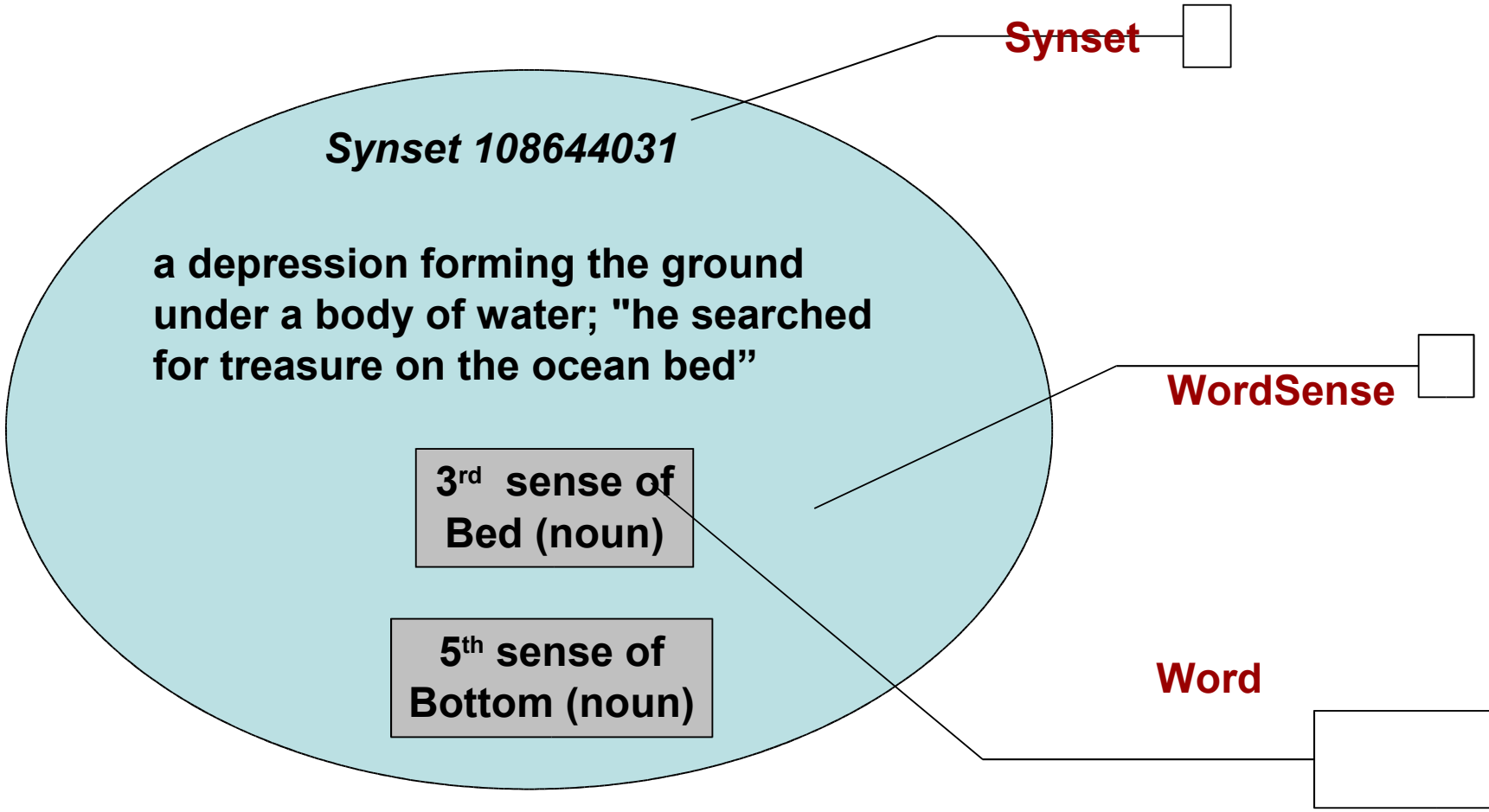
Previous version:

This is the first published version

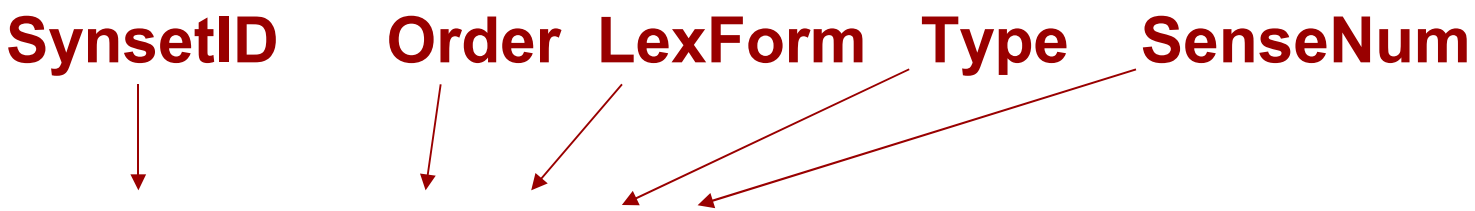
Editors:

Mark van Assem, Vrije Universiteit Amsterdam
Aldo Gangemi, ISTC-CNR, Rome

WordNet model



WordNet: internal representation



s(108644031,1,'bed',n,3,2).
s(108644031,2,'bottom',n,5,1).

s(102719813,1,'bed',n,1,51).

g(108644031,'(a depression forming the ground under a body of water; "he searched for treasure on the ocean bed")').
g(102719813,'(a piece of furniture that provides a place to sleep; "he sat on the edge of the bed"; "the room had only a bed and chair")').

WordNet URIs

- What URIs should be chosen?
  - SynSet, WordSense, Word
- URI name:
  - ID? => difficult for human interpretation
  - Human-readable concatenation

wn:synset-bank-noun-2

synset denoted by second sense of “bank”

wn:wordsense-bank-noun-1

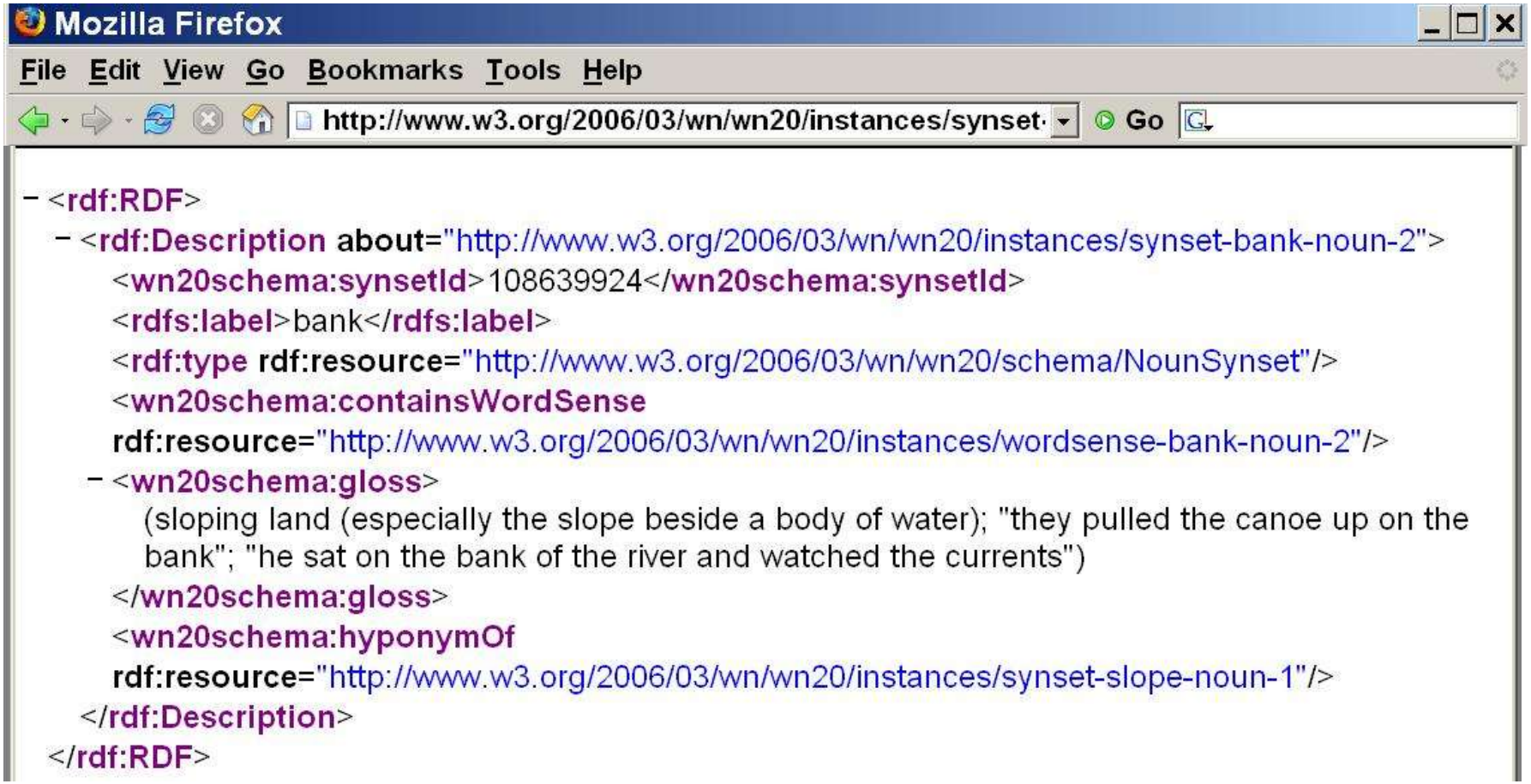
wn:word-bank

Implicit WordNet semantics

"The ent operator specifies that the second synset is an entailment of first synset. This relation only holds for verbs."

