Ontological Interoperability of Learning Objects: A Hybrid Graphical-Neural Approach

Chien-Sing Lee[1] and Ching-Chieh Kiu[2]
Faculty of Information Technology
Multimedia University
Jalan Multimedia, 63100 Cyberjaya, Selangor.
Malaysia.
http://pesona.mmu.edu.my/~cslee[1]
http://pesona.mmu.edu.my/~cckiu[2]

Abstract: This paper presents OntoShare, an automated ontology mapping and merging architecture for learning object retrieval and reuse. The architecture aims to offer contextual and robust ontology mapping and merging through hybrid unsupervised clustering techniques comprising of Formal Concept Analysis (FCA), Self-Organizing Map (SOM) and K-Means clustering incorporated with linguistic processing using WordNet. The merged ontology facilitates sharing and retrieval of learning objects from the Web or from different learning object repositories such as ARIADNE and Educause. Experimental results can be extended to other resources in databases or data warehouses.

Key-Words: Learning object, Ontology, Hybrid clustering, Formal Concept Analysis, SOM, Interoperability

1 Introduction

Ontologies enrich description of learning object metadata in digital libraries or repositories such as ARIADNE or EDUCAUSE at different levels of granularity. Among the fundamental metadata are those defined by the Dublin Core, i.e., the title of the resource, the creator or organization, subject keywords, category (abstract, advertisement etc.), the URL, unique identifier (e.g. ISBN or ISSN) and language used. These metadata can be contextualized by information such as target audience's age, mastery level or preferences and learning objectives [1], enabling adaptation of learning objects to different contexts of use. Technical implementations for these metadata are considered in content models or standards such as IEEE's LOM or ADL's SCORM.

However, the thread which binds metadata, context and content models is ontology. Ontology adds semantics by defining associations among concepts or topics and corresponding attributes, creating structural dependencies. Structural dependencies provide a macro view of interconnections among learning objects and media, creating epistemological bases for different functional combinations of learning objects and its media components when the instructor authors learning content.

2 Problem Formulation

Educational systems should be viewed from three perspectives: first, system design driven by pedagogical principles, second technological tools that enable personalization and third, standardization in terms of indexing of learning objects and content management in order to enable interoperability and reusability of learning materials [2]. This paper addresses the third aspect, i.e., interoperability.

Three problems constrain efforts to interoperate among ontologies. First, different creators use different ontologies to annotate metadata and learning object content [3].

Second, the process of ontology mapping and merging between ontologies is time consuming and tedious. Several ontology merging tools have been developed to support the ontology merging task [4, 5].

PROMPT [4] is a semi-automatic tool for system-guided ontology merging in Protégé 2000. PROMPT identifies matching class names and iteratively performs automatic updates. PROMPT also identifies conflicts and makes suggestions on means to remove these conflicts to the user.

FCA-Merge [5] is a bottom-up ontology merging approach using formal concept analysis and natural language processing techniques. Given source ontologies, it extracts instances from a given set of domain-specific text documents by applying natural language processing techniques. The concept lattice, a structural result of FCA-Merge, is derived from the extracted instances using formal concept analysis. The produced result is analyzed and merged with the existing ontology by the ontology engineer.

ODEMerge [6] is integrated with WebODE. ODEMerge performs automated supervised merging of concepts, attributes and relationships from two different ontologies using synonym and hypernym tables to generate the merged ontology. It merges ontologies with the help of corresponding information from the user. The results derived from the ODEMerge process can be modified by the user.

The approaches in [4] and [5] merge ontologies based on syntactic and semantic matching heuristics, and user interaction on the ontology merging process is requested to generate the merged ontology. We would like to fully automate the ontology merging process crucial for large repositories where the growth of learning objects can be exponential.

Third, although in [6], the merging process is automated, supervised techniques which require prior knowledge are used. Prior knowledge is sometimes not easily available.

3 Problem Solution

This paper presents the OntoShare, an ontology sharing architecture that automatically merges ontologies through a hybrid unsupervised clustering method comprising of, Formal Concept Analysis (FCA), Self-Organizing Map (SOM) and K-Means clustering incorporated with linguistic processing using WordNet.

