Technical congress proceedings as a reusable digital objects educational source

Manuel Blázquez, Manuel Castro
Electrical and Computer Engineering Department – DIEEC
Universidad Nacional de Educación a Distancia – UNED
Madrid, Spain
manuel.blazquez.merino@gmail.com ,
mcastro@ieec.uned.es

Edmundo Tovar, Fernando Pescador, Jesús Arriaga
Universidad Politécnica de Madrid, UPM
Madrid, Spain
etovar@fi.upm.es , fernando.pescador@upm.es ,
jesus.arriaga@upm.es

Abstract — **The use of technical articles in scientific and academic scopes is a well-known way to extend new creations and researching jobs throughout the world, mainly as a communication medium of knowledge among educational professionals. Nowadays, with no doubt, the Internet, is the usual way to spread any kind of documentation. Authors send the contents of their publications, based on a typical scheme composed by an abstract for a first understanding of the paper, some names, references and a further content development, using textual descriptions and graphical support elements. These parts of the document are conformed in a unique data unit. However, even the article taking part of the congress edition, the access way to it, is made through the digital media given by the congress. So, only if some congress query media are developed, it's successly possible to access to a certain article information.**

The experience shown in the content of this paper gives a solution to this access problem and bases its development on the information treatment of a congress as digital learning objects. By this way, a technical paper is treated as an information unit which is composed by a data background, essentially the metadata, which are recovered and isolated in every individual documentation.

Once a digital object and its metadata are disposed, congress edition conforms a structured data block, accessible not only in the congress scope but outside it. The use of metadata standards, as IEEE-LOM, let to information being built in digital platforms and repositories to be treated as congress information, as an individual article and even, some of its parts, as reusable digital objects

Keywords- ***learning object; reusable content; metadata; technical article***

I. Introduction

This paper introduces the procedures based on the treatment, classification and organization of the information extracted from technical articles generated in several congresses in the field of Electronics [6]. The project scope has as a primal objective, the extraction of a set of metadata [14], which define every publication in the context of a congress issue, to obtain the information related to the technical and teaching information from the publication content [5], in order to be treated as a digital object by means of its metadata.

The metadata obtained will lead to a set of references associated to every article, whose implementation let the converted digital objects to be treated as reusable learning objects [9]. Subsequently, a data-metadata set will be configured in a files structure, which permits the implementation of some actions as the creation of an homogeneous web site to access them, the dissemination of the structured contents on the Internet through repositories and digital contents platforms [16], and as a final work, to analyze the contents of the information in the congresses to study the resultant information in a tendency studies and organizations relations in Electronics and Teaching fields. To carry out the project actions, the TAEE conference (in Spanish, Applied Technologies for the Electronics Teaching) has been selected because of its increasing track as well as the variety of themes in Electronics scope.

II. The Treatment of Digital Objects

Through the celebration of TAEE conferences since 1994 to 2008, 964 documents have been published. All of them are susceptible to be converted in learning digital objects. It has been necessary to apply individualization actions in every publication in order to maintain a homogeneous appearance, given the different formats of creation and publication, as well as classification methodology to keep the papers in a common context.

Every document has been converted in electronic format and has been deeply analyzed to extract from it every component to be useful in a reusable learning context. These actions have taken in place 4441 additional component digital objects. All of them conform an heterogeneous set, i.e., their types as MIME digital objects has been classified in texts, graphics, data spreadsheets, web pages, photografies, etc. [12]

In a parallel way to the formation of every digital object, main digital objects and component digital objects, it has been performed a file structure to keep them in. The structure is based in the biannual organizations of TAEE meetings. To be placed in the structure, every learning object has been named by means of an ad-hoc unified code system to let every digital object be associated with every metadata file in an homonym way. This code system has been based in a hierarchical system as Decimal System (DS) [17] but modifying the kind of characters given to every name in order to recognize the object belonging to a determined congress, session and order of paper. It can be said that the names codification also offers the first object metadata by itself. When finishing this stage, a files structure is disposed, inside which every digital object and every metadata definition file are placed. [13]