- Example: [breathe, inhale] entails [sneeze, exhale]
- Semantics (OWL statements):
  - Transitive property
  - Inverse property: entailedBy
  - Value restrictions for VerbSynset (subclass of Synset)

Data access



- Query for WordNet URI returns “concept-bounded description”

Overview

- Issues in conversion to RDF/OWL
  - Example: Union List of Artist Names (ULAN)
  - Example: WordNet 2.0
- Work within the W3C Semantic Web Deployment Working Group
  - SKOS model for thesauri
  - Recipes for Web access to published vocabularies
  - RDFa: embedding RDF metadata in HTML

W3C Semantic Web Deployment Working Group

Making vocabularies/thesauri/ontologies available on the Web

http://www.w3.org/2006/07/SWD/

SWD goals

- Schema for interoperable RDF/OWL representation of vocabularies
  - SKOS
- Publication guidelines
  - URI management, representation of versions
- Embedding RDF in (X)HTML pages
  - RDFa

NATIONAL AERONAUTICS
AND SPACE ADMINISTRATION

NASA Taxonomy - XML DTDs for Use with the NASA Taxonomy

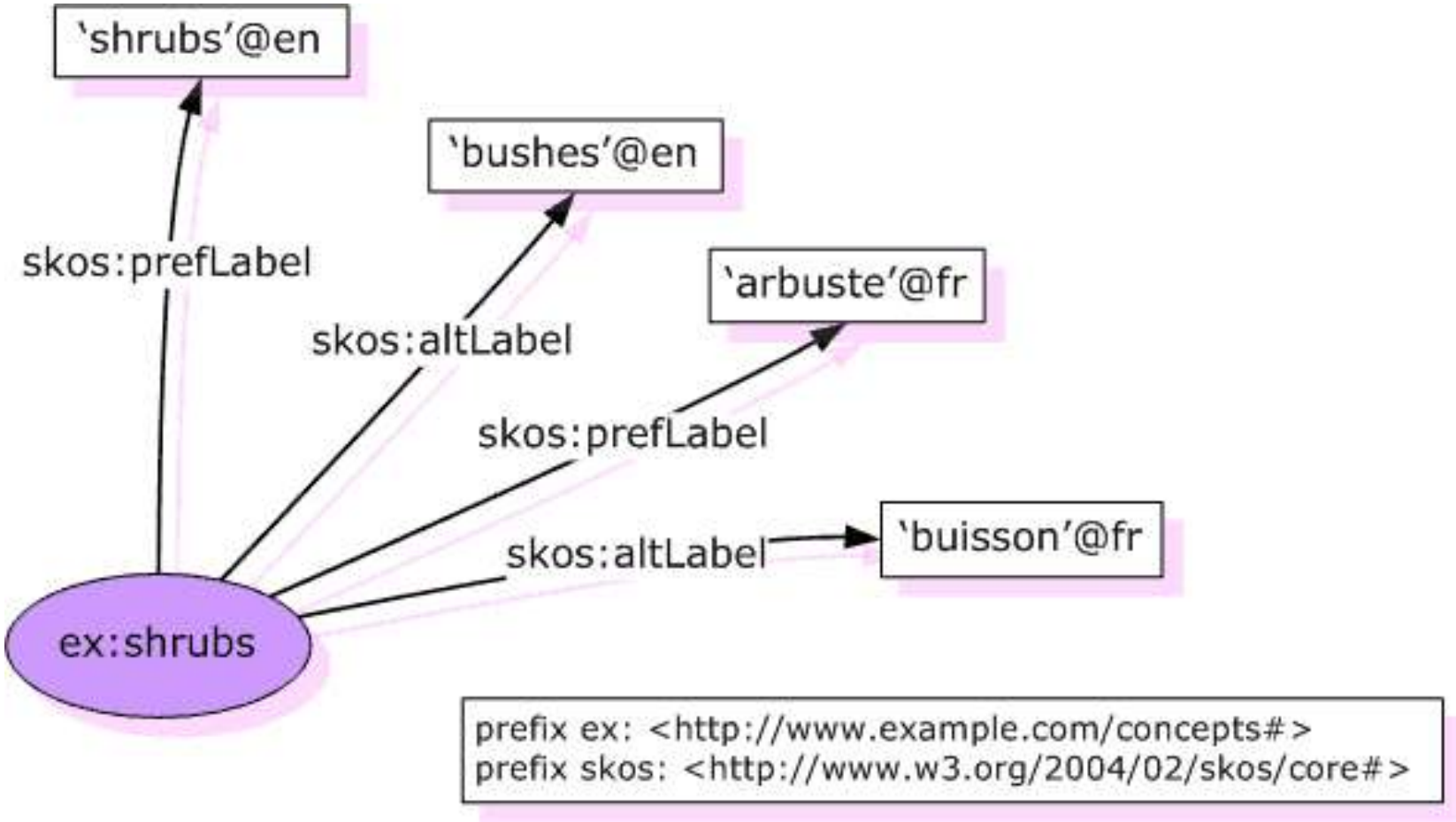
Important Update Regarding the XML format of the NASA Taxonomy - Jan 9, 2007

The next version of the NASA taxonomy will be in the SKOS format.

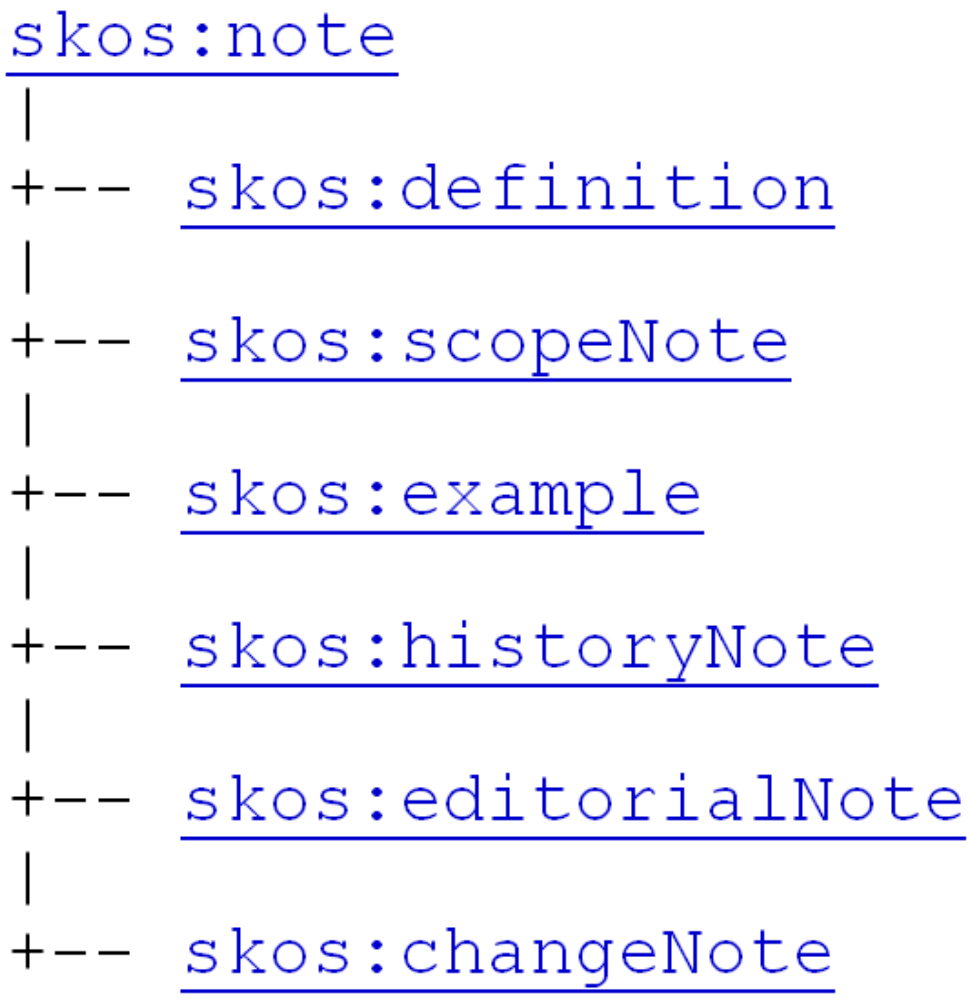
The SKOS Core is a model and an RDF vocabulary proposed by the W3C for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, other types of controlled vocabulary.

The SKOS Core Vocabulary is an application of the Resource Description Framework (RDF), that can be used to express a

