

Reshaping e-Learning Content to Meet the Standards

Giedrius BALBIERIS, Vytautas RĖKLAITIS *

*Computer Network Dept., Faculty of Informatics, Kaunas University of Technology
Studentų 50–416, LT-3031 Kaunas, Lithuania
e-mail: balbgied@pit.ktu.lt, vytaas@pit.ktu.lt*

Received: June 2002

Abstract. There is a set of e-learning standards varying on the level of complexity, detail of description and means for technical implementations. These standards open up new levels of content description, data presentation as well as have impact on reuse and interoperability of e-learning content. E-learning standards such as IMS, SCORM, IEEE LTSC METADATA, DUBLIN CORE and others have settled down in recent years. E-learning platforms started complying to the standards especially to the most comprehensive ones such as IMS and SCORM. Here comes the issue of migrating old content into standardized learning objects (as defined in SCORM standard) along with the development of easy-to-use learning objects repositories. In this paper firstly an overview of standards is given with respect to the reshaping of content to meet standards, then a practical conversion issues are examined and discussed.

Key words: e-learning, metadata, standard, xml, rdf.

Introduction and Motivation

Similarly to general internet, e-learning has gone through a number of development steps: publishing, information digitization and now is grown-out in the amounts of data and functional complexity. Of course there is a number proprietary internal standards in e-learning platforms for operation, data storage and exchange, but as functionality and data complexity grows even those proprietary standards are not longer sufficient. For example the leading e-learning suites employ up to 100 different tools for example one of of leading e-learning platforms WebCT (2002) has got 80 tools and features.

Previous tendency to create machine-readable data is transforming into machine-understandable information and becoming a major push for global standardization as in area of general technologies as well as in e-learning specifics. A suite of technologies like XML, RDF, XSLT etc. and concepts such as Semantic Web, Network Learning Objects, Shared Content Repositories come up in order to facilitate this growth of semantic web – a next generation of internet (W3C Semantic Web, 2002; W3C RDF, 1999). Stable since beginning of this century e-learning standards provide means for creating and

*Since April 2002 he is a professor at Graduate School of Information Systems, University of Electro-Communications, Tokyo.

reusing pieces of learning via so called shared content objects and distributed content repositories as described in SCORM standard (SCORM, 2002). Nevertheless the issue of converting huge amounts of inherited e-learning content into structure of learning objects and describing those in standard compliant way is a very routine, time consuming workload. We see, that it is possible to semi-automate this process via merging the knowledge, existing technologies and descriptions of standards.

This new hype of technologies and standardization originates in the concept of metadata. As defined in by Ora Lassila in W3C 'Introduction to RDF metadata' (1997), metadata is a data about data.

Why Metadata?

Looking at how IT revolution and solutions came to every facet of our lives we may find that this penetration happened and continues to happen in following steps:

- 1) routine computing;
- 2) information digitization;
- 3) automated processing.

Continually decreasing human involvement current development goes on to the last phase – automated processing. Automated processing involves many functions – from resource matching, sorting, interconnecting different objects, extending to automatic search, launching certain processes and so on. All of these decisive tasks require certain knowledge which is achieved either via direct human-expert involvement or generated via intelligent techniques. In which ever way the decisive knowledge is generated, it must be expressed in certain systematic ways recognizable by computer. This decisive knowledge is usually expressed via use of metadata.

In e-learning metadata based standards also follow the concept that *courses will always have components for which automatic semantic recognition will be impossible or too expensive*. This statement we base on the following factors:

- 1) great variety of existing file and data formats and the ones to come-up in the future;
- 2) increasing interactivity of courses via non-text based dynamic components;
- 3) increasing data volumes as well as complexity of processing diversified data formats;
- 4) diversity of subjects decrease possibilities to achieve any format generalizations and subject based standardization.

Elaborating on this concept, we see, that metadata is wrapped around actual course objects. XML based metadata descriptor references physical files of various formats and implements structural as well as semantic description aspects.

