

Connecting Connexions: Organizing and Integrating Open Learning Content with Topic Maps

Darina Dicheva¹ and Lars Johnsen²

¹Winston-Salem State University, 601 MLK Drive, Winston-Salem, NC 27110

²University of Southern Denmark, Engstien 1, 6000 Kolding, Denmark
dichevad@wssu.edu, larsjo@sitkom.sdu.dk

Abstract. In this paper we propose utilizing of topic maps as an integration tool for Connexions¹ content aiming at creating more subject-oriented, knowledge-based ways of connecting resources in the Connexions repository. We discuss to what extent and how open content, including metadata, may be extracted from Connexions e-learning modules and mapped onto topic maps using XSLT transformations and how this may be done in a topic map editor to facilitate further processing, enrichment and merging of subject-oriented learning content. The proposed approach is implemented in the educational topic maps editor TM4L.

Keywords: Digital Libraries, Open Learning Content, Topic Maps, XSLT

1 Introduction

In recent years, the notion of open educational resources has attracted a lot of attention within the area of e-learning. The idea of sharing and reusing content and tools across disciplines, departments and institutions is seen as one of the ways of addressing an increasing need for updated learning materials of high quality in many professional fields. One of the most interesting examples of open content is found in the Connexions project [1], a project at Rice University, USA, aimed at developing and distributing free and reusable content for educational purposes at all levels [2]. Teachers, college professors and even students are encouraged to supply relevant “learning modules” to a common repository and to share and reuse these modules under a Creative Commons license. The Connexions project thus encourages collaboration and sharing of content in a typical Web 2.0 fashion.

Connexions content is structured and marked up using an XML based language called CNXML. Being a document-oriented language, however, CNXML does not include elements for describing conceptual domain structures or explicit mechanisms for establishing subject identity to allow for smooth interoperability or federation of content items. Another potential shortcoming of Connexions modules lies in their inherently modular design. At worst, modules may become “information islands” where knowledge is either marooned, as it were, or only superficially connected to

¹ <http://cnx.org/>

external knowledge sources through hyperlinks or the like. Therefore, ways to organize and integrate modules and their content are obviously needed.

We suggest that open content stored in Connexions modules may be organized and integrated using Topic Maps [3] for improved navigation and more extensive reuse. In this paper we discuss to what extent, and how, content and metadata may be extracted from Connexions modules and mapped onto topic maps and how this may be done in a topic map editor in order to support further processing, enrichment and merging of subject-oriented learning content, including metadata. More specifically, we propose harvesting specific module content and mapping it onto topic maps through the use of XSLT transformations (also known as XSLT style sheets) [4]. The proposed approach is implemented in the educational topic maps editor TM4L [5].

2 Learning Modules in Connexions

Content in Connexions is created and distributed as learning modules, self-contained text units about one topic. These modules consist of components fulfilling various communicative roles: metadata, text, multimedia objects, links and so on. Modules are structured and marked up using CNXML, a proprietary XML-based language. A number of editing and validation facilities are available to authors who want to write modules in CNXML and publish them on the Connexions web site. CNXML comprises tags necessary to mark up various text elements in modules: quite a few structural tags (section, paragraph, link and so on), some semantic ones (example, definition, meaning, etc.) and few presentational ones (like emphasis). Furthermore, the CNXML schema incorporates elements to mark up different media types, internal and external links, as well as common metadata categories (date, version, keywords, etc). Learning modules are published on the Connexions web site as “structured XHTML”, that is to say as ordinary “presentational” web pages but endowed with a certain amount of mark-up revealing their underlying XML structure. This is typically done through `<div>` and `` elements carrying a class attribute signalling a certain underlying category, for instance `...`. Each learning module is realized as two web pages, one containing the actual learning content and one conveying important metadata.

