Polythematic Structured Subject Heading System

Conversion and distribution in SKOS format

Jindřich Mynarz
j.mynarz@stk.cz
PSH/SKOS is a web application for browsing, searching and distributing PSH as linked data.
Our goals

• create a web interface for browsing and searching PSH
• offer PSH as linked data
• provide secondary access to State Technical Library collections
• link to other datasets
Chosen format - SKOS

Simple Knowledge Organization System (SKOS)
• simple format for representation, sharing and linking knowledge systems (classification schemes, thesauri or subject heading systems)
• built upon RDF and RDFS
• can be used in combination with OWL
• W3C candidate recommendation [2009-03-17]

Other possibilities included MADS (Metadata Authority Description Schema) by Library of Congress, or creating our own XML schema.
Creative Commons Rights Expression Language (ccREL)

- machine-readable metadata format for expression of copyright licensing terms
- based on RDF
- may be embedded in various formats (XHTML, XML...)

http://wiki.creativecommons.org/CcREL
Inspiration

- National Library of Sweden (http://libris.kb.se/auth/154863)
- German National Library of Economics - STW Thesaurus for Economics (http://zbw.eu/stw/)
- RAMEAU Subject Headings (http://www.cs.vu.nl/STITCH/rameau/)
- Library of Congress Subject Headings (http://id.loc.gov/authorities/)
Aleph X-Server enables connectivity to the Aleph system via a standard XML interface. Provides library data (MARC records, holdings, patron information...) as XML.

```xml
<varfield id="150" i1="" i2=""/>
  <subfield label="a">internet</subfield>
  <subfield label="x">vt</subfield>
</varfield>
<varfield id="550" i1="1" i2=""/>
  <subfield label="w">h</subfield>
  <subfield label="a">www</subfield>
  <subfield label="x">vt</subfield>
</varfield>
```
Comparison

- MARC(A)
  550 |a informační sítě |x in
- OAI-MARC (XML)
  <varfield id="550" i1="" i2="" >
  <subfield label="a">informační sítě</subfield>
  <subfield label="x">in</subfield>
  </varfield>
- SKOS (RDF+XML)
  <skos:related
  rdf:resource="http://psh.ntkcz.cz/skos/PSH6531"/>
Example: SKOS RDF+XML

<rdf:RDF>
  <skos:Concept rdf:about="http://psh.ntkcz.cz/skos/PSH12361">
    <skos:prefLabel xml:lang="cs">internet</skos:prefLabel>
    <skos:prefLabel xml:lang="en">internet</skos:prefLabel>
    <skos:broader rdf:resource="http://psh.ntkcz.cz/skos/PSH12360"/>
    <skos:exactMatch rdf:resource="http://id.loc.gov/authorities/sh92002816#concept"/>
  </skos:Concept>
</rdf:RDF>
Web interface

Embedded in the current State Technical Library website:
http://www.stk.cz/cs/katalogy-a-databaze/psh/prohlizeni-psh/
... but actually runs on:
http://psh.ntkc.cz/skos

Main page:
• 44 main subject terms
• navigation from general to more specific terms
• whole Polythematic Structured Subject Heading System can be downloaded as RDF+XML
Search

- searches preferred and non-preferred term labels
- exact match redirects directly to the subject heading
Browse

- Text version
  - preferred and non-preferred labels,
  - URI,
  - language equivalents,
  - broader term,
  - narrower terms,
  - related terms,
  - links to Library of Congress Subject Headings,
  - links to DBPedia,
  - link to RDF+XML serialization
- Visualization
displays the structure of the term neighbourhood
- Search results from State Technical Library catalog
Access to RDF+XML

- linked from the term page
- HTTP request with header containing **Accept:** application/rdf+xml - http://psh.ntkcz.cz/skos/PSH2395
- http://psh.ntkcz.cz/skos/PSH2395/rdf
Links to other datasets

Automatically harvested based on exact match.
Library of Congress Subject Headings
  • http://id.loc.gov/authorities
DBPedia
  • http://dbpedia.org
Numbers

- 99,724 RDF triples
- 13,569 subject headings
- 1,511 links to DBPedia
- 1,901 links to Library of Congress Subject Headings
What's next?

- other serializations (N3, JSON)
- links to other datasets
- SPARQL endpoint
- XML sitemap for semantic search engines

What would you like to see?

Questions?