Metadata Standardization

Electronic information is encoded into bytes and chunks of bytes – files. Usually data in files is a mixture representing few classes of information:

1. Content (semantic part).
Varyingly encoded knowledge is expressed as facts or parameters of that particular knowledge domain.
2. Syntax.
It is a format of knowledge encoding. Usually follows some technical (for example: html) or knowledge domain standard.
3. Layout, design, runtime specifics.
This extra information is related to the expression and visualization of represented knowledge. In some knowledge domains (architecture, 3d graphics etc.) it carries the most information.

Standardization as defined by Webster dictionary (Webster, 2002) provides a number of meanings, which we can group into three major parts: (1) symbolic, (2) authority over values, (3) commonalties. We can note that last two items translate into technology terms: interoperability and reuse. Quite similarly to all other areas, e-learning standards approach standardization introducing such concepts as Reusable Learning Objects, Shareable Content Objects (SCO) and SCO repositories as in SCORM standard (SCORM, 2002). These concepts extend into direct use of network resources as in Network Learning Object (NLO) as developed by NETG+ (NETg, 2002). For today there are up to thirty e-learning standards. Some of the standards (like IEEE LTSC LOM, 2002) elaborate on some specific aspect of e-learning standardization. These so-called vertical standards merge into all-covering horizontal standards. For example IEEE LTSC LOM is a sub-standard as in SCORM as well as in IMS standards. Standardization which is underway has a strong fundament of modern technologies which provides the means to achieve the reuse and interoperability goals.

The Fundament of Standards

Basic Technologic Standards

There are a number of automation techniques and standards appearing on the market. Most of them are based on general purpose XML and RDF technologies. Lets investigate these core technologies and premises of that lets review the basic features of these fundamental technologies.

XML

XML is a family of technologies for structuring data and separating all three information components: semantics, syntax and design (XML, 2001).

As a practical example lets look at the lesson1.html file in Fig. 1. We can see that in order to extract semantic information such as subtitle of the lesson (text "HTML function") we have to look for tag <H2> and extract the contents of this tag. The path for scanning to this tag in Document Object Model (DOM, 2002) and expressed in data mining agent query language (Myllymaki and Jackson, 2001) would be:

```
/html/body/h2
```

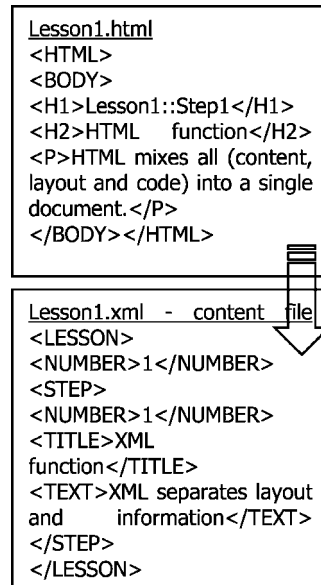


Fig. 1. Comparing HTML to XML.

As we can see this search for semantic information binds to the syntax, design and layout elements and if the designers of the course would decide to add table view, modify the layout or change some other design elements, our data extraction query would fail.

Automatic processing rules for lesson1.xml are quite different. As we can see this file contains only semantic information and a number of tags structuring the information. Thus searching for some specific data we just need to specify our query in terms taken from the tag vocabulary. By vocabulary here we name a list of all possible tags within that XML file. For example if we would look for title of the Lesson 1 Step 1 (text "XML function") we would just need to use matching XSLT expression:

```

<?xml version="1.0"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD-xsl">
  <xsl:template match="/">
    <xsl:for-each select="lesson/step">
      <xsl:value-of select="title"/>
    </xsl:for-each>
  </xsl:template>
</xsl:stylesheet>

```

As we can see from this comparison, well structured XML provides a better fundament for implementing an architecture of automation technologies: data mining, retrieval and exchange etc. From e-learning standardization point of view that provides a clear architecture for defining interfaces for automatic interconnections, data exchange and run time semantic processing of learning processes.

