Fuzzy Multidimensional Modelling for Flexible Querying of Learning Object Repositories

Gloria Appelgren Lara[1], Miguel Delgado[2], and Nicolás Marín[2,⋆]

[1] LATAM Airlines Group, Santiago, Chile
gloria.appelgren@lan.com
[2] University of Granada, Department of Computer Science and AI, 18071 - Granada, Spain
mdelgado@ugr.es, nicm@decsai.ugr.es

Abstract. The goal of this research is to design a fuzzy multidimensional model to manage learning object repositories. This model will provide the required elements to develop an intelligent system for information retrieval on learning object repositories based on OLAP multidimensional modeling and soft computing tools. It will handle the uncertainty of this data through a flexible approach.

Keywords: Learning Object Metadata, Data Warehousing, Multidimensional Model, Fuzzy Set, Fuzzy Techniques.

1 Introduction

In recent years, one of the major challenges in e-learning is the standardization of content. The Learning Objects technology allows contents that comply with certain standards, such as those indicated in the rules of the Sharable Content Object Reference Model - SCORM[3,16], to be reused in different distance learning platforms, making interoperability possible. Unfortunately, there are still shortcomings in the management and evaluation of content.

In the literature, there are many definitions for the term Learning Object (LO). One of the most accepted is established by the IEEE Standards Committee: a LO is *any digital or other entity that can be used, reused or referenced during a learning process supported by technology*[9]. The focus of LO technology is the encapsulation of content, so that it becomes an autonomous unit, i.e. a LO is self-contained and devoted to present a concept or idea. This has been established to be structured as a combination of educational content: lecture notes, presentations, tutorials, etc., and their respective metadata: title, author, rank, age, etc. These LOs and their metadata are stored in Learning Object

⋆ Work partially supported by project Representación y Manipulación de Objetos Imperfectos en Problemas de Integración de Datos, Junta de Andalucía, Consejería de Economía, Innovación y Ciencia(P07-TIC-03175).

H.L. Larsen et al. (Eds.): FQAS 2013, LNAI 8132, pp. 112–123, 2013.

Repositories (LORs), which correspond to *stores* that provide the mechanisms for searching, exchanging and reusing LOs.

The Learning Object Metadata (LOM) is a model formally approved by IEEE and widely accepted in the e-learning field[13]. LOM is based on previous efforts made to describe educational resources on projects ARIADNE, IMS and Dublin Core[6]. Its aim is to create structured descriptions of educational resources. Its data model specifies which aspects of learning object should be described and what vocabularies may be used in that description. The model consists of a hierarchical description of nine major categories that group the other fields: General, Lifecycle, Meta-metadata, Technical, Educational Use, Rights, Relation, Annotation and Classification. This standard aims to ensure interoperability between repositories from various sources.

In order to manage knowledge coming from LOs, we can take advantage of a Data Warehouse and techniques for online analytical processing (Online Analytical Processing - OLAP)[2] appropriately adapted for the management of LORs. Though OLAP systems and related intelligent management tools (Business Intelligence - BI) are really targeted to the business, in this paper, we propose their use in education, specifically to handle LO repositories.

Data Warehousing technology, due to their analytic orientation, proposes a different way of thinking within the information system area, which is supported by a specific data model, known as multi-dimensional data model, which seeks to provide the user with an interactive high-level vision of the business operation. In this context, the dimensional modeling is a technique for modeling understandable views to friendly support *end user* operations. The basic idea is that users easily visualize the relationships between the various components of the model.

In general, multidimensional models are oriented to the generation of ad-hoc reports that enable business decisions based on more accurate data[5]. However, data usually are incomplete and, in many repositories, they are expressed in natural language and are often affected by the inherent imprecision of this language. As Molina et al. [15] indicated, it is possible to model multidimensional data cubes based on fuzzy set theory and thus allow the management of uncertainty and imprecision in the data. In learning object repositories we can find the presence of uncertainty in the expressions of a LO metadata such as "Neuroscience for Kids"; in this expression we have the presence of the word *kid*, which is a categorization regarding the age of the target users of the LO, understood by any person but not easily managed by machines, that is what we do by applying fuzzy logic.

