Using Metadata For Content Indexing Within An OER Network

Ranka Stanković, Olivera Kitanović, Ivan Obradović, Roberto Linzalone, Giovanni Schiuma, Daniela Carlucci

The Fifth International Conference on e-Learning (eLearning-2014), 22-23 September 2014, Belgrade, Serbia

USING METADATA FOR CONTENT INDEXING WITHIN AN OER NETWORK

RANKA STANKOVIĆ, OLIVERA KITANOVIĆ, IVAN OBRADOVIĆ
University of Belgrade, Faculty of Mining and Geology,
ranka.stankovic@rgf.bg.ac.rs, olivera.kitanovic@rgf.bg.ac.rs, ivan.obradovic@rgf.bg.ac.rs

ROBERTO LINZALONE, GIOVANNI SCHIUMA, DANIELA CARLUCCI
Università degli Studi della Basilicata,
roberto.linzalone@unibas.it, giovanni.schiuma@unibas.it, daniela.carlucci@unibas.it

Abstract:*This paper outlines the ICT solution for a metadata portal indexing open educational resources within a network of institutions.The network is aimed at blending academic and entrepreneurial knowledge,by enabling higher education institutions to publish various academic learning resources e.g. video lectures, course planning materials, or thematic content, whereasenterprises can present different forms of expert knowledge, such as case studies, expert presentations on specific topics, demonstrations of software implementation in practice and the like. As these resources need to bediscoverable, accessible and shared by potential learners across the learning environment, it is very important that they are well described and tagged in a standard way in machine readable form by metadata. Only then can they be successfullyused and reused, especially when a large amount of these resources is reached, which makes it hard for the user to locate efficiently those of interest.The metadata set adopted in ourapproachrelies on two standards: Dublin Core and Learning Object Metadata. The aim of metadata and the corresponding metadata portal described in this paper is to provide structured access to information on open educational resources within the network.*

Keywords:*OER, Open educational resources, metadata, TEL, Technology enhanced learning*

1. INTRODUCTION

Due to intense technological development there is a growing need for reinforcing knowledge exchange between academia and industry. At the same time, the Open Educational Resources (OER) movement, aimed at providing teaching, research and learning materials under an open source licence that permits their free use, access, repurposing, reuse and redistribution by others with limited or no restrictions is rapidly gaining popularity [1].Based on these two facts, the BAEKTEL (Blending Academic and Entrepreneurial Knowledge in Technology Enhanced Learning, http://www.baektel.eu) project was initiated with the main goal of building an OER networkofferingeducational materials by higher education (HE) institutions and best practice examples by enterprise experts. The network is conceived as multilingual, which means that resources can be published in different original languages, with adequate support offered for their translation[2].

The conceptual model of the ICT solution for BAEKTEL OER framework envisages a network of nodes offering OER content and a central repository,the BAEKTEL Metadata Portal (BMP), where metadata,providing all important information on the network resources will be stored, thus enabling their centralized search and browse. The initial network consists of six nodes located at different Western Balkans (WB) universities participating in this project, with one of them hosting the BMP.

By means of metadata, or “data that describe other data” within the central BAEKTEL repository, resourceswithin the network become well described and tagged in a standard way in machine readable form.OER metadata include information such asresource title, author, subject, creation date and the like, which facilitates search, but also acquisition, use and reuse of learning objects.

In defining metadata for BAEKTEL resources existing standardization efforts has been taken into consideration. Namely,the IMS Global Learning Consortium (GLC) promotes standardization of learning object metadata vocabularies and federated search processes consistent with several different standards[3]. These standards include Dublin Core, IEEE Learning Object Metadata(LOM) and the Learning Resource Metadata Initiative (LRMI), an extension of schema.org, launched in 2011, as a joint initiative of Google, Yahoo, Microsoft Bing, Yandex and W3C. Schema.org provides a collection of schemas for HTML pages markup in ways recognized by major search providers and used for structured data interoperability [3].The main rationale for the approach fostered by IMS GLC is that the choice of a standardized learning object metadata vocabulary has valuable and beneficialinstitutional and pedagogical implications.