RDF

The Resource Description Framework (RDF) is an architecture that enables the encoding, exchange and reuse of structured metadata (Miller, 1998). RDF is an application of XML that imposes needed structural constraints to provide unambiguous methods of expressing semantics. RDF additionally provides a means for publishing both human-readable and machine-processible vocabularies designed to encourage the reuse and extension of metadata semantics among disparate information communities. From e-learning standard point of view RDF provides the means to standardize the metadata attributes (earlier mentioned vocabulary) of inherited learning resources (like html files, images, tests etc.). That way without converting them into a kind-of semantic format, external file is used to describe the semantic part.

For example the description of the author and title of the same lesson1.html (Fig. 1) in RDF framework and DUBLIN CORE metadata standard would be:

```
<? xml version="1.0" ?>
<RDF xmlns = "http://w3.org/TR/1999/PR-rdf-syntax-19990105#"
      xmlns:DC = http://purl.org/DC#
  <Description about = "lesson1.html"
    DC:Title = "Lesson 1, Step 1"
    DC:Creator = "Giedrius Balbieris"
    DC>Date = "2002-05-20"
    DC:Subject = "XML function"
  />
</RDF>
```

e-Learning Standards

There are up to thirty e-learning standards. Three major players among them are:

- DUBLIN CORE,
- SCORM,
- IMS.

From those, DUBLIN CORE is the eldest XML based standard. It is not e-learning specific standard, but it is quite often used for it's small number of basic 15 tags. SCORM standard has been developed by American department of defense and inherits the most from previous CBT AICC (AICC, 2002)

standard specifications. IMS is the standard uniting few European standardization projects. Among them are well known names such as: ARIADNE and IEEE LTSC LOM and others. IMS and SCORM standards are in stable releases for last couple of years and are used on the best practice criteria. Because of that and for compatibility reasons they share a partial sub-standards base (Fig. 2).

All of these standards are based on XML suite of technologies and are aimed at standardizing the way of describing learning resources. The standardization is achieved via structuring the concepts of e-learning area, defining different types of processes and data flows and describing this resulting structure in XML technologies.

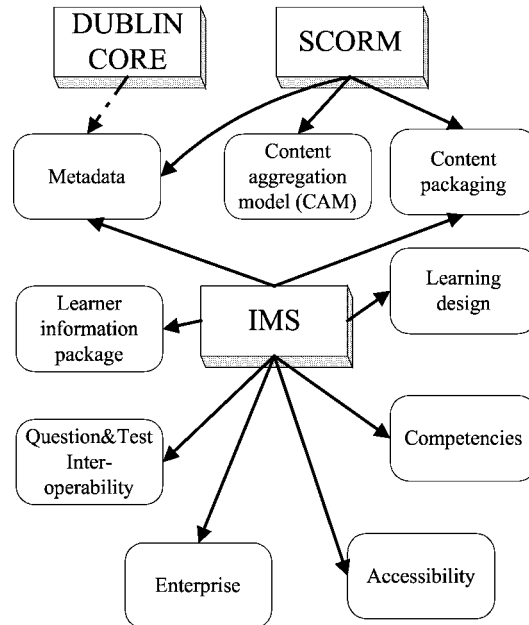


Fig. 2. Relationship between e-learning standards.

IMS

IMS – a shorthand for Instructional Management Systems. It is a set of working groups developing specifications of e-learning standards in the family of XML technologies. Project devotes a separate working group for each specific e-learning area thus totaling in eight standardization working groups:

- 1) content packaging;
- 2) metadata (merged from LOM specification, IEEE LTSC LOM, 2002);
- 3) learner information package;
- 4) question and test interoperability;
- 5) enterprise;
- 6) accessibility;
- 7) competencies;
- 8) learning design.

