# An Ontological Approach for Semantic Learning Objects Interoperability

Ming Che Lee, Kun Hua Tsai, Tung Cheng Hsieh, Ti Kai Chiu and Tzone I Wang
*Laboratory of Intelligent Network Applications, Department of Engineering Science*
*National Cheng Kung University, Taiwan*
*limingche@hotmail.com, tsaikunhua@msn.com, n9695414@mail.ncku.edu.tw,*
*n98951069@mail.ncku.edu.tw, wti535@mail.ncku.edu.tw*

## Abstract

*This paper presents a semantic-aware classification algorithm that can leverage the interoperability among semantically heterogeneous learning object repositories using different ontologies. The proposed algorithm is to map sharable learning objects, using meanings instead of just keyword matching, from heterogeneous repositories into a local knowledge base (an e-learning ontology). Significance of this research lies in the semantic inferring rules for learning objects classification as well as the full automatic processing and self-optimizing capability. This approach is sufficiently generic to be embedded into other e-learning platforms for semantic interoperability among learning object repositories. Focused on digital learning material and contrasted to other traditional classification technologies, the proposed approach has experimentally demonstrated significantly improvement in performance.*

## 1. Introduction and Motivation

Following the expeditious development of the Internet, particularly on web page interaction technology, distant e-learning has become more and more appreciated and popular. E-learning platforms emerge rapidly with their proprietary materials. In the long run, learning objects sharing and reusing among different repositories will become a trend but is still in a chaotic status. To solve the problems arising from sharing and reusing learning objects among heterogeneous repositories, a standard formats must be established. Many, including SCORM & SCORM LOM [1], IMS & IMS DRI [2], AICC [3], and etc., have been proposed by several international organizations. The Learning Object Metadata (LOM) has been approved by the IEEE-Standards Association, which is to provide structured descriptions of reusable digital contents, the "Learning Objects" (LOs). LOM is one part of the SCORM suite of standards, which is a standard for encoding general information on learning objects [1] and is meeting with universal acceptance.

The issue of semantic interoperability among learning object repositories and other learning resources on the Internet is increasingly pointing towards Semantic Web technologies in general and ontology in particular as a solution [14]. While ontology defines an explicit formal specification of knowledge of a specific domain, the effectiveness of learning objects interoperability among heterogeneous repositories are often reduced due to the use of various different ontological schemes to annotate learning objects in each learning object repository [15]. This paper will present a semantic-aware classification algorithm that can leverage the interoperability among semantically heterogeneous learning object repositories using different ontologies. The approach used is to map sharable learning objects from heterogeneous learning object repositories to local knowledge base (e-learning ontology) using their meanings instead of just keyword matching. Significance of this research is in the semantic inferring rules for learning objects classification and the self-optimizing capability. The rest of the paper is organized as follows: Section 2 describes the proposed semantic learning objects interoperability framework. Section 3 describes the details of the algorithm. Section 4 is the performance evaluation and the final section gives the conclusions.

## 2. Learning Objects Classification Framework

Figure 1 shows the proposed automatic learning objects classification framework. Domain Ontologies Repository stores domain ontologies (e.g. course ontologies, such as JAVA course ontology[4], ACM Computing Classification Ontology [5], and etc.), built by local domain experts. The framework uses formal ontology representation languages, such as RDF/RDFS [6], OWL [7], and DAML+OIL [8]. Domain experts and system managers can add, delete, modify, and update ontologies via a manager interface. Local Knowledge Repository records the classification information of learning objects.

Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)
0-7695-2916-X/07 $25.00 

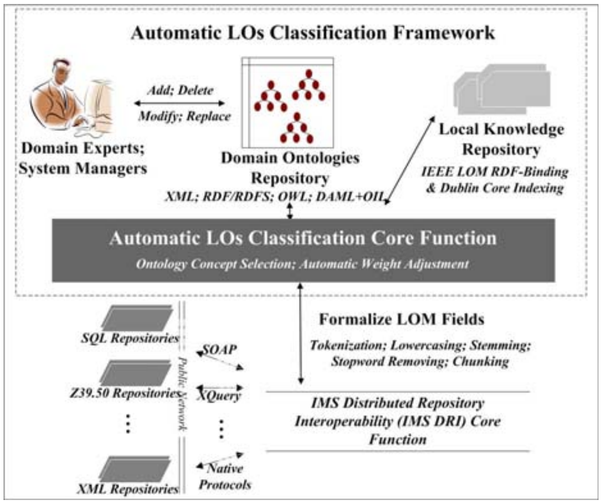


Fig. 1 The automatic learning objects classification framework

To achieve semantic interoperability of learning object repositories, the retrieved learning objects from IMS DRI [2] core function is first pre-processed by a LOM formalizing procedure that performs two functionalities:

(1) *Pre-process Words and Phrases.* This first function performs tokenization, lowercasing, stopword removing, and stemming. A learning object metadata can be treated as a simple document and there are many useless terms in classification process, such as definite article, preposition, and etc. A stop words list is referred when removing those terms. Stemming reduces inflected (or sometimes derived) words to their stem, base or root form. For example, the words ended with "ed", "ing", or "ly", are removed. This research uses the Porter's stemming algorithm [9].

