The Implementation of the Thesaurus of Greek Terms of the National Documentation Centre in SKOS

Georgia Solomou

Dimitrios Koutsomitropoulos

High Performance Information Systems Laboratory
University of Patras

Thesaurus of Greek Terms

▸ Published by the National Documentation Centre of Greece (EKT)

▸ The **first** published thesaurus in Greek

▸ Controlled vocabulary covering general terminology

▸ Aiming at use/exploitation by Hellenic Libraries and Information Centres

▸ Bilingual terms (Greek, English)

▸ Representation of both vertical (hierarchical) and horizontal associations between concepts

Terminology

- Multi-subject thesaurus
  - Aims at covering subjects for all Hellenic Libraries (various scientific fields)
- Based on:
  - the National Library of Greece subject headings
  - subject terms used by other Hellenic Libraries

  ⇨ Focusing on issues of Greek and European interest

  ⇨ Expanded with medical terminology

  ⇨ Working on expansions on more scientific fields

- Terms specialization based on the Greek translation of the 13th version of the Dewey Decimal Classification system

Standards

- Follows international standards like:
  - ◦ ISO 2788 Documentation - guidelines for the establishment and development of monolingual thesauri, 1986
    - • ΕΛΟΤ 1321:1993 (in Greek)
  - ◦ ISO 5964 Documentation - guidelines for the establishment and development of multilingual thesauri, 1985
- Other established standards for determining subjects and indexing terms were studied (e.g. ΕΛΟΤ 1312)

4

HPCLab
University of Patras

Exploitation

- A number of Public and Academic Libraries in Greece exploit the EKT thesaurus
  - e.g. **Thir@ -** Portal of the Library and Information Centre of the University of Macedonia
- Incorporation in the Hellenic Public Libraries Union Catalogue
  - A unified catalogue comprising titles from all Hellenic Public Libraries
  - Provided search service powered by EKT

Title:	**Country Insights**
Url:	http://www.intercultures.ca/cil-cai/country_insights-en.asp?lvl=8
Creator:	Centre for Intercultural Learning
Publisher:	Centre for Intercultural Learning

Description: Country facts and cultural information on the countries of the world.

en_tree:	Society and Culture
	Society and Culture -- Studies in Human Society
	Society and Culture -- Other Society and Culture
	Society and Culture -- Studies in Human Society -- Sociology
	Society and Culture -- Studies in Human Society -- Anthropology
	Society and Culture -- Studies in Human Society -- History
	Society and Culture -- Studies in Human Society -- Human Geography
	Society and Culture -- Studies in Human Society -- Indigenous Studies
	Society and Culture -- Studies in Human Society -- Studies in Human Society, n.e.c.
	Society and Culture -- Other Society and Culture -- Society and Culture, n.e.c.
Resource Type:	Data bases
Mediator:	Free
Audience:	Undergraduate
	Postgraduate
	Researchers
	Faculty
	General Interest
Language:	English
	French
Format:	Full text Resource
Chronological Coverage:	Μη διαθέσιμη
Geographical Coverage:	Διεθνής
Subject:	Κοινωνική Ανθρωπολογία
	Cultural and Social Anthropology
	Social and Cultural Anthropology
	Cross Cultural Psychology
	Διαπολιτισμική Ψυχολογία
	Folk Culture
	Λαϊκός Πολιτισμός
	Πολιτισμός
	Civilization
	Πολιτιστική Κοινωνιολογία
	Cultural Sociology
	Πολιτισμός
	Civilization
	Πολιτιστική Κοινωνιολογία
	Cultural Sociology

A record in "Thir@"

HPCLab
University of Patras
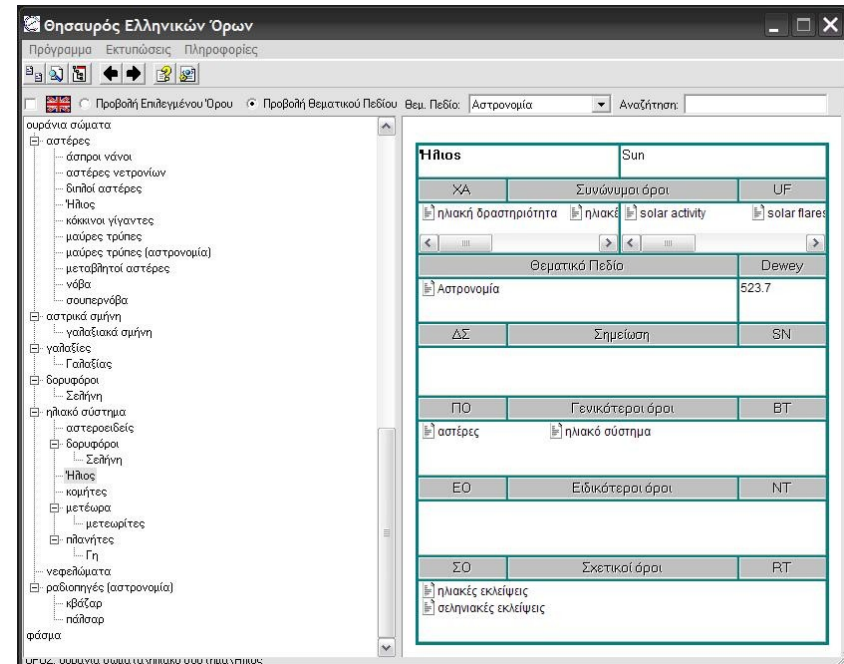
Implementations