Most automated or semi-automated mapping and merging systems are concept-based (top-down) or instance-based (bottom-up). Concept-based approaches predefine concept information such as name, taxonomies, constraints and relations and properties of concept elements. In contrast, instance-based approaches build up the structural hierarchy based on instances of concepts and instances of relations. Examples of concept-based systems are PROMPT and ODEMerge whereas an example of the latter system is FCA-Merge. The OntoShare combines the concept-based and instance-based approaches. It is concept-based at the ontological contextualization and pre-linguistic processing stage and instance-based at the contextual processing stage (if multi-valued attributes are used). This paper discusses only the concept-based aspect.

Formal Concept Analysis (FCA) provides the ontological basis for structuring associations among concepts/topics and modeling concepts/topics and corresponding attributes. Unsupervised Self-Organizing Map (SOM) and k-means do not require prior knowledge. As such, clustering results are natural. Furthermore, multi-level clustering can be implemented with these clustering techniques to improve on scalability and granularity [7]. Visualization of the merged ontology using the OntoVis visualization tool [8] enables indexing and easy retrieval of concepts and learning objects through attributes that describe each concept in FCA's formal context.

The rest of the paper is outlined as follows: Section 4 explains the OntoShare techniques, and Section 5 the OntoShare architecture. An example of a simulation result to illustrate the approach is provided in Section 6. Section 7 concludes the paper.

4 OntoShare's Techniques

4.1 Ontologies

Ontology specifies a shared conceptualization [9]. In general, ontology consists of concepts, attributes and relations. The core of the ontology is formalized as a tuple O: = (C, S_C, R, S_R, is_a, A), where *C* is Concepts of ontology and S_C corresponds to the hierarchy of Concepts. The relationship between the concepts is defined by Relations, *R* where S_R corresponds to the hierarchy of Relations. *is_a* is the hierarchical relationship between the concepts and *A* is axioms used to infer knowledge from existing knowledge.

4.2 WordNet and Similarity Measure

WordNet is the online lexical database, where each meaning of a word is represented by a synset or synonym set. WordNet organizes nouns and verbs into hierarchies of *is-a* relations [10].

Stop word filtering and tokenization are applied to transform the input for similarity measure [11]. In tokenization, the compound words are split into tokens and semantic similarity with other words will be calculated as an average over the similarity between each token and the other word.

The Leacock-Chodorow similarity measure [12] is used to discover the semantic similarity between two synsets as [13] have proven that it is the best among similarity measures. The similarity of two synsets can be defined as:

$$related_{lch}(c_1, c_2) = \max\left[-\log \frac{ShortestLenght(c_1, c_2)}{2D}\right] \quad (1)$$

The Leacock-Chodorow similarity measure relies on the shortest paths between two synsets (*ShortestLength (c1, c2)*) in an *is_a* hierarchy and this value is scaled by depth D of the taxonomy. A threshold value is set to determine acceptable semantic similarity between two synsets.

4.3 Formal Concept Analysis

Formal Concept Analysis (FCA) [14] is an unsupervised learning technique and also a conceptual clustering tool used for discovering conceptual structures of data.

A *formal context* is a triple k = (G, M, I) where G are objects, M are attributes and I is a binary relation between G and M, where $I \subseteq G \times M$.

For a set of objects $A \subseteq G$, we define A' := {m ∈ M | (g, m) ∈ I for ∀ g ∈ A} and for a set of attributes $B \subseteq M$, we define B' := {g ∈ G | (g, m) ∈ I for ∀ m ∈ B}.

A *formal concept*, a pair (A, B) is a formal context k if and only if $A \subseteq G$, $B \subseteq M$, A' = B and B' = A, where A is the extent and B is the intent of the concept (A, B). The subconcept - superconcept relation of the concepts of k is defined by (A1, B1) ≤ (A2, B2):⇔ A1 ⊆ A2 (⇔ B2 ⊆ B1). The set of all formal concepts k is called concept lattice and is denoted by βk.

