Linking and using multilanguage and multidisciplinary topic maps

2011-10-21
Motomu Naito
Knowledge Synergy Inc.
motom@green.ocn.ne.jp
Ψ http://psi.ontopedia.net/Motomu_Naito
http://www.knowledge-synergy.com/
Outline

1. Background
2. Topic Maps
3. Tool for Topic Maps: Ontopia
4. Topic maps as Structured Knowledge
5. The way to use the topic maps
6. Demo
7. Conclusion & Future work
1. Back ground

- Well structured knowledge is very attractive for human.
- It is very useful and important to use the knowledge,
  - to classify, compare, recognize and understand things,
  - to update it by newly acquired knowledge, and
  - to inherit it to the next generation.

- We should think about things from various viewpoints
  because things have various aspect.
  For example,
  Climate change has been caused because human pursued
  only economical profit.
Multidisciplinary and Multilayer Knowledge space
Things have various aspect. It is necessary to be able to choose an appropriate view according to the need.

Humanities  Environment  Philosophy  ·········

Community
Family
Individual
Back ground

- In order to understand each other and make consensus of global scale, multilanguage and multidisciplinary knowledge is expected.
- Making it by oneself is ideal.
- But making it from scratch is very laborious work and requires perspectives and knowledge in many directions.
- Fortunately many well structured knowledge already exist. (They are taxonomy, thesaurus, topic map, ontology, etc.)
- We can transform them to **topic maps** easily. In result, - they become more human and the Internet friendly, - they can be used and visualized more easily, and - they can be linked each other.
2. What is Topic Maps

Topic Maps is a series of ISO standards. It can structure knowledge and plays a role like an index in cyberspace.

The standard describes itself as follows:
"Topic Maps is a technology for encoding knowledge and connecting this encoded knowledge to relevant information resources. Topic maps are organized around
- **topics**, which represent subjects of discourse;
- **associations**, representing relationships between the subjects; and
- **occurrences**, which connect the subjects to pertinent information resources."
Standardization activity No.1 (ISO/IEC JTC1 SC34 WG3)

(1) ISO/IEC 13250: Topic Maps

- Part 1: Overview and Basic Concepts, Status: CD
- Part 2: Data Model (TMDM), Status: IS (2006.8.15)
- Part 3: XML Syntax (XTM), Status: IS (2007.3.15)
- Part 4: Canonicalization (CXTM), Status: IS
- Part 5: Reference Model (TMRM), Status: ?
- Part 6: Compact Syntax (CTM), Status: IS (2010)
- Part 7: Graphical Notation, Status: ?
Standardization activity No.2
(ISO/IEC JTC1 SC34 WG3)

(2) ISO/IEC 18048: Topic Maps Query Language (TMQL)
   Status: ?

(3) ISO/IEC 19756: Topic Maps Constraint Language (TMCL)
   Status: IS (2011)

* NP: New Work Item Proposal
* WD: Working Draft
* CD: Committee Draft
* FCD: Final Committee Draft
* FDIS: Final Draft for International Standard
* IS: International Standard
The Basic Model of Topic Maps

• Some pool of information or data
  – any type, format, or location

• A knowledge layer, consisting of:
  • Topics
    – a set of topics representing the key subjects of the domain in question
  • Associations
    – representing relationships between subjects
  • Occurrences
    – links to information that is somehow relevant to a given subject

• = The TAO of Topic Maps

P.S.
Topics, associations and occurrences have types, and all types are also topics…

(Source: Steve Pepper, “Towards Seamless Knowledge” )
Structuring information/knowledge by Topic Maps

Current situation:
A vast amount of information exists separately

Topic Maps:
Structuring information according to structured knowledge (Subjects and relationships among them)
**Published Subjects**

Published Subjects is a mechanism which enable person and computer to identify subjects (topics). And it is published on networks permanently and aims at making easy share/exchange Topic Maps.

