World Applied Sciences Journal 27 (1): 123-129, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.27.01.1614

Effective Retrieval of Text and Media Learning Objects Using Automatic Annotation

[1]V. Sree Dharinya and [2]M.K. Jayanthi

[1]School of Information Technology and Engineering,
VIT University, Vellore-632014, Tamil Nadu, India
[2]Department of Computer Science and Engineering, Netaji Institute Of Engineering
and Technology, Nalgonda District., Andhra Pradesh 508 252, India

Abstract: An emerging aspect in learning environment is to effectively retrieve learning materials from learning object repositories. An extension of the IEEE LOM standard with a domain ontology and automatic annotation for effective retrieval of learning objects is proposed. To make best use of learning objects outside the Learning Management System, we associate the learning materials with metadata pedagogic attributes such as prerequisite, type, difficulty level, importance, roles and significance. Automatic annotation of the learning object repository is developed in a semantic oriented approach. The techniques such as specific indexing for parsing the documents and using standard classification algorithms for automatic extraction with concept identification and significance for effective personalization is developed using domain ontology. The usage of Word Net based processing and the identification of learning objects based on the concept for text and media learning objects has proved effective using the domain ontology. The experimental results have proved better precision and recall using the combination of methods for effective personalization. The developed evaluating methods correlate with the scores of the topics and generally outperform the traditional term frequency inverse document frequency (TF-IDF) method. The proposed specific indexing method obtains the highest precision and recall of all.

Key words: Domain ontology · Granularity · Annotation · Learning object · Specific indexing · Metadata · Personalization · Media objects · Recall · Precision

INTRODUCTION

Learning object granularity is one of the most critical unsolved issues which are handled by many researchers Learning objects was introduced as an idea for educational resources in the form of broken modular components which are later recombined by the authors or instructors and learners. Due to its digital nature it is used and reused in different scenarios by different people unlike the traditional classroom teaching.

E-learning systems must allow users to be able to retrieve whichever content they might need from the whole repository, in order to get the maximum personalized benefit of the learning process. Many different querying methods exist, but until now the most popular method is still query by keyword. This is not an issue with LOs in the form of text, but media objects are more challenging. The media objects need to be formulated into an effective MLO, which includes annotating the videos with appropriate textual keywords to enable search by keyword. The problem is to determine which terms are the best suitable to use as annotations. This process can be done manually but has many disadvantages. The manual process is restrictive, tedious and subjective and hence an automatic annotation method is proposed.

Machine learning platforms have been increasing the demand for more learning solutions across the web. E-learning has emerged as a promising approach for effective and enhanced learning using communication

Corresponding Author: V. Sree Dharinya, School of Information Technology and Engineering, VIT University, Vellore-632014, Tamil Nadu, India.

between humans and computers. This communication technology involves the large storage of data in the form of learning objects which are stored in learning object repositories. Such learning objects are found in the gigantic web. Beyond the World Wide Web, certain specific learning materials are often created for effective and enhanced learning. Various standards exist for this established communication network. The IEEE LOM standard is one such standard for which certain attributes are being extended with domain ontology for efficient use of learning objects outside the learning management system. The attributes of IEEE LOM are used for extracting the data from the repositories. We have extended attributes like frequency, type and significance with semantic annotation for effective and personalized retrieval of learning objects as an extension work. It consists of a simple solution for integrating the semantic annotations without a special RDF LOM.

Annotation methods exist manually as well as semi automatically for the learning objects in the repository. The methods of manual annotation have been difficult and uninteresting [15]. The approach to provide automatic annotation to a set of data initially and to extend it further for a large database is proposed. The metadata tagging for learning objects of a specific domain is done using the domain ontology. The annotation and extraction is done by filtering the query based on request and is retrieved with specific concept identification and significance.

Domain ontology for semantic metadata annotation is proposed. The ontology based retrieval is used for better personalization using the user profile repository. To define the ontology based annotation model we use the semantic web technologies to propose a document model. In addition to the existing traditional methods for structure based indexing of documents we combine a mathematical concept with a classification algorithm for concept identification in learning objects [1]. This specific indexing is extended for the semantic standards for text document retrieval in an e-learning environment. Mathematical methods are used for identification of tokens in a document, the term frequency and specific indexing. The classification algorithm is proposed for the concept identification for a set of documents with specific parameters as name, type frequency and significance. Using the algorithm the experimental results have proved that better and effective personalization is attained with better precision using concept identification.

