Markup of Educational Content

Monica Langerth Zetterman
(monica.langerth@ped.uu.se)
Digital Literature, Uppsala University

The purpose of this presentation is to illustrate the implementation of methods and tools for separating markup, from the actual XML document. This is being done in the project Educational Content Markup[1] where the object is to explore and develop tools and methods for the creation and management of encoded, modularised and portable digital content to be used in educational settings.[2]

The underlying idea in the project is to enable the production, use and reuse of document content in different fields of application and to support various groups of users and end-users. Concerning the fields of application our focus is on teaching and learning in higher education and the production of academic research publications. Hence, one group of users are teachers creating e.g. collections of documents to be used in classes. We picture both students and teachers as a group of end-users. Another group of users are scholars and other content providers concerned with document creation and the production process.

In this project we have chosen to differentiate two kinds of markup in order to distinguish fields of application and groups of users. Thus, the differentiation described below guides our exploration of methods and the development of tools.

- Original/internal markup: Markup placed within the document - as opposed to stand-off markup- in this context often capturing the structure of a document or some specific features such as personal names. Either the markup was already added to the document by another part or the markup is created during the document creation process. Markup is automatically added by controlled input using templates in the DiVA system[3] or semi-automatically by the content producer.

- Additional markup: Markup added after the creation process which can be done with optional annotation techniques such as *Annotea* or the *TEI* guidelines[4]. The additional markup might be added and stored internally in the document, or, annotations might be added and stored externally in separate XML-files, not altering the any markup internally added to the document. For example *Annotea* allows annotations to be attached to resources without modifying the original/internal markup.

It is important to point out that our two views of markup can be, and often are different, but this does not mean that the methods are non-exclusive. However, a markup scheme designed to catch in-depth descriptions of a chosen aspect, rather than describing the internal structure of a document, serves a special purpose and therefore anticipates a different data model than a structured oriented scheme. That is one reason why we, in this project, think it is useful to separate the additional markup from the original/internal markup.

Other reasons to separate different kinds of markup have to do with different kinds of uses and user-groups. By various means of organising, filtering and presentation, one and the same content might be utilized in a range of subject areas and for different educational levels, from primary to higher education or in-service and in-company training. In trying to accomplish this, several different XML techniques and tools are being used for the implementation of a prototype of an educational content management system, currently consisting of three related parts.

1. In one part of the project, a prototype provides means to take the descriptive markup added to the internal XML markup and treat it as external annotations. The browser is used as the markup client, using *Annotea* protocol both on the client side and both on the server side. At the backend a RDF[5] database is used as the content store. This architecture allows usage of arbitrary application profiles and an ability to integrate the information with other tools.[6]

2. In the second part we explore tools and schemes for adding external markup - stand-off markup - to documents already annotated with TEI/XML markup of personal names, dates and places. Here we will test schemas providing uses of authority files such as the first draft of *MADS* which is designed for description of agents (e.g. people, organisations) and terms (e.g. topics, geographical places).

3. In the third part we are exploring applicable methods and tools for adding markup when authoring. The idea is to integrate the process into the author's natural workflow such as the software they use for text editing (e.g. *OpenOffice* or *MS Office*). These kinds of tools should support and facilitate work performed by the author during the production and revision of texts.

This poster will provide demonstration of a prototype tool implementation and uses, examples of markup, and a description of current status in the project progress. We are looking forward to get a chance to exchange ideas and well-informed feedback on the underlying ideas, choice of methods and implementation of tools in this project.

---

1. The project Educational Content Markup is a joint project and the partners are: The research program Digital Literature and the

Electronic Publishing Centre, both Uppsala University. The Swedish Royal Library, the National Agency for Education in Sweden, Ekelunds Publishing Company and Center for IT in Northern Sweden. The project is funded by the Swedish Agency for Innovation Systems. The author of this poster presentation is co-ordinating the project. For participants and further information, see <http://www.skeptron.ilu.uu.se/broady/dl/mu.htm>.

2. For further information on the research program Digital Literature publications and research areas, see: <http://www.skeptron.ilu.uu.se/broady/dl/>.

3. *DiVA* is the Swedish acronym for *Academic Archive Online*. For further information on the *DiVA* system and related publications, see <http://publications.uu.se/epcentre/projects.xsql> , for the technical report, see: <http://publications.uu.se/epcentre/diverse/hardware_software.pdf>.

4. See <http://www.tei-c.org/Activities/SO/sow06.html>.

5. Resource Description Framework, see <http://www.w3.org/RDF/>.

6. For a technical report on the prototype, see Engman.

Bibliography

Annotea. Accessed 2005/03/04. <http://www.w3.org/2001/Annotea/>

Engman, J. *Handling fine-meshed filtering from archive to resource level.* Master thesis, Umeå University, 2004.

Metadata Authority Description Schema. Accessed 2005-03-14. <http://www.loc.gov/standards/mads/>