In this work, the idea is to provide an intelligent system that allows users to analyze the learning object type that best fits their way of learning or the learning of their students in any field of knowledge, developing and implementing a DW OLAP technique to integrate the fuzzy sets theory to ease the extraction of knowledge [12]. The main contribution is the flexible design of datacubes for the management of a Learning Object Repository (G-LOR).

The paper is organized as follows: Section 2 briefly describes the previous concepts related to learning objects, repositories and data warehouses, and states the problem we face in this paper. Section 3 is devoted to explain the LOM IEEE standard. Section 4 develops the proposed solution to the problem using a fuzzy multidimensional model of learning object repositories. We include a use case that exemplifies this proposal. Finally, we end the paper with some conclusions and guidelines for future work.

2 Background and Motivation

2.1 LOs, LORs and LO Metadata

There are many definitions of learning objects (Learning Object, LO). As we have mentioned in the introduction, we use as reference the IEEE standard, which defines a learning object as *any digital or other entity that can be used, reused or referenced during learning process supported by technology*[9]. LOs are usually organized in Learning Object Repositories (LORs), that are repositories that allow us to store, search, retrieve, view and download LOs from all areas of knowledge. Hence the object and the repository are complementary.

In particular, the Learning Object Repositories can be classified into the object repositories containing learning to download and incorporate into a learning platform, and metadata repositories that contain the object information and a link to its location on the Internet.

The search and retrieval of LOs is guided through the use of metadata that describe these learning objects. In this sense, LOM (Learning Object Metadata) is the IEEE standard e-learning, formally approved and widely accepted [9]. LOM is based on previous efforts made to describe educational resources on projects ARIADNE, IMS and Dublin Core[6]. The aim of LOM is the creation of structured descriptions of educational resources. Its data model specifies which aspects of a learning object should be described and what vocabularies may be used in that description. The model consists of a hierarchical description of nine major categories that group the other fields: General, Lifecycle, Metametadata, Technical, Educational Use, Rights, Relation, Annotation and Classification.

2.2 Datawarehousing, OLAP and OLTP

Data Warehousing is the design and implementation of processes, tools, and facilities to manage and deliver complete, timely, accurate, and understandable information for decision making[1].

The OLAP techniques (Online Analytical Processing) develop a multidimensional analysis, also called *analysis of data hypercubes*. The data handled by this technique are imported both from external sources and from production databases. These databases are feeding production systems based on OLTP (Online Transactional Processing).

For the sake of decision-making, it is necessary to have a large amount of information organized and with specific characteristics. Thus there is a consensus that data warehouses are the ideal structure for this. We recall the definition made by WH Inmon [10], a pioneer in the field, *A data warehouse is a set oriented data by topic, integrated, time-varying and non-volatile which is used to support decision making.* There are four main categories of OLAP tools. These are classified according to the architecture used to store and process multidimensional data, these are: Multidimensional OLAP (MOLAP), Relational OLAP (ROLAP), Hybrid OLAP (HOLAP) and Desktop OLAP (Dolap, Desktop OLAP).

2.3 Fuzzy Sets Theory and OLAP

In classical data warehouses (DWH), classification of values takes place in a sharp manner; because of this real world values are difficult to be measured and smooth transition between classes does not occur. According to [14] a Fuzzy Data Warehouse (FDWH) is a data repository which allows integration of fuzzy concepts on dimensions and facts. Then, it contains fuzzy data and allows the processing of these data. Data entry with *lack of clarity* in data storage systems, offers the possibility to process data at higher level of abstraction and improving the analysis of imprecise data. It also provides the possibility to express business indicators in natural language using terms such as high, low, about 10, almost all, etc., represented by appropriate membership functions.

Different approaches have been proposed for integrating fuzzy concepts in Data Warehouses. For example, Delgado et al. [8] present dimensions where some members can be modeled as fuzzy concepts. Additionally, different fuzzy aggregation functions are used in this approach, which have been developed for different OLAP (Online Analytical Processing) operations, such as roll-up, drill, dice, among others, applying fuzzy operations [7].

Castillo et al. [4] also use fuzzy multidimensional modelling, proposing two methods for linguistically describing time series data in a more natural way, based on the use of hierarchical fuzzy partition of time dimension. This approach introduces two alternative strategies to find the phrases that make up an aggregate. In summary, it is possible to develop a DW and OLAP techniques applied to integrate fuzzy set theory to ease the extraction of knowledge[12].