▶ CD-ROM (discontinued)

▶ XML implementation
  ◦ Export from the CD-ROM

▶ On-line version (definitive)

Freely accessible at
http://public.ekt.gr/pls/thesaurus/the
saurus.main
(service available only in Greek)

EKT thesaurus in CD-ROM

EKT Thesaurus On-line

Structure

- Based on a common subject thesaurus structure (enriched with extensions):
  - Expresses three types of relationships between terms:
    - **Hierarchical** (**B**roader/**N**arrower /**M**icrothesauri **T**erm – **BT, NT, MT**)
    - **Associative** (**R**elated **T**erm - **RT**)
    - **Equivalence** for synonyms (SYN) (**U**sed **F**or - **UF**)
  - English translations included (**E**nglish **T**erm - **ET**) (some derived from Eurovoc Thesaurus)
    - ⇨ Bilingual terms
  - Correspondence to the **D**ewey **D**ecimal **C**lassification system (**DDC**)

Term Structure

XML format   **on-line form**

```
<TERM>
    <CONTEXT>αστικά δικαστήρια</CONTEXT>
    <USER>EKT</USER>
    <MT>Νομικές Επιστήμες</MT>
    <ET>civil courts</ET>
    <BT>δικαστήρια</BT>
    <NT>Άρειος Πάγος<NT>
    <NT>ειρηνοδικεία<NT>
    <NT>εφετεία<NT>
    <NT>πρωτοδικεία<NT>
    <UF>βλ. πολιτικά δικαστήρια</UF>
    <RT>πολιτική δικονομία</RT>
    <RT>αστικό δίκαιο</RT>
    <SN>some description</SN>
    <dewey>347</dewey>
</TERM>
```

Σχέσεις όρου		
αστικά δικαστήρια		
ET	1.	civil courts
DDC	1.	347
BT	1.	δικαστήρια
NT	1.	Άρειος Πάγος
	2.	ειρηνοδικεία
	3.	εφετεία
	4.	πρωτοδικεία
RT	1.	αστικό δίκαιο
		SYN: κοινό δίκαιο
	2.	πολιτική δικονομία
		SYN: αστικό δικονομικό δίκαιο
SYN	1.	πολιτικά δικαστήρια

About SKOS

- A W3C Proposed Recommendation (June 2009)
- A formal language for representing any type of structured, controlled vocabularies
  - A practical application of RDF (and RDFS)
  - ⇒ A **machine-understandable** representation framework for Knowledge Organization Systems (KOS)

W3C Document Process

Working Draft

Candidate Recommendation

Proposed Recommendation

Recommendation

HPCLab
University of Patras
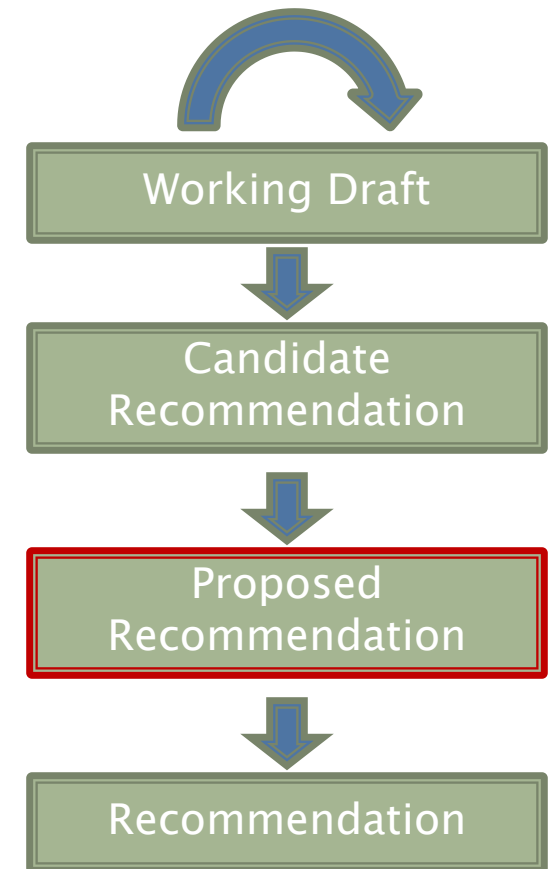
About SKOS

- Aims at bringing KOS in the distributed and decentralized environment of the World Wide *and Semantic* Web
  - An application of the library science to the Semantic Web
- A KOS is represented as a **concept scheme** comprising of a set of **conceptual resources (concepts)**
  - Description of concepts (identified by URIs)
  - Hierarchical and associative relations
- Multilingual and extensible model