Related Work

Existing Standards: Reusability is the major factor in e-learning as the contents are to be used repeatedly by different users. To employ reusability the abstractions such as learning objects and learning design are used.

Learning objects are [3] 'The main idea of learning object is to break educational content down into small chunks that can be reused in various learning environment, in the spirit of object oriented programming'. Majority of standards provide metadata specification for learning objects.

- The IEEE (Learning Object Metadata) LOM-(IEEE 2002) - (http://ltsc.ieee.org/wg12/index.html), employed for website development This was the first important standard created for defining the metadata for learning objects and now considered to be too simple. It aims to develop accelerated solutions and guidelines for learning technologies.
- The Dublin Core Metadata Initiative-DCMI-(http://dublincore.org) and IMS(http:// www.imsglobal.org) metadata are used by both academicians and corporate.
- SCORM -Sharable Content Object Reference Model (http://www.adlnet.gov/scorm/index.cfm) by Advanced Distributed Learning initiative (SCORM ADL2004). This standard is deployed by Advanced Distributed Learning initiative. This creates learning objects as instructional objects for web based learning as well. Key contributors for SCORM are AICC, ARIADNE and IEEELISC. This model is considered to be too difficult but its implementation is considered to be very consistent for metadata specification.
- *Synchronized Multimedia Integration Language (SMIL)* was developed by the World Wide Web Consortium (W3C), (2005) - This standardization was adopted by the World Wide Web Consortium and is an easy-to-learn XML-style, allowing easy design.

IEEE LOM is considered as a standard reference and many researchers have suggested the extension of IEEELOM standards[18]. We have extended a few attributes for IEEE LOM like the list of concept, significance, frequency and role of the concept.

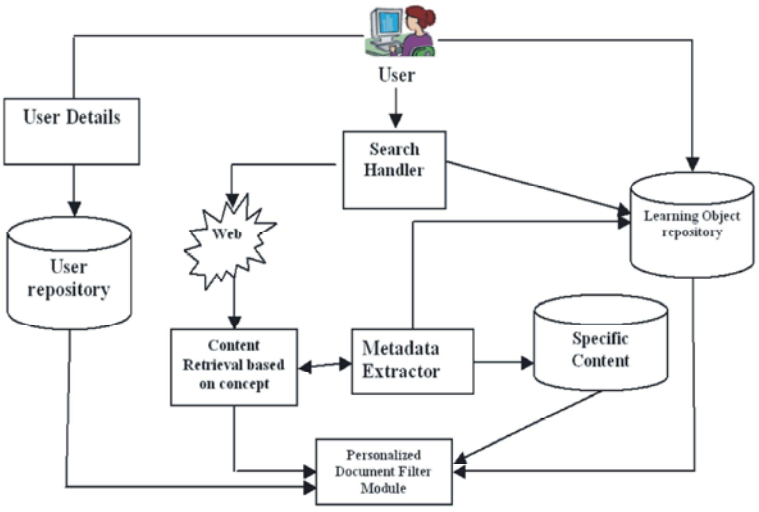

Fig. 1: Overall System Architecture

Proposed System Architecture: The system architecture gives an overview to provide automatic annotation to the documents and enable metadata to various types of documents from different sources. The annotated documents are stored in the repository and are retrieved by the user based on the user's knowledge from the repository. The system stores the intuitiveness of the user and the way of interaction along with the user request in the user repository [21, 22]. The system provides a search module which enables local and web based search for the query given by the user. The following modules are used for effective personalization and retrieval of learning content based on concept identification and significance.

Search Handler: Based on the request of the user the content is searched either in the local repository or the query is forwarded to a general purpose search engine.

Metadata Extractor Module: For the given query the content is retrieved by the content retrieval module and the metadata content is retrieved. Thereby the data is retrieved from the specific content storage.

Personalized Document Filter Module: This module filters the content according to the concept which uses domain ontology. Using the domain ontology the content is retrieved based on the prerequisite, topic, concept identification and specific terms. The documents are ranked and significance score and relevance score provides effective personalization.