The learning resources metadata specifications (released in August 20, 1999) create a uniform (i.e., XML based) way for describing learning resources so that they can be easier discovered, by different metadata aware search tools. The enterprise specification (released November 3, 1999) is aimed at administrative applications and services that need to share data about learners, courses, performance, learning process etc., across different platforms, operating systems, user interfaces, and so on. Content and packaging specification provides fundament to create reusable content objects that can be used

among all standard compliant learning management systems. Question and test specification addresses the need of sharing test items and other assessment tools across different systems. Learner profiles specification covers the organization of learner information so that learning systems can be more responsive to the specific needs of each user.

SCORM

SCORM (SCORM, 2002) translates into shared content object (SCO) repository model (RM). As a e-learning initiative from American Department of Defense the standard inherits the most from Computer Based Training standard AICC (AICC, 2002). Sharing metadata and content packaging specifications with IMS, it concentrates on AICC inherited specifics which is related to user tracking and run-time specific attributes. SCORM elaborates specification of distributed network SCO repositories. For this purpose it defines:

1. Low level content aggregation model consisting of three types of objects:
 - 1.1. Assets – smallest indivisible learning items or resources,
 - 1.2. Shared content objects (SCO) – a group of assets and other learning objects,
 - 1.3. Content aggregation – run time API for SCO interoperability and learning process tracking specification.
2. Application programming interface with typical functions for data exchange between different SCO:
 - 2.1. SCO initialization function name (LMSInitialize(...)),
 - 2.2. SCO runtime operation function names (LMSGetValue(...) etc.),
 - 2.3. SCO finish and cleanup function name (LMSFinish(...)).

Migrating learning content to the SCORM basically involves similar to IMS content migration to learning objects and description of refined objects with metadata. Extra to that SCORM provides means for describing many-SCO related logic and recording it in global course aggregation.

Dublin Core

Series of workshops on metadata starting with the first in March 1995 in Dublin Ohio (Lagoze, 1996), the USA were convened to address the metadata resource description issue and propose solutions. The Dublin workshop resulted in the Dublin Core, a set of 13 metadata elements intended to describe the essential features of networked documents. The Dublin Core metadata set is meant to be both simple enough for easy use by creators and maintainers of web documents and sufficiently descriptive to assist in the discovery and location of networked resources. Currently fifteen elements of the Dublin Core include familiar descriptive data such as author, title, and subject.

Dublin Core is a generic resource description standard, but very often due to its' simplicity is used in e-learning as well. Because of being too generic it is not sufficient

for description of all e-learning resources (for example tests etc.) and must be used only as a supplement to IMS or SCORM. From the conversion point of view DUBLIN CORE has many elements which are not included in the actual data and thus possibilities for automatic description or conversion are close to none.

e-Learning Standards and Reuse

Automation and reuse – the core concepts in the current standardization processes. In e-learning they carry out not only frameworks for resource description, but also provide means for resource interchange. Implementations of learning object repositories are still in development. Early birds like Learn eXact (LCMS, 2002) already provide multi-functional services, starting with metadata search and retrieve extending into remote usability of Network Learning Objects (NLO). Although these cutting-edge solutions provide all the fancy features, they are still pretty much proprietary and platform bound. The trend towards data exchange and XML based standardization continues to unfold via introduction of new technologies and continuous improvement of standards.

Exploring the standards in Fig. 2 we see that all of the standards contain at least metadata standardization and only next expand into other areas. This is very true from historical perspective as IEEE LTSC LOM (IEEE LTSC LOM, 2002) standardization initiative was one of the first of this kind in mid nineties.

Each of the standards defines metadata a bit differently. Nevertheless, if resource is described in any of XML based standards it is very easy to create mappings into other metadata dialect as for example the converter in DC-DOT (Metadata standard converter, 2002) project converts DUBLIN CORE into 11 other formats.

Similarly to LO use in online courses, the NLO's will be used in other different type of environments starting from personal pages, magazines extending into company intranets, content management systems etc. As an example lets consider the reuse of learning resources like in Martindale's Health Science Guide (Martindale's Health Science Guide, 2002), a resource center listing 60,000 teaching files and 129,000 medical cases. Such a resource, if standardized and made available to medical schools around the world, would greatly facilitate the creation of courses in medicine. Often, such databases much like online journalism systems will work as closed enterprise environments limited only to registered users.