The SKOSification of the EKT thesaurus

- Manually map XML elements of thesaurus to SKOS notions
- Creation of the corresponding XSL Transformation
- Application of XSLT document to XML file

| XML | → | XSL Transformation | → | RDF/XML |

provided XML format

↑

Map XML elements to SKOS

produced SKOS format

HPCLab
University of Patras
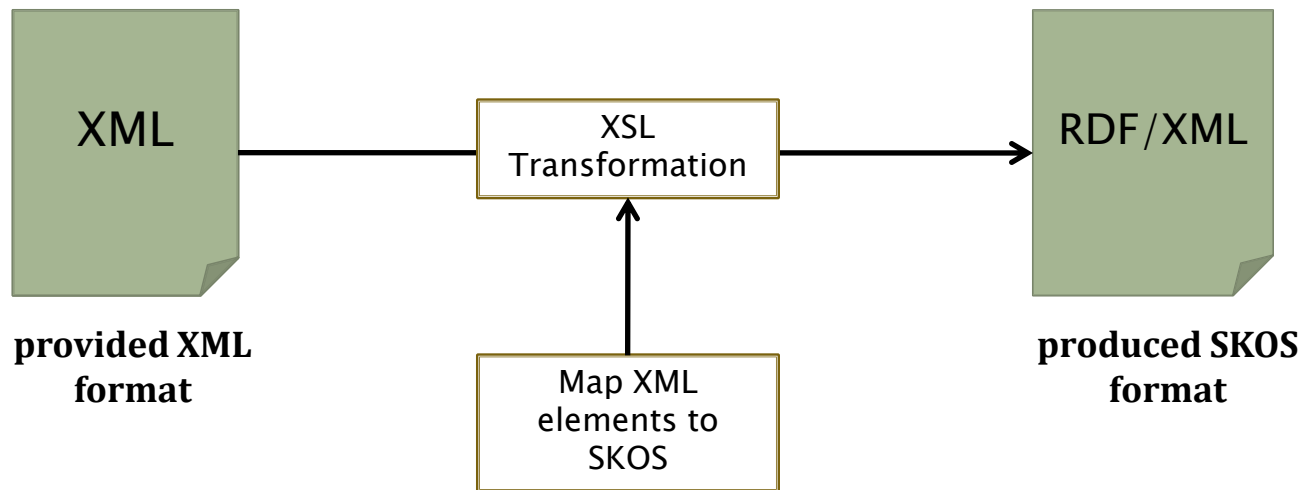
EKT XML Format Description

<TERM> - all information about a term is contained in a term element

<CONTEXT> - the name of the term

<USER> - a fixed value ("EKT")

<MT> - a broader category to which this term belongs

<ET> - term's English translation (may repeat)

<BT> - refers to a term with broader meaning (may repeat)

<NT> - refers to a term with narrower meaning (may repeat)

<UF> - a synonym (may repeat)

<RT> - refers to a term with related meaning (may repeat)

<SN> - description of the term (may repeat)

<dewey> - correspondence to dewey decimal classification system (may repeat)

The Thesaurus as a concept scheme

▸ The EKT Thesaurus is considered as a SKOS *concept scheme (an aggregation of* concepts*)*

- A concept scheme resource is described using Dublin Core (DC) metadata
- `skos:topConcept` links a concept scheme to the most general concepts it contains
- It is linked with the concepts it contains using the skos:inScheme property
- (e.g. `ekt:conceptA skos:inScheme ekt:thesaurusA`)

```
<rdf:RDF>
  <skos:ConceptScheme rdf:about="http://www.hpclab.ceid.upatras.gr/skos/EKT_THESAURUS">
    <dc:title>Thesaurus of Greek Terms</dc:title>
    <dc:description>
      The Thesaurus of Greek Terms, published by the National Documentation Centre (EKT) in Greece, in SKOS format
    </dc:description>
    <dc:creator>
      High_Performance Information Systems Laboratory (HPCLab)
    </dc:creator>
    <dc:contributor>
      High Performance Information Systems Laboratory (HPCLab)
    </dc:contributor>
    <dc:date>2009-01-10</dc:date>
    <dc:language>el,en</dc:language>
    <skos:hasTopConcept>Στρατός</skos:hasTopConcept>
  </skos:ConceptScheme>
```

HPCLab
University of Patras

Mapping of XML elements to SKOS notions

- **`<TERM>`** - becomes a SKOS **concept**

  - A base URI is assigned:

    `http://www.hpclab.ceid.upatras.gr/skos/ekt#`*`context`*

  - Concept's URI is formulated by the base URI plus `<context>` value:

  `<skos:Concept rdf:about="URI+`*`CONTEXT_value`*`"></skos:Concept>`

- **`<USER>`** - indicates creator/holder (EKT)

  - Expressed in the concept scheme's metadata

- **`<CONTEXT>`** - constitutes the preferred label in Greek ⇨

  `<skos:prefLabel xml:lang="el">`*`CONTEXT_value`*`</skos:prefLabel>`

HPCLab
University of Patras

Mapping of XML elements to SKOS notions (2)