Figure1, Illustrates the overall system architecture and the discussed modules and the relations between them to build domain ontology.

Ontology Based Approach.

Existing Ontologies: According to [18] ontology is a knowledge domain conceptualization in a communicative format which has entities, attributes and axioms. Domain ontology is an ontological structure of topics, concepts of a particular subject, the prerequisite and the relationship between them [19, 20]. The increase in cost of time and manual effort are major constraints of building domain ontology. But the accuracy level in terms of precision for retrieval of data from various concepts is relatively higher. According to [12, 13], the ontology types are based on ontology covering domain concepts, ontology based on teaching and learning strategies and ontology based on the structure of the learning object. E-learning standards are used to organize the materials in the form of modules, chapters, lessons and learning objects. The semantic data is added when a new material is added. The semantics like is Narrower Than / is Broader Than, is Alternativeto / is Less Specificthan is used by SKOS (http://www.w3.org/TR/skos-reference/)

The ontology to build a repository is proposed for the reuse of learning object. The related and relevant objects are retrieved from the repository for effective personalization [19]. The data retrieved are based on the key words according to the learner's request and selection of course based on the prerequisite of the course [17, 18]. The semantic approach builds a tagged

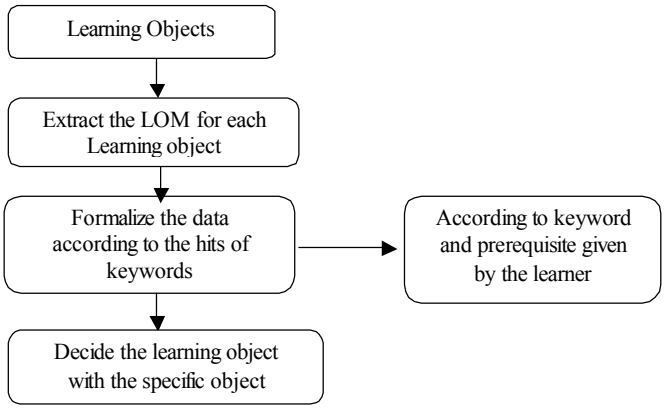

Fig. 2: Learning object repository with ontology

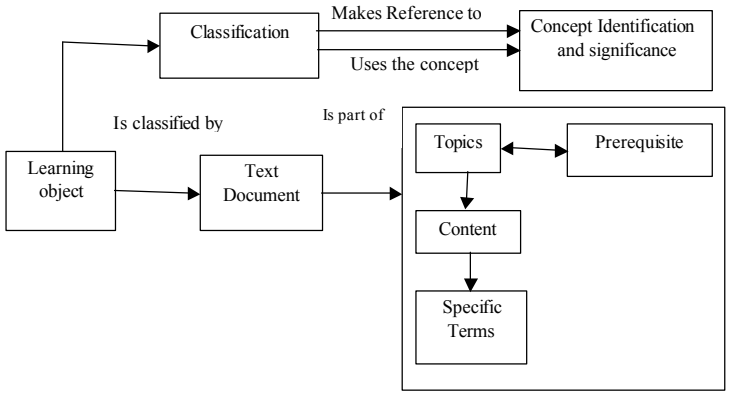

Fig. 3: Proposed Domain Ontology

metadata for the objects in the learning object repository. The appropriate term of the query search is obtained using the ontology approach for retrieval of data. The prerequisite term, frequency, significance and type which are the extended part of IEEE LOM are used for building ontology. Figure 2 represents the ontology used for retrieving the data with appropriate keyword as well as the prerequisite from a learning object repository [16].

Proposed Domain Ontology: As given by [8], the ontology should be a representation of distinct layers for different entities. In addition to the layer specified for the domain ontology we have used the prerequisite layer which consists of a list of topics as prerequisite for the given query to be searched. The ontology has a prerequisite layer mapped with the topics layer from which the content with specific concept is retrieved [17, 18]. The specific terms are mapped to the concept layer where the terms are retrieved with concept identification and significance.