- It is included in Topic Maps standard
- It allocate URI (IRI) to subject, and make possible to identify subjects.
- It provides a subject identifying mechanism to Computer and Human.
- It is applicable both information resource and things in real world.
- On the occasion of merge of Topic Maps, it is used to identify topics (subjects). The topics which have same subject are merged.
- Any one can publish published subjects.
- A good thing will survive.
Example of PSI (Subject: Lars Marius Garshol)

Ψ (PSI): http://psi.ontopedia.net/Lars_Marius_Garshol

This URI is an address of an information resource which describe the subject for human to understand

It is called Published Subject Indicator or Published Subject Descriptor

Existing PSI Server
- ONTOPEDIA (http://psi.ontopedia.net/)
- Subj3ct.com (https://subj3ct.com/)
3. Tool for Topic Maps: Ontopia

Ontopia is open source tools for building, maintaining, and deploying topic map-based applications. It consists of:

- Ontopia Topic Maps Engine
- Ontopia Navigator/Web Editor Framework
- Topic Maps Browser/Editor, Graphic Viewer
- Full-text Search Integration
- RDBMS Backend Connector
- Topic Maps Query Engine
- Topic Maps Schema Tools
- DB2TM, etc.

Ontopia can be downloaded from

http://code.google.com/p/ontopia/
4. Topic maps as structured knowledge

Making structured knowledge by oneself is ideal. As an alternative way, we can use already existing ones. For example:

1. In Agriculture: AGROVOC

2. In Library science: BSH (Basic Subject Headings), NDLSH (National Diet Library Subject Headings), LCSH (Library of Congress Subject Headings)

3. In Cultural anthropology: HRAF (Human Relations Area Files)

4. In information science: WordNet

We transformed some of them to topic maps.
AGROVOC

- AGROVOC is the world’s most comprehensive multilingual agricultural vocabulary.

- AGROVOC was first developed in the 1980s by FAO (Food and Agriculture Organization of the United Nations) and CEC (Commission of the European Communities) as a multilingual structured thesaurus for all subject fields in agriculture, forestry, fisheries, food and related domains (e.g. environment).

- Recently, organized as a concept scheme, AGROVOC contains close to 40,000 concepts in over 20 languages.
AGROVOC

• We downloaded the data at 2011/08/01 from:
  http://aims.fao.org/standards/agrovoc/functionalities/download
• We used CSV format data to transform to topic map.
• We used software tool Ontopia to make topic map and its web application.
• In result:
  - 53,630 Terms (Concepts) including non-preferred term
  - 32,971 BT(Broader Term) - NT(Narrower Term) relation
  - 22,411 RT(Related Term) relation
  - 20,771 USE(Use) - UF(Used for) relation
  - etc.
AGROVOC topic map application
Screen shots of the application
AGROVOC topic map application

Screen shots of the application

**TERM:** (c_{1665}) EN: climate

- [Ψ](http://aims.fao.org/acs/eng/agrovoc/c_1665.htm)

**LANGUAGE**

1. AR: رياض
2. CS: klima
3. DE: Klima
4. EN: climate
5. ES: Clima
6. FA: کیما
7. FR: Climat
8. HI: जलवायु
9. HU: állapot
10. IT: Clima
11. JA: 気候
12. KO: 기후
13. LC: 氣候
14. PL: Klimat
15. PT: Clima
16. RO: Clima
17. SK: klima
18. TH: ภูมิอากาศ
19. ZH: 气候

**HIERARCHY**

- (c_{330740}) phenomena
  - (c_{330740}) natural phenomena
  - (c_{1005}) climate

**BT**

1. (c_{330740}) natural phenomena

**MT**

1. (c_{1665}) Coastal climate
2. (c_{4697}) Mediterranean climate
3. (c_{14131}) Semiarid climate
4. (c_{4802}) Mountain climate
5. (c_{4802}) microclimate
6. (c_{611}) Arid climate
7. (c_{3693}) Humid climate
8. (c_{7973}) Tropical climate
9. (c_{2684}) Temperate climate
10. (c_{6090}) Polar climate
11. (c_{14666}) Subtropical climate
12. (c_{330816}) Moist sub humid
13. (c_{330849}) Semi humid
14. (c_{4759}) Mesohumid climate
15. (c_{16529}) Weather
16. (c_{5298}) Oceanic climate
17. (c_{14135}) semi humid climate
18. (c_{330769}) Dry sub humid
19. (c_{1333}) Continental climate
20. (c_{23976}) Subhumid climate
21. (c_{4822}) Monsoon climate
22. (c_{4490}) Macrocimate

