

OLAC Vocabularies and Schemas for Language Technology Fields

Baden Hughes
baden@compulinq.net
OLAC '02 Philadelphia

Language Technology (LT) Fields Needs Analysis

- Needs analysis based on ordinary end user interaction requirements
- *Possibility*: Can I use this software ?
- *Probability*: How much effort will it take for me to be able to use this software ?
- *Functionality*: Does this software do what I want ?

Language Technology Vocabulary / Schema Implications

- LT archives are often very active software resource sites (esp. open source)
- Classification and description of software has practical implications for the end user
- LT has particular technical requirements for classification and description of software resources
- LT classification and descriptions can draw on wider IT vocabularies

Draft OLAC Vocabularies and Schemas ...

- OLAC-Functionality
- OLAC-OS
- OLAC-CPU
- OLAC-Sourcecode

OLAC-Functionality ...

- status: unreviewed draft
- "Controlled Vocabulary for Functional Classification"
- currently lists 17 core categories and 98 extended functional categories for LT
- based on HLT survey version 2 (from LT-World / DFKI)

OLAC-Functionality ... cont ...

- **Functionality Divisions:**
 - Information Extraction
 - Information Retrieval
 - Authoring Tools
 - Language Analysis
 - Language Understanding
 - Knowledge Representation and Discovery
 - Spoken Language Input
 - Written Language Input
 - Natural Language Generation
 - Spoken Output
 - Multilinguality
 - Multimodality
 - Coding and Compression
 - Mathematical Methods
 - Discourse and Dialogue
 - Language Resources
 - Evaluation

OLAC-OS ...

- Status: unreviewed draft
- "Controlled Vocabulary for Operating Systems"
- currently lists 41 operating systems
- based on industry standard IT classifications
- example

OLAC-CPU ...

- status: unreviewed draft
- "Controlled Vocabulary for CPU"
- currently lists 37 CPU types
- based on industry standard IT classifications
- example

OLAC-Sourcecode ...

- status: unreviewed draft
- "Controlled Vocabulary for Programming Languages"
- currently lists 286 programming languages
- based on industry standard IT classifications
- example

Issues ...

- Community review of drafts ?
- WG for Language Technology Fields ?
- Are OLAC-Functionality descriptions applicable to more resources than just language technology ?
- Should type be revised in OLAC Metadata document ?
- Proposal for OLAC-Sourcestatus ?

Issues ... cont

- interaction of these metadata elements with other related fields eg type ?
- service provider implementations for language technology resources ?