2009 Ninth International Conference on Intelligent Systems Design and Applications

Discovering Learning Objects Usability Characteristics

Alfredo Zapata[1], Victor H. Menendez[2]
Facultad de Educación[1], Facultad de Matemáticas[2]
Universidad Autónoma de Yucatán
Mérida, Yucatán, México
{zgonzal, mdoming}@uady.mx

Manuel E. Prieto[3]
Escuela Superior de Informática[3]
Universidad de Castilla-La Mancha
Ciudad Real, Ciudad Real, España
manuel.prieto@uclm.es

***Abstract*—Metadata is the key to describe Learning Objects. Through them, we can search and reuse these resources. However, there are pedagogical and usability characteristics that metadata do not normally contain. Sources of additional information such as activity log registers in repositories can help to specify such attributes. Data mining techniques allow identifying Learning Objects usability characteristics. This paper presents the results of applying a knowledge extraction methodology to Learning Objects through the use of four data sources: metadata, pedagogical quality evaluations, user's profiles, and log files from Learning Objects management systems.**

Learning objects, data mining, knowledge extraction.

I. INTRODUCTION

Although there are many definitions, we understand Learning Object (LO), like any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning, for example, systems for computer-based instruction, e-learning systems, intelligent tutoring, etc. [9].

LO are made up of instructional content and metadata sets. These two elements form an exchange unit that can fulfill specific learning needs. Metadata are the Learning Objects differentiating elements because the natures of the object, its usage, semantics, etc, are defined through them.

The knowledge extraction in the LO has been an area explored from various point of views, for example, references [19] and [5] analyzed the text in the metadata to create groups based on the similarities calculation among learning objects.

Other important studies in the area are shown in reference [10], which focuses on ontologies development to facilitate the reuse and recovery of LO, and the work development in [12], which is based on web usage mining to discover association and sequence patterns based on the use of information for students.

The aim of this work is to extract specific usability information about Learning Objects. For this process, we present an adapted methodology for knowledge extraction and its application to obtain rules about usability characteristics of LOs. The method was applied over data sets obtained from the Learning Object Management System "AGORA".

The paper is structured in the following way: Section 2 presents the main study goals for the knowledge discovery process in Learning Objects. Section 3 describes the main elements related to this work: metadata, the Pedagogical Quality Evaluation Model and the LO Management System AGORA. Section 4 presents the knowledge extraction methodology used. Finally, conclusions and further research are outlined.

II. STUDY GOALS

The requirements of the study are specified as follows, including the goals and methodology used. The guide questions considered for knowledge discovery in Learning Objects are:

- What information sources are stored in Learning Objects?
- Which are the main characteristics of Learning Objects in relation to their search, reusability and sorting?
- Which data must be considered to be processed to discover important rules about Learning Objects usability in different environments?

III. SOME IMPORTANT CONCEPTS

Here we present some concepts related to Learning Objects and their significant data.

A. Learning Object Metadata

Ideally a Learning Object must have seven main characteristics [1]:

- Self-contained
- Interoperable
- Reusable for different contexts
- Durable and upgradeable over time
- Easy access and management
- Sequence with other objects in the same learning environment
- Concise and synthesized

Metadata contains primary and objective information about Learning Objects. They enable management, tracking and recovery [4]. Some examples of metadata fields are: keywords, learning objectives, aggregation level, prerequisites, author, date, language and version.

IEEE-LOM [8] is a standard, which describes a set of metatags used to represent metadata.

LOM-ES [6] is a Spanish application profile which contains several extensions, particularly new labels and

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.184

vocabularies. It is used in LOs classification according to a set of rules including taxonomies and thesauri that permits to specify, among others: discipline, idea, prerequisite, educational objective, accessibility restrictions and instructional or skill's levels.

These specifications are aimed to establishment LOs attributes in various aspects such as interoperability, accessibility and reusability. However, there are some pedagogical and technical features that are not covered by these standards. Examples of such uncovered aspects are: content quality, creativity, design and presentation.

