Learning Objects Automatic Generation System for Multiple Data Formats

Yushun Li

School of Educational Technology
Beijing Normal University
Beijing, China
lyshun@bnu.edu.cn

Meng Li

School of Educational Technology
Beijing Normal University
Beijing, China
limeng1988@bnu.edu.cn

Abstract—**This paper focuses on object processing digital resources to enhance their reusability and adaptability. Deep into content level of digital resources belonging to different formats, to achieve fine-grained knowledge-based learning objects automated generation, this paper presents a MPLOG framework that integrates a variety of learning resources analyzers. Under the guidance of the domain ontology, the analyzers have divided digital resources into blocks and marked them with knowledge point.**

Keywords-component; Learning Objects; Multiple Data Sources; Multiple Data Formats; Automatic Generation System; SOA

I. INTRODUCTION

With the continuous development of information retrieval technology and increasingly massive resources on the World Wide Web, problems such as how to improve accuracy and efficiency of the resource retrieval, how to achieve Content-oriented, fine-grained retrieval and meanwhile, along with in-depth Intelligent Tutoring Education System research, how to give Intelligent Tutoring Education System knowledge units of learning resources as a support at the bottom have raised concerns. Relying on teachers to promote and demanding linear knowledge acquisition, the traditional teaching way, time and location dependent, has been unable to meet the individual learning requirements of learners .If someone want to achieve this goal, they must first solve the personalized learning resources. Under existing conditions online teaching resources are abundant in general, but the efficiency of resource usage is not ideal. Users can neither fast retrieve to their desired learning resources, nor thoroughly use learning resources to search. Research on learning resources efficiently sharing, reusing and interoperability is one of the major problems during the development of e-Learning. It has reference value to solve problems in information retrieval, intelligent teaching system development and the development of personalized learning. For research and practice in this field, learning object technology is proposed. Learning object is of great significance to education unit, digital resources, teaching resources and its metadata integration into. At present, the design of learning resources, learning object metadata automatically generated, resource management and application have become an important research area in the field based on e-Learning.

In the current e-learning environment, scholars have put forward the "on-demand learning", "in-time learning" and the amount of instant "learning" needs. These can meet the personal needs of learners and satisfy the push adaptive resource to learners. Learning objects are gradually transformed into active, intelligent, which meet the specific principle of learning, knowledge learning object or the object of knowledge. Compared to the prior learning object, object of knowledge in the content structure is more focused than learning object in teaching content. Compared to the prior learning object, the object of knowledge focuses more on teaching content and teaching goal. It also contains certain teaching strategies. In use, it can better serve as teaching resources to meet the personalized learning, demonstrated in Figure 1.
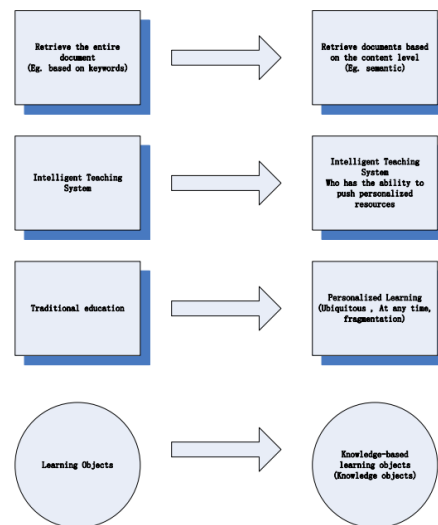

Figure 1. Changes in use of resources

II. RELATED WORK

There are two main techniques in learning object automatic annotation fields: one is based on rules, the other based on machine learning. [1]For learning object automatic generation framework and the automatic generation of metadata, the researchers have made a lot of research and exploration. Klarity and DC.dot first put forward the automatic harvesting metadata. But object format and content structure have strict request, [2] [3].KIM is a kind of modular semantic knowledge management platform, which can go through the text content of the entity and relationship

to the corresponding background knowledge, but the object is limited to text file [4]. Piggy Bank system considers resource producers's teaching methods will be reflected in his paper structure.[5] LessonMapper2 is a hybrid automatic generation system of learning object metadata. This system considers the relationship among various objectsmore closely, so their characterization of keywords coincidence degree is higher[6]. ANGRAM achieved metadata annotation of PPT document [7].He realized the metadata annotation of PPT document[8]. American researchers also proposed for word metadata annotation method [9]. CRIS UNS achieved metadata annotation method of PDF based on machine learning[10]. MARVEL achieved semantic metadata extraction of video[11]. QBIC system is a commercial video retrieval. the underlying features of the system mainly depends on the video resource retrieval, but not the deep content level. Although teaching video retrieval system designed by East China Normal University, has achieved the teaching video segmentation and retrieval based on content, the process rely heavily on manual intervention through [12]. According to availability of system framework in other research and effect of learning object generation, there are still many deficiencies. Mainly has the following several points:

a) Resource types of the learning object are various, such as text, image, video, audio and so on. And the methods of storage and content analysis for each format are different. The existing framework can only support a specific file types and formats. It cannot be extended to other formats.