A number of tools are offered at the Connexions web site to support findability and usability of learning modules and to create some cohesion among them. Tools include excellent search facilities and tools to group or link modules. Learning modules may be linked to form sequences or “courses” and so-called lenses may be applied to certain subsets of modules to group them for a particular purpose. However, as already suggested, improved navigation and more extensive reuse and integration of module content may be possible through the application of a technology like Topic Maps. Topic Maps is a standards-based model for organizing and classifying information in a wide range of structures (thesauri, taxonomies, ontologies, etc.). In particular, the Topic Maps paradigm constitutes an approach to superimposing a knowledge layer on top of distributed web resources. Thus, utilizing Topic Maps as an integration tool for Connexions content will mean a step towards creating more subject-oriented, knowledge-based ways of connecting resources within, and outside, the Connexions repository.

3 Mapping Connexions Content to Topic Maps with XSLT

In this section we briefly indicate how specific elements and element structures in Connexions learning modules map onto Topic Maps constructs like topics, base names, occurrences, associations, subject identifiers and subject locators. Topic maps generated from current modules may be said to constitute a kind of “universe of discourse” representing a network of relevant objects (modules, sections, persons, terms, keywords, etc.) as well as properties and resources intrinsic to these objects (summaries, definitions, licenses, and so forth). We suggest that mappings from Connexions learning modules to topic maps can be encoded as XSLT transformations relatively easily, due to the consistent XHTML representation of the modules.

Mapping a learning module generates a topic map containing a list of topics, or subjects. We distinguish between “addressable subjects”, such as modules, sections or multimedia objects and “non-addressable” subjects, such as authors and topics represented by terms or keywords. These topics are given *subject identity* (in Topic Maps terms): they are named and/or endowed with more specific identity markers such as *subject identifiers* (for non-addressable subjects) or *subject locators* (for addressable subjects). Sections are given their first heading as a base name; modules get their title as their base name and their web address as a subject locator; authors retain their own name as their base name while the address of their Connexions page becomes their subject identifier. In addition, we consider other potentially useful examples of subject identity assignment: some marked up elements like *term* and *foreign* may carry an underlying “source” attribute whose value may be conceived of as a kind of subject identifier (hence as a mechanism for subsequent topic merging).

The learning module mapping also results in information being attached to topics. Some information resources are mapped as occurrences: for example, definitions, if found, are mapped as occurrences of terms while metadata categories, such as summary, version, language and license, are attached as occurrences to modules.

```
1 <xsl:for-each select="//xhtml:tr[xhtml:td='Authors:']/xhtml:td/xhtml:span">
2 <topic id="{substring-after(xhtml:a/@href,/member_profile/)}">
3 <instanceOf>
4 <topicRef xlink:href="#person"/>
5 </instanceOf>
6 <subjectIdentity>
7 <subjectIndicatorRef xlink:href="http://cnx.org{xhtml:a/@href}"/>
8 </subjectIdentity>
9 <baseName>
10 <baseNameString>
11 <xsl:value-of select="xhtml:a[1]" />
12 </baseNameString>
13 </baseName>
14 </topic>
15 </xsl:for-each>
```

Fig. 1. XSLT code snippet.

Other salient information patterns in learning modules are manifested as associations between topics: authors and modules are related in an *authorship*

association, while a module and its sections are related in a *containment* association. The small snippet of XSLT code given in Fig. 1 illustrates the extraction and mapping of some author information. In this example, authors are identified via a search through `` elements in an XHTML table row containing the string “Authors” in a table cell. Information pertaining to authors, such as their names and Connexions addresses is found in the surrounding mark-up. Each *author* is transformed into a topic with a corresponding id, base name and subject identifier and the topic is assigned to the category of *person*. A simple string manipulation is carried out to ensure legal and useful id and identifier values.

4 Extracting Topic Maps from Connexions Content with TM4L

To test our hypothesis, we have designed a plug-in for the topic maps editor TM4L that implements the suggested mapping using XSLT transformations. The tool performs automatic information extraction from a specified Connexions module (its content and metadata web pages respectively) and uses the extracted conceptual structures to generate a topic map. The user specifies the URL of the Connexions module. The generated XTM file (topic map) is merged with the currently open topic map.

The implementation of this functionality involved the development of XSLT style sheets and integrating an XSLT processor in TM4L. For the latter we use Xalan-Java Version 2.7.1, the Apache XSLT processor for transforming XML documents into HTML, text, or other XML document types. It implements the W3C Recommendation XSL Transformations (XSLT) Version 1.0, and XML Path Language (XPath) Version 1.0 [6].

