Review of Challenges in Content Extraction in Web Based Personalized Learning Content Management Systems

Lakshmi Sunil Prakash  
Faculty of Computing and IT  
Sohar University  
Sohar  
00968-99620430  
lakshmi@soharuni.edu.om

N.S.Kutti  
Dept of Computer Science  
College of Sciences  
SQU,Muscat  
00968-98065648  
nskutti@squ.edu.om

ASM Sajeev  
School of Science and Technology  
University of New England  
Armidale,NSW  
+61 2 6773 2285  
sajeev@turing.une.edu.au

ABSTRACT
A preliminary work in support of our research in developing ontology driven LO extraction analyzes some of the existing Learning Content Management Systems (LCMS) with emphasis on Architecture and Design Issues. Learning Object Repositories (LOR) have great impact on retrieval and storage of contents, so performance improvement issues via content aggregation in LORs have been studied in this paper. The paper discusses the challenges that are being faced by Web-Based LCMS with regards to content extraction.

Categories and Subject Descriptors
D.3.3 [E-learning]: E-learning design models and methods

General Terms
Design,Standardization.

Keywords
WBLCMS, Personalized Learning, content aggregation, metadata, standards, Learning Object Content Models,

1. INTRODUCTION
E-learning or Web-based learning is an educational paradigm that refers to utilization of web-based technologies to direct and supervise in online the learning activities that enhance knowledge and performance, “should enable learners to participate in the creation of new knowledge as a normal part of their lives” [15].

Currently available Course/ Learning management Systems (C/LMS) mainly track deploying course-related materials, student registration, student grades, student profile and other administrative requirements of instructor-led classes. They failed to provide authoring tools and as a result they do not support any database to manage LOR. C/ LMS generally do not include ways to create and deliver dynamically new contents. Hence they are not ready for creating personalized learning environment. A comprehensive solution to fulfill the needs of personalized and adaptive e-learning along with the economic benefits of reusable learning contents has been realized with the help of Learning Content Management System (LCMS) around a dynamic LOR. The current generation of Web-based Learning Management Systems (WBLMS) allows the separation of C/ LMS and LCMS into foreground and background functionality. These systems are able to support content interoperability and sharing of learning scenarios in a seamless and dynamic way. These systems support several useful features like content authoring, version control, user information, learning context, and sequencing. The evolutionary path of e-Learning platforms has led from the monolithic (first generation) to the customizable (current generation). This continued advancement of technologies is now leading to the development of semantic and service-oriented systems (new generation).

2. PERSONALISED C/ LMS
Most C/ LMS are e-Instruction portals that deliver course contents systematically to the learner, but the role of the learner is merely a receptor of knowledge. Future working environment needs individuals who have strong self-learning skills which in turn feeds into the process of critical thinking and problem-solving skill. This cognitive skill is used as a base for creating dynamically reasoning mechanisms during domain knowledge processing. E-commerce systems have successfully implemented recommender systems [25]. This success has encouraged experts to use them in e-Learning environment to provide an automatic process to support learners in finding suitable materials as an alternative to relying on classmates, tutors and other sources [23]. The advantage of personalized learning is that every time a learner personalizes the system, the system is able to re-adjust its content to fit with the learner’s unique set of needs and tune the environment appropriately [22].

3. CHALLENGES IN CONTENT AGGREGATION
Learning Content Designers and Subject Matter Experts deal with the creation of the contents in the LOR. These contents are often dispersed in the LOR and categorized in several ways. Most LCMS have very large repositories of Learning Content which are merely static units of learning or LOs. These LOs are available in various formats and they may not be unique in contents or description about a concept or a group of concepts. This means when a Course Module is being presented to the Learners, they may be weighted down by the sheer number of LOs. They will need to wade through a large browsing space to get to the LO that best suits their learning style. But, in the worst case, a learner may...
come out of the searching, downloading and reading process without gaining anything. This is the case of the eluding LO. A LCMS that tracks a user goals and knowledge-status can support users in their navigation by limiting browsing space, suggesting most relevant links to follow, or providing adaptive comments to visible links.

4. LO AND METADATA (LOM)
IEEE LOM defines a LO as "any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning" [8]. The LOM can include information such as author, date of creation and type of resource [17]. There are several standards that define the metadata standard for a LO, some well known ones are IEEE LTSC LOM [10], ADL [2], IMS [11]. This information makes a LO accessible facilitating its reusability, updatability and retrieving.

5. LEARNING CONTENT MANAGEMENT SYSTEMS (LCMS)
The function of LCMS revolves around LOR maintained in a central database where both documents themselves and their associated metadata are stored [17]. Major characteristics of LCMS include (1) availability of authoring tools to create new LOs for the repository, (2) the ability to assemble LOs "on the fly into "learning paths" that are personalized to a learner's profile (3) ability to track who takes what, (4) ability to provide functions of LMS in communicating with the learners, getting feedback from learners, creating discussion forum and tracking their progress. More sophisticated and open LCMS allow independent storage of learning content provides a transparent environment to use shareable and re-usable LOs. Metadata provides a means to categorize and index learning documents and to make their retrieval as easy as possible (M. G. Moore, 2001).