III. DIGITAL OBJECTS METADATA EXTRACTION

To implement the digital objects metadata extraction process it has been required the achievement of a metadata structure appropriated to maintain each of the relevant datum in every article. To do this, two different aspects have been taken into account: the object belonging to a congress context and the property of reusability as learning objects. By that, the IEEE-LOM (*Institute of Electrical and Electronics Engineers – Learning Object Metadata*) metadata standard [1] has been studied to be modified in order to use it in TAEE context. The metadata structure obtained has some similar features, given the use of LOM as a metadata system focused to learning contexts. Other standard are available to perform this action, as Dublin Core (DC) [2], but authors have considered that the standard offers a partial solution to TAEE needs, given it's a common metadata system used in library funds registration, but not having enough focus to educational or learning items. However, if there's a reduced and specific metadata amount per item, both metadata system (LOM and DC) are compatible [4] between them and TAEE system, but this compatibility is compromised if the digital object has to be defined by a greater amount of elements.

In LOM, metadata are grouped in nine main blocks to maintain a hierarchical disposition: *General, Lifecycle, Meta-metadata, Technical, Educational, Rights, Relation, Annotation, Classification.* Each of them defines respectively a metadata clustering in relation to the digital object definition scope. To extend the definition of every metadata group, some qualifiers are additioned to them to offer a complete structured composed by 64 metadata. In table 1, some explanations of LOM main groups are given.

In the case of the TAEE learning object, a new organization in metadata grouping has been made, as shown in Figure 1, in a suitable way according to digital objects information in the conference context.

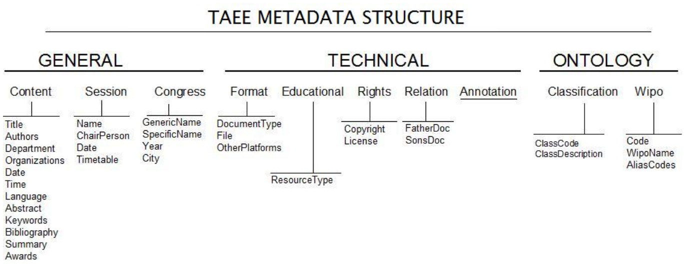

Figure 1. TAEE metadata structure

The metadata structure, therefore, offers in a first instance a division in three main blocks: *General, Technical* and *Ontology*. The first one, *General*, attends to those metadata that identify the digital object in global conference features, such as the object contents, the session and the event, establishing the nature of every document.

In the *Content* subgroup, the involved metadata are defined as "*Title*", "*Authors*", "*Departament*"and "*Organization*" to define the denomination and responsible organization; in addition, "*Date*" and "*Time*" has been created to define the exposure features in the session and congress limitations. The metadata "*Language*" let to know the language used in the paper, using the standard ISO 639:1998. "*Abstract*" and "*Keywords*" are available in the way the original texts have been written and with the information they keep it's possible to approach to the content of the full document. Other metadata as "*bibliography*", "*summary*" and "*awards*" complete the block.

The second metadata block "*Technical"* provides technical matters in relation to the original format of the document, its denomination and definition as a digital file (as done in the codification process), the kind of learning object, the owner's licenses and rights and the relation between main objects and its dependent components.

The last block "*Ontology*" is mainly based in a TAEE-specific-purpose subject structure as an electronics classification method, TAEE ontology, and it has been implemented in "*Classification*" metadata. TAEE Ontology has been created to match every subject to one or more classification families, by means of which a thematic definition can be made in the areas of electronics and teaching. In addition, it has been incorporated WIPO codes as a supplementary way of ontology, using the information given by WIPO (World Industrial Property Organization) which uses it to classify every granted patent or industrial model.

TABLE I. MAIN CATEGORIES IN LOM METADATA STRUCTURE

Group Name	Description
General	Descriptive general information of a learning object as an unit
LifeCycle	Features related with the learning object life and present state and those entities who has taken part in the creation, edition or evolution as authors, editors, etc.
Meta-metadata	Information about the creation of the learning object metadata.
Technical	Learning object requirements and technical features
Education	Conditions, properties and needs which cover the learning resource in an educational context.
Rights	Using conditions related with copyright and licenses for the resource use.
Relation	Types of relation between the learning object and others.
Annotation	Commentaries about learning uses of the digital object.
Classification	Subject description and ontologies family belongings with the specification of the Ontology used.