The integrated XSLT processor is an important extension of the overall functionality of TM4L since it supports general XML transformations, thus allowing experienced TM4L authors to import or write their own XSLT transformations for topic map generation or for rendering a topic map in a certain way. As a result, users will be able to choose an XSLT style sheet (typically containing an XTM to HTML or XTM to SVG mapping) for presentation purposes. In addition, the XSLT processor will allow converting topic maps created in XTM 1.0 to XTM 2.0 within TM4L.

Work on constructing a web crawler capable of harvesting web pages linked to the page under transformation is currently underway.

5 Conclusions

Although a fair amount of valuable information (with relatively little noise) can be gleaned from Connexions learning modules and mapped onto topic maps, the CNXML mark-up – and the way it is applied by Connexions authors – still poses some challenges for the automatic construction (and subsequent merging) of semantically rich topic maps. One problem is the lack of explicit mark-up for encoding subject identity for things like keywords, that is to say the things learning modules are “really about”. Another problem concerns the absence of tags for

marking up conceptual domain structures in learning modules, e.g. associations such as *is-a*, *is-a-kind-of*, *part-of*, *has-a*, etc. Last but not least, authors greatly differ in how extensively, precisely, and consistently they apply mark-up, especially “semantic” tags like *term*, *rule*, *example*, *definition*, and so on. A possible way to improve this would be to motivate Connexions authors to use writing methodologies like Structured Writing [7, 8]. We also suggest that a feasible direction for further development of CNXML would be its reformulation as a more subject-centric mark-up language for describing, encoding and organizing open learning resources.

TM4L already had a plug-in offering two options for automatic extraction of topic map constructs (concepts and relationships): from a website specified by the author and from Wikipedia [9, 10]. The idea was to support topic maps authors with limited knowledge of IT and ontologies by extracting concepts and relationships from web pages and using them for building “draft” topic maps for the authors to start with. The Wikipedia corpus was used as a rich source of common sense conceptualization, including reliable and sharable concept names.

The extraction of conceptual information from the Connexions web site proposed in this paper differs from the above mentioned options not only in its implementation approach but also in its goal. Instead of helping instructors build topic map-based learning content of their own, it aims at helping them build a *map* of already developed (and contributed to Connexions) open content of their interest.

We believe that Connexions will grow to become a major open content learning repository. Thus the proposed here approach and tool for building personalized conceptual maps of its content will offer valuable support to its users.

References

1. Connexions, <http://cnx.org/>
2. Henry, G.: Connexions: An Alternative Approach to Publishing, 8th European Conference on Digital Libraries, ECDL 2004, Springer, 421-431 (2004)
3. Park, J., Hunting, S.: XML Topic Maps: Creating and Using Topic Maps for the Web. Boston: Addison-Wesley (2003)
4. XSL Transformations (XSLT), <http://www.w3.org/TR/xslt>
5. Dicheva, D., Dichev, C.: TM4L: Creating and Browsing Educational Topic Maps. British Journal of Educational Technology - BJET, 37(3), 391-404 (2006)
6. Xalan-Java Version 2.7.1, <http://xml.apache.org/xalan-j/>
7. Johnsen, L.: Information Maps Meet Topic Maps. From Structured Writing to Mergeable Knowledge Webs with XML. In: Madsen, B.N. & Thomsen, H.E. (eds.): Terminology and Content Development, Litera (2005)
8. Johnsen, L.: Open Educational Topic Maps: A Text-oriented Perspective (forthcoming) (2008)
9. Dicheva D., Dichev C.: Helping Courseware Authors to Build Ontologies: the Case of TM4L, 13th Int. Conference on Artificial Intelligence in Education, AI-ED 2007, July 9-13, 2007, LA, California, 77-84 (2007)
10. Roberson, S., Dicheva, D.: Semi-automatic Ontology Extraction to Create Draft Topic Maps. In Dicheva, D., Pauca P. (Eds) 45th ACM Southwest Conference, ACMSE 2007, Winston-Salem, NC, March 18-19, 2007, 100-105 (2007)