Creating Discoverable Learning Content Using a User-Friendly Authoring Environment

Ben De Meester[1], Hajar Ghaem Sigarchian[1], Tom De Nies[1], Anastasia Dimou[1], Wesley De Neve[1,2], Erik Mannens[1], and Rik Van de Walle[1]

[1] Ghent University – iMinds – MMLab, Belgium
[2] KAIST – Image and Video Systems Lab, Republic of Korea
{ben.demeester, hajar.ghaemsigarchian, tom.denies, anastasia.dimou, wesley.deneve, erik.mannens, rik.vandewalle}@ugent.be

Abstract. Digital learning content is becoming more and more commonplace. However, the creation of this learning content is usually done in specialized publishing platforms, that may lock in the content. This prevents the learning content from being discovered automatically, and hinders its uptake. A more general approach for authoring discoverable learning content is needed. We built a user-friendly authoring environment for creating digital publications, enriched with multimedia and interactivity features. This authoring environment outputs the created content in open formats, independent of any publishing platform. We adjusted our authoring environment to create semantically enriched learning content using the RDFa syntax, without the need for manual intervention.
In this paper, we show how well-structured and well-annotated digital learning content can be created and validated using our authoring environment. The result is a discoverable publication in an open format, ready to be distributed as is, or ready to be imported in existing publishing platforms.

Keywords: Authoring Environment, EPUB 3, Linked Learning

Introduction

Digital learning content is being widely adopted as an addition to print-based learning content. More recently, this digital content is also being used as a primary source for online education initiatives such as *Massive Open Online Courses* (MOOCs) [7]. Because of this increased interest in digital learning content, the amount of available content also increases, and it becomes more and more important to efficiently and effectively discover and (re-)use available learning content on the Web.

Digital learning content is usually created and published using specialized platforms. However, these platforms (1) lock the learning content in a custom data format and (2) offer limited support to make the learning content machine-understandable, hampering the interlinking and querying of learning content, and thus its uptake [6].

To overcome the content lock-in, we propose an authoring environment that is independent of the publishing platform, and outputs the content in an open format. This way, the authored publication can still be imported in existing learning publication platforms, but the content remains accessible by other consumer services.

To increase the discoverability of the authored publication, learning content needs to be annotated properly to achieve machine-understandability. However, manual annotation is to be avoided, as authors of learning content should not be burdened with this task. Therefore, the aforementioned authoring environment has to be capable of automatically inserting the appropriate metadata for the given content.

After giving some background information in Section 1, we describe the built authoring environment to tackle the aforementioned limitations in Section 2. Finally, we conclude in Section 3.

1 Background

The following paragraphs will give an overview of how currently learning content is authored and published, and how it can be made discoverable.

1.1 The Learning Content Publishing Chain

Learning content is usually authored in a *Learning Content Management System* (LCMS) [4]. An LCMS is an authoring environment that creates learning objects. These learning objects are mostly packaged using the *Sharable Content Object Reference Model* (SCORM) specification [2].

SCORM is an e-learning specific solution to package and transfer learning content between *Learning Management Systems* (LMSs). Furthermore, the SCORM specification also defines the communication between the client-side application that uses SCORM learning objects and the host system (e.g., to verify the answers to questions inside the learning object). As SCORM is LMS-oriented, the knowledge contained in a SCORM learning object cannot be transferred to systems with a different scope. Moreover, since the resulting learning object is a packaged format, the original learning object is not really shared, but duplicated every time it is (re-)used. When the learning content is updated, the updates cannot be disseminated automatically to the users that already have a previous version of that learning content.

An LMS is a platform where teachers can present learning content to their students [5]. It can be seen as a *Content Management System* (CMS) such as Drupal[3], extended with extra tools to, e.g., administer, track, and report the learning situation of the individual students. Currently, many systems exist that combine the functionalities of an LCMS and an LMS, thus eliminating the need to export learning content from an LCMS and import it into an LMS. One example of such an integrated system is Moodle[4].

[3] `https://www.drupal.org/`
[4] `https://moodle.org/`

Although such an integrated approach does improve the ease of use for the teacher (as he or she can seamlessly create and distribute new learning content), it prevents the adoption of this learning content by other publishing environments, which further hinders the discoverability of the created learning content.