In the metadata structure, it's permitted the use of multiple values in metadata fields. This is contrary to [2] Dublin Core or LOM standards recommendations but it's justified, due a more compact database. To distinguish every elemental data, a kind of "*flag*", determined by a separation character as ";", has been used in the multiple value fields. So every value can be differentiated. In a further process, the elemental values can be disaggregated by means of the execution of a specific-purpose text treatment software, which reads every multiple value, extracts every elemental value and keep it in reference with the code number given to the belonging document.

In order to maintain the links between the metadata, the database evolutes to a multiple relation tables, in which the main table sustain the total amount of data indexed by a key field (the code of the document) and every complementary table offers a partial group of metadata linked by the same key field. Examples of these complementary tables are the Authors table, the bibliography table, the organizations table, etc.

The metadata extraction process has used, as a means of containment, several related tables from a created "ad-hoc" ODBC means database, in which each field is a container element of everyone of the metadata, whether multiple or single type.

IV. THE CONVERSION OF TEXTS INTO METADATA FILE ELEMENTS

To carry out the conversion of metadata placed in the database to the respective metadata files associated with digital objects, a software tool has been created for generating XML files [3]. The methodology, as well as their functional features, responds to an open schema, which is exportable and adaptable to other systems for metadata creation.

As a first approach to the process of creating XML metadata files, the application that has carried out this work, named the Shuttle [10], has got the tables contents as data source to keep all the organized metadata in the database, providing a open starting solution to any type of process because the Shuttle is universally adapted to any database by means of a multiple configuration definition, regardless of their size and whatever their structure.

Basically, the Shuttle program will open the database, connect to the table where metadata are contained and extract the field values in every record. As a means of data output, the values are sequentially inserted in an XML template file, where each metadata accommodation are equipped with a denomination composed by a label and a decoy as a lure, available for the recognition of the tag where the metadatum has to be inserted.

Successively, after replacing the lure with the metadatum and after filling all the places the Shuttle closes the XML file, giving to it the same name as the digital object reference, which is really recorded in the field "*File*". To illustrate the general process of operation, a block diagram has been shown in Figure 2 where the Shuttle data flow operation is defined.

To diversify the applications and uses of the metadata contained in database beyond TAEE frontiers, such as repositories or digital platforms, it has been thought the implementation of a complementary function in the Shuttle to perform XML files under IEEE-LOM standard. The functionality doesn't vary, with an exception: the treatment of the multiple data fields. It's recommended by IEEE-LOM the use of single-data fields, so then the cited complementary function has to implement a new procedure just only with the multiple-recorded metadata.

In the case of moving multiple data, given it's not permitted, they can't be extracted from the main table fields. It's necessary to switch the input data to the complementary database tables. As a simple solution, it's has been decided to give this function to the templates. Instead of doing a one-way data direction between the Shuttle and the template, it has been inserted some specific procedures in the templates, so then every time the Shuttle tries to insert a multiple metadata content in a single field, the template make an interruption to switch to the respective single-data content table. Then, this table is the only data source during this partial procedure.

The performance of this data flow double implementation generate XML file with both coexisting single-data and converted multiple-to-single data.