4.4 Self-Organizing Map

An unsupervised neural network clustering tool, the self-organizing map (SOM) [15] is used to compress complex and high-dimensional data to lower-dimensional data (2-dimensional grid) according to similarity. More similar data are grouped together in the same cluster.

The SOM is trained recursively until it converges to a stable state. For each input vector x, it is compared with the entire model vectors, where the distance between the model vectors and input vector x is computed. The nearest model vector to the input vector x is the best-matching unit (BMU) on the map, which is denoted as

$$c = \| x - m_c \| = \min_i \{\| x - m_i \|\} \quad (2)$$

Consequently, the model vectors are updated.

4.5 K-Means Clustering

A simple unsupervised clustering technique, k-means clustering [15] divides a data set to a number of clusters. It is defined as

$$E = \sum_{k=1}^{C} \sum_{x \in Q_k} \| x - c_k \|^2 \quad (3)$$

where C is the number of clusters, x is a data point, and c_k is the centroid of the data points k.

To compute the optimal number of clusters C for the data set, the Davies-Bouldin validity index [16] is used to validate each of it. The optimal k is derived from,

$$\frac{1}{c} \sum_{k=1}^{C} \max_{l \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\} \quad (4)$$

where C is the number of clusters, S_c is the average distance of all data points from the cluster to their centroid and $d_{ce}(Q_k, Q_l)$ is distance between centroids.

5 The OntoShare Architecture

The prototypical implementation of the automated mapping and merging framework as illustrated in Fig. 1 as explained in this section.

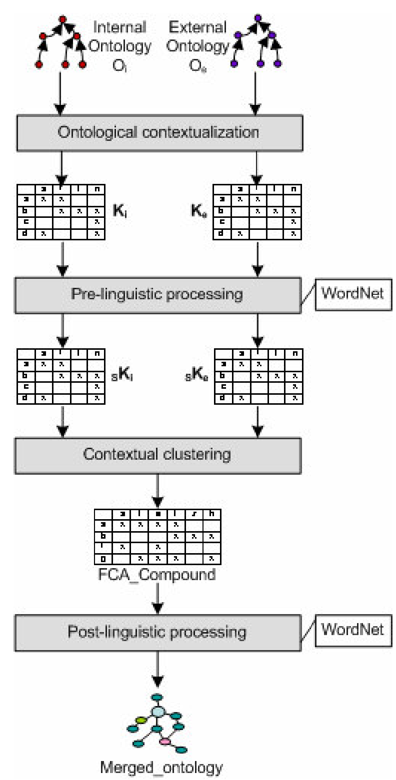

Fig. 1. OntoShare framework for ontology mapping and merging

O_i represents the internal ontology of local learning object repository and O_e the external ontology of non-local learning object repository. The overall process for mapping and merging ontology is outlined from steps 1 to 4.

Step 1: Ontological contextualization – O_i and O_e are contextualized using FCA with respect to the formal context for each ontology, K_i and K_e. Given an ontology $O := (C, S_C, R, S_R, is\_a, A)$, the ontological concepts C is denoted as G (objects) and the rest of the ontology elements, S_C, R, S_R, is_a and A are denoted as M (attributes). The binary relation $I \subseteq G \times M$ of the formal context denotes the ontology elements, S_C, R, S_R, is_a, and A corresponds to the ontological concepts C.

Step 2: Pre-linguistic processing – WordNet based on Leacock-Chodorow measure is applied to discover the semantics between both formal contexts' intents, K_i and K_e to standardize the intents of the formal contexts. Stop words filtering and tokenization are applied to transform the input for semantic similarity measurement between intents. This is followed by standardization of the ontological attributes whereby the attributes or intents are rearranged in order. Duplicated intents in each formal context are pruned. The standardized formal contexts ${}_{s}K_i$ and ${}_{s}K_e$ are computed as inputs for the next step.