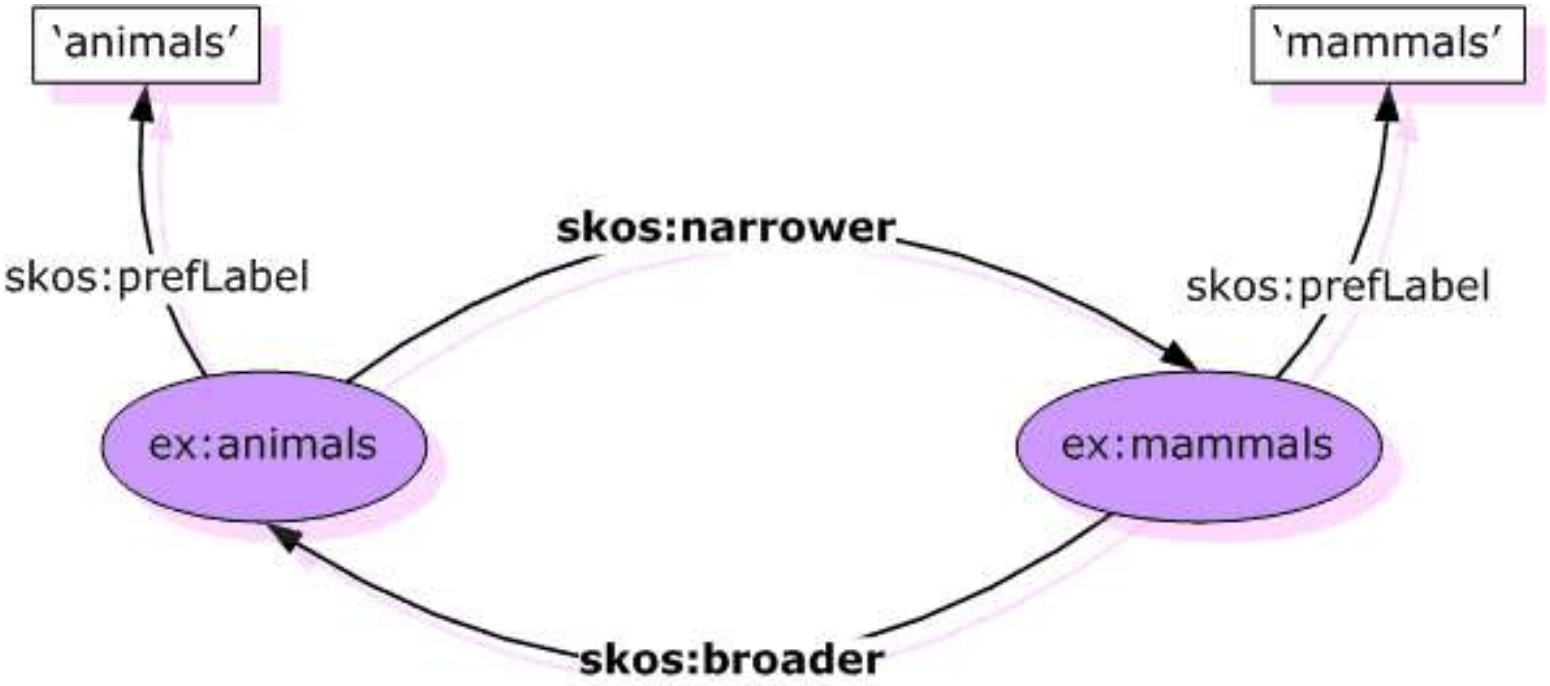
Multi-lingual labels for concepts



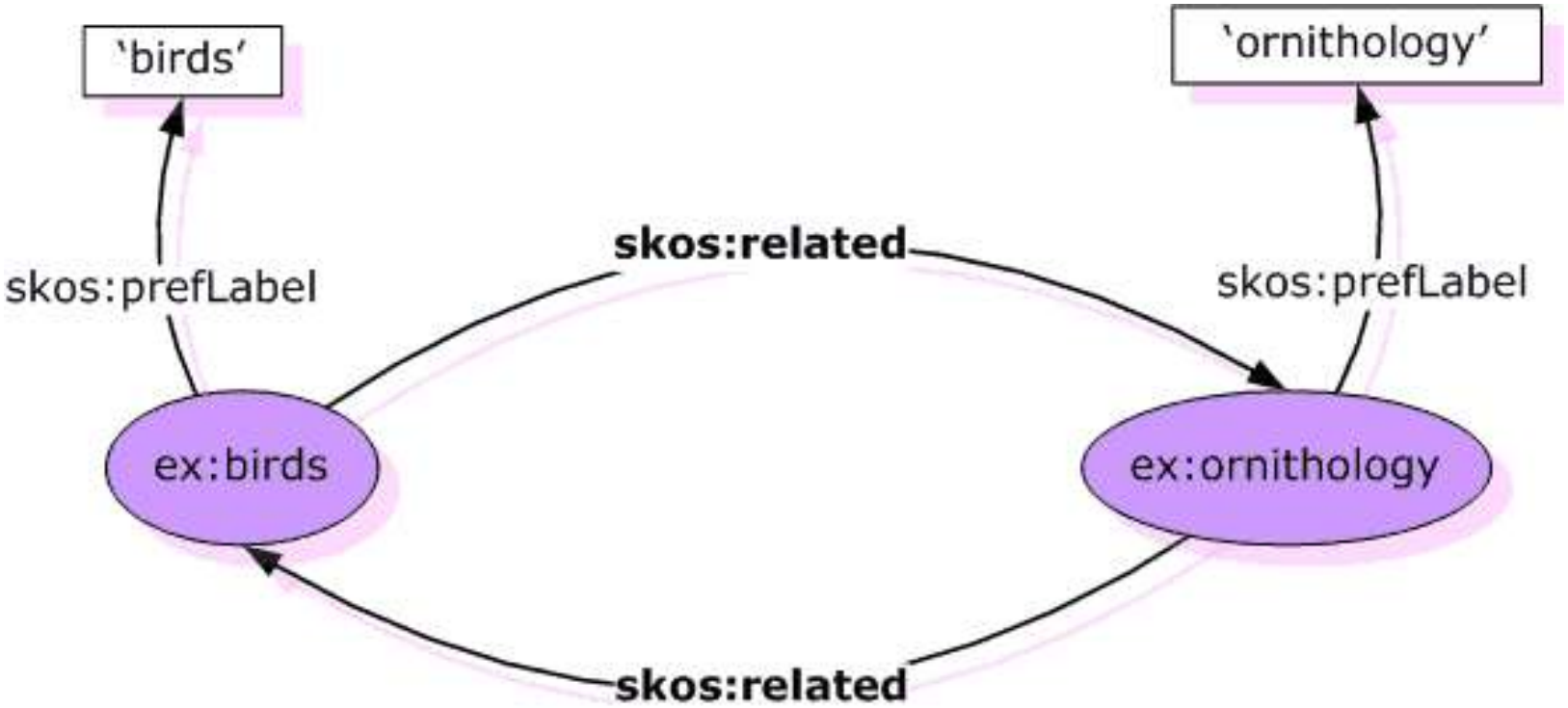
Documenting concepts



Semantic relation: broader and narrower

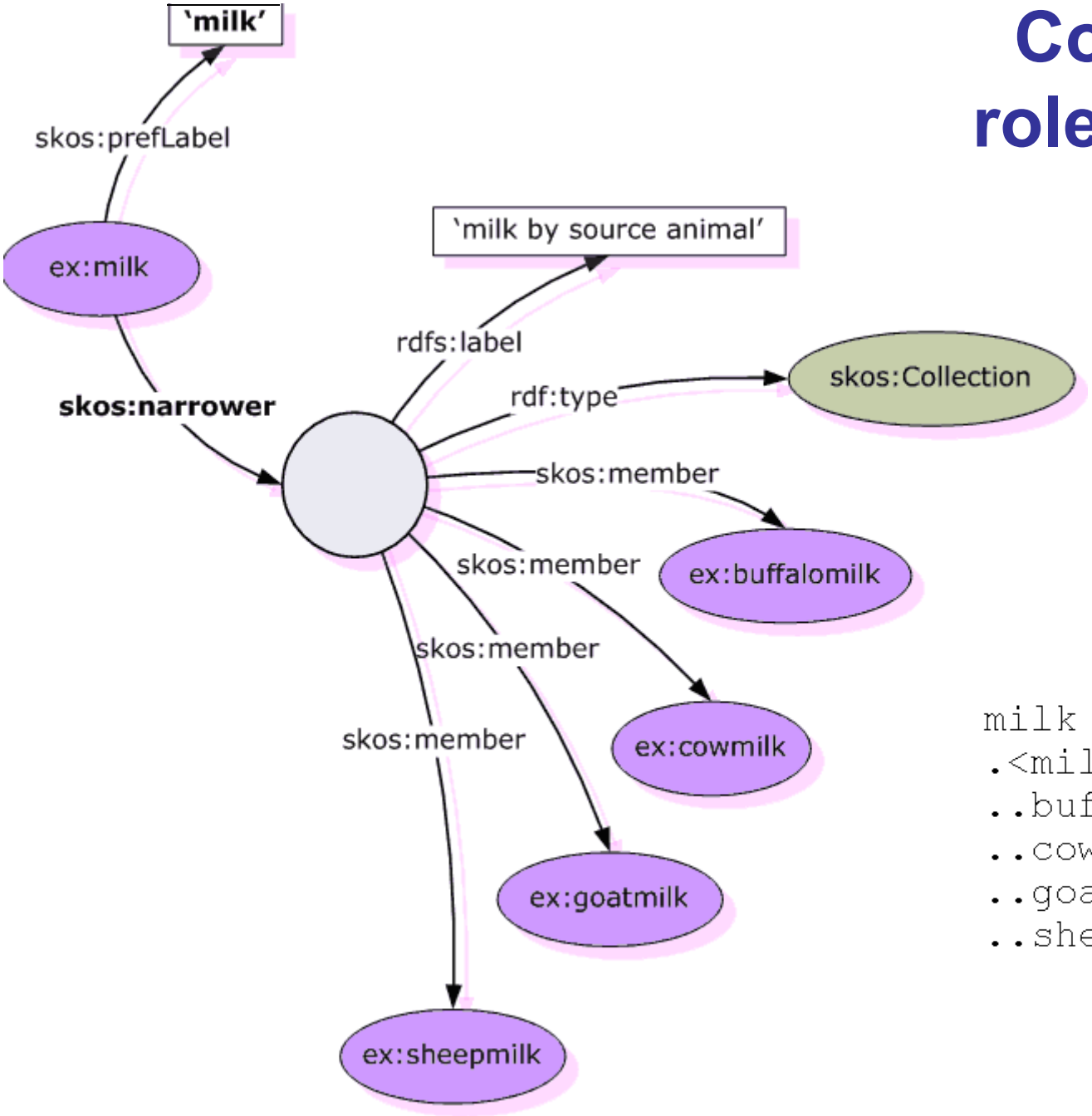


Semantic relations: related




```
prefix ex: <http://www.example.com/concepts#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
```