Meanwhile, *Open Web standards* such as *HTML* are being used as the de facto publication format on the Web to publish any kind of (possibly enriched) content. These HTML publications can be easily distributed and can be made accessible offline by packaging them as *EPUB 3*[5], an open packaging format that is capable of packaging any kind of HTML web page as an e-book. As learning content can easily be created in HTML, publishing learning content in HTML and packaging it as EPUB 3 could thus be an open alternative to the current publishing chain of creating learning content in an LCMS, packaging it as SCORM, and distributing those packages among LMSs.

Furthermore, the EPUB 3 publications can be semantically enriched with provenance information[6]. This provenance information can be used to compare versions of the same resource, and different publications of the same learning content can be compared and discovered using the included provenance information. This allows for the automatic dissemination of updates of the original learning content to already distributed publications.

1.2 Discoverable Learning Content

To make digital publications discoverable, machine-understandable metadata is needed [6]. The *Learning Object Model* (LOM) [4] is the standardized open metadata specification for learning content, also used by most currently available LMSs. This LOM specification is mapped to the *Learning Resource Metadata Initiative* (LRMI) tagging schema[7], which in turn is accepted as a part of *Schema.org*[8], a widely used tagging schema on the Web. This metadata can be inserted into HTML publications using the *microdata* format or the *RDFa* format [1]. Using these standards, and because of the fact that EPUB 3 also uses HTML to publish its contents, EPUB 3 packaged files can be annotated with LOM metadata.

We can conclude that EPUB 3 can be used as a discoverable open format for publishing learning content. In fact, there are recent efforts by the *World Wide Web Consortium* (W3C) and the *International Digital Publishing Forum* (IDPF) on a specification for using EPUB 3 in an educational environment, called *EDUPUB*[9].

[5] `http://idpf.org/epub/301`
[6] `http://www.w3.org/TR/prov-o/`
[7] `http://www.lrmi.net/the-specification`
[8] `http://schema.org/`
[9] `http://www.idpf.org/epub/profiles/edu/spec/`

2 Overview of the Authoring Environment

We developed an authoring environment for creating discoverable digital learning content, making use of Open Web standards (HTML5, EPUB 3) [3]. The usage of these standards implies that learning content can be made accessible without users having to make use of a specific L/CMS.

To improve discoverability and overall usage, the resulting learning content is automatically semantically annotated in the form of RDFa (Lite). Indeed, these annotations facilitate machine understandability, which makes it easier to query and (automatically) interlink learning content. Furthermore, the authoring environment natively handles all multimedia and interactivity features currently available on the Web. This means that adding audio, video, and interactivity widgets to learning content is as easy as adding textual content. The result is a user-friendly authoring environment that makes it straightforward to create well-formed learning content that can be published (possibly to an existing L/CMS), discovered, and interlinked.

The authoring environment[10] is a Web Application that can export well-annotated HTML5 or EPUB 3 publications with advanced multimedia and interactivity features. When authoring new learning content, the user can manually choose between a number of pre-defined, well-annotated snippets that can be inserted into the learning content and arranged by simple drag-and-drop actions. These snippets follow (among others) the EDUPUB specification, and are annotated with Schema.org terms in RDFa. Different snippets, that follow different specifications, can be created as well.

The plug-in structure of the editor assures that custom actions can be performed for certain types of snippets. For example, when inserting the intended age for a particular exercise, the user gets a dropdown menu of available age ranges (Figure 1), instead of having to manually type the age range (and accompanying semantic annotation) in the learning content. This improves user-friendliness while also streamlining the workflow for authoring digital learning content. During and after authoring, the learning content can be validated using an *XML Schema Definition* (XSD) document to verify whether the content has been structured properly.

When reviewing the machine-understandability of the learning content by making use of an RDFa inspector[11], we can verify that the learning content is well-structured and that it contains multiple semantic annotations, without having the user to edit any HTML5 code. As a result, the machine-understandable learning content can be used as a base for interlinking. For instance, by querying the Web for all exercises with the keyword `Algebra`, and with a typical age range of `11-14`, a user can retrieve all Middle School algebra exercises.