Proposed Automatic Annotation Method: Adapting to the existing standards is an important factor for annotating learning objects. The ontology concepts should be integrated in the structure of the standards. To enhance the retrieval of documents using concept identification we use attributes such as makes 'ReferenceTo' and 'usesTheConcept'. The domain ontology (Figure 3) identifies the topics with a relation to the prerequisite of the topic given by the user and uses the 'isOnTopic' annotation for retrieving the specific topic. Semantic annotations can be done by methods as given by [3, 13].

Manual annotation- The difficulty in annotating the learning objects increases as the content is demandingly increasing in different domains. Existing tools such as PhotoStuff, gonTogle, Vannotea are used for manually annotating learning objects. The difficulty of manual annotations is the usability of the tools. Manual tagging to learning objects is uninteresting and sometimes not done satisfactorily.

Automatic annotation- Automating parts of large learning object repositories reduces the difficulty of its usage. This is done by automating the tagging of learning objects. The huge data on the web is available in the form of documents and exist in different sources. To use such materials one has to filter the learning materials for building the learning object repository. It is costly and difficult to create metadata for large set of potential materials on the web.

Using the domain ontology the authors can provide annotations for the entire learning object and extraction of metadata can be done to an acceptable level. Some work on automating annotations is done by Dublin Core metadata. The specific indexing method discussed provides the metadata and extraction of documents is done with concept identification and the significance.

Specific Structural Indexing Method

Methodology 1: The document identified by the user after the execution of the search query is analyzed from the domain storage repository. The identified document is parsed and the tokens (splitting sentences into words) identified are weighed using stop word removal and stemming methods. The term weight calculation is done using term frequency (TF). TF(ti,dc) is the number of times the term 't' occurs in the document 'dc'. The popular method of calculation of weight is given as TFxIDF weighting, where the weight is calculated to be proportional to the frequency of the corresponding term 't_i' in the document 'dc' and inversely proportional to the number of documents |D|

$$Wi = \frac{TF(ti,dc)\ \log |D|}{DT\ (ti)}$$

where DT(ti) is the number of documents in the collection D which has the terms ti. The tokens in the document are identified by parsing techniques and stemming methods to compare the words[29]. The similarity is obtained by cosine similarity

$$sim\ (d,q) = cos\ (d,q) = d \bullet q\ /\ |d||q|$$

which is equal to the cosine of the angle formed by the two vectors d and q in the n dimension vector space.

Methodology 2

Proposed Classification Algorithm for Concept Identification: The terms identified may have multiple meanings and to retrieve a document term from a document belonging to a particular domain. We annotate them with a list of concepts. The significance of a concept is identified if more number of related concepts of the particular term occurs in the document. The proposed algorithm is used to retrieve the concept based term along with the significance of the term.

For every term identified from the document D the concept Cij is obtained and the significance is also computed CijSig. The CijSig is taken as the normalized frequency term. For every concept Cij the related concept rc in the document is identified. The associated concept is then normalized by the term frequency to the corresponding terms t and the related concept rc.

Significance Cij=ti freq+α* tp freq, where we assume α value to be ½, where tp =term corresponding to related concept rc. For the particular term we identify the concept with a maximum significance value. Following are the steps involved in the proposed algorithm to get the input of terms in the document D and to give the output of concepts with their significance.

First we normalize the frequency of the terms and then match the term frequency to the concept significance. Then we find the related concept rc and its occurrence in the document D by using Significance Cij=ti freq+α* tp freq where we assume α value to be ½.

Finally we select the final concept by identifying its significance with a highest significance score which is to be greater than the threshold value. The algorithm returns the list of the related concepts and the significant value. The combination of both these methods is used for filtering and extracting the query based on the domain ontology.

Results and Findings: The above said techniques as well as evaluation methods were used for a set of 25 documents. The documents were stored locally and the list of documents for a specific domain is stored. Using the domain ontology and the annotation to the list of documents the topics are filtered without concept identification. Later the documents use the specific indexing methods to filter the set of documents with the concept identification and are retrieved. The various queries are given in Table 1. The domain ontology for a set of documents locally stored is retrieved according to concept and the significance score is also calculated. The precision percentage is obtained by filtering terms with concept identification and without concept identification and also the recall is identified. The graph reflects the better precision and recall for a set of documents in a specific domain for 25 documents where ten related terms were retrieved. The effective personalization is attained in terms precision and accuracy