The focus of this paper is on metadata and their management in the context of BAEKTEL OER framework, which is described in more detail in section 2. Section 3 of this paper outlines the key aspects of metadata management, including standards for describing educational resources and the approach to BAEKTEL metadata definition. In section 4a review of the main open source Digital Asset Management (DAM)systems for

metadata managementis given. The model and description of the proposed metadata set are described in Section 5, while section 6 is dedicated to its use case model, followed by conclusions in Section 7.

2. THE BAEKTEL OER FRAMEWORK

The basic structure of the BAEKTEL OER framework isillustrated by the deployment model in Figure 1.The initial framework consist of a network of six nodes at WB universities, namely University of Belgrade (UB), University of Kragujevac (UNIKG), University of Niš (UNI) from Serbia, University of Banja Luka (UBL) and University of Tuzla (UNTZ) from Bosnia and Herzegovina and University Mediterranean (UNIM) from Montenegro. All of them develop and publish their OER independently using edX, an open-source online learning platform offered by a massive open online course (MOOC) non-profitprovider (https://www.edx.org/about-us). In addition to its own OER in edX,UB hostsBMP, the central repository with metadata for all published OER within BAEKTEL network.

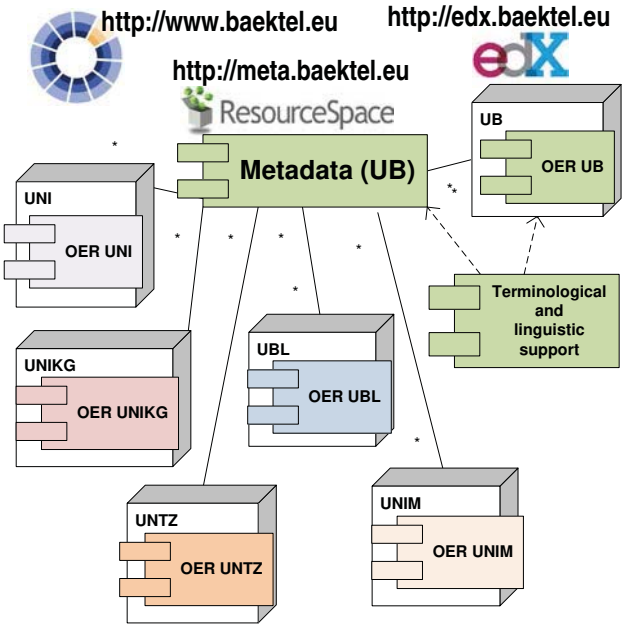

Figure 1: The BAEKTEL framework

BMP features a web application for management, browse and search of metadata, butalso web services for terminological and linguistic support. Since OER content within the network can be published in different languages, the web application and web services support the network multilinguality, but also offer various features related to query expansion, information retrieval, OER indexing, and the like.

The basic aim of the ICT solution for BAEKTEL OER framework is to support a distributed OER system. The framework is not limited to the current six nodes, but allows effortless expansion.More nodes at other institutions, academic or entrepreneurial equally, which might join the BAEKTEL network in the future, can be easily integrated in the current network. In addition, the framework can integrate OER that are not created by institutions within the network, the only condition being that those resources are registered and described in the central metadata repository.

3. METADATA MANAGEMENT

The Schema.org initiative was the result of the exponential growth of data on the web and large intranets, which made the location ofweb pages containing data of interest more and more difficult. A solution to this problem was found in introducing metadata, withthe goal to improve the display of search results, thus making it easier for users to find the right web pages. To that end, content publishers insert machine readable information into the code of web pages, which helps search engines interpret the sense of the text on those pages. One example of such tagging for a page containing a research paper is:

```
<div itemscope itemtype="http://schema.org/ScholarlyArticle">
<h1 itemprop="name">Raster georeferencing</h1>
<p itemprop="author" itemscope itemtype="http://schema.org/Person">
<span itemprop="name">Ranka Stanković</span>,
<span itemprop="affiliation">University of Belgrade–Faculty of Mining
and Geology</span></p>
</div>
```