b) Structured metadata generation tools at present are discrete, independent. Each tool can only generate certain metadata. Metadata results cannot be complete direct. And the metadata standard is not consistent, cannot be extended to other specification.It also can cause conflict and duplication of metadata.

c) In the current frame, ontology used to characterize the knowledge space or other semantic technology is too simple. It is only defined by the characterization methods of knowledge space or the content and structure of the storage format. Learning object's Knowledge space field is not detailed representation, which includes the knowledge boundary and details of the relationship between knowledge points.

d) In the generation of content metadata, the current frame basically is through text mining method that simply produces the document keywords.

The development of learning objects hopes that smaller content annotation, division of internal knowledge points, this enables the learning object free assembly, so the text mining method alone is not enough. How to combine the generated metadata to annotate on the deep learning objects with knowledge division is an urgent problem to be solved.

At the same time, with the learning object research and development，the focus lies in its metadata and content packaging. Now, after being given a semantic information and knowledge needs, learning objects become intelligent learning object or the object of knowledge. Different from the prior concept of learning objects, knowledge learning objects have active, intelligent, meet the specific learning theory etc. The realization of these features contributes to the learning object in the field of knowledge space, generating the education oriented reusable, reconfigurable fine-grained learning object.

In summary, the research focuses on the basic goal which include active, intelligent, adapted to the teaching principle. To realize the semantic annotation of knowledge learning object, a study of knowledge of automatic generation of object may further support learning system based on learning object. In this study, taking Java courses as an example, the evolutionary algorithm to achieve effective educational ontology; the knowledge learning object labeling algorithm base on the weight of the relation between concept and concept of educational evolutionary ontology. On this basis, the ontology technology combined with the teaching principle, realizes the multi-data source (PPT, word, PDF, audio and video etc.) knowledge learning semantic object annotation. And the generation of knowledge is educational unit of learning objects as well as the formation of knowledge learning object automatic generation model.

III.   EASE OF USE

The prototype system structure model of MPLOG (Metadata Generation Platform for Learning Objects) is composed of database module. Resource analysis module consists of two parts, demonstrated in Figure 2:
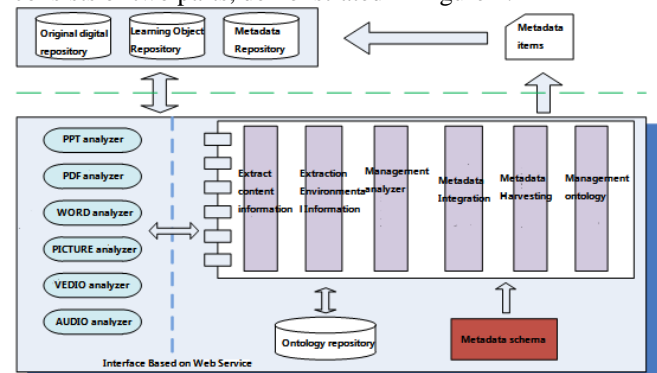

Figure 2.   Overall framework

Based on the analysis of the existing problems, this study proposes a flexible, extensible framework for automatic metadata generation platform, as shown in figure 6.

Metadata generation platform contains:the data storage module: it is responsible for the storage and management of learning object repository, meta database and the original digital repository. Meta database stores all the metadata of the learning object confirmed, for retrieval and use of learning objects. The original digital resource database, stores unprocessed media resources.

The content of information extraction module: based on metadata extraction technique, responsible for learning object content analysis. It is used for non-semi structure metadata generation, one of the core technologies of the metadata generation system.

Extraction of environmental information module: this module is to explore and mining metadata information in learning environment. This part of the metadata does not exist in the learning content, requiring learn object management system support.

Metadata harvesting module: using the metadata harvesting techniques and integration of existing metadata harvesting tools. Collecting metadata content already existed in learning objects. This module contains two functions. On the one hand, the extraction technology of metadata generated in the process of learning object content production, on the other hand, it extracts the existing non-embedded metadata.