Table 1: Query Search Based on Domain Ontology

S.No.	Search Query	Without Concept identification & significance	Precision%	Filtered Documents	Filtered with Concept Identification	Precision% using concept identification & Significance	Recall% using concept identification&Significance
1	Light	17	68	12	12	100	70.59
2	Wave	12	48	12	12	100	100
3	Wave Length	8	32	8	7	87.50	100
4	Interference	9	36	7	5	71.43	77.78
5	Intensity	8	32	4	4	100	50.00
6	Scattering	12	48	9	7	77.8	75.00
7	Atom	10	40	5	5	100	50
8	Electron	15	60	13	7	53.85	86.67
9	Motion	17	68	15	14	93.33	88.24
10	Reflection	19	76	17	12	70.59	89.47

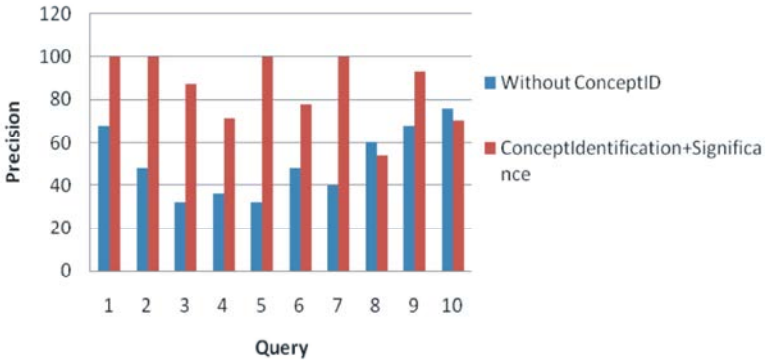

Fig. 4: Results for concept identification.

for a specific term with concept identification and significance from the domain ontology. The improvement in precision with concept identification and significance is given in the graph Figure 4. The precision percentage is shown along the Y axis and the query search in a set of documents of a specific domain for a specific concept is shown in the X axis. The results have proved that better precision and recall is obtained by concept identification and significance score using the domain ontology and annotation of learning objects by the specific indexing method.

DISCUSSION

The resources available should be connected and accessible by users based on the demand. The resources available are annotated semantically and are retrieved based on the query request from the learning object repository and the user request. The approach presented in this paper develops annotation and extracts the query based on the domain ontology proposed in this paper. The result shows that the approach identifies the strong and semantic related terms to the given query along with the concept identification and significance. Our future work will concern the extraction of learning objects for large data as well as fully automating the annotations. The precision for automatic annotation can also be worked on for large data sets on the World Wide Web.

REFERENCES

1. Advanced Distributed Learning Initiative (2002). Sharable Content Object Reference Model
2. Aliza Ali, Zahara Aziz and Rohaty Majzub, 2011. Teaching and Learning Reading Through Play, World Applied Sciences Journal, pp: 14.
3. Afiza Mohamad Ali, 2013. A Comparative Study on Reasoning Strategies in L1 and L2 Critical Reading-Thinking Tests.' World Applied Sciences Journal, 21(Special Issue of Studies in Language Teaching and Learning).
4. Aroyo, L. and D. Dicheva, 2001. AIMS,' Learning and Teaching Support for WWWbased Education. International Journal of Continuing Engineering Education and Life-Long Learning, 11: 152-164.
5. Brut, M. and Buraga, 2007. A Web Service-Based Solution for E-Learning ResourcesAnnotation and Retrieval-Conceptual Architecture, Proc. Int'l Joint Conf. Computer, Information and Systems Sciences and Engineering.