**RT**

1. (c_{2594}) environmental factors
2. (c_{7087}) site factors
3. (c_{6311}) seasons
4. (c_{3654}) Ozone depletion
5. (c_{1611}) Climatology
6. (c_{5774}) Phenology
7. (c_{1585}) Climatic change
8. (c_{34540}) Rainfall patterns
9. (c_{22954}) climatic factors
10. (c_{22942}) Atmospheric circulation

**UF**

1. (c_{36880}) Climatic effects
2. (c_{12778}) Natural climate
3. (c_{52810})

**Scope Note Reference**

1. (c_{1005}) climatic zones

**Referenced in Scope Note**

1. (c_{1665}) climatic zones
HRAF

Human Relations Area Files, Inc. (HRAF) is an internationally recognized organization in the field of cultural anthropology.

The mission of HRAF is to encourage and facilitate worldwide comparative studies of human behavior, society, and culture.

Founded in 1949 at Yale University, HRAF is a financially autonomous research agency of Yale.

* It seems very useful for us, but availability of HRAF collections is limited to members of the HRAF consortium.

Source: http://www.yale.edu/hraf/index.html
Subject Headings

Wikipedia redirects “Subject Headings” to “Index term” and define the term as

“An index term, subject term, subject heading, or descriptor, in information retrieval, is a term that captures the essence of the topic of a document. Index terms make up a controlled vocabulary for use in bibliographic records.”

(http://en.wikipedia.org/wiki/Index_term)
Subject Headings

- There have been many Subject Headings (ex: NDLSH, BSH, LCSH, etc.)
- Fortunately some Subject Headings (SHs) are published on the web and we can use them (ex: NDLSH, LCSH)
- We are working on the following SHs at the moment: NDLSH, BSH and LCSH
- Probably we can find much more SHs in various countries
NDLSH

- We downloaded the data at 2011/08/02 from http://iss.ndl.go.jp/ndla/download/
- We used TSV format data to transform to topic map.
- We used software tool Ontopia to make topic map and its web application.
- In result:
  - 19,407 Terms (Subjects)
  - 32,391 non-preferred term
  - 14,899 BT(Broader Term) - NT(Narrower Term) relation
  - 10,531 RT(Related Term) relation
  - 32,859 USE(Use) - UF(Used for) relation
  - etc.
Some part of NDLSH

Subject Headings around “ビール: Beer”

(*)&: Existence of corresponding Wikipedia article
5. The way to use the topic maps

Topic maps structure knowledge intuitively and friendly for human and connect them to relevant information resource. The well structured knowledge is very valuable and useful itself.

In addition, topic maps can make it usable as follows:

- Navigating topic maps using tolog (Query Language)
- Linking topic maps using TMRAP
- Linking topic maps using merge function
- Multi-indexing to information resources (Multidisciplinary index)
Navigating topic maps using tolog (1)

Tolog makes possible to navigate topic maps according to its structure.

- **Ontopia's topic map query language**
  - Basically Datalog adapted for use with topic maps
  - Has been in use for over 7 years
  - Drives a number of commercial applications
  - Also implemented in the open source TM4J engine

- **Basic features**
  - Graph matching (Datalog clauses)
  - AND/OR/NOT support in query body (and rule bodies)
  - Inference rules
  - Aggregate functions (like SQL)
  - Sorting (like SQL)
Navigating topic maps using tolog (2)

Query examples:

1: Query narrower term of "climate"
using psi1 for i"http://aims.fao.org/aos/agrovoc/"
using psi2 for i"http://www.topicmaps.org/xtm/1.0/core.xtm#"
psi2:superclass-subclass(psi1:c_1665 : psi2:superclass,
  $SUB1 : psi2:subclass)?