Analyzer management module: analyzer management module is mainly responsible for the management of analyzer, including loading, deleting and configurating tasks. The system provides different analyzers for different file formats and automatically determines the learning object's file types. It then calls the corresponding analyzer to access the learning object content, providing information for metadata generation process.

Ontology management module: to meet the needs of unstructured semantic metadata generation, using ontology engineering approach to the management of course ontology, structure, education ontology etc. The module has access interface of course ontology system.

Metadata integration module: according to user specified metadata specification, metadata content extraction module generated will be assembled in this module. If the user is not satisfied with the results of metadata generation, system through the conflict resolution strategy, manual modification should obtain the final correct results.

IV. CALL RELATIONS BETWEEN MODULES

Learning object metadata generation platform is based on Figure 3, the refinement of the call relations module parts. As shown below:
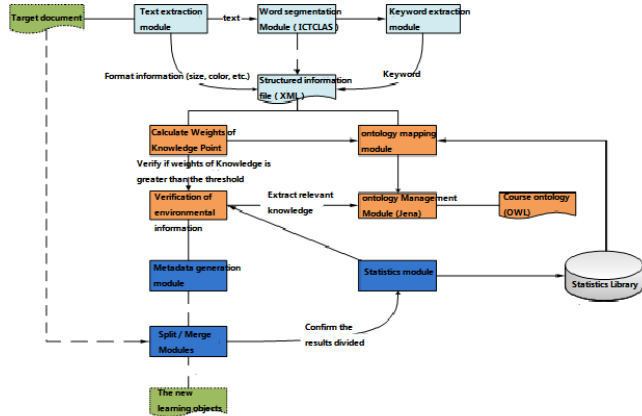

Figure 3. Call relations

A. Web Service analyzer

After the analyzer function has been achieved, using Web Service package it. Then it is released to the platform, and can be used for educational metadata generation. This paper uses Sun's JAX-WS2.0 framework to develop the Web Service. JAX-WS (Java API for XML-Based Web Services) is a Java application program interface based on XML. The client can choose RPC remote call and also choose the message mechanism based on SOAP. The programmer does not need to manage their SOAP messages. JAX will automatically convert the application calls for SOAP messages. In Server terminal, platform provides Web Service interfaces and implementations, and publish based on the Web server (e.g.: Apache). In Client terminal, you only need to specify the Web Service WSDL URL address, creating a local agent Proxy, and it can realize remote calls to the server. Of course, JAX-WS also provides a set of API calls operation of the underlying messages. You can directly through Dispatch using SOAP messages or XML message request or use Provider SOAP or XML messages. Based on WSIT (Web Services Interoperability Technology) provided by the web service, JAX-WS can be used to realize the cross platform calls between J2EE architecture and Microsoft's Net architecture.

B. Analyzer management

Analyzer management module is the core module in the platform. It is responsible for the issue of query analyzer. In the platform, all parsers use Web Service Technology, therefore the MGLOP platform using Sun's JAXR (Java API for XML Registries) technology to be responsible for the registration and administration of Web Service. JAXR provides several types of API used in the registration service (ebXML and UDDI) interaction. In general, the registration service is always in the form of running Web Services. JAXR supports three types of registration service type: JAXR Pluggable Provider, Registry-specific JAXR Provider and JAXR Bridge Provider (UDDI Registry and ebXML Registry/Repository etc.). The JAXR specification describes information model and UDDI registration standard binding description. JAXR has two parts, client and Provider:

1) The client is using the JAXR API to visit a registered Client terminal program through the JAXR provider access.

2) The fulfillment of Provider is a specific provider registration access or on the basis of a general standard (such as UDDI) typical provider registration to access .
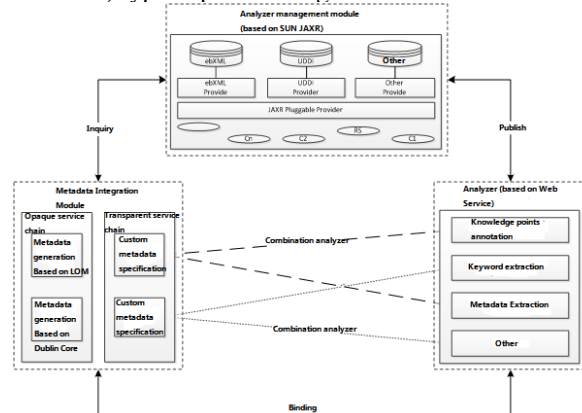

Figure 4. Example of Call relations

Figure 4 shows the call relationship between analyzer management module and metadata integration module. First of all, use the technology of JAX-WS metadata generation unit to get high cohesion into the analyzer (Web Service). Then use the analyzer to register on the management module. After that, metadata integration module query metadata generation services needed by analyzer management module (analyzer) and analyzer of info model information (including the WSDL description etc.). Finally, the metadata integration module can bind and invoke the need analyzer based on the related information.

