Structured Metametadata Model to Augment Semantic Searching Algorithms in Learning Content Repository

Amirah ISMAIL
Faculty of Information Science & Technology,
Universiti Kebangsaan Malaysia, Bangi
43600 Selangor , Malaysia

Mike JOY
Department of Computer Science,
University of Warwick,
CV4 7AL, Coventry,
United Kingdom

Jane SINCLAIR
Department of Computer Science,
University of Warwick,
CV4 7AL, Coventry,
United Kingdom

And

Mohd Isa HAMZAH
Faculty of Education,
Universiti Kebangsaan Malaysia,
Bangi 43600 Selangor, Malaysia

ABSTRACT

This paper discusses on a novel technique for semantic searching and retrieval of information about learning materials. A novel structured metametadata model has been created to provide the foundation for a semantic search engine to extract, match and map queries to retrieve relevant results. Metametadata encapsulate metadata instances by using the properties and attributes provided by ontologies rather than describing learning objects. The use of ontological views assists the pedagogical content of metadata extracted from learning objects by using the control vocabularies as identified from the metametadata taxonomy. The use of metametadata (based on the metametadata taxonomy) supported by the ontologies have contributed towards a novel semantic searching mechanism. This research has presented a metametadata model for identifying semantics and describing learning objects in finer-grain detail that allows for intelligent and smart retrieval by automated search and retrieval software.

Keywords: Metadata, Metametadata, Semantic, Ontologies.

1. INTRODUCTION

World Wide Web is the *raison-d'être* for the hypertext format that the Internet supports. The current growth of the Internet has enabled access to very large amounts of information resources located in different and heterogeneous systems. Hypertext systems are meticulously practical for managing and browsing through large databases or corpora that comprise of disparate types of information.

Current research into frameworks and models of hypertext has entailed both the web infrastructure and embedded link structure. The Semantic Web [3] is an imposing vision that supports conveying metadata about resources in an explicit, understandable and machine-processable way for searching and organizational purposes.

In this era of the digital world of information, there are issues regarding searching and finding relevant and potentially useful learning materials related to users" needs. Web service technology allows a consistent access via web standards to software and applications on many computer platforms, and has supported the transformation from a static document collection to an intelligent and dynamic data integration environment.

Recently, new phrases have become common in this area of research, such as "Learning Objects","Learning Object Metadata" and "Learning Object Repositories". These terms have mainly been defined and applied due to their general meaning in the Educational Technology field and this is appropriate due to the interdisciplinary nature of the subject.

In this paper, we focus on metadata instead of learning objects themselves. Metadata is "structured data which describes the characteristics of a resource" [10]. Metadata can be described as structured information that describes resources or learning materials to support the searching, discovering and managing activities to display extracted information in some way.

Metadata created for educational elements which implicate general meaning across learning contexts and disciplines are open to explanation. For instance, in higher education, learning requires certain objectives to be achieved and the learner to be assessed. However, the main concern of instructors, designers, learners and academics is the nature of interactivity within a digital learning situation.

A learning object metadata file may include certain types of information or pedagogical attributes about the learning objects such as the creators' name, organizational connection, learning objectives, prerequisites and keywords. It may be based on the IEEE LOM schema on metadata and content packaging.

Metadata can be categorized depending on certain functions such as administrative, descriptive, technical usage, nature, technique of creation, category, structure, and semantics levels [6]. This also means that a few issues relating to Learning Objects, such as learning object management, creation, quality and granularity, will not be regarded as main topics for discussion, although certain requirements for handling the learning process and instructional theories in the field of E-Learning may be addressed.

This research work may be regarded as a test bed for presenting meta modelling languages, metadata sets, metadata organization and searching mechanisms with the help of *ontologies* for educational purposes.Ontologies outline the vital infrastructure of the Semantic Web [3].

This means that, as "ontology", any formalism will be considered within a well identified mathematical framework which supports user-defined relations and concepts and a subconcept taxonomy [4].

2. AIM AND RESEARCH NOVELTY

Research Aim

The major research question has been designed as follows:

- How can pedagogic metadata adaptation be handled effectively?

The aim of the current research is to explore, design and evaluate a model for describing and identifying the pedagogic semantic relationships of learning objects by using tagged metadata.

