Sharing open courseware content through learning objects standards

Sergio Freschi, Rafael Calvo
Web Engineering Group
The University of Sydney

A number of universities around the globe have decided to share their learning materials, making them available for use and modification by learners and other institutions. This initiative, started by the Massachusetts Institute of Technology (MIT), has opened up great opportunities for new ways of reusing content and for collaboration amongst teachers. It has also created new challenges, including the fact that there are thousands of courses available which cannot be easily modified by authoring tools or integrated into the courses managed by Learning Management Systems, due to the fact they are not in a standards compliant format. Regrettably, making them standards compliant is time-consuming and expensive, requiring a lot of effort from academics and institutions. This paper reports on a project to automatically structure learning materials and package them so they can be reused. We present a tool that produces IMS-CP compliant courses, with embedded metadata, in an automatic way by combining custom-built information extraction techniques and open e-leaning standards. Extensive testing was carried out on different learning object repositories.

Keywords: open courseware, e-learning, standards, metadata, content packaging, learning management systems, learning objects repositories

Introduction

The new trend of Open CourseWare (OCW), has been increasingly documented in the research literature (Baldi et al., 2002; Materu, 2004; Yue et al., 2004). It is an effort to share knowledge and make the best educational use of learning materials. Educators from around the world may share the content and the design of their courses, improving them through collaboration. Materu (2004) described how students may enhance their coursework or pursue self-study and how the general public may glimpse the depth and breadth of what leading universities are offering and benefit from reading lists and lectures.
The very first OCW project was carried out by Massachusetts Institute of Technology (MIT) in 2001. Since then, MIT (MIT OCW website, 2006) has published 1250 (Dec/2005) of its 1800 university courses. Most initial efforts have been of 'converting' (cloning) university face-to-face courses into 'digital' learning materials, and making them available to any educational institution, students and self-learners round the world. MIT's OpenCourseware materials have been translated to German, Hindi, Mandarin and Spanish, and the translations made available. Universities around the globe have followed, and many more courses have been made available. Our project tries to use these valuable resources in new ways to support student learning. Other projects have used learning resources for extracting metadata (Sonntag, 2004) or for adaptive learning experiences, as in the ELCAT system (Clements & Xu, 2005). However, these projects haven't made use of educational standards and/or the course data haven't been placed in databases, making it hard to reuse in other contexts or in standard Learning Management Systems.

The most relevant characteristic of these projects is that the learning activities are 'passive', in the sense that students do not get to interact with teachers or other students. This is an important difference with distance learning initiatives and all these projects made clear that they do not replace face-to-face learning. Students are not assessed so these are not degree-granting activities and students do not get formal credits. All course materials are free of charge and users (individuals and institutions) can modify and distribute the content as long as they adhere to its copyright license. The Creative Commons license (Creative Commons website, 2006) is the most commonly used and provides a flexible licensing approach for creative works. It is an alternative to the totally restrictive "All Rights Reserved" copyright license, so it has become a legal mechanism for many organizations to contribute their work to the 'commons' while clearly stating on what conditions the work is distributed.

Our project aims at helping people and institutions get the most out of the many resources that are available under these (CC) licenses.

The OCW repositories contain courses with more or less the same structure: a syllabus, that describes the course, its aims and expected outcomes, a calendar and lecture notes. Few courses include multimedia resources such as audio files and PowerPoint presentations. These repositories are in many ways static, and since they cannot be easily integrated into systems where students can engage in discussions, or participate in activities, their learning potential is limited. In order to reuse the materials in innovative ways, the content must be integrated into an educationally sound design. But this has a number of technical challenges (Boyle, 2003).