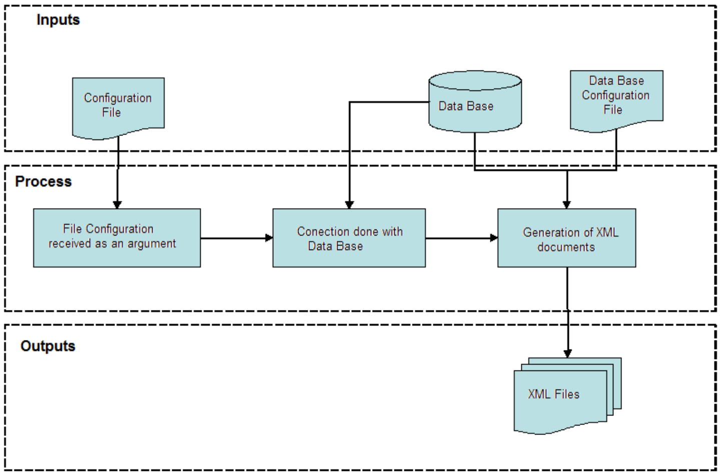

Figure 2. Block diagram of the operation of the application Shuttle

The execution of the Shuttle application has been implemented for the generation of 964 XML files that comprise the main learning objects, having made 34 metadata insertions per document. In various tests carried out under these conditions, it has been obtained an average running time of 2 minutes and 2 seconds, which means that a XML document has been created in 126.56 milliseconds average. In a subsequent set of tests, the same operations have been conducted on 4441 metadata files from the additional components.

Under similar conditions, the average running time has been 7 minutes and 55 seconds, giving a period of 107.96 milliseconds per XML file. This divergence between time values is due to latency and delay times outside the program execution. Applying linearization calculations in the Shuttle application behavior, it has been determined a performance average delay time of 6.03 seconds for each set of 1000 generations and an estimated average time of 101.52 milliseconds per XML file.

V. THE APPLICATIONS DERIVED FROM THE FILE STRUCTURE

The achievement of the processes set up to this point, a files structure of objects and associated metadata has been obtained. It can be said that at this point, the first goal of unification, standardization and labeling of TAEE documentation has been overcome, giving way to the second phase consisting of the dissemination of contents. It has, therefore, been necessary to create a file structure that can be adapted to both consequential actions: the creation of a Web that contains the information hosted on TAEE metadata and digital objects themselves and secondly, the bulk of the structure in institutional repositories and educational platforms.

To perform the first action on the Web site creation, and given the number of objects to handle, it has been taken the decision to create a specific Java software, which from the data contained in the database, will be able to generate static web pages to host the metadata and lead the user to each of the digital objects. The production process has led to an amount of web pages, one per congress, which provides information not only general but structured according to congressional sessions as divisions of every congress. Thus, by navigation through the site, it’s possible to access in a few steps to both digital object as well as the defining tab of each learning object.

Figure 3 shows a representation of the different actions performed by the application made for automatic generation of pages, on which insertion of formats and attributes has been later applied.

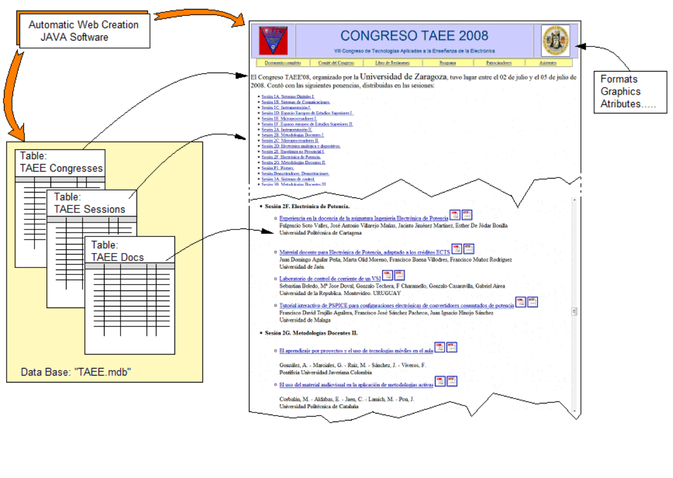

Figure 3. Website of congress and actions of its formation

In this figure 3, it can be seen that accessing the metadata central database, different page sections have been put as the header information, the list of sessions equipped with navigation and internal links, the list of documents that compose the session with the author information, the publication date and the figures which let to access to the digital object (PDF) and to a digital object form created as a result of the application of an XSL transformation exerted on each XML document. The implementation of the XSL transformation over XML files results in an XHTML page.

In order to spread the digital objects furthermore, a process of dissemination of digital objects from TAEE has been carried out through institutional repositories. Due to the amount of objects, a kind of bulk has been performed in our university's institutional repository, eSpacio-UNED [15], located in the web site http://e-spacio.uned.es as a means for the storage and distribution of digital objects, using open protocols like OAI-PMH [11]. The integration of the resulting objects has generated in the environment of the repository a number of elements making up every digital object structure. Thus, the components associated with the digital object are: a component of LOM metadata, a component of Dublin Core metadata derived from the LOM description, a component linked to the URL of the digital object itself and a component of the relationship of that object to other TAEE objects. This means that every digital object will have an own configuration collection, whose handling and visualization is possible by assigning a persistent URL to every one of the components of the object and the object itself. This allows the access to both learning object and its components, not only from the environment of the repository, but so foreign to it. By "*foreign*", it’s meant to be incorporated into the elements likely to be found by the searching engines available on the web as Google.