Collections: role-type trees




```
milk
.<milk by source animal>
..buffalo milk
..cow milk
..goat milk
..sheep milk
```

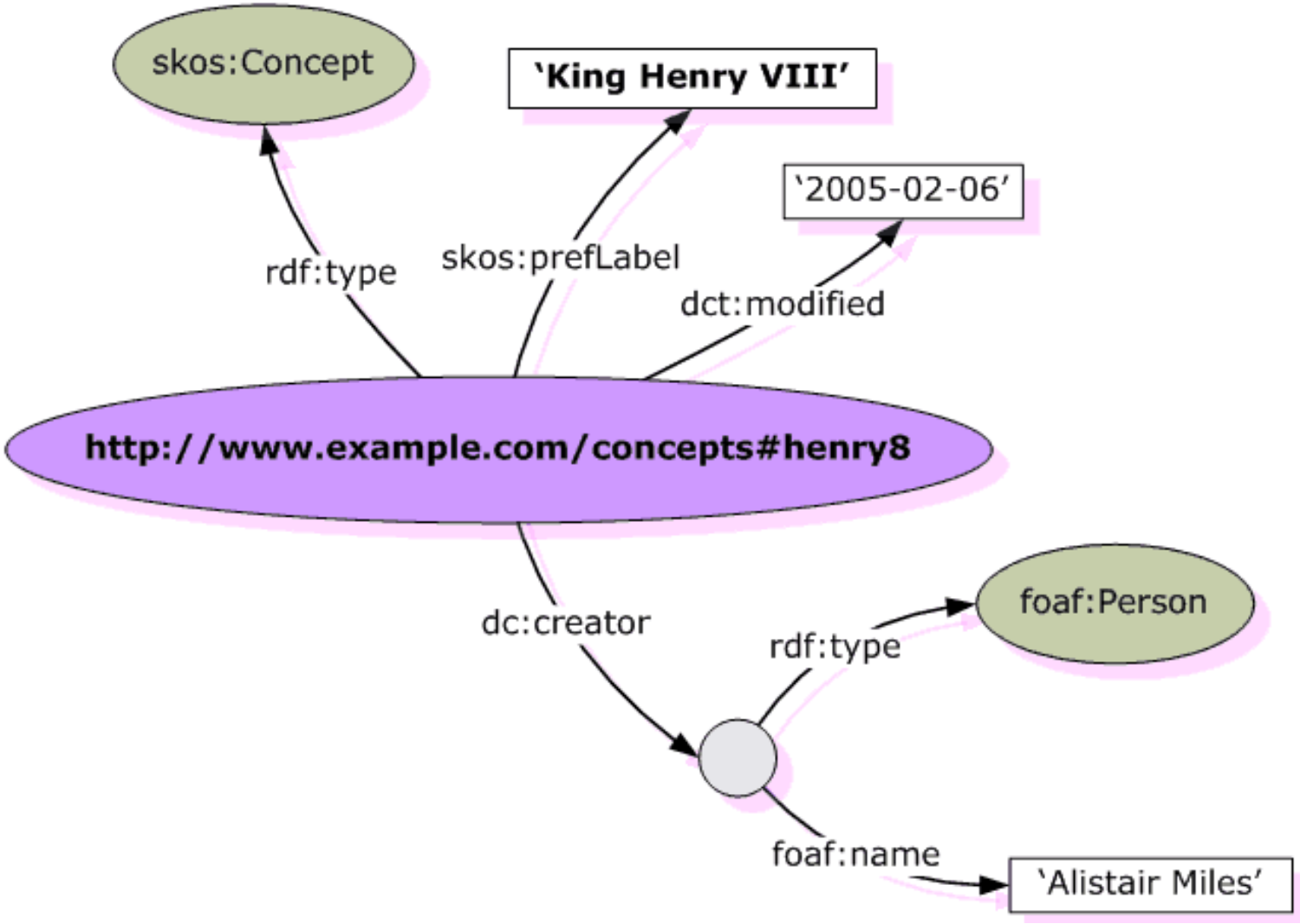
Adding semantics

- Adding OWL statements
  - **skos:related rdf:type owl:SymmetricProperty**
  - **skos:broader owl:inverseOf skos:narrower**
- Inference rules
  - Collection membership rule

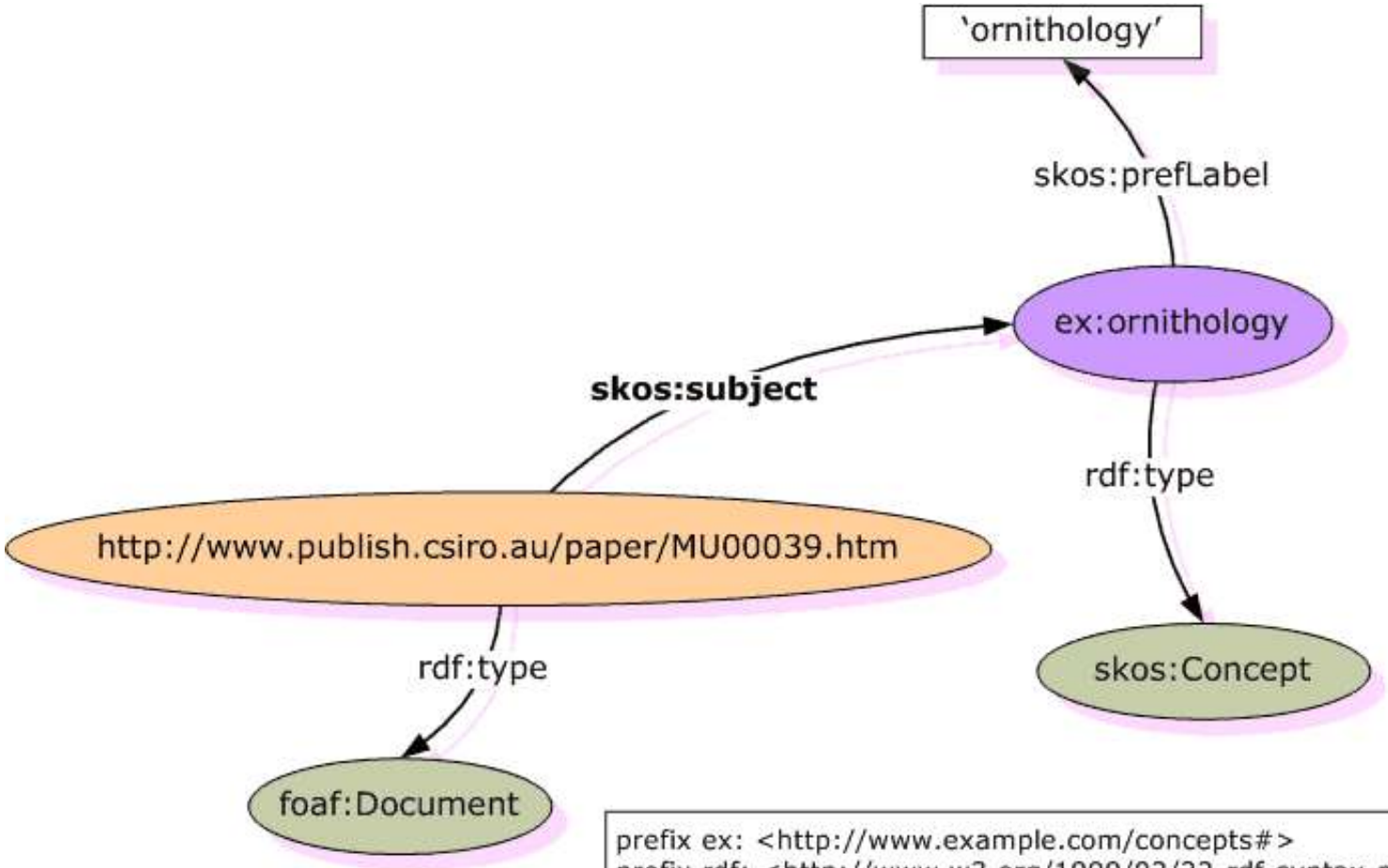
    (?s skos:narrower ?c) (?c skos:member ?t)

    → (?s skos:narrower ?t)
- Interpreting thesaurus relations such as **broader** as subClassOf can be useful but is often imprecise

SKOS semantics: concepts are not the real things

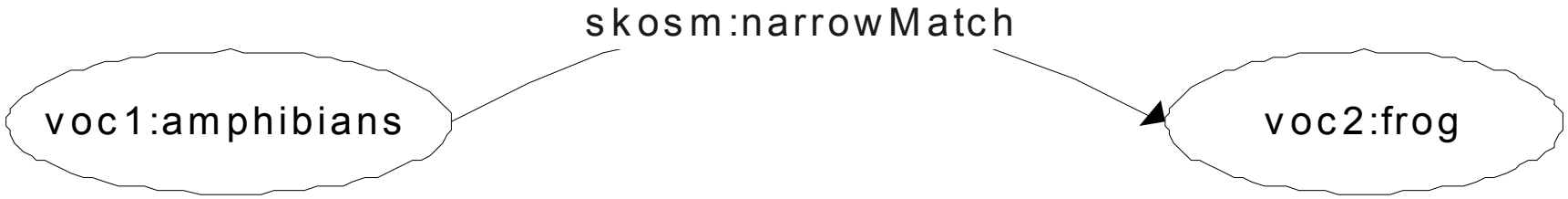


Indexing a resource with a SKOS concept



Semantic alignment links

- Learning relations between thesauri is important form of additional semantics
  - Example: AAT contains styles; ULAN contains artists, but there is no link
  - Availability of this kind of alignment knowledge is extremely useful
  - Cf. demo



Warning: unstable part of SKOS!

