УДК 519.7:007.52

N. LESNA, prof., V. REPKA, T. SAKALO

XML APPLICATIONS FOR STUDY MATERIAL MARKUP IN DISTANCE LEARNING

Introduction

The need for structured text data appeared in the Internet in the context of highly dynamic development in the area of distance learning technologies. It is extremely hard to analyze a continuous unstructured text. A parsing program cannot distinguish logical fragments within an unordered text flow; it can neither define data elements of such a document, nor their relative position.

To solve this problem a Hypertext Markup Language, a simplified version of SGML (Standardized Generalized Markup Language, ISO 8879), was introduced. It does not require laborious syntax definition, it is easy for memorizing, and it is absolutely open for application programmers. HTML is aimed at the document structure design for its subsequent visual presentation in the global network. The existing ideology defined a technology for development of information resources for the Web.

Kharkiv National University of Radio Electronics (KNURE) actively participates in the development of concepts, principles and models of distant learning to meet conditions of the Ukrainian multilevel education system. KNURE is a member of the Ukrainian Distance Learning Alliance.

Researchers in KNURE have used the hypertext for informative educational server development, and have designed HTML templates for hypertext learning materials.

1. The preliminary analyses

An essential disadvantage of the developed information resources consists in an unbreakable association of resource contents and visual representation.

The Cascading Style Sheets standard provides solution for that problem. It defines parameters for any HTML tag representation making it possible to divide the logical structure of a document and its visual representation [1–4].

As for now plenty of markup language specifications are available. These include document markup definitions for internet-related application in various areas, including science, education, finance and commerce. Let us consider the most important specifications, which can be used for development of the educational system for distant learning.

The most common specification is the Mathematical Markup Language (MathML), associated with representation of specific documents containing mathematical formulae and other expressions. It provides rules for very specific tasks and does not touch on any of other needs when developing distant learning courses. It can be suitable for description of the contents portions only.

The Synchronized Multimedia Integration Language (SMIL), recommended by W3C standard, provides a way for creation of documents containing synchronized multimedia information. Such documents contain a set of directives describing textual, video and audio data and define a playback sequence. SMIL can help to create real-time web presentations and reuse the existing multimedia objects. But much like MathML this language is quite specific in nature and cannot cover the whole range of needs for the distant learning.

A group of companies, including AT&T, CANADIAN IMPERIAL BANK of Commerce, CyberCash, DigiCash, Fujitsu and some others, has agreed to design the Internet Open Trading Protocol (IOTP), formerly the Open Trading Protocol (OTP), a specification that provides protection when performing an interoperable electronic payoff in the Internet commerce. It is optimised for the case where a buyer and a merchant are not previously acquainted and the payment system is independent. It addresses security during the online payment process. This protocol can be applied to the payment system, if it is included in the distant learning educational resource.

The first language designed specifically for educational projects is the Tutorial Markup Language (TML) [5]. TML is an interchange format designed to separate the semantic content of a question from its screen layout or formatting. The language is designed to support several different types of questions within the same content model. TML 4.0 is essentially the super-set of HTML with new elements added to describe information in question.

TML version 4.0 has been specified using SGML, the ISO standard language for the formal document types description. Future versions of TML are likely to be represented using the Extensible Markup Language (XML), a simplified subset of SGML, and structured using the data modelling facilities of the Resource Description Framework (RDF).The work is currently in progress on the TML version 5.0, re-named as *Tutorial Modelling Language* to mark the shift to the RDF-based approach.

Moreover, it facilitates search, index and procession of test results. This language defines e.g. an admissible number of attempts to answer, test questions, correct answers as well as permitted hints.

But the language is not completely suitable for our needs. It also can markup only a quite restricted set of study materials and it was designed without taking into account some country-specific options and features.

Another educational language,the Learning Material Markup Language (LMML) is dedicated to the adaptive learning. It is a meta-description of educational materials divided into modules and deployed on distributed web sites [7]. This markup framework is the XML framework and it is based directly on the Passau Teachware Model like all its LMML languages. It uses modularity of XML to include the blocks coded with other languages like MathML, SMIL etc. LMML provides a student with an opportunity to select necessary courses. Consequently, it also provides an individual approach to studies. However, this method does not give any way for dividing educational materials based on the difficulty level, or for marking out key definitions etc.

The similar approach was used in TeachML. Each document written in TeachML is a reusable unit called a module. It contains several specific sublanguages dedicated to write different document sections like a structure, model integration, cross references, etc. This language additionally introduces markup of didactic sections in study course, which forms different sequences of educational materials supply. Nevertheless, this language does not give an opportunity for adaptive way of providing learning contents depending on the course difficulty.