These could be expressed by generating educational metadata using a semantic search engine and novel reference mechanisms for semantic relationship metadata, later known as *Metametadata*, by using SCO (*Sharable Content Object*) to represent the learning objects, according to the SCORM (*Sharable Content Object Reference Model*[1].

The novel aspects of the research have been motivated by these essential points such as:

- The needs to extend the educational metadata elements to identify the semantic relationships between metadata tags for each learning objects or pedagogical resources.
- The needs to enhance and extend vocabularies specifically designed for pedagogical resources purposes to tailor the users' need.
- The requirement to design such a semantic taxonomy that contains metatagging instances by selecting the nearest similarity term based from the LOMv1.0 vocabulary in order to assist the semantic interoperability.
- To bridge the peculiarity within the LOM hierarchical conceptual data schema to present, manage and maintain the learning object repository in order to maintain the semantics of a LOM metadata instance.

Novelty of The Research

The novelty of this research is as follows:

- Novel Metametadata taxonomy has been developed which provides the basis for a semantic search engine to extract, match and map queries to retrieve relevant results.
- Search algorithms have been developed which include semantic search of capturing metadata instances which determine the relevancy of the retrieved results when measured against the search criteria.

The use of ontological views is a foundation for viewing the pedagogical content of the extracted metadata by using the pedagogical attributes from the metametadata Taxonomy. We classify the research of semantic search into five categories in accordance with their objectives, methodologies, and functionalities.

New scheme for presenting the metametadata is offered by applying the metadata development tool suitable for searching learning objects in the learning repository and the design and development of a novel taxonomy, a metametadata taxonomy.

Specifications

A principal motivation for using metametadata in the context of a pedagogic architecture which uses learning objects is that if the designer or administrator wishes to integrate metadata from various repositories or sources, the format and content of the metadata may vary considerably. A "high-level" view of the metadata, in the form of metametadata, will assist the process.

Metametadata are data about metadata which represent semantic relationships between items of metadata and between the metadata and one or more semantic domains. The relationships may be structural (physical and logical organization of metadata), behavioral (static or dynamic - change, view, modify semantics) or environmental (creator, revision history).

Metametadata will use higher-level definitional associative keywords, or vocabularies from documents describing content, to capture those relationships.

Metametadata are structured descriptions about a set of metadata which intelligently describe and capture relevant identified characteristic properties and relationships between metadata to aid locating, managing and retrieving data.

Metametadata are useful for the following purposes:

- providing sufficient information about metadata to enable intelligent searching;
- implementing flexible dynamic semantic mappings between metadata vocabularies;
- processing and displaying different explicit and implicit characteristics of the stored data sets;
- associating sets of related data by identifying semantic relationships between the associated metadata;
- providing consistent semantics and structures for metadata in the repositories or database schemas, browsing interfaces and presentation of content [8].

We should note that the distinction between metadata and metametadata may not always be simple. For example, a keyword may be used to tag a learning object, and if that keyword is unchanging it is clearly metadata. However, if a set of keywords might change (perhaps as a result of the use of the learning object), then they may reasonably be considered metametadata.

This is because the changes to the metadata are information about the metadata and about the context of the learning object, which may be categorized as environmental changes to the description of the original metadata but not information about the learning object itself. Figure 1 shows the changes in the environmental information for the original metadata.

```
Reference: 12345
 Contributor
        role: Creator
entity: Wiley, J. Contributor
        role: Validatorentity: Meta project
        date: 2007-08-08
 Contributor
       role: Publisherentity: Western
       University date: 2007-04-06
Educational
Intended end-user role: Learner; Author;
Teacher Learning context: Higher Education
       Typical age range: 17-25
Metadatascheme: IMS-IEEE LOM
Language: en
```

Fig. 1: Environmental Metametadata record

A classification scheme for pedagogic metametadata has been designed in order to provide a strong foundation for the future implementation of a pedagogic architecture supported by metametadata.

A pedagogical context for behavioral metametadata may be considered as a semantic structure or network whereby pedagogical entities are assembled. A pedagogical document contains a pedagogical context together with links such as prerequisites by connecting and describing the metadata and the metadata sources. For example, in Figure 2, behavioral metametadata may identify connectedness relations between certain learning objects and the contexts of those learning objects.

```
Reference: 12345
IsAPrerequisiteFor: 67890
UsedBy
  University: Warwick
  Module: CS456
UsedBy
  University: Birmingham
  Module: CS/200813
```

Fig. 2: Behavioral Metametadata

Behavioral metametadata can be considered as knowledge about the metadata itself, and can be used to express similarities between items of metadata. Metametadata formats are supported by IMS as CORBA and XML bindings, and in RDF. Structural metametadata can be used to specify the types of metadata for a particular information source.