- **<MT> - Microthesauri Term:** refers to a top (broader) term
  - ◦ Top concept in the EKT concept scheme ⇨

    `<skos:hasTopConcept>MT_value</skos:hasTopConcept>`
  - ◦ Expressed using an indirect hierarchical (broader) link to the described term ⇨

    `<skos:broaderTransitive rdf:resource="URI+MT_value"/>`

- **<ET> - English Term** used for the term's English translation ⇨

  `<skos:prefLabel xml:lang="en">ET_value</skos:prefLabel>`

  If more than one **<et>** terms exist:

  `<skos:altLabel xml:lang="en">ET_value</skos:altLabel>`

HPCLab
University of Patras

Mapping of XML elements to SKOS notions (3)

(Hierarchical relationships)

- **<BT> - Broader Term:** asserts a direct (broader) link between the described term and the <bt> value (term) ⇨

  `<skos:broader rdf:resource="URI+`*`BT_value`*`"/>`

- **<NT> - Narrower Term:** asserts a direct (narrower) link between the described term and the <nt> value (term) ⇨

  `<skos:narrower rdf:resource="URI+`*`NT_value`*`"/>`

(Associative relationships)

- **<RT> - Related Term:** refers to a term that relates to the described one ⇨

  `<skos:related rdf:resource="URI+`*`RT_value`*`"/>`

HPCLab
University of Patras

Mapping of XML elements to SKOS notions (4)

(Equivalence)

▸ **<UF> - Used For**: indicates a (quasi-)synonym ⇨

  <skos:altLabel lang="el"> URI+*MT_value* </skos:altLabel>

(Documentation)

▸ **<SN> - Scope Note:** explanation of the term's intended meaning ⇨

  Corresponds to `<skos:Note>`

  • Subproperties: `<skos:scopeNote>,<skos:example>,` `<skos:historyNote>,<skos:editorialNote>,` `<skos:changeNote>,`**`<skos:definition>`**

  `<skos:definition xml:lang="el">`*SN_value*`</skos:definition>`

Mapping of XML elements to SKOS notions (5)

- **`<DEWEY>`** - A unique identifier based on the Dewey Decimal Classification system

  - Represented by **`<skos:notation>`**:

    - used with a typed literal

    - datatype URI assigned:

      http://dewey.info/schema-terms/Notation

  `<skos:notation rdf:datatype="http://dewey.info/schema-terms/Notation">387.733 48</skos:notation>`

HPCLab
University of Patras

Mapping Summary

XML element	SKOS notion
`<TERM>`	`<skos:Concept>`
`<USER>`	-
`<CONTEXT>`	`<skos:prefLabel lang="el">`
`<MT>`	`<skos:broaderTransitive>`
`<ET>` (first)	`<skos:prefLabel lang="en">`
`<ET>`	`<skos:altLabel lang="en">`
`<BT>`	`<skos:broader>`
`<NT>`	`<skos:narrower>`
`<RT>`	`<skos:related>`
`<UF>`	`<skos:altLabel lang="el">`
`<SN>`	`<skos:definition>`
`<DEWEY>`	`<skos:notation>`

HPCLab
University of Patras

Problems with the EKT theasurus

▸ Inconsistency among hierarchical/associative relations
  ◦ identical values for
    • the `<bt>` and `<rt>` node, or
    • the `<nt>` and `<rt>` node ...

    in the same `<term>` !

▸ Cause
  ◦ both `skos:broader` and `skos:narrower` properties are disjoint with `skos:related`

HPCLab
University of Patras

Example and Solution

⇨ *Solution*

- only one type of relation was kept
  - According to the description of the problematic term (referred as broader or narrower)
- Manually fixed 15 terms