The above analysis of the existing markup languages shows that any of them are neither efficient nor sufficient. The above-mentioned languages are based on the meta-description of educational materials only and they do not structure contents of the provided information for lectures, practical exercises etc. A specific markup language for study material should be designed to simplify its processing with software.

2.The problem

Analysis of the existing principles in the area of information resources development defined a range of tasks to be solved for development of the distant learning educational resource. Probably, the most important challenge is to create a markup language being enhancement of the XML specification. This language must successfully structure all necessary learning materials and it should be compatible with the existing specifications.

The resulting language specification will make it possible to organize the unified access to the educational resource. It will feature a restricted access for unauthorized users and will give a possibility for integration with the global structure of Semantic Web network.

The work is apparently to be done taking into account the best achievements and efficiently implemented features of other languages discussed above. However, it needs to include some elements specific for the given goals.

3. The solution

As the previous analysis indicated, most of the learning tools are designed using the standardized markup languages. So, the problem solution will obviously be based on the XML standard.

The eXtensible Markup Language (XML) is a multi-purpose technology for development of customized document markup languages, which are used in the contemporary manner of informational web sites development. The eXtensible Hypertext Markup Language (XHTML) designed for compatibility between HTML and XML is a direct descendant of XML. XHTML is an XML subset and it is destined to convert the existing web pages into XML which is a recently adopted unconventional standard for Internet sites design [6].

Development of the language specification is based on the Document Type Definitions (DTD). DTD is a set of rules defining instructions being transferred to a parser for document procession. Also DTD includes a set of tag and attribute definitions, entities, notation conventions and comments defining how the document should be structured.

DTD files define a set of valid tags, identify tag tree and define valid attributes for each tag, if applied for XML documents. DTD syntax is quite a specific in nature and demands additional developer's efforts. The SGML standard demands DTD for each document, and such a complexity leads to that fact that SGML is not so widely used in comparison to HTML. However, XML documents do not require DTD for a proper parsing. The XML standard defines a means for parsers to process documents even without any DTD once they meet requirements for well-formed XML documents.

The DTD for XML provide following rules:

– tag and attribute rules;
– category (macro-definition) descriptions;
– binary data formats descriptions.

All of them describe basic language units like tags, attributes, symbolic constants, external binary data files etc.

4. Education material DTD description

The DTD developed for the above-mentioned informational educational Internet resource contains a hierarchic structure consisting of three nesting levels. Those are as follows:

– *top level* defines study plan;
– *middle level* defines study course;
– *low level* defines course elements.

The top and middle levels contain meta-information about the educational material while the low level contains lecture materials, details on practical exercises, seminars, etc. as well as information concerning the control of study success.

According to this structure the design of the markup language for educational materials includes development of the «course elements» structure. The analysis of the existing researches on the educational materials structuring for the distant learning applications along with the analysis of the existing demands of the educational system defined the following items in the «course elements» structure for DTD:

<lecture> — lecture material;
<practice> — materials for practical exercises;
<seminar> — seminar materials;
<laboratory> — materials for laboratory works;
<test> — testing materials.

The next stage of DTD development is a detailed design of the structure, taking into account descendant tags and tag attributes. This will allow the software agent running on the resource to process the document contents and gain additional information about document structure.

The Table contains all tags used in the learning material structuring. Here is a short description of the introduced tags

<course> tag is the root XML tag in the document. It should appear once in the whole document. It contains all the contents of the course, including modules like lecture, practical exercises, etc.
<lecture> is used to define a lecture contents. It can appear inside the root document tag as many times as necessary. The tag contains title and body tags, actually common sub tags for the most top-level structure. The semantics reflected is the lecture materials.
<practice> tag is supposed to be used for description of practical exercises available for the course. Like <lecture>, it is a top-level tag and it contains similar common tags.
<seminar> defines contents to be provided for the seminars. The tag structure is the same as described for <lecture>.
<laboratory> tag purpose is to describe laboratory exercises; its structure is the same as described for <lecture>.
<tutorial> is a tag which can appear in the body section of any of the above-mentioned top-level tags. Its aim is to provide an end-user with a test questionnaire on the module contents. It can contain as many <question> and <answer> tags as necessary.
<test> tag defines a set of test questions. This question collection is free to be chosen by a person developing the learning material.
<task> tag defines a task for the module where it appears. It cannot contain any other tags and appears once inside each <body>.
<title> describes a title of the module where it appears. It is also a low-level tag and it should appear once inside each <body> tag.
<body> tag is the tag separating the meaningful contents in each module from other additional information. It should appear once inside each of the top-level tags within the document and can contain any of low-level tags.
<section> is the logical division within contents of the top-level module. It can appear only in <body> tag.
<headline> is an optional topic of a section.
<subhead> is an optional subtopic of a section.
<para> wraps every single paragraph in a section. It can appear as many times as necessary inside <section> tag only and does not contain any sub tags.
<question> is a low-level tag which doesn't appear or appears repeatedly in <test> or <tutorial> and defines a question.
<answer> defines an answer to the question in the <test> or <tutorial> module. It is a low-level tag and can appear as many times as necessary

