Metadata translation between LOM and CRM based on ontology mapping

Susumu Sawai[1], Toshiko Wakaki[2] and Masakazu Tachikawa[3]

[1]Contents Reseach Department, Gakujoken(Information Reseach Center for Learning)

23,Nangen-Cho,Shinju-KuTokyo,160-0012,Japan

[2,3] Department of Electronic Information Systems, Shibaura Institute of Technology

307 Fukasaku,Saitama-City,Saitama, 330-8570 Japan

E-mail: [1]nac02440@nifty.ne.jp, [2]twakaki@sic.shibaura-it.ac.jp , [3]p02093@sic.shibaura-it.ac.jp

ABSTRACT: In this paper, We emphasize to realize a syllabus reasoning system using Tagging and Metadata. We have been able to obtain reasoning/search results using metadata translation between LOM and CRM based on ontology mapping.

1. Introduction

In this decade, authors have been coordinator of several Japanese Government ICT projects.

GENES Project was launched in August 2001, in cooperation with Fujitsu and Ministry of Education(MEXT) . GENES is a Social Information Organization , named by Gakujoken network studying group. Gakujoken is the public organization managed by MEXT. The roles of GENES is to collect and to offer the multimedia contents to learners,as a “digital Cultural Genes”.

From technical points of view, there is growing demand for information retrieval and natural language processing utilizing ontology in various fields such as Web portal sites and e-learning, and the effectiveness of the usage of such is becoming increasingly acknowledged.

Currently, systems such as the Science Museum Net (S-Net) and the NICER search system are being operated. These systems are based on CRM compliant sample data metadata or LOM indexed to sample data or learning material archives that are being collected at National Science Museum and NICER. There is a need to mutually utilize the archives between these two systems, however since the metadata concept differs in both systems, it is necessary to address technical issues such as the occurrence of data loss on metadata translation between these systems. Meanwhile, W3C recommendations were made on technologies to express metadata and ontology on the Web, RDF and OWL [3] of Semantic Web. On the other hand, recently efforts have been made for ontology mapping, in other words, methods to relate and translate differing ontologies. In this study, we have related the concepts of each metadata in both systems using ontology expressed by RDF of the Semantic Web, as well as uniquely building a biological taxonomy ontology expressed by OWL and conducting ontology mapping using this based on knowledge supplement; and realized interconversion between the two systems, and have done a usability assessment of such.

2. Ontology Mapping Between LOM And CRM

2.1 RDF expression of LOM and CRM

In this study, we consider two types of metadata, that is, LOM and CRM compliant data. LOM (Learning Object Metadata) is a data model that defines the structure of metadata and vocabulary specification of a learning object, standardized by IEEE in 2002. Currently the standard has 96 elements defined. We have proposed the RDF expression of such in [1] and [4]. CIDOC CRM (Conceptual Reference Model) [2] is a model established by the International Council of Museums, which expresses object information related to the management of museum collections , that was proposed in the OWL expression. The National Science Museum has established metadata compliant to the CIDOC CRM, which consists of 70 elements [7].

Title	URL	Abstract	・・・
“TORAFUGU”	http://～	This fish ～	・・・

DB URL	Japanese Name	Kingdom Name	・・・
http://～	“TORAFUGU”	Animalia ・・・	・・・

Table 1:LOM (above) and CRM (below)

2.2 Ontology Mapping

Since LOM and CRM are metadata based on different concepts, they are not compatible on their own. Therefore, we conducted ontology mapping, that is, expressing both metadata by ontology and translating them into equivalents. One data does not necessarily correspond to one data, and there are the following three possibilities:

1) One LOM data corresponds to one CRM data

2) Multiple LOM data correspond to one CRM data

3) One LOM data corresponds to multiple CRM data

In this study, majority of the incidents correspond to 1 and 3 above. A representative example of case 3 is the CRM data of information related to the scientific name, Japanese vernacular name and introduction equivalent to an LOM keyword. Figure 1 shows the concept of ontology mapping.