2.4 Need for DW Techniques in LORs

Learning Objects are developed and stored in various repositories on the web. They involve an enormous potential for the benefit of e-learning. However, there are many technical issues to be considered so that they can be reusable, interchangeable or manageable.

The most widely used standard for learning object repositories is called LOM (Learning Object Metadata) defined by the IEEE (Institute of Electrical and Electronics Engineers, Inc.). This standard, has different sections and fields to describe in detail a learning object and its characteristics through a series of metadata grouped into categories. The proposed metadata are of generic type:

title, description of the subject, format, language. Other LOM model elements, education related, are the least densely populated: the duration of the learning activity, difficulty, structure, granularity, etc. [11]. Therefore we are in a scenario with incomplete data, usually affected by imperfections of different types such as imprecision, inconsistencies, etc.

The information associated with these fields is usually stored in natural language. This feature provides *good expression* but does not facilitate or enable a setting for the automatic inference and reasoning on the metadata records[18]. In conclusion, current models suggest very rigid structures for the representation of the domains[15].

The storage and retrieval of learning objects can benefit from the use of Data Warehouse System and OLAP techniques [2] that allow flexible intelligent management on learning object repositories. The multidimensional model is highly appropriate to represent complex metadata models and OLAP operations could serve as basis for a friendly querying of the repositories. Additionally, the use of fuzzy subsets theory together with these tools, permits to handle data imperfections produced by the use of natural language when inserting LOs metadata.

This paper seeks to add flexibility and uncertainty management by designing a multidimensional model where we will apply fuzzy sets for that goal in fact and dimensions tables. The objective is to design a flexible model for the intelligent management of learning object repositories (G-LOR).

3 Learning Object Metadata Standards

As we have commented before, LOM (IEEE Learning Object Metadata) is the standard for e-learning formally approved and widely accepted [5]. LOM is based on previous efforts made to describe educational resources on projects ARIADNE, IMS and Dublin Core [2]. The aim of LOM is the creation of structured descriptions of educational resources. Its data model specifies which aspects of a learning object should be described and what vocabulary may be used in that description.

This model proposes a hierarchical description in nine major categories that group the other fields: General, Life Cycle, Meta-Metadata, Technical, Educational, Rights, Relation, Annotation and Classification. Table 1 describes each category of the LOM model.

We have used the LOM model to design a database to store the learning objects. The database is described in figure 1 presented below, where we can see that the general entity which controls access to the data for each learning object based in LOM.

This database was generated in the Database Management System (DBMS) Oracle express 11g. This is a repository of LO as proof of concepts. And that is the base of our proposal for fuzzy multidimensional modelling of learning object repositories.

Table 1. Categories of the LOM model

Category	Description
General	Information that describes the learning object as a whole. It describes the purpose of education. Identifier includes fields such as IT, title, description, etc.
Life Cycle	Characteristics related to the history and present state of the learning object and those who have affected this object during its evolution.
Meta-Metadata	About the metadata themselves, not regarding the learning object being described. It contains information such as who has contributed to the creation of metadata and the kind of contribution he has made.
Technical	Technical requirements and technical characteristics of the learning object, such as size, location or format in which it is located. Additionally, this element stored potential technical requirements to use the object referred to metadata.
Educational Uses	Policies educational use of the resource. This category includes different pedagogical features of the object. Typically, includes areas such as resource type - exercise, diagram, figure - and level of interactivity between the user and the object-high, medium, low-, or the context of resource use - college, primary education, doctorate- among others.
Rights	Terms rights that concern the resource exploitation. Details on the intellectual property of the resource. It also describes the conditions of use and price when applicable.
Value	Value of the resource described with other learning objects. Explains the type of relationship of the learning resource to other LOs. It has a name-value detailing the name of the LO-related and type of relationship, is part of, is based on, etc.
Annotation	Includes comments on the educational use of the learning object, as well as its author and date of creation.
Classification	Description subject of the appeal in a classification system. It reports if the LO belongs to some particular subject. For example, physics or history. It allows as much detail as you want by nesting of topics.