5.1 Learning Object Content Models
Learning Object Content models provide a framework for defining the structure namely the aggregation and granularity of LOs. The Reusable Learning website (2004), sponsored by the U.S. National Science Foundation (NSF), defines the aggregation level as "the degree to which a digital learning resource is made up of other digital learning resources. The higher the aggregation level the deeper the hierarchical structure." The term "granularity" is defined as "the size, decomposability and the extent to which a resource is intended to be used as part of a larger resource. More granular digital learning resources are larger and composed of smaller pieces" (Reusable Learning, 2004). Following is a review of Content Models with respect to aggregation levels and the metadata that they contain.

5.1.1 CISCO Systems with RLOCM
This model proposes 5 aggregation levels for structuring learning contents: Subtopic (small chunks of information), Topic (a self-contained Reusable Instruction Object), Lesson (RLO), Module (set of lessons) and Course (set of modules). A lesson can be reusable in multiple courses and learning contexts. A lesson or RLO (set of 5 to 9 RIOs) consists of a single learning objective, an overview, a summary, and a collection of topics (or RIO), as well as practice activities, assessment, and metadata. Topics are grouped into five category types, including concepts, facts, procedures, processes, and principles. Although Cisco treated a lesson as a LO, each aggregation level can be used as a LO [6]. This somehow makes possible to deliver different logical size LOs to train learners at different categories. This is a step forward towards personalized learning.

5.1.2 DNER & LO model
In this model [7] there are seven levels of aggregation of learning content, like Information Objects (raw data), Information resources (collection of raw data), LOs (aggregation of information resources), Unit of Study, Module. This model deals mainly with the creation, sharing, reusability, re-purposing and interoperability of learning content and object in the context of UK Higher Education. In this model the components are categorized as objects that serve information (information, objects, information resources and collections) and those having learning objective (LOs, units of study, modules and courses). The aggregation in this model is managed with the help of mere metadata in a static manner.

5.1.3 IEEE LSTC LOM Aggregation Model
In this model the learning contents are aggregated according to the following aggregation levels i.e. fragment, lesson, course, courses. However as mentioned in [9] the lack of clarity with regard to the use of aggregation levels and corresponding LOM terms, in this model, makes it difficult for the specific metadata elements and values to be used by metadata creators.

5.1.4 Learnativity Content Model
This model was developed by [13]. It represents five different levels of aggregation which are Content Assets, Information Objects, LOs, Learning Components (Lessons or Chapters) and Learning Environment (Courses or Collections). Though this model defines several layers of Metadata including content assets (Raw Media), Information Object, LO (Application Objects), Learning Component (Aggregation Assemblies) and Learning Environments (Collections) it does not have the ability to assemble LOs into "a learning paths" dynamically that is personalized to a learner's profile.

5.1.5 LCMS Model
This model proposed by [26], defines three aggregation models that include assets, LOs and learning units. In this model, the LOs are considered to be small enough for realizing effective reuse of them. The granularity of LOs here may allow reassembling suitable "learning paths".

5.1.6 NETg LO Model
This model is composed of four levels, organized by Topic (NETg LO), Lesson, Unit and Course. Each Topic consists of learning objective, activity item and assessment item. A Lesson contains many topics and a unit consists of several lessons. A course consists of a collection of independent units [12]. The advantage of this model is that each component (Topic, Lesson, Unit or Course) in this model are independent structures. Therefore topics, lessons and units can be easily embedded in different units. However, only the abstract structural model of these units is provided in this model. Again this model also does not have the ability to assemble LOs into "a learning paths" dynamically that is personalized to a learner's profile.
5.1.7 SCORM content model
This model contains three components [16] assets; sharable content objects (SCOs) and content organizations [2] SCORM components are not bound to a specific sequencing or navigation rules, but the activities are embedded with sequencing rules. The various SCORM components are incorporated using the IMS content packaging specification within a single content aggregation or unit of learning. [26] mentions that the SCORM content Models do not provide a clear aggregation level in that for example an asset component like a web page component may contain another asset component. However each component in the SCORM Content Model has a metadata i.e. the Asset Metadata, SCO Metadata, and Content Organization Metadata, Activity Metadata and Application Metadata. SCORM Metadata shares the same metadata as the IEEE LO. LORs employ these models to structure the individual components in the LORs. The Metadata layer in each LOR gives information of each LO. This facilitates searching and reusing these components. Searching for content across LORs requires approaches such as federated search and Metadata aggregation model. The above mentioned approach and its implications are to be investigated in future work.

5.1.8 Knowledge Puzzle Content Model
The model [1] aims to reproduce learning material from annotated content. Annotations are performed semi-automatically and are stored as knowledge fragments. These knowledge fragments are used as a knowledge base for a training environment which may be an e-Learning environment. The main objective of these annotations is to enable the automatic aggregation of Learning Knowledge Objects (LKOs) guided by instructional strategies, which are provided through SWRL rules. From these LKOs SCORM-compliant LOs will be generated. This model is quite different from the previous models in that ontology is used to extract and construct the learning contents.