B. The LOs management in AGORA

The AGORA project (from a Spanish acronym that means Help for the Management of Reusable Learning Objects) [14] aims to provide an infrastructure that supports the development of instructional design activity. Particularly it provides solutions for Learning Objects management. The main components of AGORA are:

- A knowledge base consisting of instructional ontologies. The initial operative version includes models of instructional design and a methodology for the population, editing and refinement of instructional engineering ontologies.
- A system for automatic discovery of knowledge about instructional design based on KDD techniques.
- A module for Learning Objects management. Its purpose is to catalogue, compose and process LOs based on accepted international standards. It includes mechanisms for automatic metadata generation and a repository management system.
- A meta-search engine specialized in e-Learning resources available in multiple repositories through a semantic approach that improves the chances to get relevant results according to the teacher's instructional needs.
- A method and a tool for pedagogical quality evaluation of digital learning resources MECOA (in Spanish means Quality Evaluation Model for Learning Objects) [13]. It is a model oriented to determine LOs quality from a pedagogical perspective. The purpose of this model is to evaluate Learning Objects through five main pedagogical dimensions: content, performance, competence, self-management, meaning and creativity. Each dimension in the model comprises a set of features defined through fuzzy labels.
- A Recommender System capable of supporting LOs design, search, recovery and reuse activities, based on the teacher´s resource development requirements and their own profiles. To meet this goal it is necessary to have a rules set that establish or suggest actions or elements. Works such as this provide an initial rules set and conditions that can be used to assist the teacher in their resources management. This module is still in development because it is an integrating component of the previous modules.

IV. KNOWLEDGE EXTRACTION METHODOLOGY

For knowledge extraction from the user activity registers in AGORA, we adapted the work referred by C. Romero et al. in [17]. The resulting methodology consists of several stages which are described below (Fig. 1):

A. Data recollection

We used a data set obtained from metadata stored in the AGORA repository. Educational aspects were covered with the quality evaluation indictors provided by experts using MECOA model also included in the AGORA platform.

Other important data source was the AGORA's registration log files. This files stores evidence of teachers' activities in using the various services provided by the platform: item generation, searches, evaluation, download and display of stored resources. The user's profiles which define characteristics and background in terms of their knowledge, skills and experience in various fields are also considered.

B. Data preprocessing

Data obtained in the previous phase undergo a transformation process in order to restructure them and to discard redundant or not useful information. This phase consists of several activities that are described below:

1) Select data: 300 Learning Objects of a total of 860 were chosen for this report. This selection was based on the metadata completeness and their nature, with the purpose of

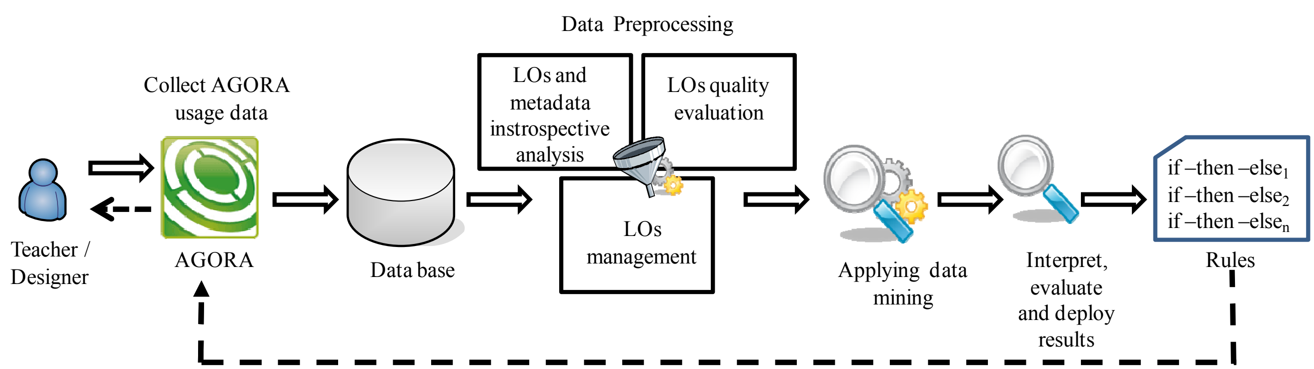

Figure 1. Learning Objects KDD Scheme

having uniform Learning Objects sets. These objects were published in the AGORA platform by 170 teachers from various universities.

2) Create summarization tables: All available information for each of the selected Learning Objects was retrieved and three main information sources were selected:

a) Metadata: whose values describe Learning Objects and are based on LOM and LOM-ES specifications.

b) Quality evaluation from the pedagogical perspective: Data were provided by assessments carried out by teachers and experts following MECOA model.

c) User activity: obtained from the integration of use and navigation patterns and the teachers profiles. All above information was included in tables which registers describe each Learning Object.