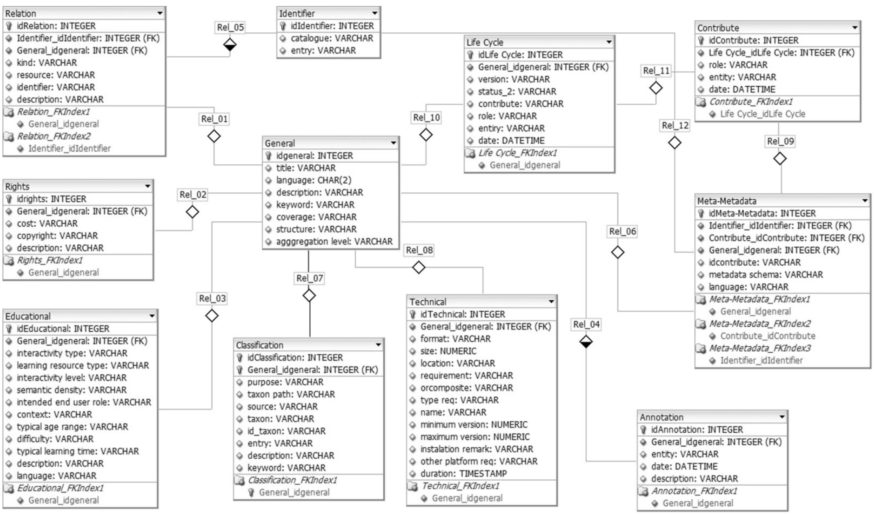

Fig. 1. The LOM model

4 A Fuzzy Multidimensional Model of Learning Object Repositories

The database described in the previous section is the basis for the design of a multidimensional model whose objective is to allow the flexible querying of learning objects from this repository.

A dimensional model can be expressed as a table with a composite primary key, called the fact table, and a set of additional tables called dimension tables.

In our application domain, the fact table is called Learning Objects (LO) and we propose a set of 12 dimension tables, namely: typical range of age, degree of difficulty, length, creation date, type of interactivity, level of aggregation, contribution and life cycle, localization, path taxonomic classification, context or scope, level of interactivity and finally, semantic density. Figure 2 describes this star model.

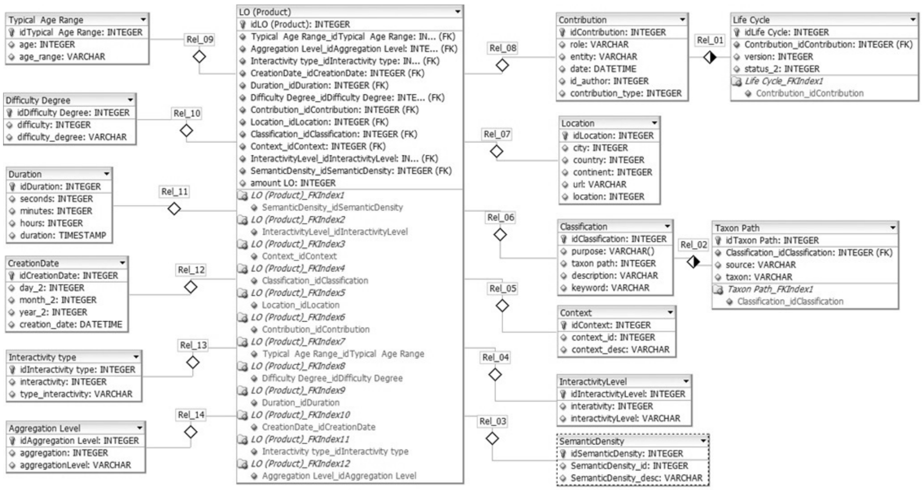

Fig. 2. M-LOR. Learning Object Repository Multidimensional Model.

4.1 The Fuzzy Dimensions

Some of the dimensions of the model depicted in figure 2 are related to fuzzy concepts. In this work, in order to be able to model this kind of fuzzy dimensions, we have considered to use the fuzzy multidimensional model introduced by Molina et al. [15].

The model proposed by Molina et al. is founded on the use of dimensions where the hierarchies of members are defined through the use of a fuzzy kinship relation. The use of this type of hierarchies make the modelling of dimensions related to fuzzy concepts possible, because membership functions of labels can be used to set out the mentioned kinship relation.

