Final Study Report on Semantic Metadata Mapping Procedure

June 22, 2009
Tae-Sul Seo and Tae-Hoon Lim

tsseo@kisti.re.kr
taehoon@dpc.or.kr
Brief History

- 2007-07-19/New York, USA
  - WG2N1042(SC32N1658) Request for a new study period

- 2007-12-06/Seoul, Korea
  - WG2-N1077, WG2-N1078 Study report

- 2008-05-28/Sydney, Australia
  - WG2N1122 Request for extending the study period

- 2008-11-19/Vilamoura, Portugal
  - WG2N1216, WG2N1217 Request for project sub-division

- 2009-04-12/Summary of Voting
  - SC32N1856 Proposed SC32 resolution for sub-division is adopted.
  - Project Number 1.32.16.01.05.00
Progress Summary

- The WD of SMMP is being prepared at the moment.
  
  1. Some Korean experts have reviewed the WD. Their comments should be considered.
  2. Canadian comments on the letter ballot should be also considered.
  3. The strategy or plan for linking to MDRs based on ISO/IEC 11179 should be considered. (Bruce Bargmeyer’s comment at the Open Forum 2009)
  4. It should be converted to the Technical Report format. (Canadian comment on the letter Ballot)

- WD is expected to be submitted until the next interim meeting.
Overall Procedure

Main Processes

1. Identifying metadata sets
2. Grouping data elements
3. Semantic Mapping

Sub-Processes

1. Identifying metadata sets to be mapped
2. Finding object classes
3. Grouping data elements by objects
4. Finding properties
5. Grouping data elements by properties
6. Finding data element concepts
7. Mapping by data element concepts
8. Recommending data element names
9. Giving notes about heterogeneities
Types of Semantic Heterogeneities

- **Hierarchical Difference**: due to different level of detail
  - Generalization: Specialization, Composition: Decomposition

- **Domain Difference**: due to different context and culture

- **Lexical Difference**: is different appearance
  - Synonyms, Abbreviation, Acronyms, Case sensitivity, Language, Variation

- **Syntactic Difference**: due to different arrangement of parts
  - Ordering, Delimiters, Missing

- **Complicated Differences**: due to different policies
Relationship to ISO 11179 MDR

ISO/IEC 11179 MDR

Metadata Mapping Agent

Read only user
Context
Obj. Class
Property
DEC
DE

Submitting Organization
Submitter
Obj. Class
Property
DEC
DE

Identifying Namespaces
Finding Object Classes
Grouping DEs by Obj. Classes
Finding Properties
Grouping DEs by Properties
Finding DECs
Mapping by DECs
Recommending DEs
Giving Notes

2009-06-22
SC32/WG2 Meeting in Jeju
Canadian Comments on Letter Ballot

1. We find the name ‘Semantic Metadata Mapping’ unclear. What is ‘semantic metadata’? We suggest changing the title to e.g. ‘Metadata Concept Mapping Procedure’.
   - We will consider it in the future

2. This should be a type 3 Technical Report.
   - Thanks for information.

3. The WD provided as WG2 N1217 needs to be converted to the Technical Report format.
   - We will convert as soon as we get the format.
Canadian Comments on Letter Ballot

4. Examples in WD are of a library & information standards nature. From a "linguistic equivalency" perspective ISO 5964 - Guidelines for the establishment and development of multilingual thesauri would be relevant to this project.

- We will refer to ISO 5964 in the course of preparing the WD.
Canadian Comments on Letter Ballot

5. The approach to "metadata element set" seems to involve that of a "controlled vocabulary". Here SC32/WG1 - Business standard development of (FDIS) ISO/IEC 15944-7 Business operational View- Part 7:e-Business Vocabulary" may assist in the development of this project (especially its Clause 5.2 "Principles and rules for the development of a definition for a concept and assignment of an appropriate term in a "controlled vocabulary" [See document JTC1/SC32 N1826.

- We will refer to ISO/IEC 15944-7 clause 5.2 in the course of preparing the WD.
Canadian Comments on Letter Ballot

6. Standards development work in this area, which may be useful for this project, is already underway in JTC1/SC36 e-Learning via its WG7 Culture, Language and Individual needs via the development of ISO/IEC 24751-8 Information technology - Individualized adaptability and accessibility in e-learning, education and training - Part 8: Language Accessibility and Human Interface Equivalencies (HIEs) in e-Learning Applications: Principles, Rules and Metadata Elements". Here Canada is willing to provide the necessary liaison with SC36/WG7."

- Some principles of 24751-8 might be related to the types of semantic heterogeneities, especially “Domain Deference”.
- We will consider liaison with SC36/WG after looking into ISO/IEC 24751-8 in the future.
Thank you!