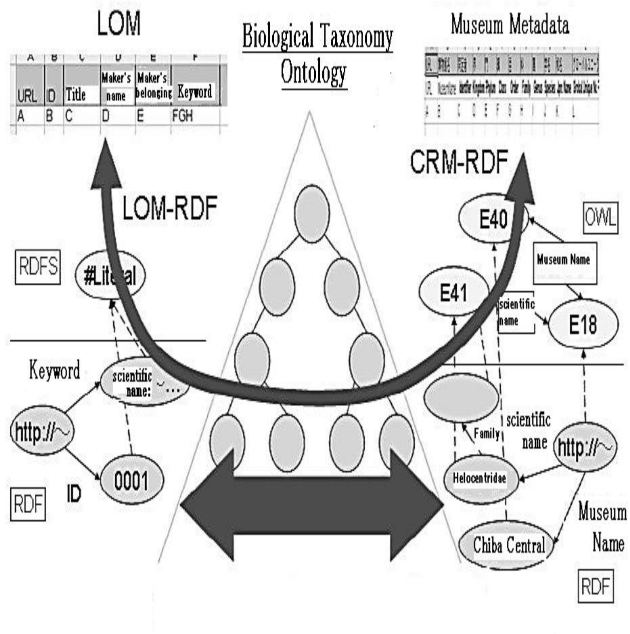

Figure 1: The Concept of Ontology Mapping

2.3 Supplement of Lacking Knowledge using the Biological Taxonomy Ontology

When conducting ontology mapping, there are certain data that need not only conversion but also require translations estimated from its value, and if the mapping is not done accurately, information loss will occur. In this study, it is necessary to supplement biological taxonomy knowledge that has been lost for some reason, such as a missing value, or its corresponding item being indefinite. A biological taxonomy is a method of systematically classifying living organisms by a class structure starting from Kingdom, Phylum, Class, Order, Family, Genus and Species, in descending order. For example, a “takifugu rubripes” would be classified as a “Metazoa Chordata Osteichthyes Tetraodontiformes Tetraodontidae Takifugu Takifugu rubripes”. In this study, we took notice of this class structure, and used the characteristics of this structure to express and establish the ontology for supplementing the lacking knowledge by the OWL language.

- Information such as in which classification from Kingdom to Species the biological classification belongs to, or whether the name is a scientific name or a Japanese vernacular name, shall be expressed as an OWN class. As each classification name holds information on the classification and language, the product of the OWL classes that express the classification and language shall be newly expressed as a class (Figure 2 (1)).

- Each classification name shall be expressed as an entity that belongs to the class newly created by the product. For example, a “FUGU-KA (Tetraodontidae in Japanese)” is a “family” as well as a “Japanese vernacular name”, therefore shall be expressed as an entity belonging to a class created by this product (Figure 2 (2)).

- The hierarchical relationships between each classification shall be associated using the super Classification property, and as such hierarchical relationships are inclusive relationships, the property shall hold transitivity (Figure 2 (3)).

- A scientific name and a Japanese vernacular name shall be associated using the Japanese Name, scientific Name properties (these properties are inverse properties) (Figure 2 (4)).

- As redundancies shall be seen for a Japanese vernacular name in different classes, we avoid such redundancies by numbering a Japanese vernacular name, and associate a common keyword using the word property (Figure 2 (5)).

The supplement procedure shall be done in the following manner:

1) The lowest ranking scientific name or Japanese vernacular name among the known values shall be obtained from LOM-RDF or CRM-RDF. If the name obtained was a Japanese vernacular name, then the corresponding scientific name shall be obtained using the biological taxonomy ontology.

2) Checks will be made whether or not there are any instances in the classification one rank above the obtained scientific name. If an instance exists, no knowledge supplement shall be conducted, and the subject shall be changed to the checked instance. If there are no instances one rank above the obtained scientific name, the upper class of the subject shall be searched using the biological taxonomy ontology, and the lacking information shall be filled in.

3) This method of supplementing lacking knowledge shall be repeated until the highest class.