Step 3: Contextual clustering – Initially, the standardized formal context, ${}_{s}K_i$ of the internal ontology O_i is presented to SOM to discover the intrinsic relationship between ontological concepts. Subsequently, k-means clustering is applied on the trained SOM to reduce the problem size of the SOM cluster to the most optimal number of k clusters based on the Davies-Bouldin validity index. Finally, the standardized formal context, ${}_{s}K_e$ of the external ontology O_e is clustered by SOM's BMU into its appropriate cluster without need for prior knowledge of internal ontological concepts. The outcome is a compound of both standardized formal contexts, *FCA_compound*.

Step 4: Post-linguistic processing – Semantic similarity measure using WordNet based on the Leacock-Chodorow measure is applied to discover the semantics between the extents in *FCA_compound*. Stop words filtering and tokenization is firstly used to transform the extents of the formal context into suitable representation for semantic similarity measure. The duplicated extents in the formal context are automatically pruned by maintaining the internal ontological concepts and structure. The binary relations $I \subseteq G \times M$ of duplicated extents are merged and the inheritance of superconcept-subconcept relations are updated.

Lastly, the merged ontology is computed from the context.

6 Simulation Results

For simulation purposes, the ontological concept *publication* from the two ontologies to be merged is illustrated in Fig. 2 (internal ontology) [17] and Fig. 3 (external ontology) [18]. The ontologies are visualized as concept lattices using the FCA tool, ConExp [19]. The internal ontology consists of 12 ontological concepts with 14 ontological properties. Meanwhile the external ontology consists of 12 ontological concepts with 30 ontological properties.

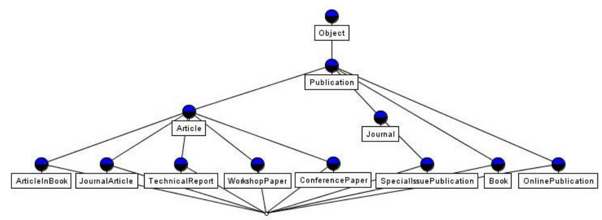

Fig. 2. Internal ontology

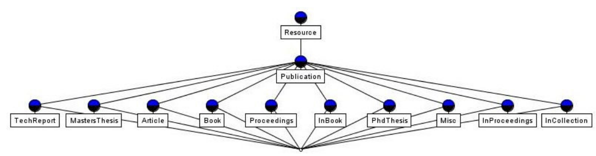

Fig. 3. External ontology

Initially, at the *ontological contextualization* phase, the ontologies, O_i and O_e are conceptualized using FCA into K_i and K_e. Pre-linguistic processing using WordNet based on Leacock-Chodorow measure at threshold is used to standardize the intents of the contexts.

At the *contextual clustering* stage, the standardized formal context, ${}_{s}K_e$ of the internal ontology, O_i is fed into SOM and k-means. The optimal number of clusters k is 3 as validated by the Davies-Bouldin validity index with db-index, 0.8238 and the sum of squares error, 2.3230. The clustered ontological concepts for the internal ontology are depicted in Fig. 4.

Subsequently, the standardized formal context, ${}_{s}K_e$ of the external ontology, O_e is fed into the trained SOM to discover semantics between internal ontological concepts and external ontological concepts and new ontological concepts. The ontological concepts are clustered into the most similar cluster by SOM's BMU as illustrated in Fig. 5. The external ontological concepts are clustered in cluster 1 and 2 as boxed in Fig. 5. Most of the new concepts are clustered in cluster 2. FCA compound is constructed from clusters 1 and 2.