3) Data discretization: Generally, each of the attributes in the table must be converted to discrete values in order to facilitate the analysis and results interpretation. In our case the information was already described by labels sets, so it was not necessary to carry out this process.

4) Data transformation: Data are converted to a format facilitating portability and use for data mining algorithms. In our case the ARFF format (Attribute-Relation File Format) is required.

The data set generated in this phase consists of 300 instances and 38 attributes. Among them 13 belong to LOM and LOM-ES standard, 12 belonged to the pedagogical quality evaluation model and 13 to the users' activities. Some examples of these attributes are: interactivity type, technical requirements, file type, resource balance and usability.

C. Applying data mining techniques

From the preprocessing phase, a data set has been obtained. Data mining algorithms were applied to these results with WEKA [20] and KEEL [3] systems:

- Clustering Algorithms. For clustering testing, the following algorithms were used: SimpleKmeans [11] and EM (Expectation Maximization) [7].
- Classification algorithms. We considered some of the attributes that define the clusters as a class. This is achieved using ID3 (Induction Decision Trees) [15] and C4.5 [16] algorithms. These tests are intended to verify the effectiveness in the

TABLE I. SOME OF THE BEST RULES OBTAINED WITH THE ID3 AND C4.5 ALGORITHMS

Algorithm	**Considering The Format as Classification Attribute**		**Considering The Learning Resource Type as Classification Attributes**	
	Rule-generated	***Rule interpretation***	***Rule-generated***	***Rule interpretation***
ID3	initiative = no apply; learning resource type = slide; level = training professional => PDF	The LO has a *PDF* format if it does fulfill the next requirements: the interface does not allow user interaction and is used as a presentation at training level.	format = SWF; structure = atomic; interactivity type = expository; generalization = demonstration => slide	A LO is an educational resource classified as *presentation*, if it does fulfill the next requirements: has a SWF format, contains an explicit pedagogical purpose, exposed concepts and transfers knowledge through demonstrations.
	requirement = internet connection; initiative = navigation path; structure = atomic => SWF	The LO has a *SWF* format if it does fulfill the next requirements: Internet connection, allows the user to choose the navigation path and the structure is atomic.	Format = DOC; interactivity level = high; semantic density = high => questionnaire	A LO is an educational resource classified as *questionnaire*, if it does fulfill the next requirements: has a DOC format, contains a high interactivity with the user and contains too much information.
	requirement = specific software installation; initiative = Problems solution; information about the objective = enough; learning resource type = slide => PPT	The LO has a *PPT* format if it does fulfill the next requirements: requires Internet connection, allows the user to solve problems, contains objectives and is used as a presentation.	Format = PPT; interactivity Type = active; Semantic Density = medium => exercise	A LO is an educational resource classified as *exercise*, if it does fulfill the next requirements: has a PPT format, contains a medium interactivity type and level information medium.
C4.5	requirement = internet connection; interactivity level = high => SWF	The LO has a *SWF* format if it does fulfill the next requirements: requires Internet connection and the interactivity level with the user are high.	interactivity type = expository; format = SWF; result = knowledge => slide	A LO is an educational resource classified as *presentation*, if it does fulfill the next requirements: exposed concepts, has a SWF format and transmit knowledge.
	requirement = internet connection; interactivity level = high; interactivity type = active => SWF	The LO has a *SWF* format if it does fulfill the next requirements: Internet connection contains a high interactivity with the user and activities for the user.	interactivity type = expository; Format = WMP; context = languages => exercise	A LO is an educational resource classified as *exercise*, if it does fulfill the next requirements: exposed concepts, has a WMP format and is used in learning language.
	initiative = no apply; result = knowledge ; learning resource type = exercise => PPT	The LO has a *PPT* format if it does fulfill the next requirements: the interface does not allow user interaction, transmit knowledge and is used as a exercise in class	interactivity type = expository; format = PDF; result = knowledge; problematic = questions => slide	A LO is an educational resource classified as *presentation*, if it does fulfill the next requirements: exposed concepts, has a PDF format, transmit knowledge and suggests questions to the user