Migrating Content to Standards

Standard Course Package

The diagram in Fig. 3 represents the migrated course with course content package on the right and metadata descriptor on the right. Course content package as defined in by IMS Content packaging specification is any kind of three most popular archives: zip, jar or tar.gz. Files in the package are described and organized into a set of SCO and one NLO. As we can see on the right of the figure some SCO may contain other SCO or just

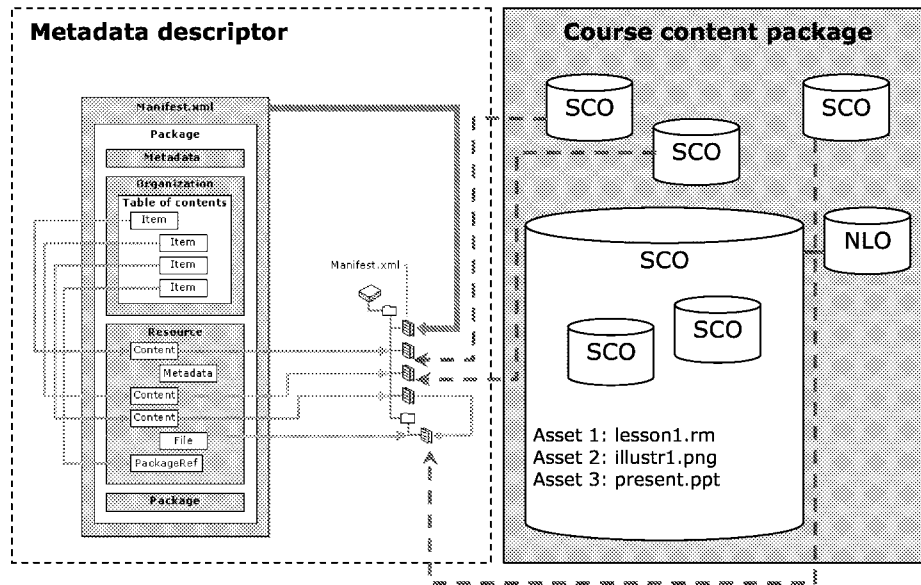
Standard course package

Fig. 3. Standardized course package.

physical files so called assets. All structure and dependency is bound in manifest.xml file which contains three sections:

1. **Metadata.**
Describes content aggregation, facilitate its' reuse and discoverability within content repository. Covers various aspects of learning process starting with costs of the unit, who is creator, prerequisites extending into all course specifics and administrative areas.
2. **Organization.**
Describes internal structure of learning objects. Can be compared to a mind map or navigation map.
3. **Resources.**
Describes SCO, external NLO and physical files. It may also contain references to sub-manifests of stand-alone SCO as well as items covered by other IMS standardization working groups such as QTI and others.

Possibilities of Automated Migration to e-Learning Standards

Provided we have a course without description in standards, there are following possibilities for automation of migration to standards:

1. **Automatic learning object identification (Fig. 4).** The concept of structured learning material has always been among the best principles and recommendations for preparing learning materials, especially in e-learning. Although not very clearly separated, the structured independent learning objects can be identified in content

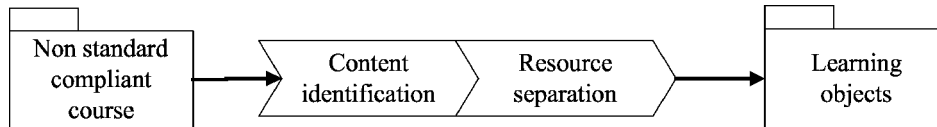


Fig. 4. Migration model. Phase 1: learning object identification and separation.

structure page existing. This course contents page exists in most learning management systems and is pretty similar in all of them. After learning objects identification process (Process 1 'Content identification' in Fig. 4), physical resources can be separated and packaged in the second process. Resource separation is achieved according to dependency of resources to identified Shared Content Objects. The resource separation need not to be executed on physical file structure level (there may be some hidden relative file dependencies), but is recorded into the resulting aggregation and learning object metadata descriptor files.