To improve the access to objects, like through TAEE web, it has been developed an user interface to access TAEE data located in the repository. In that interface, XML data has been used as site backbone. By means of “*adhoc*” style sheets transformation and cascade style sheets, it has been possible to conclude the interface attending to the repository philosophy. An illustration of user interface homepage has been shown in figure 4.

Figure 4. Main page from TAEE interface in the institutional repository in UNED

To access into the repository and have a view to the contents, it can be reached through the site http://espacio.uned.es:8080/fedora/get/taee:Congreso/demo:Collection/view . From the main page, the navigation links can redirect the user to any content in a specific congress, as well as, to invoke a query by means of a searching tool to find any TAEE object. As an alternative, the eSpacio-UNED list of services can be used, so in TAEE limits as in any other kind of object in any other collection.

VI. ANALYSIS DEVELOPED IN RELATION OF METADATA AND TAEE DIGITAL OBJECTS

By using the data obtained throughout the described phases in this paper, a study and subsequent analysis of metadata [7] has been carried out, in order to get statistics to highlight various aspects. In the horizon of the analysis, it can be found the performance and TAEE usefulness as a meeting place for organizations focused on researching and teaching electronics. Therefore, the issues arised by the analysis can be describe as:

a.- General aspects related to conferences organization, regarding proposed themes, number of presentations and reports, etc,

b.- An study of the thematic issue in TAEE, focusing it on two fronts, the analysis and use of ontology inside TAEE and the analysis of the keywords in documents,

c.- Temporary changes in subjects from three viewpoints: thematic evolution over time, abandonment of different obsolete topics and emergence of new themes and issues with the use of new devices, techniques, methodologies, etc,

d.- Study of participating authors and organizations and their relationships to each other, deepening their relationship with common themes, the degree of collaboration between them, and so on.

Therefore from the analysis of the suggested points, the main objective of the works is achieved by means of a detailed knowledge of the data generated from TAEE reusable learning objects. This analysis is related to a meta-analysis [8] made together this project as a comprehensive study of the contents with the purpose, additional statistical analysis, identification of the path along TAEE conferences, but mainly specializing in a relationships study among participating organizations. Thus, among the obtained outcomes, it was found some organizations which serve as connecting node between gateways of knowledge, so as other ones have been identified as organizations which operate as islands or terminals. This central organizations are the most prolific and matches with those who have taken part of some TAEE organizing committees.

In a first approach to the subjects treated in TAEE, a general study has been made of the digital objects dedication, using the TAEE ontology, using the data extracted from the themes in every TAEE publication. A general data distribution can be observed in figure 5, where a circular graphic has been exposed to define the dedication of publication in nine main blocks. Each sector in the graphic represents the total amount of papers in relation to every subject.

From figure 5 picture, a first interpretation from data can be extracted: there are two main thematical blocks in which, at least, are grouped the 66% of the whole documentation. These blocks are Learning Software Tools and Educational Topics. The remaining ontology families has a much lower ontological resulting dedication value, with the exception of the family that brings together analog, digital, communications and energy systems.