TABLE

Tag name	Tag descendants	Tag Parent	Occurrence rate	Location
1	2	3	4	5
<COURSE>	Root tag contains all elements	Not available	Once per document	Wraps all tags in the course description document
<LECTURE>	<TITLE>, <BODY>	<COURSE>	Repeatedly	Inside <COURSE>
<PRACTICE>	<TITLE>, <BODY>	<COURSE>	Repeatedly	Inside <COURSE>
<SEMINAR>	<TITLE>, <BODY>	<COURSE>	Repeatedly	Inside <COURSE>
<LABORATORY>	<TITLE>, <BODY>	<COURSE>	Repeatedly	Inside <COURSE>
<TUTORIAL>	<QUESTION>, <ANSWER>	<BODY>	Once per each element <BODY>	Inside <BODY> after <SECTION> or <TEST> or <TASK>
<TEST>	<QUESTION>, <ANSWER>	<BODY>	Once per each element <BODY>	Inside <BODY>
<TASK>	Not avaliable	<BODY>	Once per each element <BODY>	Inside <BODY>

TABLE (CONTINUED)

1	2	3	4	5
<TITLE>	Not available	<LECTURE>, <PRACTICE>, <SEMINAR>, <LABORATORY>	Once per each parent element	Inside <LECTURE>, <PRACTICE>, <SEMINAR>, <LABORATORY> before <BODY>
<BODY>	<SECTION>, <TUTORIAL>, <TEST>, <TASK>	<LECTURE>, <PRACTICE>, <SEMINAR>, <LABORATORY>	Once per each parent element	Inside <LECTURE>, <PRACTICE>, <SEMINAR>, <LABORATORY> after <TITLE>
<SECTION>	<HEADLINE>, <SUBHEAD>, <PARA>	<BODY>	Repeatedly	Inside <BODY>
<HEADLINE>	Not available	<SECTION>	Repeatedly	In the beginning of <SECTION>
<SUBHEAD>	Not available	<SECTION>	Repeatedly	After <HEADLINE>, if any
<PARA>	Not available	<SECTION>	Repeatedly	After <HEADLINE>, <SUBHEAD>
<QUESTION>	Not available	<TUTORIAL>, <TEST>	Repeatedly	Inside <TUTORIAL>, <TEST>
<ANSWER>	Not available	<TUTORIAL>, <TEST>	Repeatedly	Inside <TUTORIAL>, <TEST>

The designed markup language efficiently covers user's needs for development of teaching courses with separated structure and contents of educational materials from their visual presentation displayed to the end user.

5. Conclusion

The designed language specification makes it possible to solve the problem of separation into a visual presentation and a document structure; this provides an opportunity for software to select terms, definitions and other structure units of the educational material as well as to estimate the document difficulty using attributes of certain tags. Such an approach to the educational material markup gives a possibility to solve many of the distant learning problems, including the adaptive learning problem.

References: 1. Christian Süß. Adaptive Knowledge Management: A Meta-Modeling Approach and its Binding to XML. *// In: H.-J. Klein (Ed.), 12. GI-Workshop Grundlagen von Datenbanken, Plyn, TR 2005*, Christian-Albrechts-University Kiel, Germany, 2000. 2. Clark. James Clark. Comparison of SGML and XML. *http://www.w3.org/TR/NOTE-sgml-xml-971215.* 3. IETF RFC1808 (Internet Engineering Task Force). *RFC 2141: URN Syntax*,ed. R. Moats. 1997. 4. Mathematical Markup Language 1.01 Specification, *W3C Recommendation*, revision of 7 July 1999, *http://www.w3.org/TR/REC-MathML/.* 5. TML Language Specification, *NetQuest online publication, http://www.ilrt.bristol.ac.uk/netquest/about/lang/.* 6. XML: Extensible Markup Language. E.R. Harold, *Hungry Minds, Inc.*, 1998. 7. LMML schema., Institute for Information Systems and Software Engineering University of Passau online publication, *HTTP://WWW.LMML.DE/.*

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 04.01.2004