One of the main tasks within the development of the ICT solution for the BAEKTEL metadata portal was to define anappropriate metadata schema, drawing its data elements from one or more namespaces, that is, containers for sets of identifiers. Namely, the BMP schema contains elements taken from standard namespaces with guidelines for metadata creation. In metadata specification, the standard nomenclature was used, enabling learning resources to be described and shared in a common way, and thusenhancingtheir accessibility from other OER portals.

The need for metadata

OER need to be shared, accessible and discoverable by potential users across the learning environment. They should be annotated in such a way that the users can understand what specific learning objectsare about, what is their learning content and prerequisites for their use,without even seeing them [5].

When a large amount of OER is reached, it is even more important that they are well described and tagged in a standard way in machine readable form. In that case, results returned by search engines are more relevant, and both educators and learners can find and compare learning materials that best suit their current needs more efficiently.

OER or the related metadata standards are often stored in the so called Learning Object Repositories (LOR). Different LORs address different needs and therefore have different metadata schemas. Chan & Zeng emphasize thatmuch effort has to be devoted to achieving or improving interoperability among metadata records in order to enable federated searches and facilitate metadata management [6].

The following subsection describes some of the common standards used in educational settings.

Standards for describing educational resources

Koutsomitropoulos et.al [7] point out that, although generic metadata specifications, such as the Dublin Core(DC) [8,9], seem to fulfil the need for documenting web-distributed objects, educational resources demand a more specialized treatment and characterization. They propose a mapping of the IEEE 1484.12.1-2002 LOM Standard elements to DC, as a basis for delivering web services for educational resources. Namely, LOM Standard,provided by the Institute of Electrical and Electronics Engineers, is the leading educational metadata specification[10].The standard groups data elements to describe a learning resource into the following nine categories: general, lifecycle, meta-metadata, technical, educational, rights, relation, annotation and classification. LOM Standard has more than 70 possible elements, andFriesen points out that it is widely used in educational context and applied in several learning object repositories [11].

Along the same lines, the Education Working Group of the Dublin Core Metadata Initiative developed DC terms to describe educational resources. They also proposed a number of LOM elements to be added to enhance the DC record. Several metadata initiatives follow the recommendations provided by the DC Education Working Group.

The International Standards Organization (ISO) sub-committee on Information Technology for Learning, Education and Training (ISO/IEC JTC1 SC36) is also involved in metadata standards for learning resources. They focus on existing standards and technical reports and conducted a survey on the use of LOM. The group published a first draft of a standard towards the end of 2005, but an ISO International Standard for metadata has not yet been released.

As mentioned before, metadata help users find relevant resources and enable them to make informed decisions as to whether or not a particular resource is relevant to their purposes. When metadata are shared with external portals, the visibility of the resources is additionally increased. One such important federated search project is the Global Learning Objects Brokered Exchange (GLOBE) [12], an international consortium that strives towards making shared online learning resources available to educators and students around the world.

The approach to BMP metadata

Focusing on DC and LOM Standard an analysis was performed of the strengths and weaknesses in order to select metadata that will best improve the search and browse functions of the BMP.

LOM, asthe leading, widely used, educational metadata specification is recommended by the Sharable Content Object Reference Model (SCORM), a collection of standards and specifications for web-based e-learning. The major drawback of LOM is that it has too many elements, which makes itoverly complicated. In practice, most communities use just a few elements from the LOM schema. Consequently,the resulting metadata schemasdo not differ muchfrom simpler standards such as DC.

On the other side, DC is compact, well explained and widely used, but it lacks elements for a comprehensive description of learning resources.

The approach to defining metadata within BAEKTEL draws fromtheFAO Learning Object Resources Metadata Application Profile (FAO),which combines DC and LOM Standard [13].In development of the BMP model, compliance withthese standards was obligatory,asthe BMP metadatahadto providefor sharing with otherOER repositories. At the same time, the number of mandatory elements had to be carefully selected,thus preventing metadata from becomingthe bottleneck of the whole system.