Example

Term: government
BT: central administration
RT: central administration

Term: central administration
NT: government
RT: (no related term "government")
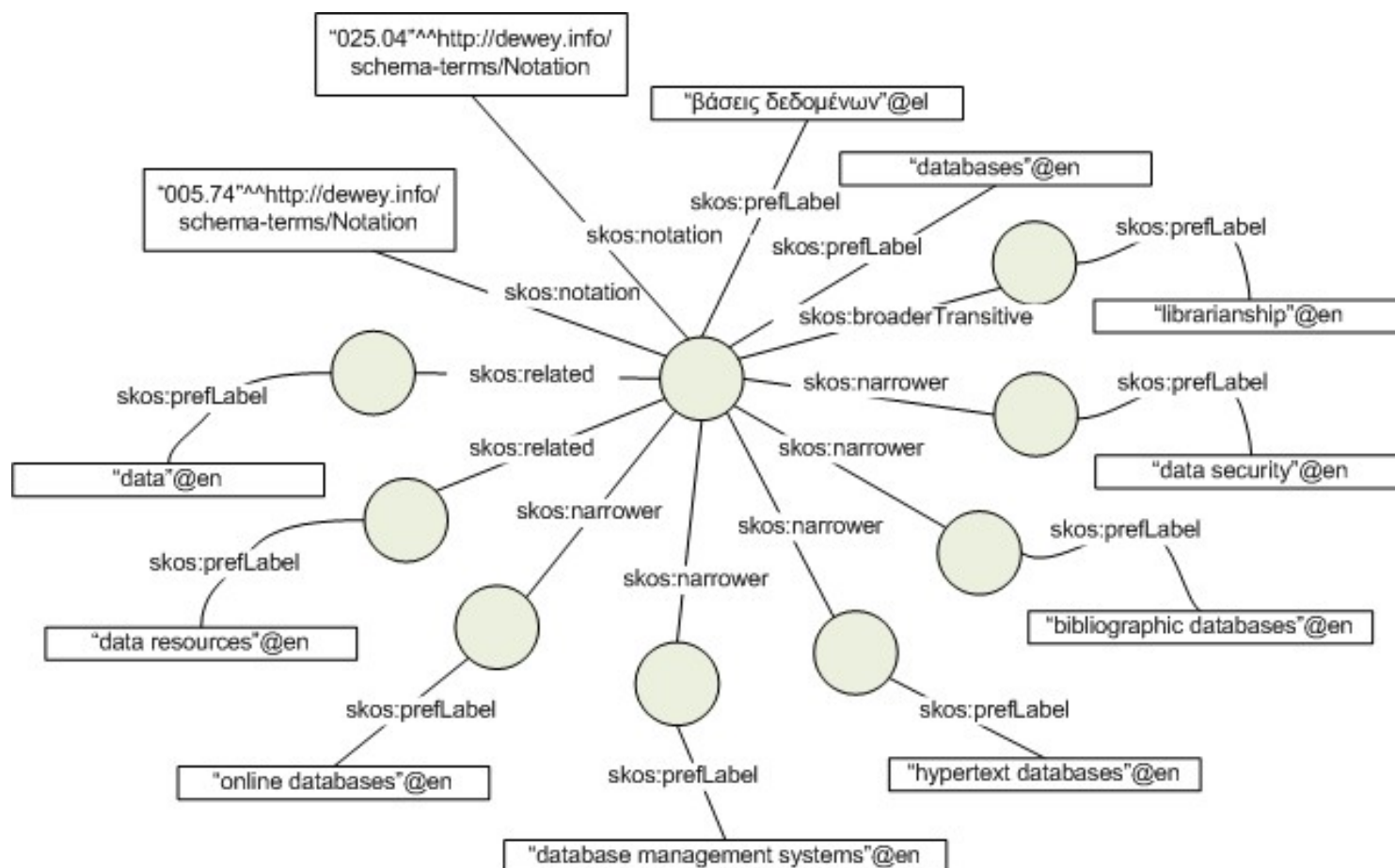
deletion of **RT** "central administration"
from term "government"

HPCLab
University of Patras

The RDF/XML Format of the SKOS Theasurus

Term: "databases"@en
            "βάσεις_δεδομένων"@el

```
<skos:Concept rdf:about="http://www.hpclab.ceid.upatras.gr/skos/ekt#βάσεις_δεδομένων">
   <skos:prefLabel xml:lang="el">βάσεις δεδομένων</skos:prefLabel>
   <skos:prefLabel xml:lang="en">databases</skos:prefLabel>
   <skos:broaderTransitive rdf:resource="http://www.hpclab.ceid.upatras.gr/skos/ekt#Βιβλιοθηκονομία"/>
   <skos:narrower rdf:resource="http://www.hpclab.ceid.upatras.gr/skos/ekt#ασφάλεια_δεδομένων"/>
   <skos:narrower rdf:resource="http://www.hpclab.ceid.upatras.gr/skos/ekt#βιβλιογραφικές_βάσεις_δεδομένων"/>
   <skos:narrower rdf:resource="http://www.hpclab.ceid.upatras.gr/skos/ekt#συστήματα_διαχείρισης_βάσεων_δεδομένων"/>
   <skos:narrower rdf:resource="http://www.hpclab.ceid.upatras.gr/skos/ekt#βάσεις_δεδομένων_υπερκειμένου"/>
   <skos:narrower rdf:resource="http://www.hpclab.ceid.upatras.gr/skos/ekt#επιγραμμικές_βάσεις_δεδομένων"/>
   <skos:related rdf:resource="http://www.hpclab.ceid.upatras.gr/skos/ekt#πηγές_πληροφόρησης"/>
   <skos:related rdf:resource="http://www.hpclab.ceid.upatras.gr/skos/ekt#δεδομένα"/>
   <skos:notation rdf:datatype="http://dewey.info/schema-terms/Notation">005.74</skos:notation>
   <skos:notation rdf:datatype="http://dewey.info/schema-terms/Notation">025.04</skos:notation>
</skos:Concept>
```
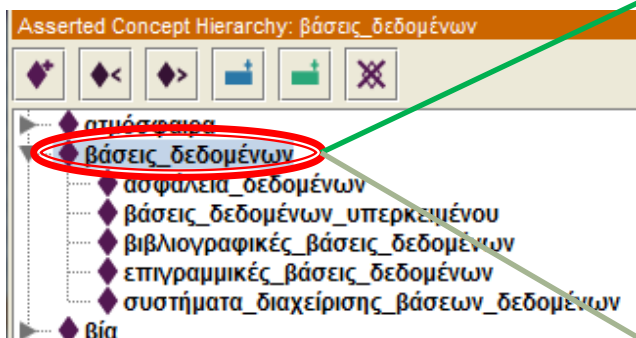
Visualization of a Term