Metadata can be extracted from template information sources, using structural metametadata instances. These information sources or learning objects are selected from a repository, according to a URL expression for each template source. The structure of the instantiated strongly typed metadata classes, along with their equivalent XML representation, is specified within the metametadata.

We therefore propose a taxonomy of metametadata in order to provide a common framework containing semantic definitions together with further contextual expression.

The work on the Metametadata taxonomy is focused on the identification of the required metadata elements consisting of *Class, Property* and *Representation*.

Metametadata Element Concept (MeMeC) = ObjectClass + PropertyMetametadata element (MeMe) =Metametadata Element Concept + [Representation]

Figure 3 presents the Metametadata Element Concept to view the relationship between metametadata element, *representation, object classes, property* and *value domain*. A class is a set of clearly defined ideas, abstractions, or "things" in the real world which have common behaviour and properties.

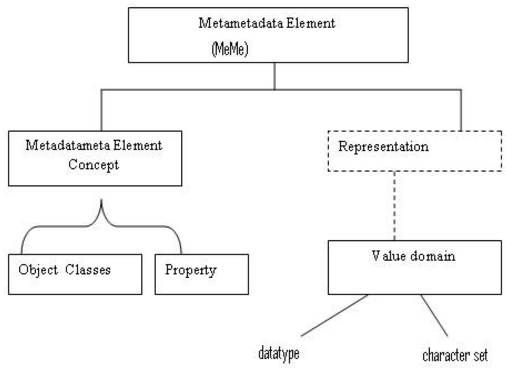

Fig.3:Metametadata Element Concept

A property is an attribute common to all members of a class. A representation of data describes a value domain, data type, and a character set. *Object classes* can be described as the entity (the thing") for other objects specialization. Specialization may permits object classes to be grouped and subtyped to help users browse and locate relevant object classes.

A property describes the particular characteristic or attribute of that entity. Examples for broadly defined object classes include *Person*, *programmer* and *organization* or specific object classes example such as *Client* or *Child. An object class* can be related with a *single parent object class*. A child object inherits features of its parent object class which may contain unique features.

The metametadata concept is based on pedagogical selection by having type- based logical representations that will be used as vocabularies the common kinds of learning object features. However, the educational category does not describe the significant connections or relationships between each of the following metadata elements: Interactivity type, Learning resource type, Interactivity level, Intended end user role, Context, Difficulty, Typical learning time, Description and Language of the typical intended user [7].

There is a need to improve the semantic relationships between metadata under the educational metadata category in LOM in order to improve learning object reusability. Therefore, it is necessary to find a semantic definition by describing each metametadata type that would link pedagogical aspects of chosen learning objects.

3. METAMETADATA TAXONOMY

We propose a taxonomy as shown in Figure 4 for pedagogic metametadata which uses the IEEE LOM metadata specification elements, together with key pedagogic characteristics, and metametadata elements for relational and classification purposes. The distinction between data and metadata is well understood, and metadata models may be described by classes, relationships and properties, known collectively as *types*. Our proposed taxonomy consists of a collection of types of metametadata, analogous to types of metadata, which we refer to as *connector.*

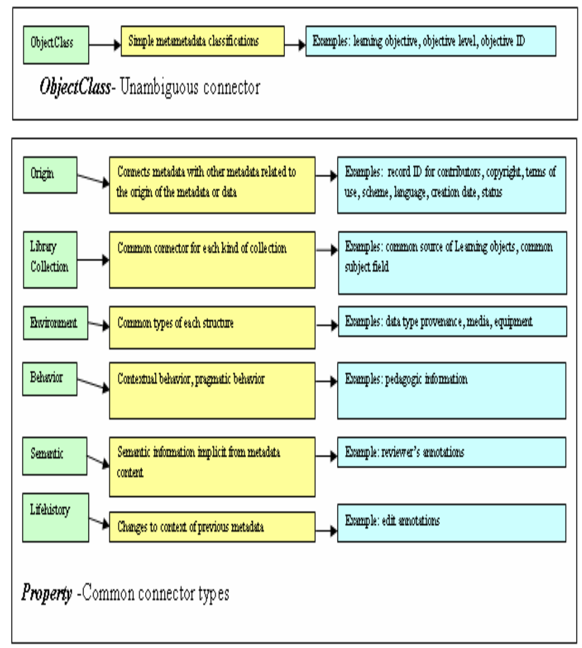

Fig. 4: ***Metametadata Element Concept (MeMeC)***

Figure 4 shows a proposed *Metametadata Element Concept (MeMeC)* to show the element commonalities that are able to provide an organized structure for interactors, and are by subdivided into two distinct categories, *unambiguous connectors* and *common connectors.*