4. METADATA PLATFORM

The metadata platform for the BAEKTEL Metadata Portal was selected bearing in mind that the main goal of BMP is to providefacilities to learners for metadata search and direct access to learning resources, such as courses, training materials, guidelines, case studies, best practices and the like,on any media that supports educational material, as well asOER metadata management facilities to OER creators.

As the approach to BMP development was based on adaptation of one of the existing open source software solutions, a review of the main open source Digital Asset Management(DAM) systemswas performed, which set aside three possible platforms: NotreDAM, ResourceSpace and DSpace.

NotreDAM (http://notredam.org/) has an impressive set of features (http://notredam.org/overview/) and looks very promising, but it is still in the development stage, with modest documentation, which is mostly unfinished, and versions operating under specific operating systems. Documentation for the latest version of NotreDAM package is currently under construction, but the instructions for previous versions are also incomplete. It is working properly on Ubuntu 10.04, but for Ubuntu 12.04 it works only partially. The system is developed in Python.

DSpace(http://www.dspace.org/) is a full featured, open-source solution for storing, indexing and retrieving digital resources. It is highly configurable and can support any metadata schema. DSpace is academically oriented to a great extent, with numerous features, but hard to master by ‘ordinary’ users. It has an unpleasant user interface and the overall user experience is poor.

ResourceSpace(http://www.resourcespace.org/) is an open source DAM system released under a BSD-style license. It requires PHP, MySQL, and the GD Graphics

Library, and works with most web server software and any operating system. Some of its main features are:

- Intelligent search ordering by scoring resources against keywordson basis of user search activity
- Preselected groups of resources
- Resource access level permissions by user group
- Multilingual,allowingthe user to change the languagewith most major languages supported
- Automatic thumbnail creation for resources
- Minimal hosting requirements

After comparing and analysing the three DAM systems, ResourceSpace has been selected as the most suitable platform for the BAEKTEL metadata portal.

5. THE PROPOSED METADATASET

The cross-comparison of themetadata requirements for BMP and analysis of existing standards resulted in the metadata set based on DC with some elements taken from LOM, which describe theresources in a way that facilitates exchange with other OER systems. Figure 2 provides an overview of the elements included in BMP.