The EKT thesaurus through Protégé 4 and SKOSed plugin

Asserted data properties *(SKOS labels, etc)*

SKOS Data Property Assertions: βάσεις_δεδομένων

SKOS preferred label ⊕

- prefLabel "βάσεις δεδομένων"@el
- prefLabel "databases"@en

SKOS notation ⊕

- notation "005.74"^^http://dewey.info/schema-terms/Notation
- notation "025.04"^^http://dewey.info/schema-terms/Notation

Concept Hierarchy for the term "Βάσεις Δεδομένων"@el (*Databases)*

Asserted Concept Hierarchy: βάσεις_δεδομένων

- ατμόσφαιρα
- βάσεις_δεδομένων
  - ασφάλεια_δεδομένων
  - βάσεις_δεδομένων_υπερκειμένου
  - βιβλιογραφικές_βάσεις_δεδομένων
  - επιγραμμικές_βάσεις_δεδομένων
  - συστήματα_διαχείρισης_βάσεων_δεδομένων
- Βία

Asserted object properties *(broader, narrower…)*

Other property assertions ⊕

- narrower ασφάλεια_δεδομένων
- narrower βιβλιογραφικές_βάσεις_δεδομένων
- narrower επιγραμμικές_βάσεις_δεδομένων
- narrower βάσεις_δεδομένων_υπερκειμένου
- broaderTransitive Βιβλιοθηκονομία
- narrower συστήματα_διαχείρισης_βάσεων_δεδομένων

HPCLab
University of Patras
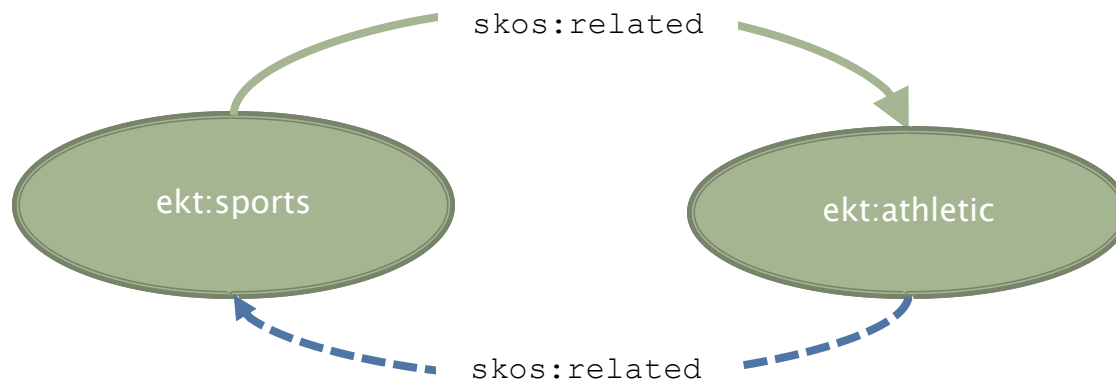
Worth SKOSifying?

- ▸ It is a W3C Proposed Recommendation (June 2009)
  ⇨ (almost) a **standard**!
- ▸ A data model producing descriptions in a **machine readable** format
- ▸ A means to achieve **interoperability**

  "*SKOS... enables data and technology sharing across diverse applications*"

- ▸ SKOS data model implemented in OWL
  - ◦ Rich semantic relations, constructs and *axioms*
  - ◦ Provide for **reasoning** and **inferences**
  - ◦ Brings KOS to the Semantic Web reality

HPCLab
University of Patras

OWL axioms for SKOS

- `skos:broader`(`skos:broaderTransitive`) **is inverse of** `skos:narrower`(`skos:narrowerTransitive`)

- `skos:topConceptOf` **is inverse of** `skos:hasTopConcept`

- `skos:broaderTransitive/skos:narrowerTransitive` **are transitive**

- `skos:broader`(`skos:narrower`) **is subproperty of** `skos:broaderTransitive`(`skos:narrowerTransitive`)