W3C standardization process

- Input: draft specification
- Collect use cases
- Derive requirements
- Create issues list: requirements that cannot be handled by the draft spec
- Propose resolutions for issues
- Get consensus on amended spec
- Find two independent implementations for each feature in the spec
- *Continuously*: ask for public feedback/comments

(YES, YOU!)

W3C Working Draft

SKOS Use Cases and Requirements

W3C Working Draft 16 May 2007

This version:

http://www.w3.org/TR/2007/WD-skos-ucr-20070516/

Latest version:

http://www.w3.org/TR/skos-ucr

Previous version:

This is the first public Working Draft

Editors:

Antoine Isaac, Vrije Universiteit Amsterdam, aisaac@few.vu.nl
Jon Phipps, Cornell University, jphipps@madcreek.com
Daniel Rubin, Stanford Medical Informatics, dlrubin@stanford.edu

Example use case and requirement

- **2.3 Use Case #3 — Semantic search service across mapped multilingual thesauri in the agriculture domain**

  "This application coming from the AIMS project […] includes some more specific links […] String-to-String relationships …"

acronym	Food and Agriculture Organization	FAO
spelling_variant	organisation	organization
translation	vache	cow

"Requires: […] R-RelationshipsBetweenLabels"

Example issue: relationships between lexical labels

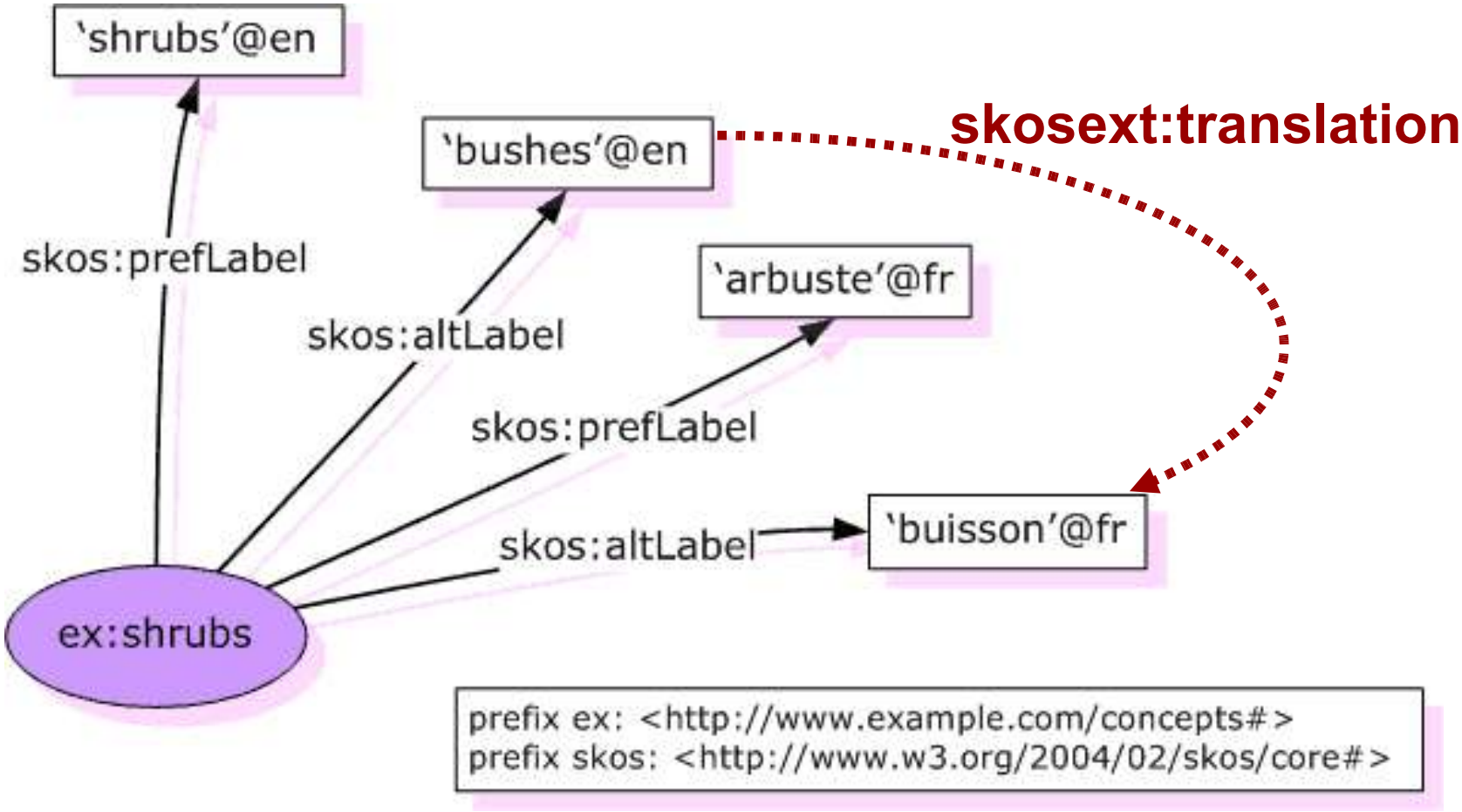
"R-RelationshipsBetweenLabels

Representation of links between labels associated to concepts

The SKOS model shall provide means to represent relationships between the terms associated with concepts. Typical examples are […]"

- In current SKOS spec labels are represented as literals
- This is a problem because literals have no URI, so cannot be subject of an RDF property
- Possible resolutions:
  - Labels/terms as instances of a new class
  - Relaxing constraints on label property

Example issue: relationships between lexical labels



SWD goals

- Schema for interoperable RDF/OWL representation of vocabularies
  - SKOS
- Publication guidelines
  - URI management, representation of versions
- Embedding RDF in (X)HTML pages
  - RDFa

Recipes for vocabulary URIs

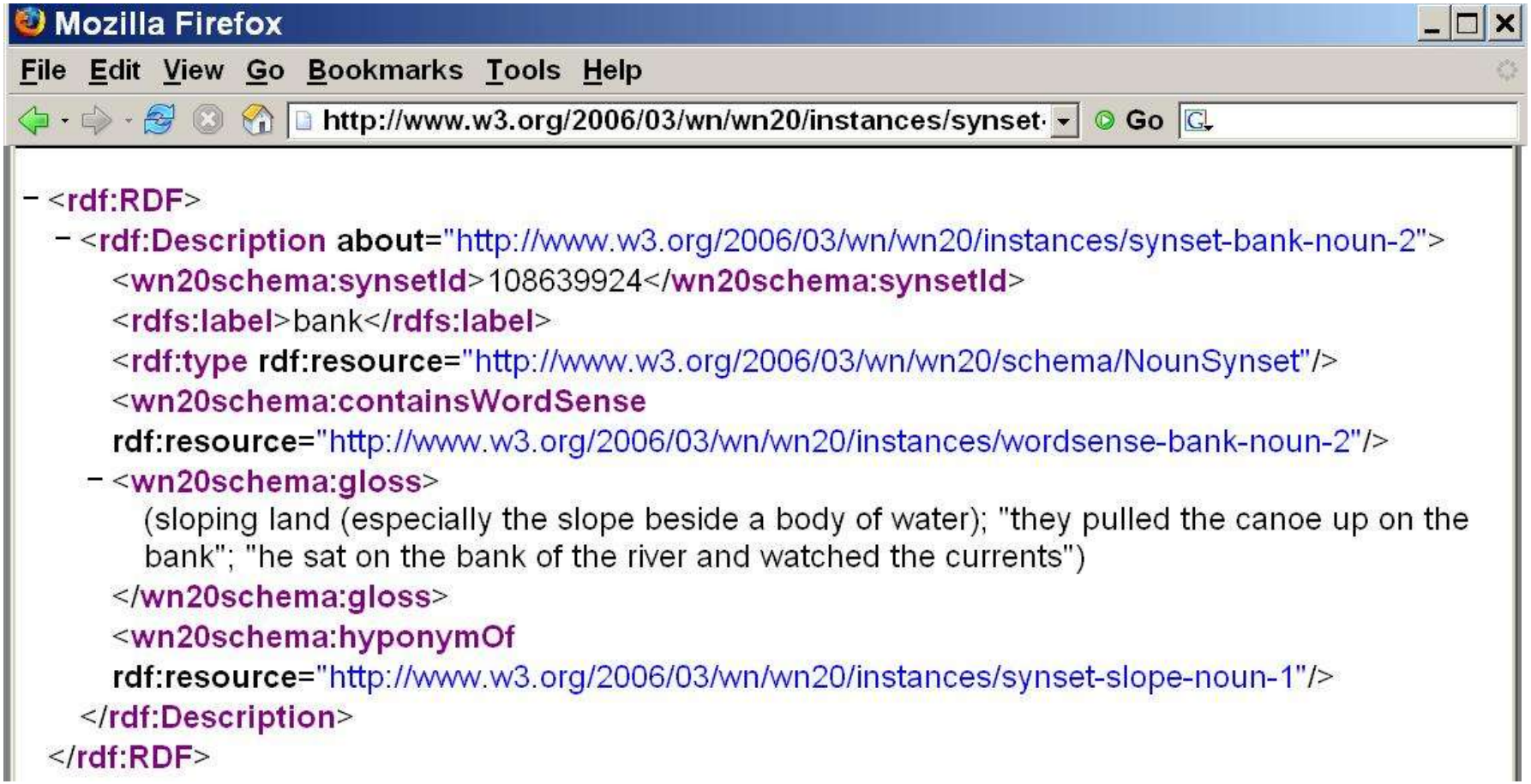
- Simplified rule:
  - Use “hash" variant” for vocabularies that are relatively small and require frequent access

    http://www.w3.org/2004/02/skos/core#Concept

  - Use “slash” variant for large vocabularies, where you do not want always the whole vocabulary to be retrieved

    http://www.w3.org/[...]/instances/synset-bank-noun2

Data access



- Query for WordNet URI returns “concept-bounded description”