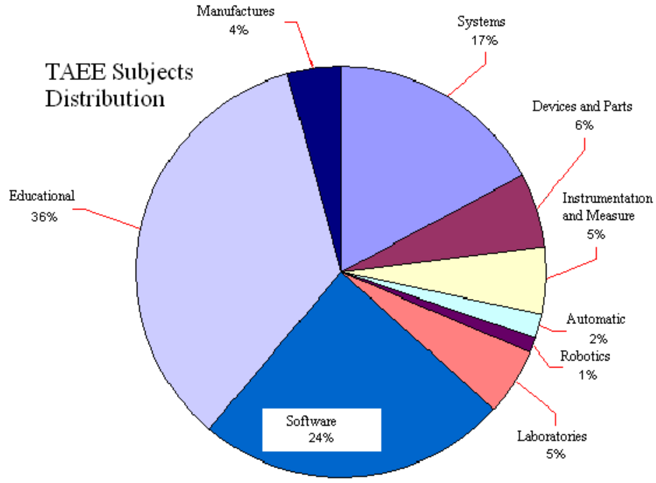

Figure 5. Thematical distribution about TAEE presented papers

Generalizing in TAEE subjects by a triple-block grouping, i.e., clusters which define thematic dedication in Technical, Educational and Software blocks, it's observed an almost exactly division in overall numbers. Using this division, it can be done an evolutionary study over time. This has been shown in figure 6.

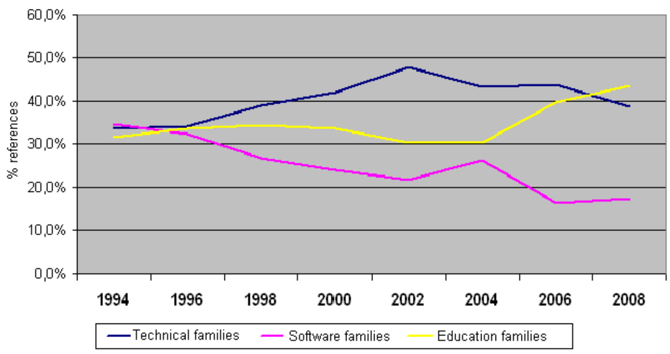

Figure 6. Temporal evolution of thematic division in TAEE

In that figure, it can be observed a similar beginning dedication in 1994 congress. Every one of the thematic blocks evolutes in a different way through time. In the first track, between the 1994 and 2000 meetings, the Technical branch has an increasing behaviour, driven primarily by publications with reference to digital systems and laboratory utilities. The Educational branch maintains the numbers of editions, mainly related to learning methodologies and learning based in practises. In a third place, Software family, tends to decrease about ten points, despite the growth in the use of related Internet tools.

Since 2002 conference, there's a new scenario and it can be observed an oscillating behaviour in the dedication of papers subjects. A growing tendency starts in relation to learning subjects, mainly driven by changes in legislation and education management system, which requires an adjustment in the departments organization in order to create adapted materials in new learning environments. On the other hand, in families dedicated to learning software and technical systems there is a downward trend. The final situation after the conclusion of the last meeting in 2008 leaves asymmetrical figures from the first congress in 1994. In the last conference, the rough division in the dedication is subject to 40% respectively for the families of teaching and technical systems, while reducing the dedication of the family of software tools slightly below 20%.

To access to further conclusions and data, the authors encourage to those interested in deepening reading the complete information from the annalysis, which is available in the Electrical and Computer Engineering Department Web site in UNED.

VII. Conclusion

The content of this document offers the work done on the processing of information generated at TAEE congresses. The general objectives consist in the creation of a unified and structured information environment from the documentation generated in TAEE. XML files have been created in connection with every digital object extracted from the general documentation, and it has been used in further utilities. These utilities have been defined as ways for disseminating TAEE information and documentation. In a first instance, it has been developed a common Web environment, in which TAEE services and links have been centralized and which sustain searching tools that favors the thematic consultation of the documentation.

To spread TAEE documentation beyond its frontiers, it has been necessary to make an adaptation of TAEE information to metadata standards in order of promoting online presence through hosting on documentation in repositories. This documentation hosted in educational platforms permits the creation of taylor-made courses, by means of TAEE objects converted in reusable objects.

In a parallel way, it also has been studied the performance of TAEE through its conferences by means of a content analysis of its digital objects through which it's likely to identify the path TAEE has taken, the relations between the organizations and agencies who has taken part of TAEE, as a social network and the methods and procedures carried out in greater profusion in the field of education.

The achievement of the objectives aims TAEE to enhance as Internet social network by adopting the measures applied to documentation and adaptation to the appropriate standard formats. Thus, it aims to provide the public and interested academic, the digital learning handled, allowing the reuse of materials generated as a source of knowledge.