2. Recognition and description of the resources via adaptive data mining solutions (as defined by Myllymaki and Jackson, 2001). This kind of approach produces highly accurate description of learning resources, but requires participation of expert at least in production of customized discovery rules (Step 2 in Fig. 5) for each set of templates for campus courses.
3. Automatic description with keywords (Step 4 in Fig. 4). After identification of learning objects metadata description can be automated via using advanced searching techniques (such as multi-search engine clustering (Hannappel *et al.*, 2001)) coupled with the knowledge of the standards. These advanced search techniques involve not only word counting, but also language dependent heuristic analysis tightly coupled with specifics of the particular standard and course design templates.

Indeed standardization is beneficial. It provides not only reusability and interoperability, – it extends current LMS with new features implemented in standards. These features implement more interactivity and include more logic components inside. Identification of such logic components as in Fig. 6, requires course redesign and can not be automated. An example of such logic component would be quiz performance or acquired knowledge based learning path designs. Nevertheless the techniques described above still can help much at least in suggestive ways.

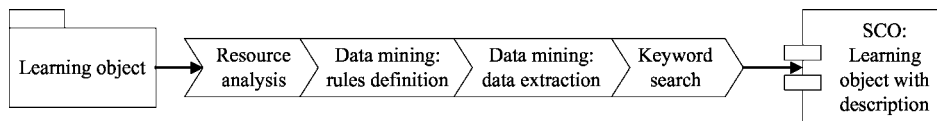


Fig. 5. Migration model. Phase 2: learning object semi-automatic description process.

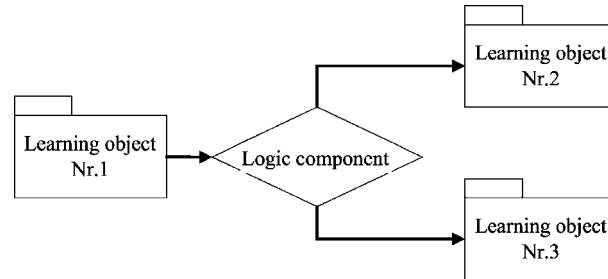


Fig. 6. Standard based logic components.

Summary

In this paper we have reviewed current trends of reuse and automation in e-learning. Reviewing of current e-learning standards we discovered that they are:

- 1) based on metadata;
- 2) standardize vocabularies covering various e-learning aspects;
- 3) all of them are dialects of XML and because of that are convertible among themselves.

We examined different possibilities for automating the migration process and found that migration of existing e-learning content to standardized semantic web networks will involve a synergetic effort comprising of:

- 1) physical files, file formats and means to operate on them: open, access information, describe, change;
- 2) XML suite of technologies and practical use in describing learning resources;
- 3) intelligent data mining and semantic analysis science and technologies;
- 4) knowledge description standardization (like knowledge trees and ontologies);
- 5) IMS, SCORM, DUBLIN CORE and other metadata based standards.

Overall we see that there are certain possibilities to achieve semi-automated migration to standards although this is not a panacea and expert involvement is still needed (at least for selecting rules proposed by the system).

References

- AICC (2002). Aviation Industry CBT Committee Standard Specifications, <http://www.aicc.org>
- DOM (2000). Document Object Model (DOM) Level 2 Core Specification, <http://www.w3.org/TR/DOM-Level-2-Core/>
- Hannappel, P., R. Klapsing, G. Neumann (2001). MSEEK – A Multi Search Engine with Multiple Clustering, <http://citeseer.nj.nec.com/362175.html>
- IEEE LTSC LOM (2002). Draft Standard for Learning Object Metadata, <http://ltsc.ieee.org/doc/wgl2/scheme.html>
- IMS (2002). Instructional Management Systems Standardization Initiative, <http://www.imsproject.org>
- Lagoze, C. (1996). The Warwick Framework: A Container Architecture for Diverse Sets of Metadata, <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>