(2) *Classify the terms in the LOM fields into two sets – the Important Term Set (**ITS**) and the Assistant Term Set (**ATS**).* In LOM, each field records specific information of learning objects, for example the "*Rights (category 6 in LOM V1.0)*" records the copyright information and the "*Description (category 1.4 in LOM V1.0)*" gives a brief content description. It is hard to say which field records more important information than the others. According to the purpose of this research, some fields indeed provide more information while classifying them into the course ontology. In the proposed framework, **ITS** includes *"Title (1.2)"*, *"Description (1.4)"*, *"keywords (1.5)"*, and *"Classification (9)"*, while **ATS** includes *"Lifecycle (category 2)*", *"Meta-metadata (category 3)"*, *"Technical (category 4)"*, *"Educational (category 5)"*, *"Rights (category 6)"*, *"Relation (category 7)"*, and *"Annotation (category 8)"*.

In the classification, extracted terms of each LOM field are constructed into a vector for subsequent evaluation, in which the **ITS** has a higher weight than the **ATS**. A vector is a two part structure, including *Term* and *Frequency*. The "*Term*" records the term that match a concept in the ontology. The "*Frequency*" records how many times a term appears in the field. Figure 2 shows a sample of a *Description Term Vector* and part of the metadata of a learning object for object oriented programming.

```
<description>
    <langstring>
     often an object of one class "is-an" object of
     another class as well. A rectangle certainly is a
     quadrilateral (as are squares, parallelograms and
     trapezoids). Thus, class Rectangle can be said to
     inherit from class. Quadrilateral. In this context,
     class Quadrilateral is a superclass, and class
     Rectangle is a subclass. A rectangle is a specific
     type of quadrilateral, but it is incorrect to claim
     that a quadrilateral is a rectangle (the
     quadrilateral could be a parallelogram).
    </langstring>
</description>
```

| Term | object | inherit | class | superclass | subclass |
|---|---|---|---|---|---|
| Frequency | 2 | 1 | 6 | 1 | 1 |

*Description Term Vector*

Fig. 2 A sample of the Description Term Vector

In this sample, "*object*" matches an ontology concept and appears two times in the LOM, "*inherit*" matches and appears one time, and etc. The information recorded in the vectors is used to calculate the matching degree of an ontology. The detailed process will be described in the following section.

## 3. Ontology Concept Selection & Automatic Weight Adjustment Core Function

The proposed approach includes two main functions: (1) Ontology Concept Selection, and (2) Automatic Weight Adjustment. The former determines which concept in the ontology a learning object should be associated with, i.e. as an instance to the concept. The latter adjusts the weight of terms of a concept via a learning mechanism. In the ontology, each concept is described by a set of terms, and each term would have different weight in the process of classification. Fig. 3 illustrates a part of the Introductory Java Course Ontology [4]. As can be seen, the concept "*Class*" is described by the set of terms {*class, member, field, instance*}, the concept "*Superclass and Subclass*" is described by the set of terms {*class, final, super, subclass, superclass*}, and etc. The formal definitions are given as follows.

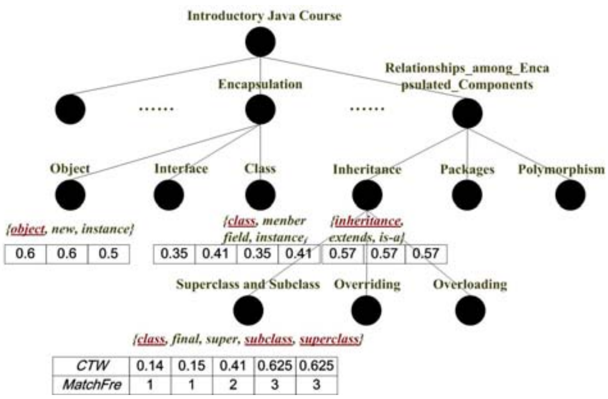


Fig. 3. A sample of Java course ontology and weight evaluating.