Recipes for serving RDF

- Persistent URIs and version-specific content

  HTTP 303 redirection

  – Client asking http://example.org/voc#myClass

  – Client redirected to

  http://example.org/voc-files/voc-version3.rdf#myClass

- For more information and other recipes, see:

  http://www.w3.org/TR/swbp-vocab-pub/

SWD goals

- Schema for interoperable RDF/OWL representation of vocabularies
  - SKOS
- Publication guidelines
  - URI management, representation of versions
- Embedding RDF in (X)HTML pages
  - RDFa

A RDFa sample

Regular HTML

```
<h1>Photo Album #12345: Vacation in the South of France</h1>
<h2>created by Mark Birbeck</h2>
```

HTML with RDFa

```
<h1 property="dc:title">Vacation in the South of France</h1>
<h2>created by <span property="dc:creator">Mark Birbeck</span></h2>
```

Resulting RDF statements

```
<> dc:title "Vacation in the South of France"^^XMLLiteral .
<> dc:creator "Mark Birbeck"^^XMLLiteral .
```

Linking to other resources

Regular HTML

```
This document is licensed under a
<a href="http://creativecommons.org/licenses/by-nc/2.5/">
  Creative Commons Non-Commercial License
</a>.
```

HTML with embedded RDF

```
This document is licensed under a
<a rel="cc:license"
   href="http://creativecommons.org/licenses/by-nc/2.5/">
  Creative Commons Non-Commercial License
</a>.
```

Statements about other resources: photo example

```
<li> <img src="/user/markb/photo/23456" />,
  <span about="/user/markb/photo/23456" property="dc:title">
    Sunset in Nice
  </span>
</li>

<li> <img src="/user/markb/photo/34567" />,
  <span about="/user/markb/photo/34567" property="dc:title">
    W3C Meeting in Mandelieu
  </span>
</li>
```

RDFa demo

- *Having time, feeling lucky and online?*
- *Slides*

More information

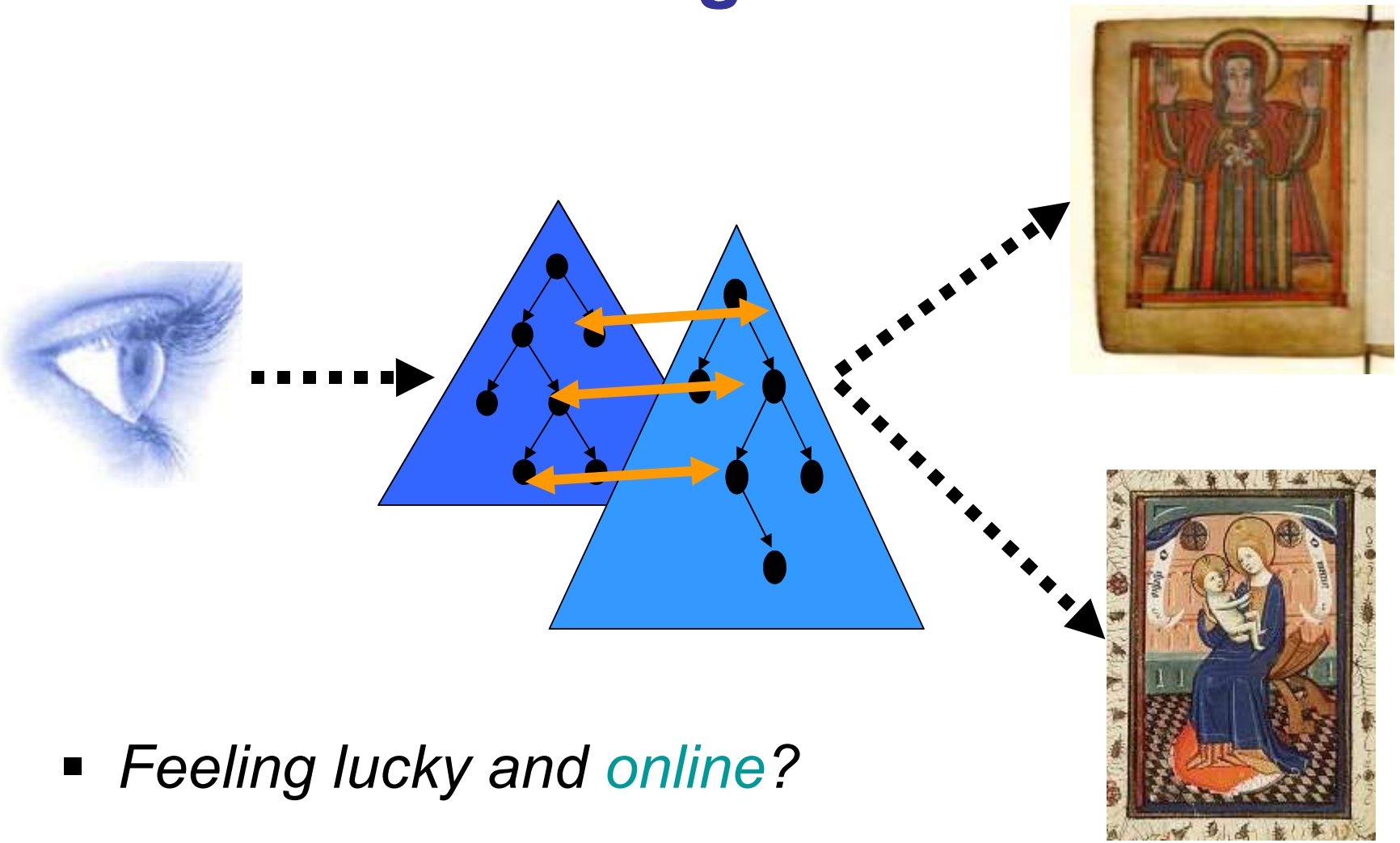
Thanks

- Reminder: we ask for feedback!
  - Questions and comments highly welcome

- aisaac at few.vu.nl
- schreiber at cs.vu.nl

- Continue for demo?