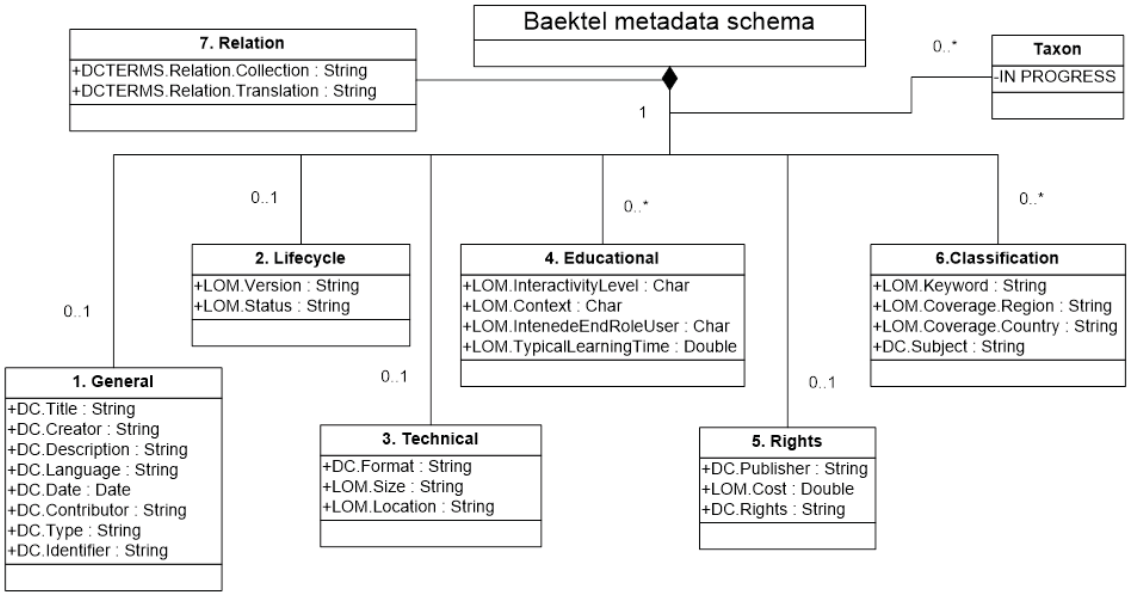

Figure2: Metadata model for BAEKTEL portal

General data are taken from the DC standard. They are: title, creator, description, language of the content of the resource, date when the resource was made available to the public, contributor and type of resource. Also, for each of the resources an identifier is created, as a unique code that provides unambiguous access to the resource.

Title is a name given to the resource. Creator could be a person, group of people or organizations responsible for producing the content of the resource. Description is the abstract, a concise description of resources.Contributor is a person, organization, or service responsible for making contributions to the resource.

Type identifies the nature of the content of the resource, such as"best practice", "case study", "exercise", "guidelines", "lesson", "module", "monitoring"and "evaluation techniques", "policy brief", "portal", "promotional material", or "reference material".

The Lifecycle category describes the history and current state of a learning object. Lifecycle fields, version and status are taken from the LOM Standard. Version indicates the edition of a learning object. Status indicates whether the resource development is completed and ready for publication, for example,"pending submission", "pending review", "active", "waiting to be archived", "archived", or "deleted".

Technical data are format, size of the digital resource in bytes and location (web address).Format is the layout of the resource in terms of how the information contained in the resource is organized. It indicates whetherit is an electronic document, paper only document, slide(s), website, cd-rom/dvd, audio, or video.

Educational data, taken from the LOM standard, suggest the auditorium the resource is intended for, the environment for learning, estimated duration of the course and degree of interactivity.

Interactivity level indicates the degree to which the learner can influence the aspect or behaviour of the resource. Value for this field can be "very low", fora document intended for printing; "low",a video clip with play and pause controls; "medium",a hypertext; "high", a lesson with multiple-choice exercises providing feedback; "very high", a virtual 3-D environment that enables exploring.

Context is the principle environment within which the learning process, that is, the use of the learning object is intended to take place. By selecting the audience and level for the material("school", "higher education", "training", "other"), users conducting searches will be able to narrow in on the appropriate resources.

Intended end user role represents the principal user for whom the resource was designed ("learner", "teacher", "manager", "supervisor").

Typical learning time is the approximate or typical time it takes to work with or through this learning object.

The attributesin the Rightscategory are publisher, rights and cost.Publisher is the individual, group, or organization named in the document as being responsible for that document's publication, distribution, issuing, or release. Rights includes information about various property rights associated with the resource, including intellectual property rights (e.g. creative commons license). Cost indicates whether use of this learning object requires payment.

Classification allows for systematic arrangement and browsing of resources, by grouping them into classes, according to common characteristics.Classification categorycontainsattributes: subject, keywords and coverage. Subject is the topic of the resource, while keywords are used in indexing and information retrieval. Coverage is the spatial characteristics of the intellectual content of the resource,a region and/or countryindicating the jurisdiction under which the resource is relevant.

6. BMP USE CASE MODEL

The BAEKTEL platform makes OER materials freely available to anyone, anytime via the internet. At that, OER learners from universities and enterprises are able to watch the lectures at their own pace in order to better prepare themselves for class or work activities. BAEKTEL accessibility services are supposed to support formal learning description methods, as well as methods for describing cognitive student and teacher workload. Means for easy integration of learning content from different sources have alsobeen provided[14].

In order to ensure the abovementioned functionalities, BAEKTELframework implementsthree user profiles or roles:resourcecreators,course participantsandsystem administrators.