**Definition 1. Basic Concept Score - BCS**

$$BCS^i = \sum_{j=1}^{n} (TW + KW + DW + CW) \times CTW_j^i \tag{1}$$

where

$$TW = \alpha \times \text{Frequency in Title Term Vector}$$
$$KW = \beta \times \text{Frequency in Keyword Term Vector}$$
$$DW = \gamma \times \text{Frequency in Description Term Vector}$$
$$CW = \delta \times \text{Frequency in Classification Term Vector}$$

As mentioned, a concept in the ontology is described by a set of terms with different significance. For the LOM of a learning object, if at least one of the terms in ***ITS*** matches a term describing concept **i**, then concept **i** is referred to as a Basic Concept. The Basic Concept Score given by Eq. (1) thus represents the degree of the semantic similarity between the LOM and concept **i**. The values of α, β, γ, and δ parameters will be optimized in later experiments. The $CTW_j^i$ in Eq. (1) is the "**C**oncept **T**erm **W**eight" of concept **i**, which represents the significance of term j in concept **i**. Its value is obtained as follows by Eq. (2-4).

$$MTF_{CT_j^i} = \frac{MatchFre_{CT_j^i}}{\sum_{x=1}^{n} MatchFre_{CT_x^i}} \tag{2}$$

$$ICF_{CT_j^i} = \log \frac{\#\text{ of concepts in ontology}}{\#\text{ of concepts that have } CT_j^i} \tag{3}$$

$$CTW_j^i = \frac{MTF_{CT_j^i} \times ICF_{CT_j^i}}{\sqrt{\sum_{x=1}^{n} (MTF_{CT_x^i} \times ICF_{CT_x^i})^2}} \tag{4}$$

The CTW is a modified TF-IDF measure [10], referred to as the MTF-ICF measure. The corresponding idea is that the learning objects having been classified are treated as a training set, and concept terms' weights are increased if the learning objects are classified according to the terms.

A **Match**ed **Fre**quency Array, the **MatchFre**, is used to record the above-mentioned information for each concept. The initial value of each element in MatchFre is set as 1. For a concept, into which if a learning object is classified according to any concept term, the corresponding value in MatchFre is increased. Eq. (2) is to calculate the Matched Term Frequency (MTF) for a term **j** in a concept **i** and normalizes the value space to between 0 and 1, via dividing the MatchFre by the summation of all MatchFre of concept i. Eq. (3) calculates the Inversed Concept Frequency (ICF) for term **j** in concept **i**. Higher MTF means that a concept term has higher weight in the concept while Lower ICF means that a concept term appears in many concepts and results in a lower significance for the specific concept. Eq. (4) is the normalized MTF-ICF weight and the value is also limited to between 0 and 1.

The followings illustrates the calculations of a real case in the Introductory Java Course Ontology [4] that has 170 concepts. As seen in Fig. 3, the MatchFre of "*superclass*" and "*subclass*" are both 3. The Matched Term Frequency (MTF) of "*superclass*" and "*subclass*" is thus:

$$MTF_{CT_j^i} = \frac{MatchFre_{CT_j^i}}{\sum_{x=1}^{n} MatchFre_{CT_x^i}} = \frac{3}{10} = 0.3,$$

the Inversed Concept Frequency (ICF) is:

$$ICF_{CT_j^i} = \log \frac{\#\text{ of concepts in ontology}}{\#\text{ of concepts that have } CT_j^i} = \log \frac{170}{1} = 2.23$$

and the Concept Term Weight (CTW) is:

$$CTW_j^i = \frac{MTF_{CT_j^i} \times ICF_{CT_j^i}}{\sqrt{\sum_{x=1}^{n} (MTF_{CT_x^i} \times ICF_{CT_x^i})^2}} = \frac{0.3 \times 2.23}{1.069} = 0.625$$

**Definition 2. Assistant Term Match Score - ATMS**

ATS uses Eq. (5) to calculate the matching score of ATMS, which represents the matching degree of assistant term set of inputted LOM.

$$ATMS^i = \sum_{j=1}^{n} ATW \times CTW_j^i \tag{5}$$

where $ATW = \sigma \times \text{Frequency in Assistant Term Vector}$

Similar to α, β, γ, and δ of Eq. (1), the parameter σ is to adjust the weight of ATS, and the value of σ will be determined by experiments.