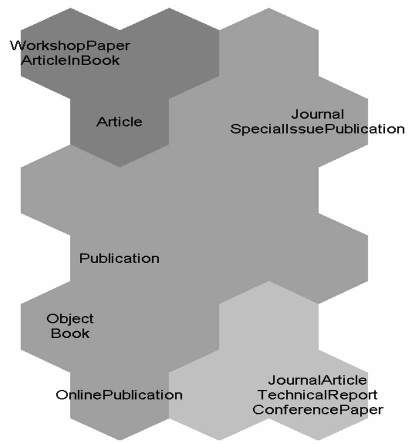

Fig. 4. Clusters of internal ontology

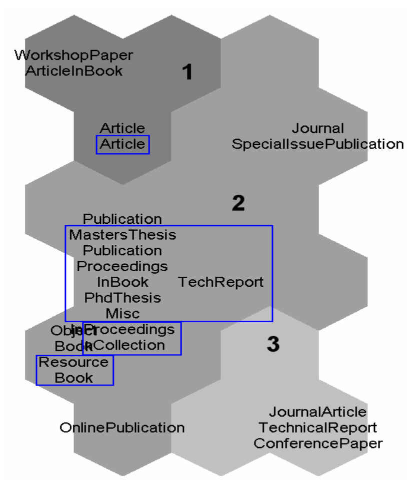

Fig. 5. Clusters of external ontological concepts

Post-linguistic processing is applied to prune the duplicated ontological concepts in the *FCA_Compound*. There are four duplicated ontological concepts: *Article, Book, Publication* and *TechReport*. SOM discovered eight new ontological concepts -- *InBook, InCollection, InProdeedings, MastersThesis, Misc, PhdThesis, Proceedings* and *Resource*. For the new ontological concepts, they will be dynamically updated into the internal ontology with their ontological properties. The *is-a* relationship and superconcept-subconcept relationship between the concepts are updated, whereas the subconcepts will inherit new ontological properties from the superconcepts. The merged ontology is illustrated in Fig. 6.

It is noted that obtaining a suitable threshold value requires multiple trials. Future work will involve the use of the Takagi-Sugeno fuzzy model [20] to model the dynamic linguistic on-line processing based on the ontology engineer's defined rules to function as a validity index for similarity measurement in WordNet.

7 Conclusion

We have proposed the OntoShare architecture for automated ontology mapping and merging and for dynamic update of the new ontological concepts to enable interoperability between ontologies for learning object reuse and sharing. The advantage of using unsupervised clustering techniques has been shown to increase robustness in the ontology mapping and merging process. Furthermore, prior knowledge is not required. Results can be extended to ontology mapping for resources in databases and data warehouses.

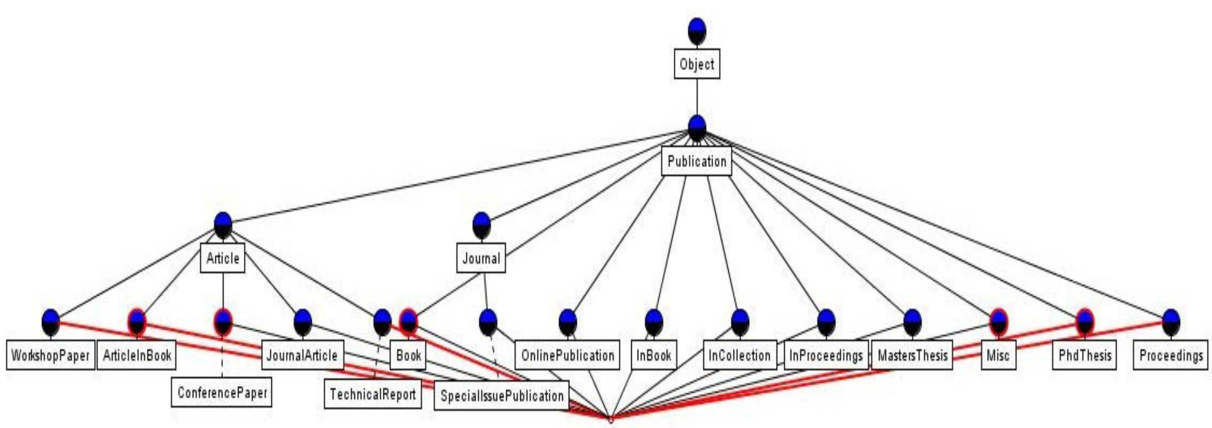

Fig.6. Merged Ontology

References:

[1] J. Qin and C. Finneran, Ontological representation for learning objects, *In Proceedings of the Workshop on Documents Search Interface Design and Intelligent Access in Large-Scale Collections*, *JCDL'02,* Portland, OR, 18 July 2002.