C. To achieve a variety of metadata standard based on service chain

Launched in ISO/TC211 and OGC service architecture specification (ISO 19119), service chain to achieve a variety of metadata specifications based on the concept of service chain is proposed. In simple terms, the service chain is a sequence of services. It is composed of one or more services. Service function should meet the needs of requests. In ISO19119, according to the degree of control for different user service chain, service chain is divided into three types: opaque, transparent and translucent type. Opaque service chain is also called integrated service chain, where all services together and in the form of a single service. To hide the cooperation among all the individual service, the user does not need to know the details of the metadata generation process. This study uses opaque service chain technology. According to the metadata specification different implementation details (such as Dublin Core or LOM), the definition of service process and service chain in advance for each metadata standard (analyzer) calling sequence, and then use the method of aggregation service based on workflow, the application logic and process logic separation, in the logic of the services in heterogeneous leads to the problem of shielding, provides a flexible and reliable way to call. Opaque service chain reduces the user's requirements. Reducing the workload can satisfy the metadata generation tasks quickly.

V. CONCLUSION

This study is based on the service oriented architecture, designing a set of flexible, open, scalable learning object metadata generation platform -MGPLO (Metadata Generation Platform for Learning Objects). The platform not only provides the learning object metadata generated results, also publishes metadata generation service. Any user can create metadata service that our platform provided and integrate that into its own system. In addition, using the standard Web Service interface, the platform supports convenient, fast integrated traditional metadata generation tools, which used metadata and metadata extraction technology in the platform, to the maximum extent to achieve the openness and flexibility. In addition, for various forms of learning object characteristics, concept of resource analyzer has appeared in the platform. Face resources in different formats, respectively using a special analyzer analysis, annotate learning object content resources in different formats. Analyzers packaged as Web Service based on JAX-WS technical standards. With the increase, replace and modify managed analyzer in system, to meet the needs of Scalable System. As a standard Web Service specification, as long as the client can support standard WSDL interfaces, the analyzer and the platform provided can achieve cross-platform calls. This avoids the use of heterogeneous systems to face the re-development of new Web Service trouble. Finally, in order to support learning objects multi-level, fine-grained reuse, the platform introduces a semantic technology platform for knowledge learning object annotation.

[1] Kovacevic A, Ivanovic D, Milosavljevic B, et al. Automatic extraction of metadata from scientific publications for CRIS systems[J]. Program: electronic library and information systems, 2011, 45(4): 376-396.

[2] Greenberg J. Metadata Extraction and Harvesting: A Comparison of Two Automatic Metadata Generation Applications[J]. Journal of Internet Cataloging, 2003, 6: 4.

[3] Noufal P P. Metadata: Automatic generation and extraction[J]. 2005.

[4] Popov B, Kiryakov A, Kirilov A, et al. KIM–semantic annotation platform[M]//The Semantic Web-ISWC 2003. Springer Berlin Heidelberg, 2003: 834-849.

[5] Huynh D, Mazzocchi S, Karger D. Piggy bank: Experience the semantic web inside your web browser[M]//The Semantic Web–ISWC 2005. Springer Berlin Heidelberg, 2005: 413-430.

[6] Motelet O., Baloian N., Pino J. A. Learning Object Metadata and Automatic Processes [J]. Issues and Perspectives, DCC - Universidad de Chile, 2006.

[7] Jovanovic J, Gasevic D, Devedzic V. Ontology-based automatic annotation of learning content[J]. International Journal on Semantic Web and Information Systems (IJSWIS), 2006, 2(2): 91-119.

[8] Cardinaels K, Meire M, Duval E. Automating metadata generation: the simple indexing interface[C]. Proceedings of the 14th international conference on World Wide Web. ACM, 2005: 548-556.

[9] Baird K. Automated Metadata: a review of existing and potential metadata automation within Jorum and an overview of other automation systems[J]. Edina, Glasgow, 2006.

[10] Motelet O, Baloian N. Hybrid system for generating learning object metadata[C]//Advanced Learning Technologies, 2006. Sixth International Conference on. IEEE, 2006: 563-567.

[11] Flickner M, Sawhney H, Niblack W, et al. Query by image and video content: The QBIC system[J]. Computer, 1995, 28(9): 23-32.

[12] Jiang Xin. Instructional video retrieval platform construction and application [D]., East China Normal University, 2012.