1. *ObjectClass – Unambiguous connectors.* These areclassification metametadata, such as identifications for types of metadata which might be used for cataloguing purposes. There is only one type of unambiguous connector, which we refer to as the *Class* type of metametadata.
2. *Property – Common connectors:* These represent any instances of relationships between selected metadata and other metadata, for example, instances of all classes that may be connected by a generic form interface for displaying object data. We can subdivide these into six generic abstract classes that we refer to as *types* (based on the IEEE LOM educational metadata elements), as following.

- *Origin Type*: an attribute of the origin of the records. For example, two documents sharing a common author might use origin metametadata to store that relationship.
- *Library Collection Type*: information about commonality of a group of metadata. For example, the fact that a set of learning objects is sourced from a common repository might be represented by library collection metametadata.
- *Environment Type*: information about commonalities in the administrative or technical metadata. For example, a set of learning objects which share a common type of interface, which could be identified by the authoring tools (as specified in their metadata), would be linked by environment metametadata.

- *Behavior Type*: information about metadata behavior, such as contextual or pragmatic. For example, a set of learning objects which contains metadata indicating the cognitive abilities of the target students might be identified through behavior metametadata.
- *Semantic Type*: information about semantic content of metadata. For example, if a set of learning objects contains metadata which are reviews of each object, then a subset of those objects with positive reviews might be identified through semantic metametadata.
- *Lifehistory Type*: information about changes in metadata. For example, two Learning objects whose metadata had been edited at a similar time might be linked using lifehistory metametadata.

4. SYSTEM FRAMEWORK

The design of architecture, OMESCOD, is shown in Figure 5. The process and development of Metametadata commences with parsing the data that are the stored learning objects (documents), and metadata from the documents.

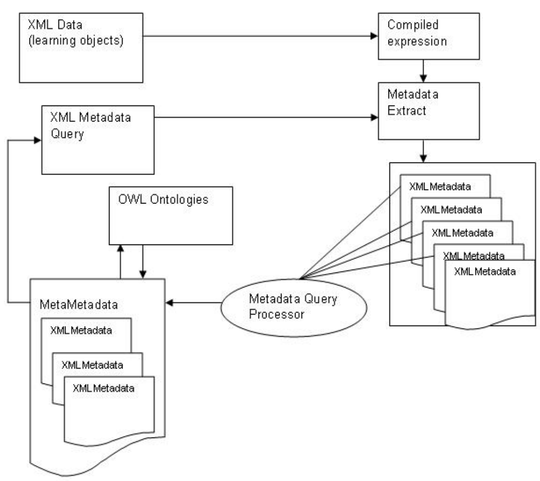

Fig. 5: The OMESCOD architecture work flows

Metadata are stored as XML, and correlate with data elements by matching the attribute ID in the data element, <metaRef> with specified <metaID>.

To increase the level of interoperability, it is possible to declare search types as shared OWL resources on the Web, which includes the required search Properties, and other possible relationships with other types of search. The ontology of types of queries may contain *machine-understandable* information for the processing of the results or for sending search elements.

Each identified relationship within the XML metadata is matched with the ontologies using Protégé-2000[9] as an ontology editing environment used to manage domain models and knowledge-modelling structures with ontologies.

This can be accomplished by firstly, identifying the domain and scope of the ontology by developing an initial small ontology of classes and slots. The classes and the class hierarchy of the can then be defined, followed by the learning objects content (domain) and the properties of classes by describing the internal structure of concepts as shown if Figure 6.

