UPDATE ON TOPIC MAP AND LINGUISTIC STANDARDS.

DAVID GARCÍA MARTUL
UNIVERSITY CARLOS III
1. WHAT’S AN INDEX?
2. KNOWLEDGE ORGANIZATION SYSTEMS IN DIGITAL LIBRARIES
3. LINGUISTIC ONTOLOGIES
4. KNOWLEDGE MANAGEMENT ONTOLOGIES: TOPIC MAPS.
5. IDENTITIES IN SUBJECTS
6. SEMANTIC RELATIONSHIPS
7. TOPIC MAPS: A CALL CENTRE TAX MAP.
8. TOPIC MAPS: MERGING OR COMMUNICATION.
9. CONCLUSIONS

NKOS 2004 UPDATE ON TOPIC MAP AND LINGUISTIC STANDARDS.
WHAT'S AN INDEX?

- TOPICS
- TOPIC NAMES
- TOPIC CLASSES
- OCCURRENCES
- ASSOCIATION

NKOS 2004
SUBJECT BASED CLASSIFICATION: CONTROLLED VOCABULARIES

LIST OF TERMS SORTED ALPHABETICALLY WITHOUT RELATIONSHIPS BETWEEN EACH ITEM.
SUBJECT BASED CLASSIFICATION: TAXONOMIES

- TAXONOMIES:
  LIST OF TERMS
  SORTED BY
  HIERARCHICAL RELATIONSHIPS

**BUBL LINK** Catalogue of selected Internet resources

Home | Search | Subject Menus | A-Z | Diary | Countries | Types | Updates | Random | About

400 Language

- 400 Language: general resources
- 401 Language: philosophy and theory
- 402 Language and language acquisition
- 403 Language: education and research
- 409 Languages spoken around the world
- 410 Linguistics
- 411 Writing systems
- 414 Phonology and phonetics
- 415 Grammar
- 419 Sign language
- 420 English language
- 421 German language
- 429 31 Dutch language
- 439 7 Swedish language
- 439 02 Norwegian language
- 440 French language
- 449 Italian language
- 452 Polish language
- 459 Spanish language
- 462 Portuguese language
- 470 Latin language
- 469 Greek language
- 469 Other languages

NKOS 2004

UPDATE ON TOPIC MAP AND LINGUISTIC STANDARDS.
SUBJECT BASED CLASSIFICATION: THESAURUS

1. HIERARCHICAL
2. EQUIVALENCE
3. ASSOCIATIVES

NKOS 2004
LEXICAL DATABASE composed by lexical unities and their relationships.

It organizes around a words collection with same root which can be interchanged in a specific context: SYNSET.

WordNet 1.7.1 Search

Search word: 
Find senses

Overview for "taxonomy"

The noun "taxonomy" has 3 senses in WordNet.

1. taxonomy -- (a classification of organisms into groups based on similarities of structure or origin etc)
2. taxonomy -- (biology) study of the general principles of scientific classification
3. taxonomy -- (practice of classifying plants and animals according to their presumed natural relationships)

Search for: Synonyms, ordered by estimated frequency of senses
Show glosses
Show contextual help
Search

Return to WordNet home

NKOS 2004
SEMANTIC NETWORKS: FRAMENET

- LEXICAL DATABASE which provides a corpus of sentences noted syntactically and semantically that it evokes their different means or scopes of one term.
- It’s a script-like structure of inferences linked to meanings of linguistic units.
- Each frame identifies a set of frame elements (FEs) which are frame specific semantic roles.
- The description of each lexical item identifies the frames which underlie a given meaning and the ways in which the (FEs) in structures headed by the word.
- It documents range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses, through manual annotation of example sentences and automatic summarization of the resulting annotations.
- It has been translated to RDF and OWL.

NKOS 2004
Criminal_process

Definition:

A Suspect is arrested by an Authority on certain Charges, then is assigned as a Defendant. If at any time the Defendant pleads guilty, then the Defendant is sentenced, otherwise the Defendant first goes to trial. If the Verdict after the trial is guilty, then the Defendant is sentenced. In the end, the Defendant is either released or is given a Sentence by a Judge at the sentencing. The core roles in this frame include Jury, Judge, Defendant, District Attorney, Defense Attorney, an Authority such as the police/sheriff. Non-core roles include Witness, Accomplice and Victim.