The learning content needs to be converted to new standards-compliant formats as described later. But since this process is very time consuming even for a small number of courses, the conversion must be done automatically. OCW websites were built using standard templates but the structure they have is very coarse, probably because further structuring the content would require additional work by academics. What is more, each organization uses different template formatting (e.g. some courses contain two-dimension tables while many others use just a single paragraph). The materials are not labeled with the necessary metadata that describes the topic, level of difficulty or knowledge required. Due to these limitations, teachers cannot reuse, repurpose and/or search or easily import their courses into an LMS where they could add learning and assessment activities.

Today, the courseware provides the same curriculum structure and content to different learners despite individual differences such as knowledge background, learning style, learning speed, etc. In the future, these materials would be better used in learner-adapted environments.

To overcome the above problems, this article introduces a multidisciplinary approach to combine information extraction techniques with open educational standards, specifications and learning objects repositories and by doing so, contribute to the understanding on how these specifications are applied in real implementations, and hopefully provide a benchmarking collection of learning materials that helps:

- enhance reusability and interoperability of digital learning resources
- enrich learner's online experiences
- teachers and instructional designers get the most of their actual learning materials or courses by adapting and personalizing content in a more efficient way; they can complement or supplement their local universities courses as well
- develop courses that are self-contained and can be shared easily.
- improve our ability to search and repurpose content.

Methodology

A number of technologies were used to download, parse and package the courses available at OCW websites. The MIT group coordinating OCW could have packaged the materials, but due to technical difficulties have not done it yet. Since the tool described here can be easily adapted to other similar repositories by changing the configuration files used by the parser, our approach is of a more general use. Our approach will allow non-OCW repositories to be packaged, distributed and reused in standard compliant formats.

OCW Spider, a custom-built Web crawler, was developed to download each course. A strategy was devised for gathering each part of a course, including all the attached resources, and building a new course structure in a reasonably efficient manner.

Tools were developed (e.g. a parser and a wrapper) to perform course structure analysis and extract details from each course such as: title, description, authors, and keywords among others. A wrapping tool was also developed, which produces a valid XML manifest file as output, ready to be uploaded into any IMS-CP (IMS Global Learning Consortium website, 2006) conformant LMS. This manifest file includes metadata as well.

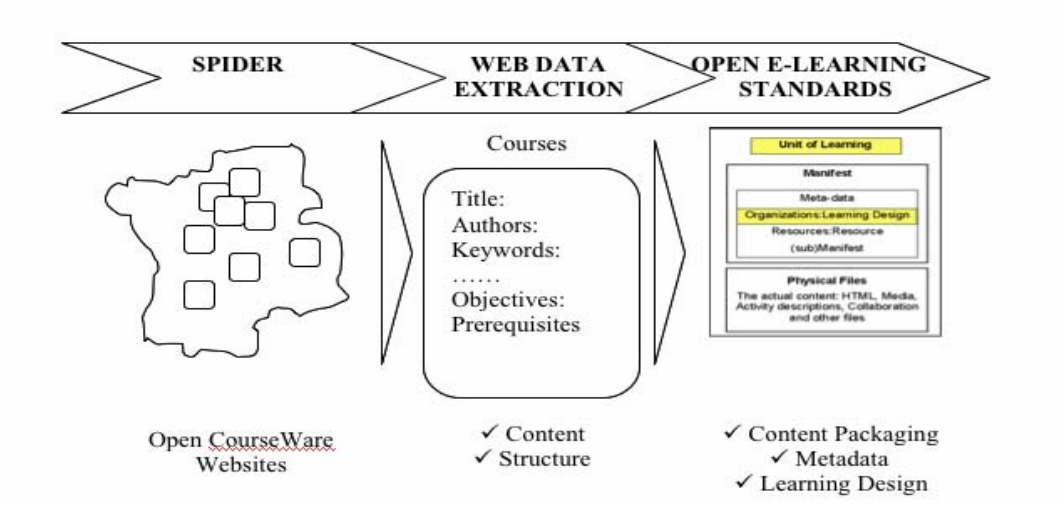

Figure 1: System implementation

Implementation design

Architecture

Figure 1 shows the system architecture of the proposed tool. The system consists of three modules: Spider, Parser and a Wrapper. These modules work as follows.