6. Currier, S., 2008. Metadata for Learning Resources: An Update on Standards Activity for 2008, ARIADNE, pp: 55.
7. Wiley, D.A., 1999. Learning Objects and the New CAI: So what do I do with a learning object?, 1999.
8. Devshri Roy, Sudeshna Sarkar and Sujoy Ghose, 2005. Automatic Annotation of Documents with Metadata for use with Tutoring Systems. Indian International Conference on Artificial Intelligence (IICAI), 20-22: 3576-3592.
9. Dicheva, D. and C. Dichev, 2004a. A Framework for Concept-Based DigitalCourse Libraries. Journal of Interactive Learning Research, 15(4): 347-364.
10. Faizah Mohamad and Nuraihan Mat Daud, 2013. The Effects of Internet-assisted Language Learning (Iall) on the Development of Esl Students' Critical Thinking Skills. World Applied Sciences Journal, Volume Issue 21 (Special Issue of Studies in Language Teaching and Learning).
11. IMS Global Learning Consortium (GLC), "IMS Application Profile Guidelines Overview," http:// www.imsglobal.org/ ap/apv1p0/ imsap_oviewv1p0.html, 2005.
12. Kanninen, E., 2008 Learning Styles and E-Learning. Master's thesis, Tampere University of Technology.
13. McCalla, G., 2004. The Ecological Approach to the Design of E-learning Environments: Purpose-based Capture and Use of Information about Learners. Journal of Interactive Media in Education (Special issue on the Educational Semantic Web), pp: 7.
14. Mihaela M. Brut, 2011. Florence Sedes and Stefan Daniel Dumitrescu, A Semantic-Oriented Approach for Organizing and Developing Annotation for E-learning, IEEE transactions on Learning Technologies, 4(3): 239-248.
15. Mohan, P. and C. Brooks, 2003. Learning Objects on the Semantic Web, Proc. International Conference on Advanced Learning Technologies.
16. Noor Lide Abu Kassim, Nuraihan Mat Daud and Nor Shidrah Mat Daud, 2013. Interaction Between Writing Apprehension, Motivation, Attitude and Writing Performance: A Structural Equation Modeling Approach', World Applied Sciences Journal, Volume Issue 21 (Special Issue of Studies in Language Teaching and Learning).
17. Popescu, 2009. E. Learning Styles and Behavioral Differences in Web-based Learning Settings. In Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies.
18. Qiu, Y., G. Guan, Z. Wang and D. Feng, 2010. Improving news video annotation with semantic context. In Proceedings of the 2010 International Conference on Digital Image Computing: Techniques and Applications, DICTA 10, 214–219, Washington, DC, USA. IEEE Computer Society.
19. Shang, Y., H. Shi and S. Chan, 2011. An Intelligent Distributed Environment for Active Learning, ACM – 1-58113-348-0/01/0005.
20. Ruhizan M. Yasin, Saemah Rahman, Ramlee Mustapha and Kamarudin Tahir, 2011. Development of Generic Employability Skills Through Peer Interaction and Contextual Teaching and Learning in Community Colleges.', World Applied Sciences Journal, 15(Innovation and Pedagogy for Lifelong Learning).
21. Sree Dharinya, V. and M.K. Jayanthi, 2012. An Approach Towards Redefining Granularity Of Learning Objects for Effective and Adaptive Personalization, Journal of Theoretical and Applied Information Technology, 41(1): 98-108.
22. Syed Mohammad Syed Abdullah, Ahmad Zamri Khairani, Nordin Abd. Razak, Jamalsafri Saibon and Azlinda Mohd. Ariff, 2011. Teaching Efficacy among College Student-Teachers of Diverse Background, World Applied Sciences Journal, pp: 14.
23. SMETE, The National Science, Mathematics, Engineering and Technology Education Digital Library, http:// www.smete.org
24. Taniar, D. and J.W. Rahayu, eds, 2006. Web semantics Ontology, Idea Group Publishing.
25. Morita, T., N. Fukuta, N. Izumi and T. Yamaguchi, 2006. DODDLEOWL: A Domain Ontology Construction Tool with OWL, Proc. Int'l Semantic Web Conf., pp: 537-551.
26. Ullrich, C., 2004. Description of an Instructional Ontology and its Application in Web Services for Education. Workshop on Applications of Semantic Web Technologies for e-Learning.
27. Zahara Aziz, Shurainee Hanim Mohamad Nor and Rozalina Rahmat, 2011. Teaching Strategies to Increase Science Subject Achievement: Using Videos for Year Five Pupils in Primary School',World Applied Sciences Journal, pp: 14.
28. Zapata- Rivera, J.D. and J.E. Greer, 2004b. Inspectable Bayesian Student Modeling Servers in Multi-Agent Tutoring Systems. International Journal of Human-Computer Studies, 61(4): 535-563.