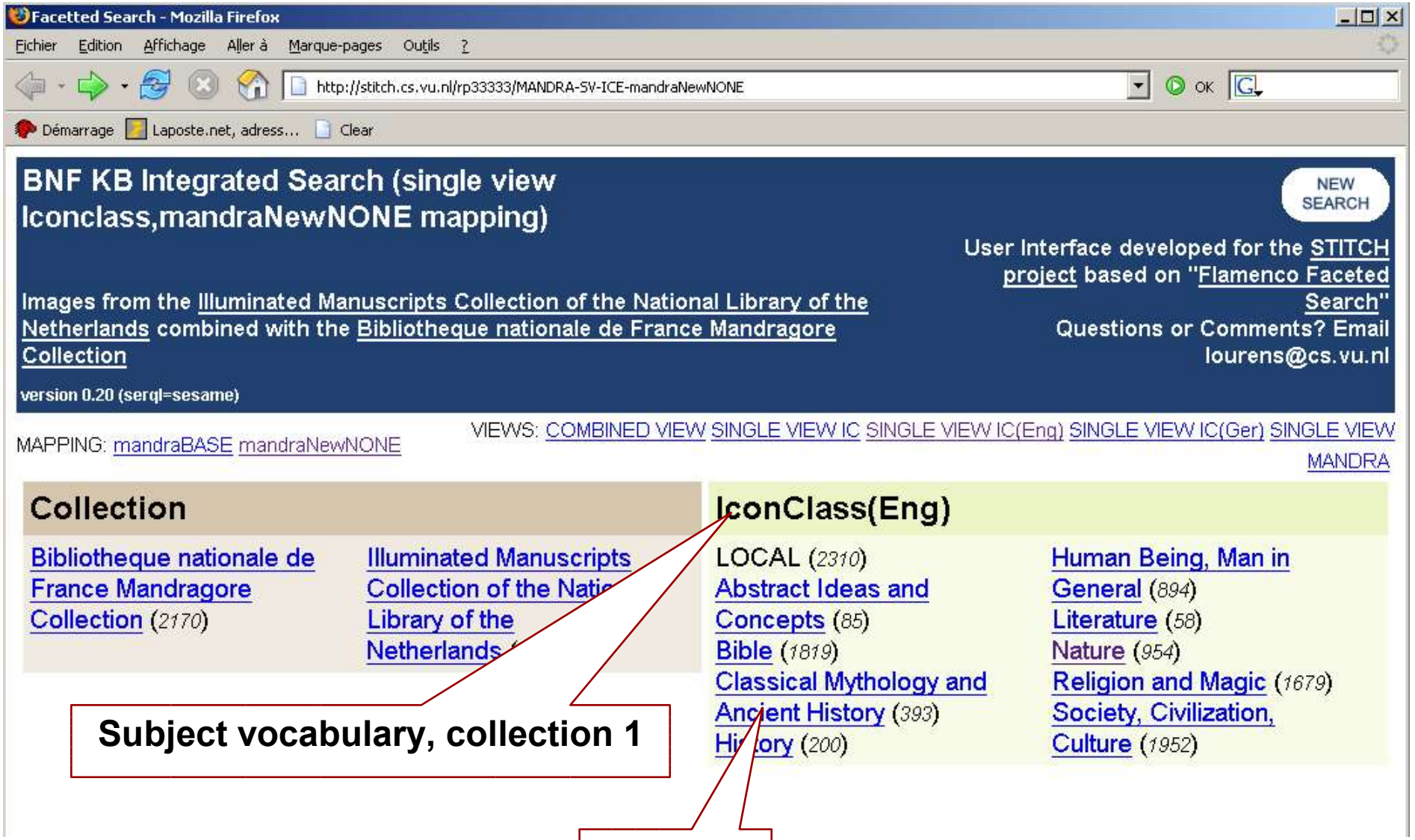
SKOS Demo: browsing and alignment



- *Feeling lucky and online?*

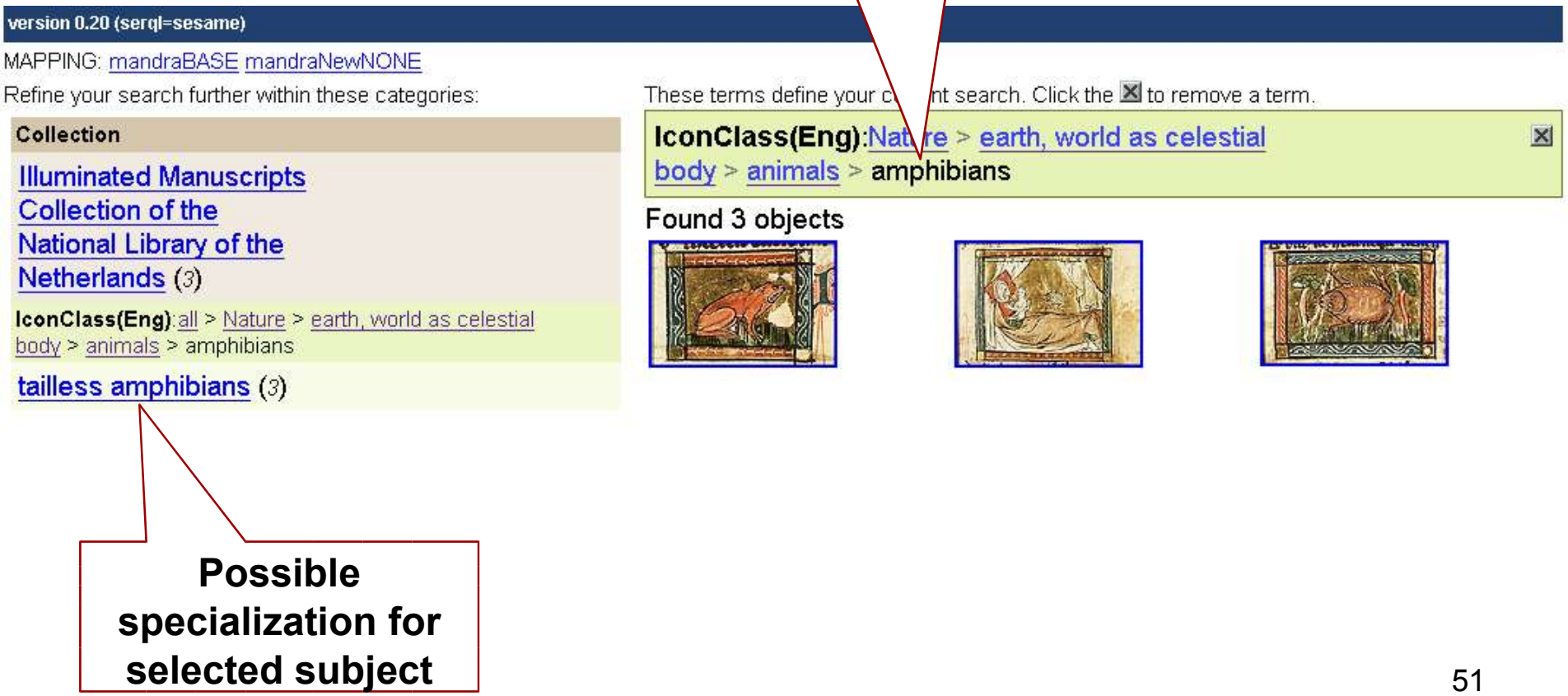
Back

Demo: SKOS, browsing and alignment



Demo: SKOS, browsing and alignment

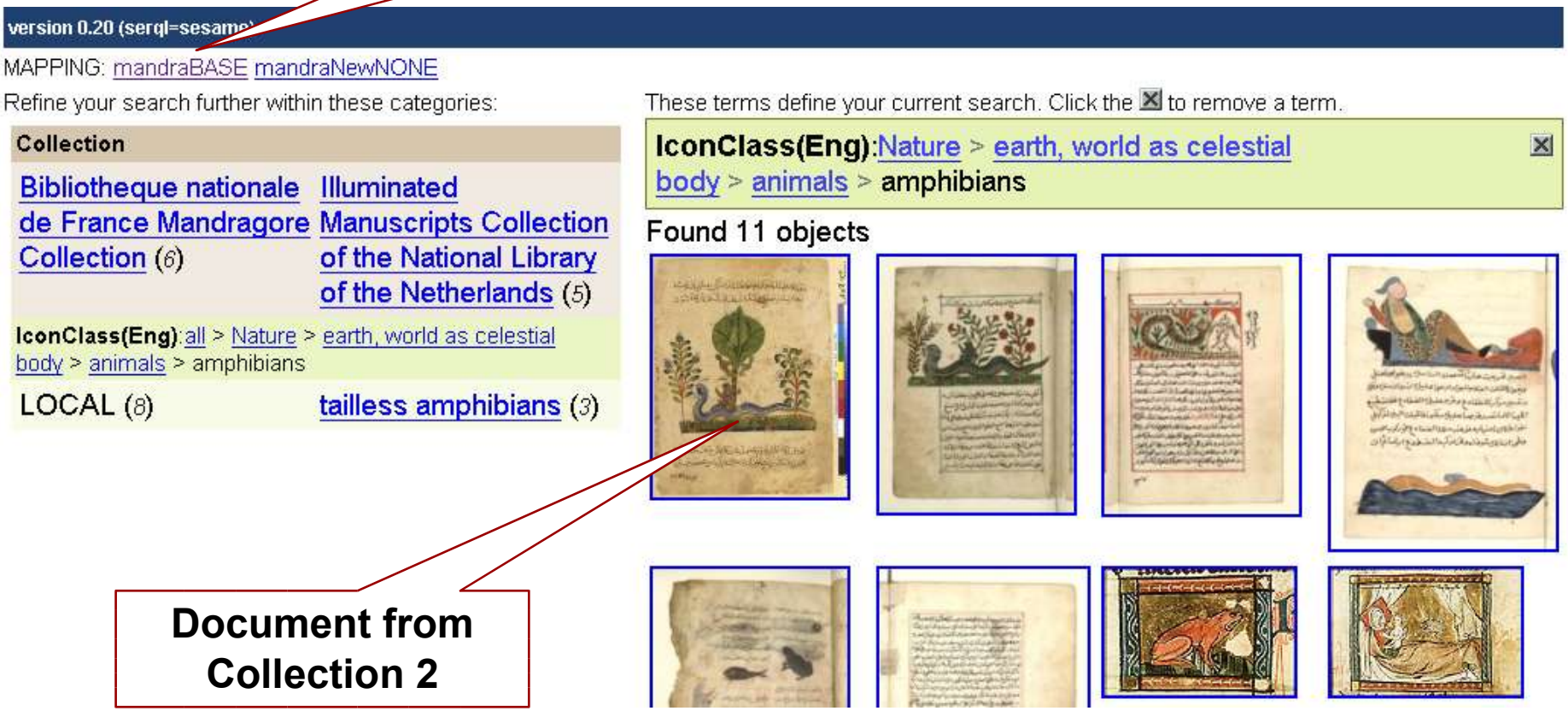
Hierarchical path from root to selected subject



Possible specialization for selected subject

Demo: SKOS, browsing and alignment

Semantic alignment of subjects activated



Document from Collection 2

Demo: SKOS, browsing and alignment

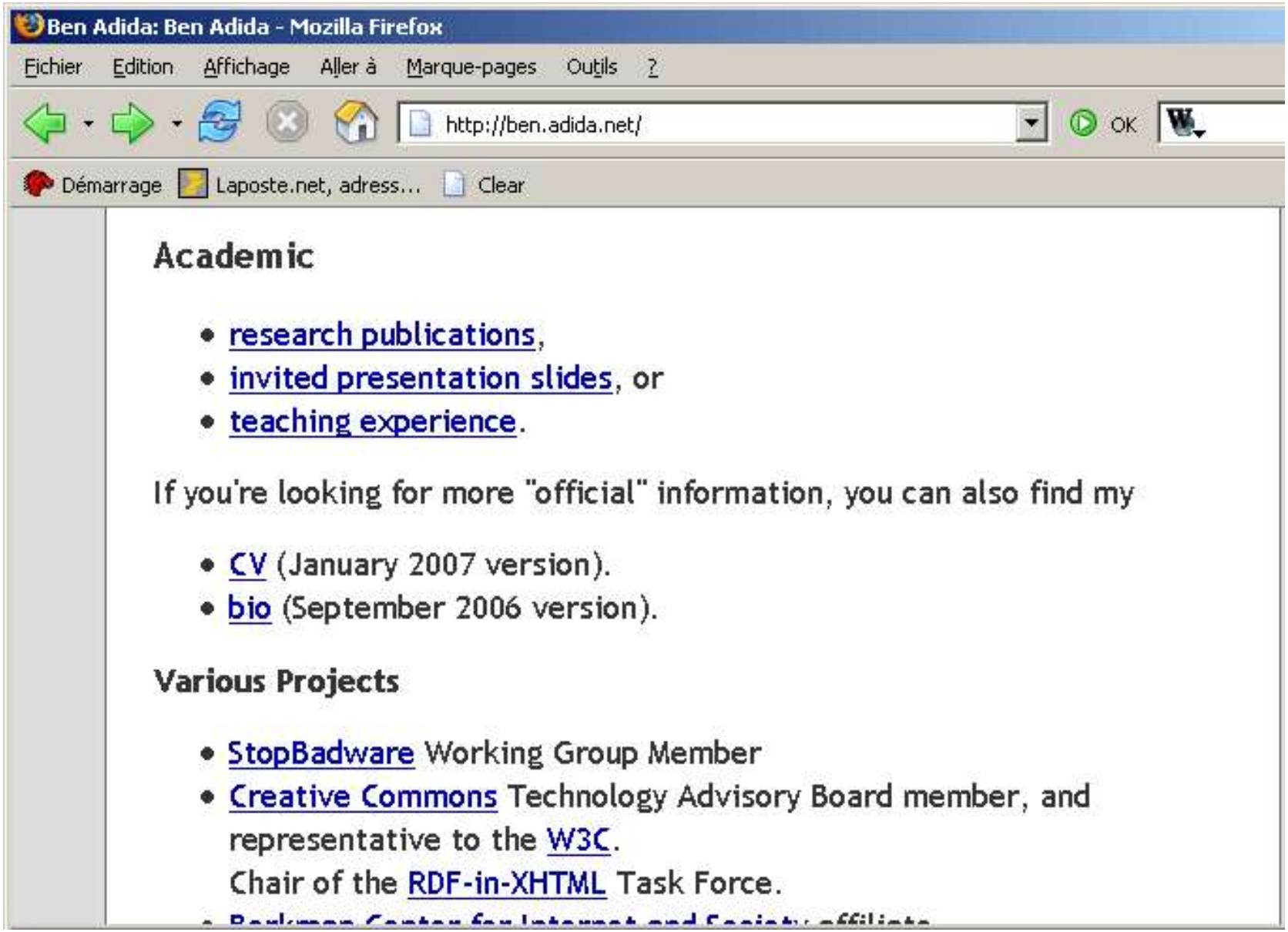
version 0.20 (serql=sesame)