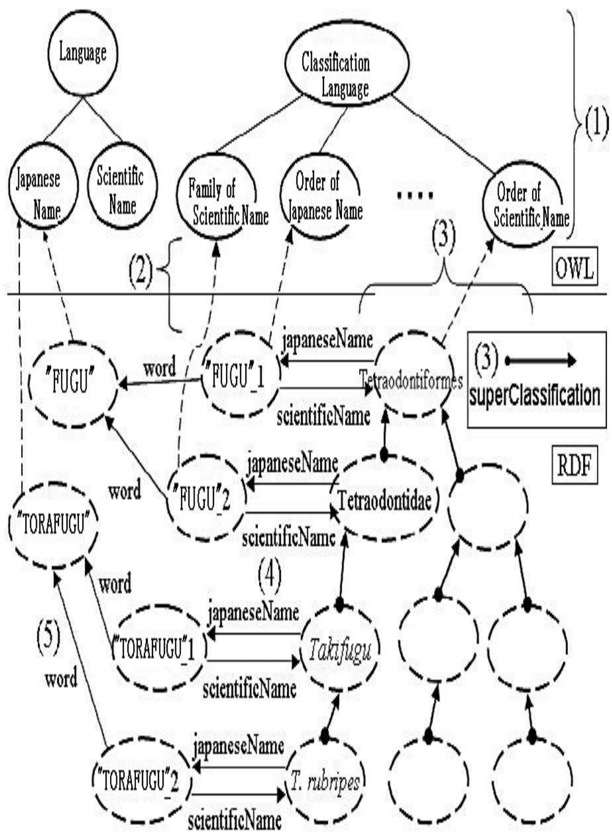

Figure2: The Concept of Biological Taxonomy Ontology

3. Implementation And Assessment

A translation tool using Java was implemented for the mapping between LOM and CRM. Jena (Java API) [5], which enables OWL assumption, was used for the handling of RDF on translation, or value estimations and knowledge supplement using the biological taxonomy ontology built by OWL.

For assessment, we used 1,000 entries of CRM data at the Fisheries Research Laboratory of Mie University to assess knowledge supplement.

Table 2 shows the number of entries before and after the entries of scientific names related to the biological taxonomy ontology were supplemented. It can be seen that knowledge was supplemented using the biological taxonomy ontology in entries in which initially no data was included for Kingdom, Phylum, Class, Order, with only one exception in which no ontology was established.

	Kingdom	Phylum	Class	Order	Family	Genus	Species
Before Supplement	0	0	0	0	1000	999	999
After Supplement	999	999	999	999	1000	999	999

Table2: A Comparison of Before and After Supplement using data at Mie University

Subsequently, we assessed the knowledge supplement. Figure 3 shows the results after searching and obtaining the name of the Kingdom to Species of a fish called "Benitsuke Ginpo", a specimen data at Mie University. Prior to the supplement, we used the sample data search system of S-net, and after the supplement, we used a tool called Sesame [8], a tool that enables SQL-like searches against RDF called SeRQL. At S-net, several data could not be retrieved since no data had been registered in several entries, however it can be seen that after supplementing the lacking knowledge, all class names have been obtained.