1 The Spider is in charge of downloading each course from the specified OCW website in an automatic way, including all the attached resources. It also creates a custom directory structure to hold each part of the course and where the rest of the process will take place.
2 Each course is then parsed in order to analyse each part of a course (e.g., syllabus, calendar, assignments, etc). Then, it extracts all available metadata (e.g., course title, authors, keywords, course size, discipline it belongs to, etc).
3 This module wraps up each part of the course producing an XML manifest file. Then all course resources such as: PowerPoint presentations, lectures (PDF files), audio files and so on are zipped together making this course a self-contained course.

We have used the Perl programming language, so our implementation is platform independent.

Conclusions

In this paper, we have outlined the motivations for sharing learning materials in order to promote the exchange of knowledge among teachers and learners. We have developed tools that can perform the task of automatic downloading, metadata extraction, parsing and standard compliant packaging of these materials.

We successfully processed 900 courses, about 4 GB of digital learning materials, of MIT OCW courses into an IMS-CP format. One of our evaluation criteria involved randomly choosing half of the total courses within each discipline (32 disciplines in total) and importing them in four different LORs in an attempt to evaluate and test our system implementation. Courses could be successfully imported, modified and then exported in the same format. This allows for a course (or part of it) to be reused.

Our three-step extraction process improves the reusability, interoperability and searchability of learning resources.

Some implementation issues have become apparent while running the modules. The most important one is the need for better information extraction techniques. Due to the fact that every OCW project has developed its own website structure and formatting styles, and our custom built extraction techniques

were developed based on MIT OCW website, when applied to a different website, they don't perform in the same way without applying some minor changes to the configuration files. However, we still get reasonable accuracy in the extracted data.

We have combined different technologies in one by implementing a tool that makes use of Web Information Extraction (WIE), World Wide Web Perl libraries, open educational standards and Learning Objects Repositories Systems.

The work done so far could lead to further research and development work using the IMS Learning Design Standard, which allows sequencing of activities within classes or sessions as well as synchronous activities such as chat, brainstorming sessions, etc.

References

Baldi, S., Heier, H., & Stanzick, F. (2002). *Open Courseware vs. Open Source Software – a Critical Comparison*. 10th European Conference on Information Systems, 1375–1383.

Boyle, T. (2003). Design Principles for authoring dynamic, reusable learning objects. *Australasian Journal of Educational Technology*, 19(1), 46–58.

Clements, I., & Xu, Z. (2005). ELCAT: an e-learning content adaptation toolkit. *Campus-Wide Information Systems*, 22(2), 108–122.

Creative Commons website. http://creativecommons.org/ [viewed 10 Aug 2006]

IMS Global Learning Consortium. *IMS Content Packaging Best Practice and Implementation Guide* v1.1.4 [Electronic Version]. Inc. http://www.imsglobal.org/ [viewed 18 Sep 2006].

Materu, P. N. (2004). *Open Source Courseware: A Baseline Study* [Electronic Version]. The World Bank from http://www.worldbank.org/afr/teia/open-courseware.pdf.

MIT OpenCourseWare (OCW) website. http://ocw.mit.edu/index.html [viewed 18 Sep 2006].

Sonntag, M. (2004). *Metadata in E-Learning Applications: Automatic Extraction and Reuse*. Paper presented at the 12th Interdisciplinary Information Management Talks, Universitätsverlag Rudolf Trauner, Linz.

Yue, K.-B., Yang, T. A., Ding, W., & Chen, P. (2004). Open courseware and computer science education. *Journal of Computing Sciences in Colleges*, 20(1), 178–186.

Author contact details

Sergio Freschi, The University of Sydney, School of Electrical and Information Engineering, Building J12, Sydney NSW 2006, Australia. Email: sfreschi@ieee.org