FEs:

Core:

- Charges [Chrg]: This FE identifies the Charges brought against the Defendant.
- Court [Ct]: This FE identifies the court involved in a trial.
- Defendant [Def]: The Defendant is charged with an offense.
- Defense [Def]: The Defense represents the interests of the Defendant.
- Judge [Judge]: The Judge heads the court where arraignment occurs and the case is tried.
- Jury [Jury]: The FEs identifies the people who are charged by the Court to listen to the testimony, evidence, and arguments, and come to a consensus about the guilt or innocence of the Defendant.
- Offense [Off]: This FE identifies the Offense which the Defendant is accused of committing.

Prosecution [Proc]: This FE identifies the attorney(s) prosecuting the Defendant.

Inherited From:

- Inherited By:
- Subframe of Crime Scenario:
- Has Subframes: Arraignment, Arrest, Sentencing, Trial
INDEX OF LEXICAL UNITS

- It is an index to alphabetical lists of the names of the lexical units (LUs).
- Each LU name is followed by the part of speech, the name of the relevant frame, and its status.
- The lexical entry provides two tables with information about the LU: Frame Elements and their Syntactic Realizations; and Valence Patterns.

NKOS 2004
LEXICAL ENTRY

- It is the classical entry of a dictionary
- It defines his morphological form
- It gives their frame elements, number of texts annotated with this word in each realization.
- Valence Patterns

**probe.n**

*Frame: Criminal investigation*

**Definition**

COD: an investigation.

<table>
<thead>
<tr>
<th>Frame Element</th>
<th>Number Annotated</th>
<th>Realization(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accused</td>
<td>8</td>
<td>N Mod 4 PP[into] Comp 4</td>
</tr>
<tr>
<td>Investigator</td>
<td>4</td>
<td>N Mod 2 PP[into] Comp 2</td>
</tr>
</tbody>
</table>

**Valence Patterns:**

These frame elements occur in the following syntactic patterns:

<table>
<thead>
<tr>
<th>Number Annotated</th>
<th>Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 TOTAL</td>
<td>resident investigator</td>
</tr>
<tr>
<td>1</td>
<td>N Mod PP[into] Comp</td>
</tr>
<tr>
<td>5 TOTAL</td>
<td>resident investigator</td>
</tr>
<tr>
<td>3</td>
<td>N Mod</td>
</tr>
<tr>
<td>2</td>
<td>PP[into] Comp</td>
</tr>
<tr>
<td>1 TOTAL</td>
<td>investigator</td>
</tr>
<tr>
<td>1</td>
<td>PP[into] N Mod</td>
</tr>
<tr>
<td>3 TOTAL</td>
<td>investigator</td>
</tr>
</tbody>
</table>

NKOS 2004
It acts as a dictionary and a thesaurus. Like a thesaurus words are linked to the semantics frames in which they participate, and frames are linked to wordlists and to related frames.

It provides all texts used for analyzing frame elements of each entry.

Each sense of a polysemous word belongs to a different semantic frame.

A script-like structure of inferences that characterize a type of situation, object or event.
NKOS 2004 UPDATE ON TOPIC MAP AND LINGUISTIC STANDARDS.

RELATIONSHIP FRAMENET & SEMANTIC WEB

Need to describe semantic relationships between different information resources for recalling by subject context in computers.

In a Domain of Knowledge we can implement recognizing of nuances with semantic networks but we need use ontologies implementation languages (OWL) for computers can do it.

NKOS 2004
STANDARD ISO/IEC 13250.
A set of information resources regarded by a topic map application as a bounded object set whose hub document is a topic map document conforming to the SGML architecture defined by this International Standard. Or it’s any topic map document conforming to the SGML architecture defined by this International Standard.