System administrators (Figure3) manage and maintain theBAEKTELMetadata Portal and OER platforms. Administrator managesuser accounts, opensnew accounts for teachers and assignsappropriateprivileges to users.Modification ofthe initialsetof metadata is alsoperformedby the system administrator.

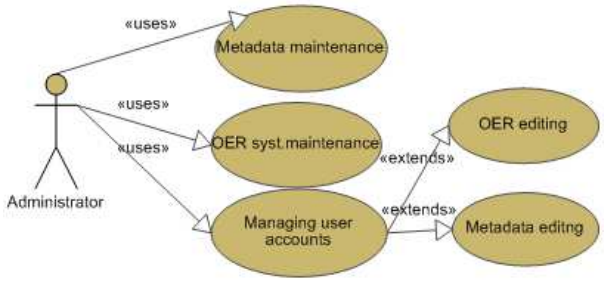

Figure3: Administrator as the user of BAEKTEL

The resource creator (Figure 4)has to own an account with privileges for teacher profile. Sincethe framework iscomposed of differenttypes of software and several repositories, it was desirable to providesingle sign-on (SSO). Single sign-on is a feature of access control of multiple related, but independent software systems. It allows the user to log in once and gain access to all systems within the network without being prompted to log in again at each of them[15].

After setting up a new OER, resource creators are required to fill metadata. They can use the offered terms or add new ones using the custom terminological web application.In the terminological dictionary, a definition is givenfor each term, with its synonyms and translation in English, Russian and other languages[16].If the resource is HTML based, an additional possibility is to link key terms in the text with dictionary entries via web services, thus providing the learner with additional explanations and translations to other languages. Furthermore, textual resources can be tagged, annotated and classifiedusing the bag-of-wordsapproach [17].

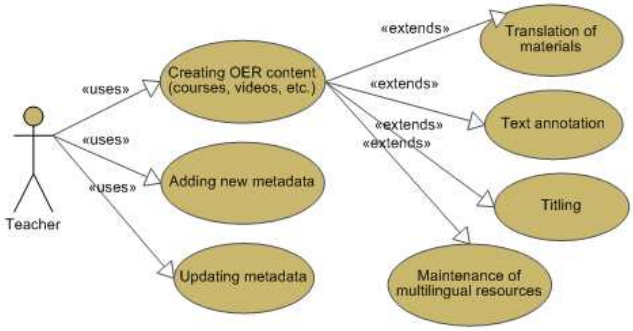

Figure4: Teacher as the user of BAEKTEL

Metadata search and browse is publicly available without log-in, but for accessing OER contentuser registration is required. The learners(Figure 5)will also have the SSO possibility,namely to log-in once and follow all the courses that are offered, regardless of particular physicalOER location.

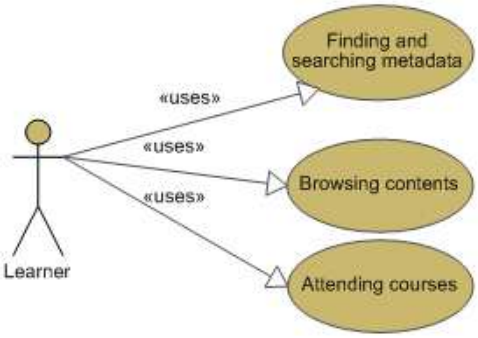

Figure5: Learner as the user of BAEKTEL

7. CONCLUSION

The ICT solution for BAEKTEL metadata portal outlined in this paper enables efficient search and browse of OER content and provides the infrastructure for successful blending of two major sources of engineering knowledge:

the academia and the enterprise. It thus contributes to the important task of preparing university students for their future jobs, but also enables them to improve their academic knowledge after graduating, by offering them a live-long learning opportunity.

Given the vast variety of content and the expected growth of the number of resources, as well as different profiles of potential users, indexing of resources that enables their efficient location within the network became a critical issue. To that end a metadata vocabulary and data structure syntax based on DC and LOMwere implemented withinResourceSpace,to offer a flexible and robust mechanism for indexing OER content and enabling the user to easily locate the resources of interest.