2: Query related term of "climate"
using psi1 for i"http://aims.fao.org/aos/agrovoc/"
using psi2 for i"http://www.h-gis.org/psi/agrovoc/"
psi2:relate(psi1:c_1665 : psi2:rt, $RT2 : psi2:rt)?
Navigating topic maps using tolog (3)

3: Query narrower term of "climate" in Thai using psi1 for i"http://aims.fao.org/aos/agrovoc/"
using psi2 for i"http://www.topicmaps.org/xtm/1.0/core.xtm#"
using psi3 for i"http://www.h-gis.org/psi/agrovoc/"
select $LTNAME from
psi2:superclass-subclass(psi1:c_1665 : psi2:superclass,
$SUB1 : psi2:subclass),
topic-name($SUB1, $TNAME),
scope($TNAME, psi3:TH),
value($TNAME, $LTNAME)
order by $LTNAME?
Linking topic maps using TMRAP (1)

TMRAP makes possible to link separate topic maps.
• TMRAP (Topic Maps Remote Access Protocol) is a web service interface
  ( http://www.ontopia.net/topicmaps/tmrap.html )
• It makes possible to retrieve and modify Topic Maps fragments from a remote Topic Maps server
• Two protocols (HTTP or SOAP) are usable
• Consist of the following methods
  - get-topic
  - get-topic-page
  - get-tolog
  - add-fragment
  - delete-topic
Linking topic maps using TMRAP (2)

Topic map fragments providing web service using TMRAP

App1 (Client)

- Topic Maps
  - Web Application
    - JSP Page
- Ontopia
  - Navigator Framework
  - Query engine
- Topic map 1

Request (get-topic) with ID
Return (TM fragments)

These topics have same ID (PSI)

“climate” topic

App2 (Server) Web service

- Topic Maps
  - Web Application
    - JSP Page
- Ontopia
  - Navigator Framework
  - Query engine
- Topic map 2

“climate” topic
Linking topic maps using TMRAP (2)

Example of request to server (get-topic):

Return from server
Linking topic maps using merge function

- Topic maps can be merged according to ISO standard.
- In that case topics which have same subject are unified.
- Existing tools support the merge function.
  - Omnigator in Ontopia
  - command-line utility in Ontopia
    “net.ontopia.topicmaps.cmdlineutils.Merger” utility

Example:
java net.ontopia.topicmaps.cmdlineutils.Merger
merge-input1.xtm merge-input2.xtm merge-output.xtm
Multi-indexing to information resources

- Information resources are indexed from various perspective.
- By using topic maps from various discipline, human can be supported his intellectual activity.
6. Demo

I will do short demo if I have enough time
7. Conclusion & Future work No.1

- Multidisciplinary knowledge is necessary to understand things comprehensively.
- Multilanguage knowledge is important for people to communicate and understand each other.
- Making structured knowledge by ourselves is ideal though, making it from scratch is very laborious work.
- As an alternative way, we can use already existing ones.
- We are using Topic Maps for encoding those knowledge and representing them human friendly.
- It made us possible to navigate and link those knowledge and construct Multi-index for related information resources.
- We think that Topic Maps has high affinity to express and visualize them.
Conclusion & Future work No.2

We have a lot of future works:
- To make more topic maps for various discipline
- To develop PSI server and publish it on the web
- To develop good tagger for information resources
- To provide web services for providing topic map fragment
- Others (Many, Many, Many, ….)

Next slides show “Big picture of our system” and “Ultimate goal”
Big picture of our system

Multilanguage and Multidisciplinary knowledge

Information resources (DB, Archive, Library, Collection, Gallery, etc.)
Ultimate goal

Topic maps represent concept relations in human brain and connect them related information. We can navigate and link these knowledge and easy to find related information resources.
ありがとうございます。

Thank you!

Any suggestion?