After the BCS and ATMS are calculated, two more scores, the Candidate Concept Score (CCS) and the Normalized Candidate Concept Score (NCCS) are calculated for each concept by Eq. (6) and Eq. (7). CCS is the summation of BCS and ATMS, and NCCS is the normalized value of CCS, which is normalized to 0~1.

$$CCS^{i} = BCS^{i} + ATMS^{i} \tag{6}$$

And

$$NCCS^{i} = \frac{CCS^{i}}{\sqrt{(\sum_{j=1}^{n} CCS^{j})^{2}}} \tag{7}$$

**Definition 3. Hierarchy Impact Score - HIS**
To reduce the polysemy problem (for example, in figure 3, the term "class" appears in different concepts and have different meanings), if two Basic Concepts are in a parent-children relation, they will weight each other by the following formula:

$$HIS^{i} = \sum_{k=1}^{m} \frac{NCCS^{i} \times NCCS^{k}}{\text{number of hops from i to k}} \tag{8}$$

The purpose of this step is to emphasize the semantic integration of an ontology structure. The normalization of CCS (i.e. NCCS) is to limit the increasing level while multiplying two NCCSs.

**Definition 4. Total Concept Score - TCS**

$$TCS^{i} = NCCS^{i} + HIS^{i} \tag{9}$$

Finally, we can determine the Total Concept Score (TCS) for each concept. The concept with highest TCS is chosen for the input learning object.

## 4. Performance Evaluation

### 4.1 The Domain Ontology

An ontology designed for general purpose is called an upper ontology, such as the WORDNET and the ACM Computer Classification System (ACM CCS). While an ontology designed for a specific domain or for solving a specific problem is called a domain ontology. The ACM CCS is chosen as our computer science upper ontology. The 1998 version of the ACM CCS scheme, probably the most comprehensive classification and valid to date, is the key resource in the information and computer technology domain. Although many classifications are no longer used in the latest update, the full version that published in 1998 is still used in the evaluation for completeness.
An introductory java programming ontology, the JLOO, is used as the domain ontology in the experiments for the proposed framework [4]. This ontology is based on the Computing Curricula CC2001 of the ACM and IEEE/CS [11], and is designed for providing a Java learning object knowledge base. It is focused on atomic knowledge units of introductory Java programming, thus the Java APIs are not included. For a same reason, other advanced learning concepts, such as swing, networking, multi-thread, are also absent from it. This ontology includes totally 170 concepts and 94 relations currently. Please refer to [4] for more details.

### 4.2 Learning Objects

To evaluate the proposed approach, totally 926 learning objects are collected from various repositories and web sites (only metadata and links instead of physical files are collected). For evaluating interoperability, their metadata are transformed to conform to the format of IEEE LTSC 1484.12.1 RDF Binding of Learning Object Metadata version 1.0. The implementation of the IEEE-LOM RDF Binding used in this research was supported by [12]. The RDF representation in [12] is almost fully Dublin Core RDF compatible, in the sense that Dublin Core metadata constructed according to this binding can be understood directly by a Dublin Core-aware parser.

### 4.3 Experimental results

**Experiment 1. System Parameters Tuning**
Because of the lack of ATS in LOMs, this experiment is to evaluate and tune the parameters α, β, γ, and δ in ITS. To tune α, others are set as 0.5, and to tune β, α is set to the optimal value, others remain 0.5, and so on. Figure 4 shows the experimental results.

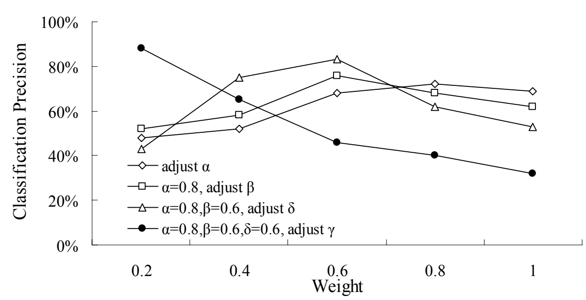


Fig. 4. Experimental results of experiment 1.

This experiment shows that setting α = 0.8, β = 0.6, γ = 0.2, and δ = 0.6 will get the best performance. This result suggests the field "title" has the most critical information describing a learning object.

**Experiment 2. Comparing with a traditional keyword-based classification approach and another ontology-based approach**
This experiment compares the proposed approach with a traditional vector space model (VSM) approach [10] and the Latifur's ontological approach [13]. A VSM builds all possible terms in the system and calculates the similarity

of two input documents by the following formula:

$$\text{Similarity}(D_i, D_j) = \frac{\vec{D}_i \bullet \vec{D}_j}{\left|\vec{D}_i\right| \times \left|\vec{D}_j\right|}$$