Vocabulary	Class / Property	Definition
Java_programming_introduction	Class	Introduction of java programming
Java_programming_origin	Class	The origin of java programming
Java_programming_basic_concept	Class	Basic concepts of java programming
Learning_arrays_functionalities	Class	Java programming functionalities
Methods_in_programming	Class	Input and output
Loops_in_programming	Class	Introduction to loop
Selection_in_programming	Class	Introduction to selection
Context	Class	Pedagogic attributes context
Level_difficulty	Class	Pedagogic attributes for level of difficulties
Learning_Objective	Class	Pedagogic Attributes for learning objective
Interactivity_level	class	Pedagogic Attributes for different level of interactivity
Prerequisite	class	Pedagogic Attributes for prerequisite
Loop	class	introduction of Loop
while-loop	class	while-loop
do-while-loop	class	do-while-loop
for-loop	class	for-loop
conditional_expressions_if	class	If
conditional_expressions_if-else	class	if-else
conditional_expressions_if-else-if	class	if-else-if
has Attribute	Object property	Based on Attribute

ventional parser

Fig. 6: The OMESCOD architecture work flows

5. QUERY SEARCH

Semantic Search

We present the semantic search method to evaluate the performance of metametadata and ontology searching by looking into two scenarios to utilize the semantic relationship between tagged metadata based on the Metametadata taxonomy(refer to Fig. 4).

The data set consists of XML documents that are used for querying by using keyword controlled vocabularies. A typical document may be a list of elements stored in specific domains.

Queries can be made through a simple keyword based search form, or can be submitted as SPARQL queries, optionally containing extensions that can specify the degree of confidence required for each term in the query.

Keyword based queries are expanded into SPARQL queries, so all searches use the same internal process. The most basic search is for a set of keywords, where the results will list ontologies containing all the keywords. The query can be made more specific by adding search directives to the query.

Our metametadata notion may support the process of creating such a semantic and implicit meaning for a tagged metadata by capturing the relationship between one item of metadata and another by their attributes to identify the relevant learning resource. Metadata can describe other metadata which may be used to view the semantic relationships between keywords or vocabulary concepts.

The use of metametadata definitions within our scope of research also focus on the method of supporting semantic algorithms which will exploit the terms or controlled vocabularies from the ontologies within the domain of learning Java Programming identified from the repository to search for relevant learning objects.

Each of a set of learning objects should contain tagged identifiers defined from the development of a novel metametadata taxonomy, so as to support a search based on the semantic relationship between each tagged metadata item for the learning objects, and using pedagogic attributes which would disclose the pedagogical contexts of the learning objects to the user.

Much recent works in educational technology area are more towards designing framework for adapting metadata while still lacking on the needs to augment the pedagogical values for metadata. This paper has focused on the designing and implementation of the novel Metametadata framework as part of solution to retrieve and achieve better relevant result for learning materials in computer science domain.

6. REFERENCES

[1] ADL. (2004). "Sharable Content Object Reference Model (SCORM)." Accessed on 10th August 2007. Retrieved from http://www.adlnet.org

[2] Berners-Lee, T. (1999). "The Semantic Toolbox: Building Semantics on Top of XMl-RDF." Accessed on 9th April 2007. Retrieved from http://www.w3.org/DesignIssues/Toolbox.html

[3] Berners-Lee, T., Hendler, J. and Lasila, O. (2001). "The Semantic Web." **Scientific American**. **284**(5): 35-43.

[4] Garshol, L. M. (2002). "Topic Maps in Content Management : The Rise of the Itms." Accessed on 10th November 2007. Retrieved from http://www.ontopia.net/topicmaps/materials/itms.html

[5] Garcia, R., Gil,R., Delgado,J. (2007).“A Web Ontologies Framework for Digital Rights Management.” Journal of Artificial Intelligence and Law, 15:137–154.

[6]Gilliland-Swetland, A. (2000). "Introduction to Metadata, Pathway to Digital Information."Accessed on 28th May2007.Retrieved from http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.html

[7] IEEE. (2002). "Standard for Learning Object Metadata." Accessed on 7th April 2007. Retrieved from http://ltsc.ieee.org/wg12

[8] Ismail, A. and Joy, M. (2007). "A Novel Approach for Metametadata Views in Learning Domains". ED-MEDIA - World Conference on Educational Multimedia, Hypermedia and Telecommunications, Vancouver, AACE

[9] Protege. (2007). "Protege." Accessed on 21st December 2007. Retrieved from http://protege.stanford.edu/

[10] Taylor,C. “An Introduction to Metadata”.(2003). Accessed on 11th January 2007. Retrieved from http://www.library.uq.edu.au/iad/ctmeta4.html