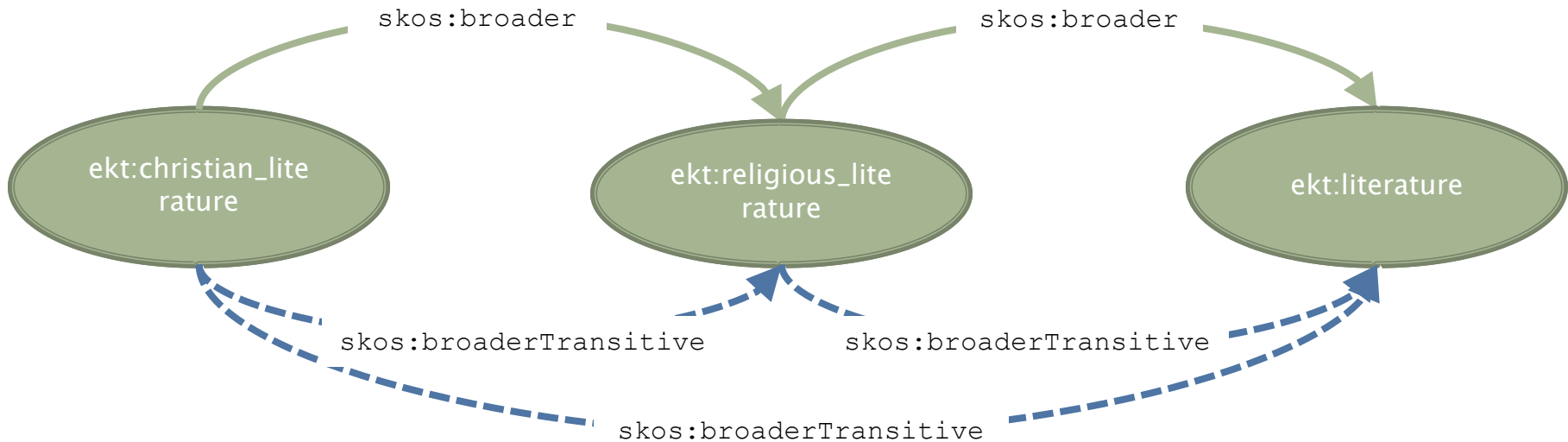- `skos:related` **is symmetric**

Inference Example

Example 1 –symmetric relation `skos:related`

Inference example (2)

Example 2 – transitive relation `skos:broaderTransitive`

Worth SKOSifying?

- A widely adopted standard

  e.g. Library of Congress Subject Headings are expressed in SKOS

  (on-line at http://id.loc.gov/authorities/)

Library of Congress in SKOS

Term: "Ad hoc networks (Computer networks)"@en

Ad hoc networks (Computer networks)

URI: <http://id.loc.gov/authorities/sh2007004723#concept>

Type: Topical Term

Alternate Labels: MANETs (Computer networks); Mobile ad hoc networks; Wireless ad hoc networks

Broader Terms:

- Computer networks
- Wireless communication systems

Narrower Terms:

- Cognitive radio networks
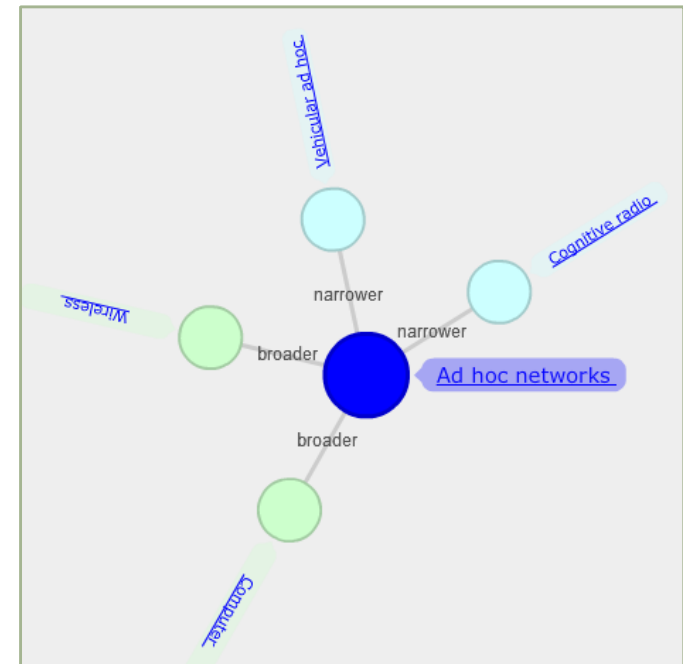- Vehicular ad hoc networks (Computer networks)

Sources:

- Work cat.: 2007026336: Policy-driven ad hoc network management, c2008: ECIP galley (Mobile ad hoc networks; MANETs)
- EI, viewed June 28, 2007 (uncontrolled identifiers: Mobile ad hoc networks; MANET)
- Applied Science, viewed June 28, 2007 (Ad hoc networks)
- Inspec, viewed June 28, 2007 (controlled indexing: Ad hoc networks)
- LC data base, June 28, 2007 (in titles: Wireless ad hoc networks; MANETs; Mobile ad hoc networks; Ad hoc networks)