This experiment uses VSM to calculate the similarity of inputted LOMs and ontology concepts. For example, if all possible terms set of the ontology is {black, cup, coffee, green, tea}, the input $LOM_i$ is : {a cup of coffee and a cup of break tea} and concept-j is : {green tea}, then the "frequency vector" of $LOM_i$ and concept-j are {1,2,1,0,1}, and {0,0,0,1,1}, respectively. Thus, the similarity between $LOM_i$ and concept-j is: $\frac{1}{\sqrt{7} \times \sqrt{2}} = 0.267$

Latifur et al. [13] use an ontology base on the sport directory of the web portal Yahoo with audio files as instances. Conventional keyword-concepts mapping is used to determine candidate concepts. Confusions on terms like athlete names "Bryant Kobe", "Bryant Mark", and "Reeves Bryant", may happen in the classification. Relationships between these mapped concepts and others are consulted and a static disambiguation algorithm is proposed to prune those irrelevant concepts and allow relevant ones to associate with inputted documents. For comparison, this experiment also implements Latifur's approach that uses single LOMs as documents.

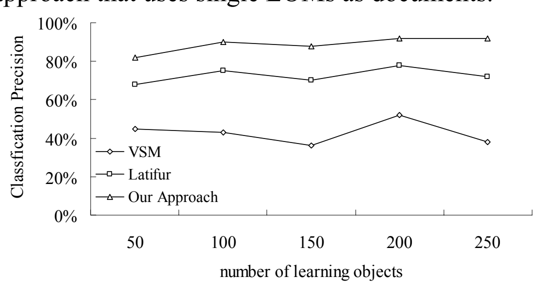


Fig. 5. Classification Precision of different approaches.

Figure 5 shows the experiment results. The classification precisions are determined manually. The results show that ontological approach can increase the precision between 30% and 50% averagely, and our approach performs still 20% better than Latifur's. This experiment demonstrates that using the uniqueness of the LOM's characteristics is a practical way to achieve learning object interoperability and the proposed approach can effectively remove much of the semantic ambiguities between LOs and the local knowledge base.

## 5. Conclusions

This paper presents a new semantic-aware classification algorithm that can leverage the interoperability among semantically heterogeneous learning object repositories using different ontologies. The proposed algorithm maps sharable learning objects collected from heterogeneous repositories into a local knowledge base (an e-learning ontology) using their inferred meanings instead of just keyword matching. Compared with other ontology-based approaches, the significance of the proposed approach lies in the semantic inferring rules for classification, the full automatic process and the self-optimizing capability that can adjust the weight of concept terms to optimal settings automatically. The power of the approach also comes from utilizing the LOM's unique characteristics instead of just handles document contexts. Experimental results also show that this approach can significantly increase the precision of classification and, in the mean time, remove semantic ambiguities.

## References


[1] SCORM (Sharable Courseware Object Reference Model), http://www.adlnet.org

[2] IMS Global Learning Consortium, Inc., "Instructional Management System (IMS)", http://www.imsglobal.org.

[3] Aviation Industry CBT Committee (AICC), "Computer Managed Instruction (CMI)", http://www.aicc.org/.

[4] Ming-Che Lee, D. Y. Ye, T. I. Wang, "Java Learning Object Ontology", The 5th IEEE International Conference on Advanced Learning Technologies, pp538-542, Kaohsiung, Taiwan, 5-8 July 2005.

[5] ACM Computing Classification System, https://www.acm.org/class/

[6] Lassila, O., and Swick, R. W3C Resource Description framework (RDF) Model and Syntax Specification. http://www.w3.org/TR/REC-rdfsyntax/

[7] OWL Web Ontology Language Reference. http://www.w3.org/TR/2004/REC-owl-ref-20040210/

[8] DAML+OIL Reference Description, http://www.w3.org/TR/daml+oil-reference

[9] The Porter Stemming Algorithm, http://www.tartarus.org/martin/PorterStemmer/.

[10] G. Salton, M. McGill, "Introduction to Modern Information Retrieval," McGraw-Hill New York, 1983.

[11] CC2001, http://www.sigcse.org/cc2001/

[12] IEEE-LOM. IEEE-LOM RDF Binding. http://kmr.nada.kth.se/

[13] Latifur, Khan, Dennis McLeod, Eduard, Hovy, Retrieval effectiveness of an ontology-based model for information selection, *The VLDB Journal*, 2004, pp. 71–85.

[14] T. B. Lee, J. Hendler, and O. Lassila, The semantic web. Scientific American, 2001.

[15] Aroyo, L., Dolog, P., Houben, G-J., Kravcik, M., Naeve, A., Nilsson, M. & Wild, F. (2006). Interoperability in Personalized Adaptive Learning. Educational Technology & Society, 9 (2), 4-18.




0-7695-2916-X/07 $25.00