Title:	Fable : le cobra et les grenouilles
Picture:	
Image taille réelle:	http://visualiseur.bnf.fr/Visualiseur?Destination=Mandrag
MANDRAGORE:	plante
MANDRAGORE:	naja
MANDRAGORE:	grenouille
MANDRAGORE:	fable
MANDRAGORE:	arbre
DATE:	13e siècle

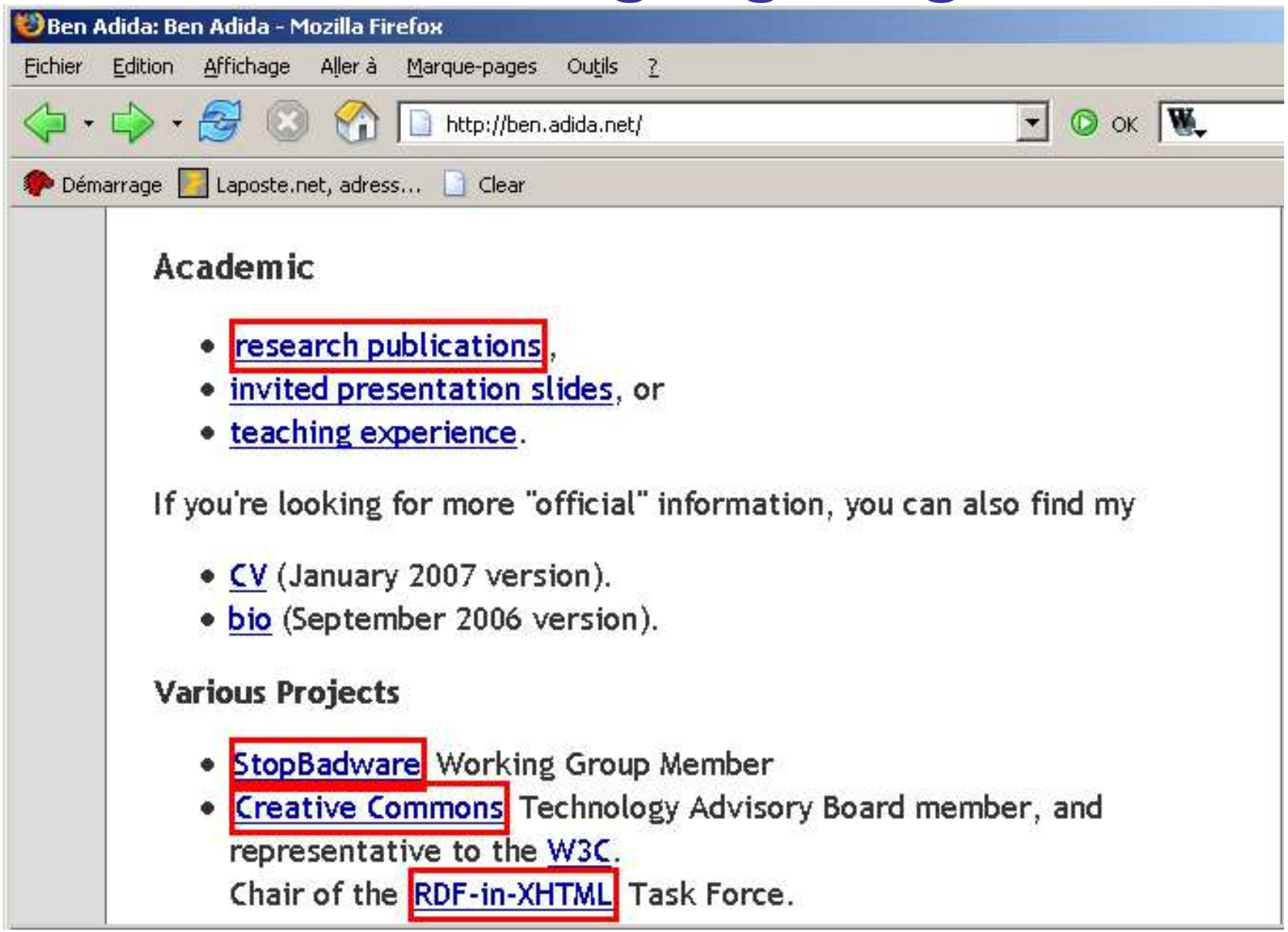
Subject from voc2 aligned to voc1:amphibians"

Back

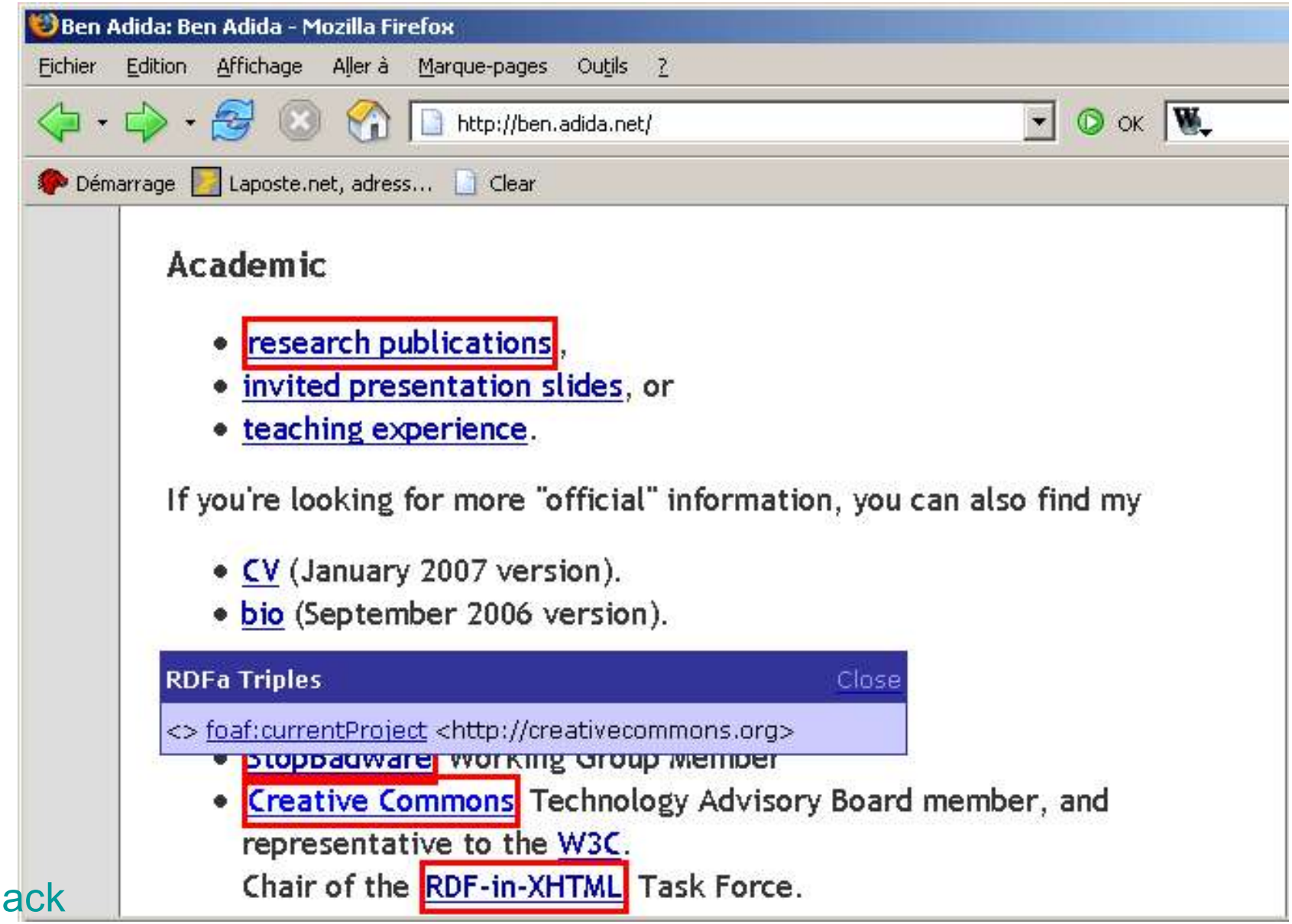
RDFa demo: a page with RDFa



RDFa demo: highlighting RDFa



RDFa demo: displaying triples



Back