TABLE II. SOME OF THE BEST RULES OBTAINED WITH THE A PRIORI AND PREDICTIVE A PRIORI ALGORITHMS

Algorithm	Rule-generated	Rule interpretation
Apriori	result = knowledge; cognitive process = access to information; self-knowledge = no activities => choice between alternative problems solutions = no solution	If the LO consists of concepts and it has no tasks and activities then it doesn't provide problem-solving alternatives.
	interaction components = integrated; choice between alternative problems solutions = don't have => structure = atomic	If the LO have integrated components and doesn't have alternative to solving problems then its structure is atomic.
	interaction components = integrated; result = knowledge => structure = atomic	If the LO have integrated components and consists of concepts then its structure is atomic.
Predictive Apriori	topology = conceptual; interaction components = integrated => level = professional training	If the LO proposed concepts and its components are integrated then its use is oriented to professional training level.
	topology = conceptual; self-knowledge = no activities; discipline = social sciences => choice between alternative problems solutions = no solution	If the LO proposed concepts, no activities and is oriented to social sciences area then it doesn't provide problem-solving alternatives.

classification rules generation from both systems and thus provide corroboration if rules are similar.

- Association algorithms. For the association rules generation we have executed the A priori [2] and Predictive A priori [18] algorithms. For both algorithms, we determined the generation of 100 rules, which have a minimum support of 0.3 and minimum confidence of 0.9 as parameters.

D. Interpret, evaluate and deploy results

Finally, results were interpreted, analyzed and compared to determine the rules Among them, there were considered only those rules providing relevant information.

1) Clustering results: The best results were obtained with the EM algorithm in which 3 clusters were generated. Clusters can be described in terms of LOs grouped as follows:

a) Cluster 1 (59%): Most of the Learning Objects in PPT format (presentation files) have medium semantic density and are used by professionals who have experience in the use of technological tools and work in the area of social sciences and administration, and mainly used for presentations.

b) Cluster 2 (14%): Learning Objects based on web pages (HTML) are geared toward reading and problem solving and are used by professionals in technological area.

c) Cluster 3 (23%): Learning Objects based on SWF (animation files) are used for quizzes and exercises for professionals in education and humanities that have a good technological background.

2) Classification results: We obtain a set of IF-THEN-ELSE rules from decision trees that show interesting information about the LOs classification considering two classification's attributes:

a) The format, whose labels are: mpeg, doc, xls, ppt, avi, wmp, swf, gif, flv, html, pdf, other.

b) The learning resource type whose labels are: exercises, readings, simulation, questionnaire, diagrams, exams or experiments.

After a first analysis, we eliminate rules little in depth, or with irrelevant information. The table 1 shows some of the best rules obtained.

3) Association results: We obtain a set of IF-THEN rules from the algorithms. After an analysis, we eliminate those rules that were with irrelevant information. The table 2 shows some of the best rules obtained.

Analyzing the results of each of the classification and association algorithms, it was observed that the best rules generated have similar characteristics.

V. CONCLUSIONS AND FUTURE WORK

In this paper we presented an adapted methodology to produce rules about Learning Objects usability characteristics.

Data sets were obtained from log files generated with the use of the AGORA Learning Objects management system; by means of the activities carried out by the AGORA users, as well as the results of MECOA pedagogical evaluations and the LOM and LOM-ES based metadata.

Tests provided relevant information about the attributes that define Learning Objects indicators.

One application of the rules obtained, is implemented in the AGORA platform environment, since we allow elements which are crucial for sorting, suggest or recommend actions about Learning Objects.

Currently, the recommendation is implemented into the Learning Objects generation process, specifically in editing metadata. Values for the missing metadata are suggested using the generated rules set.

Clusters generated define relationships between Learning Objects and usability characteristics. This allows the development of classifiers to improve the search mechanisms in AGORA. For example, information relating to user profiles and their needs may be considered as filters for the recommendation of resources or users with similar needs.

As future work is to analyze the information source on the Learning objects publishing activities for each user. This will allow us to generate rules in a personalized way; through them, is possible to several extract individuals characteristic.

These new rules mixed with general rules will get better recommendation schemes in AGORA, such as metasearch resources and assisted metadata filling.

All these developments are aimed at generating a recommended system to operate in environments from the Learning Objects repositories.

The rules that were generated and the data set are available at http://www.kaambal.com/agora/.

ACKNOWLEDGMENT

This work is partially supported by AECID A/016625/08 Project (Spain) and YUC 2006-C05-65811 project FOMIX CONACYT (México).

REFERENCES

[1] Advanced Distributed Learning, SCORM Specification. Retrieved on February 2, 2009 from http://www.adlnet.gov/Technologies/scorm/.

[2] R. Agarwal, T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases," Proc. ACM Int. conf. Management of Data (SIGMOD 93), ACM Press, May. 1993, pp. 207-216, doi:10.1145/170035.170072.