LC Classification: TK5105.77

Created: 2007-07-18

Last Modified: 2007-07-19 07:54:25

Alternate Formats: RDF/XML, N-Triples, JSON

HPCLab
University of Patras

Library of Congress in SKOS

Term: "Ad hoc networks (Computer networks)"@en

```xml
- <rdf:RDF>
  - <rdf:Description rdf:about="http://id.loc.gov/authorities/sh2007004723#concept">
      <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2007-07-19T07:54:25-04:00</dcterms:modified>
      <skos:narrower rdf:resource="http://id.loc.gov/authorities/sh2008006628#concept"/>
      <skos:narrower rdf:resource="http://id.loc.gov/authorities/sh2008002568#concept"/>
      <skos:altLabel xml:lang="en">Mobile ad hoc networks</skos:altLabel>
      <skos:altLabel xml:lang="en">MANETs (Computer networks)</skos:altLabel>
      <skos:altLabel xml:lang="en">Wireless ad hoc networks</skos:altLabel>
      <owl:sameAs rdf:resource="info:lc/authorities/sh2007004723"/>
    - <dcterms:source xml:lang="en">
        Work cat.: 2007026336: Policy-driven ad hoc network management, c2008: ECIP galley (Mobile ad hoc networks; MANETs)
      </dcterms:source>
    - <dcterms:source xml:lang="en">
        Applied Science, viewed June 28, 2007 (Ad hoc networks)
      </dcterms:source>
    - <dcterms:source xml:lang="en">
        Inspec, viewed June 28, 2007 (controlled indexing: Ad hoc networks)
      </dcterms:source>
    - <dcterms:source xml:lang="en">
        LC data base, June 28, 2007 (in titles: Wireless ad hoc networks; MANETs; Mobile ad hoc networks; Ad hoc networks)
      </dcterms:source>
    - <dcterms:source xml:lang="en">
        EI, viewed June 28, 2007 (uncontrolled identifiers: Mobile ad hoc networks; MANET)
      </dcterms:source>
      <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
      <skos:inScheme rdf:resource="http://id.loc.gov/authorities#topicalTerms"/>
      <skos:inScheme rdf:resource="http://id.loc.gov/authorities#conceptScheme"/>
      <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2007-07-18T00:00:00-04:00</dcterms:created>
      <skos:broader rdf:resource="http://id.loc.gov/authorities/sh85029513#concept"/>
      <skos:broader rdf:resource="http://id.loc.gov/authorities/sh92006740#concept"/>
      <skos:prefLabel xml:lang="en">Ad hoc networks (Computer networks)</skos:prefLabel>
  </rdf:Description>
```

About SKOS and OWL 2

- The SKOS is itself defined as an OWL ontology
  - ◦ Each SKOS vocabulary is an instantiation of the SKOS ontology/schema (`skos.rdf`)
- SKOS in OWL 1.0 ➪ produces OWL Full ontologies
  - ▪ Metamodeling
  - ▪ Allows for hierarchies of annotation properties
- (SKOS in OWL DL ➪ a proposal comprising a subset of the original OWL vocabulary)

HPCLab
University of Patras

About SKOS and OWL 2

- Important Issues
  1. There are still a number of semantic conditions that are not expressed formally in this schema, e.g.
     - `skos:prefLabel`, `skos:altLabel` and `skos:hiddenLabel` are pairwise disjoint properties
  2. Some semantic conditions cause SKOS ontology to be OWL Full , e.g.
     - `skos:changeNote`, `skos:definition`, … are each sub-properties of `skos:note` (hierarchy of annotation properties)

- Solution

  Using constructs allowed only in OWL 2 such as:
  - role-chains
  - `DisjointObjectProperties, subAnnotationPropertyOf`
  - *Punning*

  Attempts at creating OWL2 axioms for SKOS
  > http://apollo.hpclab.ceid.upatras.gr:8001/hpclabwiki/Skos2Owl2
  > http://lists.w3.org/Archives/Public/public-swd-wg/2009Mar/0043.html

  Relevant discussion between the OWL WG and the SKOS WG
  > http://lists.w3.org/Archives/Public/public-swd-wg/2008Oct/0059.html

HPCLab
University of Patras

Raising Awareness

SKOSified version of the EKT Thesaurus:

▸ Already announced to the SKOS WG mailing list

Publicly accessible at:

http://werg.hpclab.ceid.upatras.gr:8000/ekt/ekt_to_skos.rdf

▸ Included in the on-going SKOS implementation report (April 2009)

http://www.w3.org/2006/07/SWD/SKOS/reference/20090315/implementation.html

Implementations

Issue	Type	URL	Notes
194: UMBEL Vocabulary	Vocabulary	http://umbel.org/documentation.html	
195: UMBEL Web Services	Service	http://umbel.zitgist.com/	
196: STW Thesaurus for Economics	Vocabulary	http://zbw.eu/stw	
197: The Thesaurus of Greek Terms in SKOS	Vocabulary	http://public.ekt.gr/pls/thesaurus/thesaurus.main http://www.ekt.gr/products/thesaurus/	
198: STW Thesaurus for Economics Service	Service	http://zbw.eu/beta/stw-ws	

HPCLab
University of Patras