Acknowledgment

The authors acknowledge the Spanish Science and Innovation Ministry and the CYTED-508AC0341 "SOLITE – FREE SOFTWARE IN TELE-FORMATION" Project.

In memoriam of Tomas Pollan a good friend and a best colleague.

References

[1] Learning Technology Standards Committee of the IEEE. "*Draft Standard for Learning Object Metadata*" IEEE 1484.12.1-2002. 15 July 2002

[2] DCMI Web Site. Available in URL: http://dublincore.org/index.shtml Last Visited: July 27, 2009.

[3] Juan Diego Gutiérrez Gallardo. "*XML – Manual imprescindible*". Anaya Multimedia. ISBN 84-415-1576-X

[4] Juan Ignacio Rouyet, Victor Martín. "*A comparative study of the metadata in SCORM and Dublin Core*". Universidad Pontificia de Salamanca

[5] Arriaga, J., Carpeño, A., Gordillo, T. "*Del objeto de aprendizaje a la implementación de una asignatura. Un caso práctico*". Universidad Politécnica de Madrid. Congreso TAEE 2006.

[6] Manuel Blázquez, Miguel Latorre, Gabriel Diaz, Manuel Castro, Jesus Arriaga, Edmundo Tovar, Fernando Pescador, Cesar Sanz, Tomás Pollán. "*Creación semiautomática de objetos educativos y metanálisis de TAEE*". CITA 2009 – Gijón, Spain.

[7] Electric and Computer Engineering Departament Web Site. Industrial Engineerieng Superior Technical School. Spanish University for Distance Education - UNED "*Análisis estadístico de la documentación generada en TAEE*". July 2009. Available in URL: http://www.ieec.uned.es/Investigacion/RedOber/index.htm

[8] Edmundo Tovar, Jorge A. López, Manuel Blázquez, Miguel Latorre, Manuel Castro, Gabriel Díaz, Fernando Pescador, Cesar Sanz, Tomás Pollán Oskar Casquero, Tomás Pollán y Jesús Romo "*Metaanálisis de la documentación generada en los congresos TAEE*". Electric and Computer Engineering Departament Web Site. Industrial Engineerieng Superior Technical School. Spanish University for Distance Education - UNED. July 2009. Available in URL: http://www.ieec.uned.es/Investigacion/RedOber/index.htm

[9] Miguel Latorre et al. "*A good practice example on Learning Object Reutilization*". DIEEC, UNED, 2008.

[10] Manuel Blázquez, Pablo Blázquez. "*Proceso automático de formación de archivos de metadatos*". IEEE Electrronic Journal of the UNED Student Branch. July 2009. ISSN: 1989-2195. Available in URL http://www.uned.es/ieec-uned/

[11] Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Technologies Report. October 13, 2006. Available in URL: http://xml.coverpages.org/oams.html. Last visited: March 7, 2009.

[12] Miguel Ángel Marzal García-Quismondo, Javier Calzada-Prado, Aurora Cuevas Cerveró. "*Desarrollo de un esquema de metadatos*" Scientific Documentation Spanish Magazine nº 29. Oct-Dec pag.551-571, 2006. ISSN 0210-0614

[13] A.M. Bianco, M. de Marisco, M. Temperini. "*Standards for e-learning. Quis Team*". Available in URL: http://www.tosip.no/QUIS/ . 2005. Last visited March 7, 2009

[14] Marie-Claude Côté, (TBS), Cpt. Peter Hope (DND). "*Presentation to the GOL Metadata Working Group*". Government of Canada E-learning Metadata Application Profiles. January 20, 2004

[15] e-Spacio UNED Project. Institutional Repository for digital contents in Spanish University for Distance Education - UNED. Descriptive Paper. UNED Library Services. Available in URL: http://e-spacio.uned.es/fez/index.php. Last visited: October 20, 2009.

[16] OpenCourseWare Webpage – Universia. Available in URL: www.universia.es/index.htm. Last Visited: August 14, 2009.

[17] Miguel Benito."*Sistemas de clasificación. Manual de aprendizaje de la Clasificación Decimal Universal y breve introducción a la Clasificación Decimal de Dewey*". Editorial Taranco, 1999