[3] J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesús, , S. Ventura, J.M. Garrell, , C. Romero, J. Bacardit, V. Rivas and F. Herrera, "KEEL: a software tool to assess evolutionary algorithms for data mining problems", Soft computing: A fusion of foundations, methodologies and applications, vol. 13(3), Oct. 2008, pp. 307-318. doi: 10.1007/s00500-008-0323-y.

[4] H. S. Al-Khalifa and H. C Davis, "The evolution of metadata from standards to semantics in E-learning applications," Proc. ACM Int. conf. in Seventeenth Conference on Hypertext and Hypermedia (HT'06), ACM Press, August 2006, pp. 69-72, doi: 10.1145/1149941.1149956.

[5] H. Ayad and M. Kamel, "Clustering Learning Objects Collections Using Cluster Ensembles," The 3rd Annual Scientific Conference of the LORNET Research Network (I2LOR 06), Nov. 2006.

[6] J.J. Blanco, A. Galisteo del Valle, A. García et al., Perfil de aplicación LOM-ES V.1.0. Asociación Española de Normalización y Certificación (AENOR). Retrieved on December 16, 2008 from http://www.educaplus.org/documentos/lom-es_v1.pdf.

[7] A. Dempster, N. Laird and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm" Journal of the Royal Statistical Society, vol. 39(1), 1977, pp. 1-38, doi: 10.2307/2984875.

[8] IEEE. Learning Technology Standards Committee, Draft Standard for Learning Object Metadata (LOM). Retrieved on Dec. 15, 2008 from http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.

[9] IEEE Learning Technology Standards Committee, WG12: Learning Object Metadata, Retrieved on november 18, 2008 from http://ltsc.ieee.org/wg12/.

[10] C. Kiu and C. Lee, "Learning Objects Reusability and Retrieval through Ontological Sharing: A Hybrid Unsupervised Data Mining Approach," Proc. IEEE International Conference 7th Advanced Learning Technologies (ICALT 2007), IEEE Press, Jul. 2007, pp. 548-550, doi: 10.1109/ICALT.2007.177.

[11] J. MacQueen, "Some methods for classification and analysis of multivariate observations," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California press, 1967, pp. 281–297, doi:10.1234/12345678.

[12] Y. Ouyang and M. Zhu, "eLORM: Learning Object Relationship Mining based Repository" Proc. IEEE Conference 4th Enterprise Computing, E-Commerce, and E-Services (EEE 2007), IEEE press, Jul. 2007, pp. 691-698, doi:10.1108/14684520810879863.

[13] M. E Prieto et al. Agencia Española de Cooperación Internacional para el Desarrollo, Proyecto AECI A/8172/07: Metodología y herramientas para la evaluación de la calidad de los recursos para tele-aprendizaje en la formación de profesores, 2008. Rep.Int.

[14] M. E. Prieto, V. Menendez, A. Segura and C. Vidal, "A Recommender System Architecture for Instructional Engineering" in M.D. Lytras et al. (eds.) Emerging Technologies and Information Systems for Knowledge Society, Springer-verlag Berling Heidelberg, 2008, pp. 314-321.

[15] J. R. Quinlan, "Induction to Decision Trees" in Machine Learning 1, Academic Publishers, vol. 1, Mar. 1986, pp. 81-106, doi:10.1023/A:1022643204877.

[16] J. R. Quinlan, C4.5: Programs for machine learning, Morgan Kaufmann Publishers Inc, 1993.

[17] C. Romero, S. Ventura and E. García, "Data mining in course management systems: Moodle case study and tutorial" Computers & Education, vol. 51(1), Aug 2008, pp. 368-384, doi: 10.1016/j.compedu.2007.05.016.

[18] T. Scheffer, "Finding association rules that trade support optimally against confidence" in Principles of Data Mining and Knowledge Discovery, vol. 2168, J. Komorowski, J. Zytkow, Springer, Heidelberg , 2001, pp. 424-435.

[19] K. Shaban, O. Basir and M. Kamel, "Learning Objects Clustering based on Semantic Understanding of Text" 3rd Annual E-learning Conference on Intelligent Interactive Learning Object Repositories (I2LOR 2006), Nov. 2006.

[20] I. H. Witten and E. Frank, Data mining: Practical machine learning tools and techniques, Morgan Kaufman Publishers inc, 2005.