5.2 CHALLENGES WITH FUTURE CONTENT EXTRACTION
The major challenges faced by Web Based Learning Content Management System are summarized in the following points. (1) In WBLMS we come across the following scenario where a learner is enrolled in a particular unit/course, he/she will need to complete a stipulated number of units within a certain period of time. If so then a Course must be populated with the right number of LOs, a good variety of LOs, well described LOs and most importantly the most suitable LO that should match his/her Learning Preferences. (2) The Learning Content though existent in the LO repository cannot reach the Learner as there is no mechanism in the LCMS that can deliver this LO as the most suitable for that particular Learner [4]. (3) The LCMS require that a tool be integrated which can facilitate the aggregation of only the most suitable LOs to a learner. This is possible only if the meanings of the content embedded in the system can be understood and the content is intelligent to find some content similar to itself. (4) This intelligence needs to be built in the system to enable the system in generating query-sensitive learning path. (5) Ultimately to facilitate the process of personalized learning it is important for the learner to be able to receive the most suitable set of LOs belonging to a particular domain.

5.3 CURRENT RESEARCH IN THE DOMAIN
In this section some of the research in the area have been analyzed and presented:
- The Knowledge Puzzle [1] an ontology-based platform designed to facilitate domain knowledge acquisition from textual documents for knowledge-based systems. This platform employs natural language processing and machine learning technologies to automatically generate domain ontology from documents’ content.
- The aim of the PSO (Particle Swarm Optimization) Algorithm developed by [17] is to design, develop and test a PSO agent that performs automatic LO sequencing through learner competencies. This model employs competencies as a means for defining constraints between LOs, so that a sequence of LOs is represented by relations among LOs and competencies. This model is also able to derive new sequences if permutation operations are allowed between LOs in sequences.
- In [28] a modular framework is proposed for an adaptive learning system that can segment and transform teaching materials into modular LOs based on the SCORM standard such that subject contents can be composed dynamically according to the profile and portfolio of individual students. This adaptive system retrieves a set of LOs with ascending degrees of difficulty. This is a dynamic set which will be subject to change depending on the result of various assessments that the learner will perform during the learning process. The main deficiency of this work is that the retrieval process will work in a LOR having multiple LOs with similar learning outcomes and similar content is not explored.
- Self-regulated learning assisted mechanisms are proposed by Chih-Ming Chen [5] in a novel personalized learning system. In this approach, these mechanisms help learners enhance their self-regulated abilities efficiently and assist them in becoming life-long learners who have autonomous self-regulated learning abilities.
- O-DEST project [27] proposes an ontology-driven e-Learning system. The system comprises ontology for the e-Learning process, such as course syllabus, teaching methods, learning activities, and learning styles. Ontologies for the knowledge domain specific teaching subjects will enable the usage of the teaching contents in the Semantic Web.
- In [21] the problem is tackled from the interface level by retrieving the learning pattern / style of the learner by using a set of Adaptive LO. This approach uses a 5-tier architectural model. In this model each LO structure is adapted to the learning styles. The ALO will have a well-defined metadata structure that can be used to customize the Learning processes.
- Recent research in LOs contribute for the search of patterns for instructional content development, in order to make them adaptive, generic, portable and scalable enough to improve their potential for reusability, automated LO assembly and effective search.

6. REFERENCES
Repositories, A Semantic Lifecycle Approach to Learning Object
M. A. Sicilia
Information 3:4. Accessed online on
and Abuse of Reusable Learning Objects. Journal o
Products, Their Structure and Core Beliefs. NetG
[Internet] Available from:
Cisco Systems white paper
based society Thailand,
36(5), 8816-8829,12 Information literacy for the knowledge-based
with the learning objects. Interdisciplinary Journal of
system based on roles and activities for thai learning
Schluep, S., Ravasio, P., & Sissel
Electronic Commerce. Routledge/F
Building”, In J.W.Guthrie (Ed.), Encyclopedia of Education
bs Gutiérrez J.2008. Swarm Intelligence in e-
Learning: A Learning Object Sequencing Agent based on
Objects. Open and Distance Learning Series. London, Routledge/Falmer.
Web-based Learning Objects. Proceedings of International Conference
on Web engineering, Oviedo,Asturias,Spain
American Journal of Distance Education, 2001.
through learning styles. . Interdisciplinary Journal of
EduLearn:An E-Learning Architecture for Prototyping Web-
Based Learning System, Proc of Intl Conf on Enterprise Inf
Sys and Web Technologies, (EISWT-07)., Orlando, USA.,
Recommender Applications Applications of Data Mining to
Electronic Commerce. Data Mining and Knowledge
Implementing learning content management. in Human-
Computer Interaction: INTERACT’03, Amstad.
system based on roles and activities for thai learning
environment. Interdisciplinary Journal of Knowledge and
C., & Tsai, J. (2008), An Object-Oriented Course
Educational Technology & Society, 11(2), 171-191. (SSCI)