However, a lot of work still needs to be done before BAEKTEL enters full exploitation to the benefit of future and current university students, as well as university graduates working in enterprises. Namely, the populationof the network with resources is now crucial for bringing the BAEKTEL into full function, thus providing usability to the features outlined in this paper.

LITERATURE

[1] Atkins, D.E., Brown, J.S. and Hammond, A.L. *A Review of the Open Educational Resources (OER) Movement: Achievements, Challenges, and New Opportunities*. 2007. http://www.hewlett.org/uploads/files/ReviewoftheOERMovement.pdf.

[2] Obradović, I, Stanković, R., *Using technology for knowledge transfer between academia and enterprises* Proceedings of IFKAD, Matera, Italy, pp. 788-801, 2014.

[3] IMS Global Learning Consortium, *Application andPolicy Brief: Efficient and Descriptive Learning Object Metadata: An Essential Component of K12 Instructional Reform, Version 1.0* http://www.imsglobal.org/articles/APB2-022013.pdf 2013.

[4] Barker, P and Campbell, L.M. *Learning Resource Metadata Initiative: using schema.org to describe open educational resources*. In Proceedings of OpenCourseWare Consortium Global 2014: Open Education for a Multicultural World. http://conference.oeconsortium.org/2014/wp-content/uploads/2014/02/Paper_34-LMRI1.pdf, 2014.

[5] Velichová, D., *RLO in e-Learning Solutions.* In Proceedings of 9th International Conference Virtual University Slovakia, http://virtuni.eas.sk/rocnik/2008/pdf/aid000262.pdf, 2008.

[6] Chan, L.M., Zeng, M.L., *Metadata Interoperability and Standardization – A Study of Methodology Part I, Achieving Interoperability at the Schema Level*. D-Lib Magazine, Vol. 12(6), 2006. http://www.dlib.org/dlib/june06/chan/06chan.html

[7] Koutsomitropoulos, D.A., Alexopoulos, A.D. Solomou, G.D. and Papatheodorou, T.S., *The Use of Metadata for Educational Resources.* In Digital Repositories: Practices and Perspectives, D-Lib Magazine, Volume 16, Number 1/2, 2010.

[8] The Dublin Core Metadata Initiative (DCMI) http://dublincore.org/

[9] ISO 15836:2009 Information and documentation - The Dublin Core metadata element set.

[10] IEEE Learning Technology Standards Committee (LTSC), 1484.12.1-2002 IEEE Standard for Learning Object Metadata.

[11] Friesen, N., *Technical Evaluation Report, 40. The International Learning Object Metadata Survey*, CanCore Initiative, 2004. http://www.irrodl.org/index.php/irrodl/article/view/195/277

[12] GLOBE (Global Learning Objects Brokering Exchange) http://www.globe-info.org/ Retrieved 2014-08-20.

[13] *FAO Learning Object Resources Metadata Application Profile*, ftp://ftp.fao.org/gi/gil/gilws/aims/metadata/docs/learnap.pdf

[14] Obradović, I., Stanković, R. Prodanović J., Kitanović, O., A *Tel Platform Blending Academic and Entrepreneurial Knowledge. In* Proceedings of The Fourth International Conference on e-Learning (eLearning-2013), Belgrade, Serbia, 2013

[15] *SSO and LDAP Authentication, Authenticationworld.com.*Retrieved 2014-08-20. http://www.authenticationworld.com/Single-Sign-On-Authentication/SSOandLDAP.html

[16] Stanković, R., Obradović, I., Kitanović, O., Kolonja, Lj. *Building Terminological Resources inan E-Learning Environment*. *In* Proceedings of The Third International Conference on e-Learning (eLearning-2012), Belgrade, Serbia, 2012.

[17] Bag-of-words representation of text, https://inst.eecs.berkeley.edu/~ee127a/book/login/exa_bag_of_words_rep.html