[2] D. Sampson, C. Karagiannidis, and Kinshuk, Personalized Learning: Educational, Technology and Standardization Perspective," *Interactive Educational Multimedia*, Vol. 4, 2002, pp. 24-39.

[3] C. C. Kiu and C. S. Lee, Discovering Ontological Semantics for Reuse and Sharing of Learning Objects in a Contextual Learning Environment, *The 5th International Conference on Advanced Learning Technology (ICALT2005)*, 8-9 July 2005, Kaohsiung, Taiwan.

[4] N. F. Noy and M. Musen, PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment, *In Proceedings of the 17th National Conference on Artificial Intelligence (AAAI'00)*, Austin, TX, USA, 2000.

[5] G. Stumme, Conceptual Knowledge Discovery: Data Mining with Formal Concept Analysis, *ECML/PKDD 2002*, Helsinki, Finland, 19-23 August 2002.

[6] M. Fernández-López and A. Gémez-Pérez, The Integration of ONTOCLEAN in WEBODE, *EKAW2002 Workshop on Evaluation of Ontology-Based Tools (EON2002)*, Sigüenza, Spain, 30 September 2002.

[7] C. S. Lee and Y. P. Singh, "Student modeling using principal component analysis of SOM clusters," *International conference on Advanced Learning Technologies (ICALT 2004)*, IEEE Computer Society Press, Joensuu, Finland, August 30-September 1, 2004

[8] C. S. Lee and W. C. Lim, Ontology-based 3-d visualization of composited concepts, *World Conference on Educational Hypermedia, Multimedia and Telecommunication*, 27 June-2 July 2005.

[9] M. Ehrig and Y. Sure, Ontology Mapping - An Integrated Approach, *In Christoph Bussler, John Davis, Dieter Fensel, Rudi Studer, Proceedings of the First European Semantic Web Symposium, volume 3053 of Lecture Notes in Computer Science, Springer Verlag,* Heraklion, Greece, 2004, pp. 76-91.

[10] S. Patwardhan, S. Banerjee and T. Pedersen, Using Measures of Semantic Relatedness for Word Sense Disambiguation, *The Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, 17-21 February 2003.

[11] N. Seco, T. Veale and J. Hayes, An intrinsic information content metric for semantic similarity in WordNet, *In Proceedings of ECAI'2004. The 16th European Conference on Artificial Intelligence*. Valencia, Spain, 2004, pp. 1089–1090,

[12] C. Leacock and M. Chodorow, Combining local context and WordNet similarity for word sense identification*, In Fellbaum, C., ed., WordNet: An electronic lexical database. MIT Press*, 1998, pp. 265–283.

[13] M. Warin, H. Oxhammar and M. Volk, Enriching an Ontology with WordNet based on Similarity Measures, *2nd Workshop organized by the MEANING Project (MEANING-2005)*, Trento, Italy, 3-4 February 2005.

[14] B. Ganter and R. Wille, Applied Lattice Theory: Formal Concept Analysis, 1997, *http://www.math.tudresden.de/~ganter/psfiles/concept.ps*

[15] J. Vesanto and E. Alhoniemi, Clustering of the Self-Organizing Map, *IEEE Transactions on Neural Networks*, Vol. 11, No. 3, 2000, pp. 586-600.

[16] J. C. Bezdek and N. R. Pal, *IEEE Transactions on Systems, Man & Cybernetics – Part B: Cybernetics,* Vol. 28, No. 39, pp. 301-315.

[17] I. Horrocks, Defines concepts from academic research, *http://protege.stanford.edu/plugins/owl/owl-library/ka.owl.*

[18] UMBC eBiquity, eBiquity Publication Ontology, *http://ebiquity.umbc.edu/v2.1/ontology/publication.owl*

[19] S. Yevtushenko, System of Data Analysis Concept Explorer (in Russian), *In Proceedings of the 7th National Conference on Artificial Intelligence (KII-2000)*, Russia, 2000, pp. 127-134.

[20] N. E. Mastorakis, Modeling Dynamical Systems via the Takagi-Sugeno Fuzzy Model, *WSEAS Transactions on Systems*, Vol. 3, No. 2, April 2004, pp.668-676.