[10] Showcase: `http://users.ugent.be/~bjdmeest/lile_demo.mp4`
[11] E.g., `http://rdfa.info/play/`

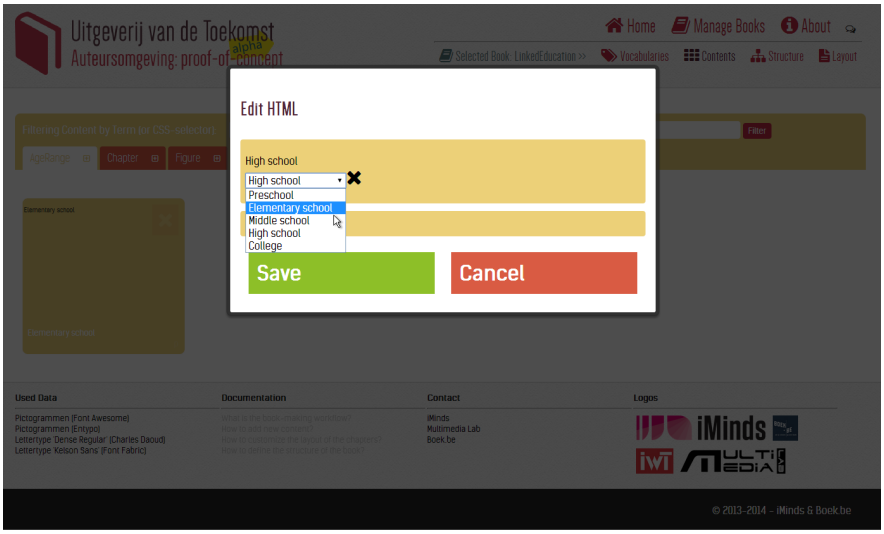

Fig. 1. User-friendly semantic annotation of learning content using the built authoring environment

3 Conclusions and Future Work

By building an authoring environment that can automatically add semantic annotations to digital learning content, we remove the cumbersome burden of having to manually insert annotations in digital learning publications. The resulting output consists of well-structured, well-annotated, and validated learning content that can be published to multiple platforms. As such, our authoring environment makes it easier for a general audience to create well-defined learning content, thus contributing to an improved production and uptake of machine-understandable learning content.

For now, the definition of the snippets needs to be done manually by an expert before well-annotated content can be created. For future work, it could be possible that, based on the XSD document or another type of specification document, the snippets can be created automatically. This would also broaden the use of the authoring environment, as any type of publication (e.g., medical or legal) could be created, as the snippets do not need to be created manually for every different use case.

Acknowledgements The research activities described in this paper were funded by Ghent University, iMinds, the IWT Flanders, the FWO-Flanders, and the European Union, in the context of the project "Uitgeverij van de Toekomst" (Publisher of the Future).

References

1. Bizer, C., Mendes, P.N., Jentzsch, A.: Topology of the Web of Data. In: Semantic Search over the Web, pp. 3–29. Springer Berlin Heidelberg (2012), http://dx.doi.org/10.1007/978-3-642-25008-8_1
2. Bohl, O., Scheuhase, J., Sengler, R., Winand, U.: The Sharable Content Object Reference Model (SCORM) - a critical review. In: Werner, B. (ed.) Proceedings of the International Conference on Computers in Education. vol. 2, pp. 950–951. IEEE, IEEE Computer Society, Auckland, New Zealand (December 2002), http://dx.doi.org/10.1109/CIE.2002.1186122
3. De Meester, B., De Nies, T., Ghaem Sigarchian, H., Vander Sande, M., Van Campen, J., Van Impe, B., De Neve, W., Mannens, E., Van de Walle, R.: A Digital-First Authoring Environment for Enriched e-Books using EPUB 3. In: Proceedings of the 18th Int'l. Conference on Electronic Publishing (ELPUB), June 19-20, Thessaloniki, Greece (2014), http://dx.doi.org/10.3233/978-1-61499-409-1-68
4. Harman, K., Koohang, A.: Learning objects: standards, metadata, repositories, and LCMS. Informing Science Press (2007)
5. Ninoriya, S., Chawan, P., Meshram, B., VJTI, M.: CMS, LMS and LCMS for eLearning. IJCSI International Journal of Computer Science Issues 8(2), 644–647 (2011)
6. Rosati, A., Mayernik, M.: Facilitating data discovery by connecting related resources. Semantic Web Journal (2013)
7. Wulf, J., Blohm, I., Leimeister, J.M., Brenner, W.: Massive open online courses. Business & Information Systems Engineering 6(2), 111–114 (2014), http://dx.doi.org/10.1007/s12599-014-0313-9