NKOS 2004
Subject Identity: To achieve a one-to-one relationship between topics and the subjects that they represent, in order to ensure that all knowledge about a particular subject can be accessed via a single topic.

When the subject is an addressable information resource, its identity may be established directly through its address. However, abstract subjects are not directly addressable. This problem is solved through the use of subject indicators.

Subject Indicator: It’s a resource that is intended to provide a positive, unambiguous indication of the identity of a subject. Because it’s a resource, a subject indicator has an address that can be used as a subject identifier.
A Published Subject Indicator is a subject indicator that is published and maintained at an advertised location for the purposes of supporting topic map interchange and mergeability. A published subject is any subject for which there exists at least one published subject indicator. A published subject identifier is the subject identifier of a published subject indicator.

This is a set of published subject indicators useful for identifying languages by URI. The identifiers are URIs based on the bibliographic three-letter codes in the ISO 639 standard (Codes for the representation of names of languages).
CURRENT STANDARD
Only accounts for explicit Equivalence relationships. Hierarchical relationship only allowed for genus-species relationship, with few exceptions. Associative relationship only allowed across categories.
TOPIC MAPS USE CASE: A CALL CENTRE TAX MAP

- The tax map was designed and implemented to increase the accuracy of information given to the public by optimizing access to relevant information for the call assistors.
- Tax map is an integrated navigation system that gives access to tax related information by topic.
- It also enables direct navigation between related topics, in addition to the traditional access to documents by document type and through tables of contents.

3 document types:
- Publications
- FAQs
- Tele-Tax Topics

Hierarchical navigation to documents.

Each pub has its own table of contents.
Topics originate from the collation of all indexed terms in the Tax Information Publications as well as from keywords used in the publications known as FAQ. Topics collated from FAQ’s are considered the most frequently used topics, and are given the type of “key topic”. The set of key topics has been enriched “by hand” by a team of tax experts. Key topics are accessible through an alphabetical index and topics in relation with forms are accessible through a specialized “Form topic” index.

Main Topics, enriched by a team of tax experts, are isolated from the total number of topics resulting from the integrated indexes of the publications as well as from the FAQs.

- **Main topics** are accessible through alphabetic indexes.
- **Form topics** are accessible through a specialized list.
- **All other topics** can be retrieved using the built-in search engine.

NKOS 2004
Each topic has its own web page which gives access to all places it occurs among the various document types. In addition each topic is linked to related topics, enabling access from one topic page to another.

When a topic has several names, they are preserved so that they can be accessed by their value in the index.

One name is chosen as the “main name” and each of the other names are displayed as synonyms on the topic page.
Occurrences of a topic are represented on its topic page and are linked to the relevant locations in the publications.

Occurrences are differently represented according to the publication type in which they are found.

The occurrences are grouped (scoped) according to the type of document in which they are found.
TOPIC MAPS REMOTE ACCESS PROTOCOL

- Service for communicating between topic maps applications
- It defines a set of operations for construction of integrated topic map portals
- It works with integration of topic pages, topic occurrences and topic map portal services.
- The key is subject-based organization of information
- It is expected that the improvement of additional services based on TMQL help the communication and networks of topic maps.

I need information about taxes in exporting products

* The question is:
What information do have about http://psi.tax.gov/taxexport in your system?

* The answer would be:
What information do have about http://psi.tax.nl/taxexport in your system?

I’ve this info

Portal A: TaxMap in USA

Portal B: TaxMap in Holland

NKOS 2004
CONCLUSIONS

1. Topic Maps are a Metaindex which properties are improved with interaction between distributed topic maps.

2. Framenets is a Lexicographyc System tied to last works in computational linguistics about terminological information retrieval and lexicography. It could provide a lexicographic database for automatic building of knowledge organization systems.

3. Framenets could help to disambiguation of topics in knowledge organization systems through treatment of marked corpora.

4. The bigger problem is the identity of subjects represented by PSI.

5. We will need measures of identity of subjects for communication distributed topic maps.

6. Progressive appearance of standards for communication and queries to different topic maps.

7. Development of topic maps will go tied to development of Portals and Digital Libraries.
THANK YOU VERY MUCH