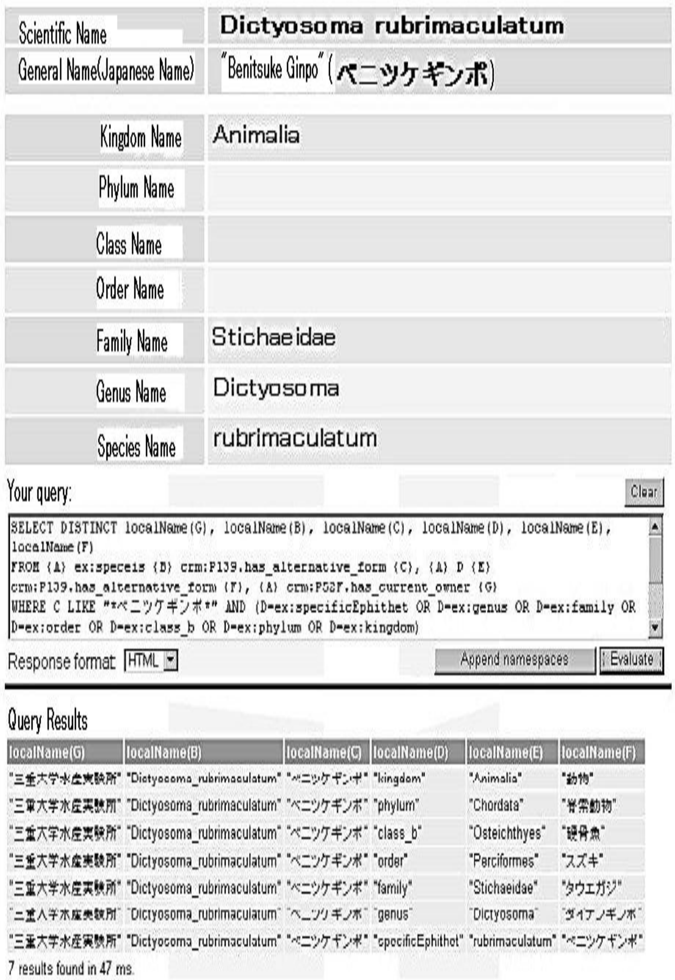

Figure3: the results after searching of Before(above) and After Supplement (below)

Lastly we show an example of ontology mapping from LOM to CRM.

Figure 4 shows the LOM on Thunnus alalunga (Japanese vernacular name “Bin naga”) and an example of this data being mapped to CRM. It is effective to map this specimen from LOM since it is not registered in CRM, however simple translation cannot be conducted since the original LOM includes no data on its taxonomy. As such, we used its Japanese vernacular name “Bin naga” and conducted knowledge supplement using biological taxonomy ontology to fill in the empty data, on which we were able to handle this data as CRM.

Title	"Bin Naga"
Abstract	Breast Viret is very long. Meat is white and quite frank . It is processed into canned food(sea-chicken). The length is around 1m. It will lay eggs around June. It fishes in the sea near Kyushu district. Summer is a season.In the Kanto district, it is called "Bin Chow", and In the Kansai district, it is called "Tombo".
Keyword	Sea-Chicken,Tombo,Bin Chow,Bin Naga,Fish

Kingdom	(Japanese Name)	Phylum	(Japanese Name)	Class	(Japanese Name)	Order	(Japanese Name)
Animalia	動物	Chordata	脊索動物	Osteichthy	硬骨魚	Perciforme	スズキ

Family	(Japanese Name)	Genus	(Japanese Name)	Species	(Japanese Name)
Scombridae	サバ	Thunnus	マグロ	alalunga	ビンナガ

Figure4: An Example of Mapping from LOM (above) to CRM (below)

4. Conclusion

It is expected that the range of contents to be searched through systems such as S-Net can be expanded, by conducting ontology mapping between LOM and CRM using the biological taxonomy ontology; enabling counter-operability and mutual searches between NICER learning material contents and museum sample contents; and on the other hand, supplementing metadata values of lacking knowledge.

References

[1] M. Tachikawa, S. Sawai, T. Wakaki: E-Learning Support System based on SemanticWeb, IEICE 5L-L, PP.221-222, 2006

[2] Ryouji Akimotoi: The problem of Metadata in collection management of museum, 25th Digital Library Workshop, 2004.

[3] G.Antoniou, F.v.Harmelen: A SemanticWeb Primer, The MIT Press, 2004.

[4] S. Sawai, T. Wakaki, M. Tachikawa, and S. Kunifuji: Multimedia Portals based on Semantic Web Technology for GENES Communities - The Implementation of LOM-RDF Conversion Tools, Proceeding of SWAMM, WWW2006 Workshop, 2006.

[5] Jena: http://jena.sourceforge.net/

[6] NICER: http://www.nicer.go.jp/

[7] S-net: http://www.science-net.kahaku.go.jp/

[8] Sesame: http://www.openrdf.org/