- Lassila, O. (1997). Introduction to RDF Metadata,
<http://www.w3.org/TR/NOTE-rdf-simple-intro-971113.html>
- LCMS (2002). Learning Content Management System, <http://www.learnexact.com/>
- Martindale99.s Health Science Guide (2002). The "Virtual" Medical Center. Web Site,
<http://www-sci.lib.uci.edu/HSG/Medical.html>
- Metadata (2000). Intro into Metadata, http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.html
- Metadata standard converter (2002). <http://www.ukoln.ac.uk/cgi-bin/dcdot.pl>
- Miller, E. (1998). An Introduction to the Resource Description Framework,
<http://www.dlib.org/dlib/may98/miller/05miller.html>
- Myllymaki, J., J. Jackson (2001). Web-based data mining,
<http://www-106.ibm.com/developerworks/web/library/wa-wbwm/?dwzone=web>
- NETg (2002). Network Learning Object Concept,
<http://www.netg.com/CorporateSolutions/LearningTechnology/Tools.asp>
- Qu, C., W. Nejdil (2002). Towards Interoperability and Reusability of Learning Resource: a SCORM-Conformant Courseware for Computer Science Education, <http://www.kbs.uni-hannover.de/change-tao/icalt24.pdf>
- SCORM (2002). Shared Content Repository Model, <http://www.adlnet.org>
- W3C Semantic Web (2002). <http://www.w3.org/2001/sw/>
- W3C RDF (1999). Resource Description Framework (RDF) Syntax,
<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>
- WebCT Inc. (2002). WebCT – a course management system. <http://www.webct.com>
- Webster (2002). Meaning of word 'standard' in Webster dictionary,
<http://www.m-w.com/cgi-bin/dictionary?book=Dictionary&va=standard>
- XML (2001). XML in 10 points, <http://www.w3.org/XML/1999/XML-in-10-points.html>

G. Balbieris, born in 1974. In 1992 graduated Biržai 2nd Secondary school as well as School of Young Programmers. In 1996 graduated Kaunas University of Technology (KUT) with bachelor diploma in informatics. In year 1998 graduated KUT with master degree in informatics. He continues his studies for Ph.D. in the IT sciences. Last 7 years he is working in e-learning area.

V. Rėklaitis has been involved in networking development at Kaunas University of Technology over the last 12 years. He headed joint projects funded by EC under TEMPUS and INCO-Copernicus programmes related to both teaching of networking and developing of e-learning applications. Since April 2002 he is a professor at Graduate School of Information Systems, University of Electro-Communications, Tokyo.

E-mokymosi turinio standartizacija

Giedrius BALBIERIS, Vytautas RėKLAITIS

Jau yra nusistovėję keletas e-mokymosi standartų. Jie skiriasi galimybėmis, aprašymu bei techninėmis realizacijomis. Standartų šerdis – XML technologijų grupė. Standartizuojamos XML ontologijos ir taip pasiekiamos naujos galimybės: nuoseklus turinio aprašymas, vaizdavimo šablonai. Taip pat išlošiama pasiekiant aukštesnę integraciją, bei mokomosios medžiagos daugkartinį panaudojimą (angl. reuse). Per pastaruosius keletą metų nusistovėjo šie stambiausieji e-mokymosi standartai: IMS, SCORM, IEEE LTSC METADATA, DUBLIN CORE. Išsamiausiai ir labiausiai išbaigtais, tokiais kaip IMS, SCORM, ėmė sekti ir e-mokymosi programų sistemos. Iki šiol sukurta daugybė e-mokymosi kursų bei įvairios mokomosios medžiagos. Autoriai iškelia šią problemą ir analizuoja įvairias esamas technologijas bei jų taikymą siekiant dalinai automatizuoti šių jau egzistuojančių mokomųjų objektų standartizacijos procesą.