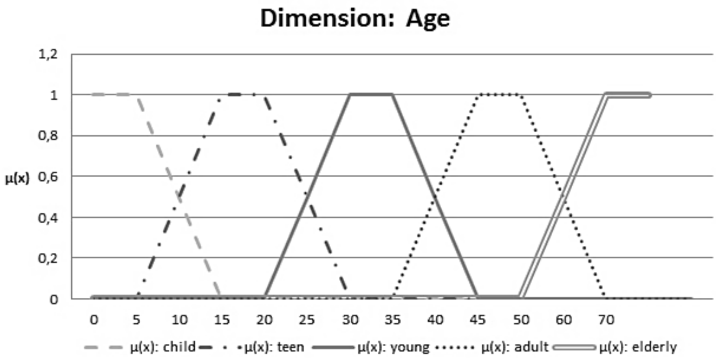

Fig. 3. Age partition

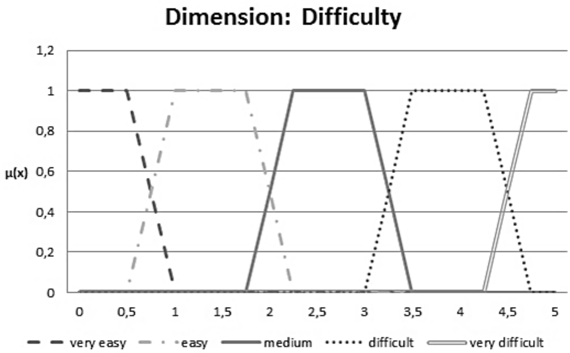

Fig. 4. Difficulty partition

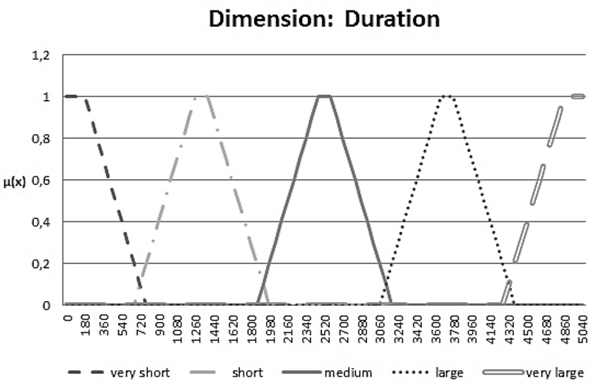

Fig. 5. Duration partition

In our fuzzy model for learning object repositories, we have used fuzzy concepts like Age, Difficulty, and Duration to define fuzzy hierarchies. See figures 3, 4, and 5.

For each of these concepts, we have developed a dimension in our model with at least three levels (basic domain, fuzzy partition and all).

4.2 An Example Datacube

We have considered a datacube to resolve questions concerning the analysis of community contributions. Figure 6 depicts a diagram that shows the star model of learning object provider.

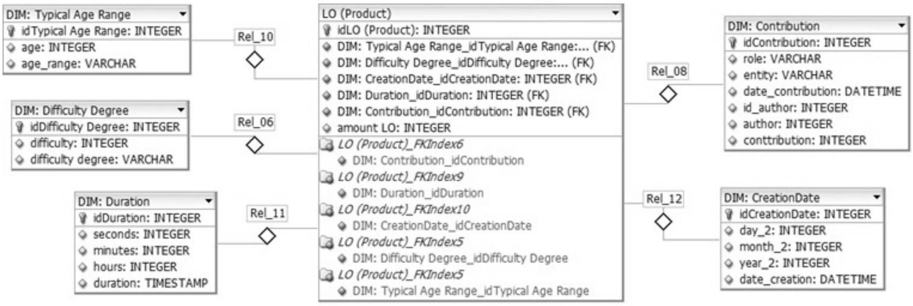

Fig. 6. Star Model about Analysis Community Contribution

The datacube is build in order to solve queries aimed at the measurement of the productivity of a community that provides learning objects and study the relationships between these four variables of the LOM model. An example query could be *to find the amount of free radicals prepared in 2010 for large communities of adults, which difficulty degree is very difficult and duration is medium.*

The fuzzy scheme of the data warehouse may be more complicated than the crisp one due to additional dimensions, fact tables, and relationships[17].

The dimensions by extension, are:

– Contribution role type: short, medium, and large community
– Creation Date: day, month, and year
– Age: child, adolescent, young, adult, and elderly
– Difficulty: very easy, easy, medium, difficult, and very difficult
– Duration: very short, short, medium, long, and very long

See Figure 7 in order to see the levels of the hierarchy developed for each dimension.

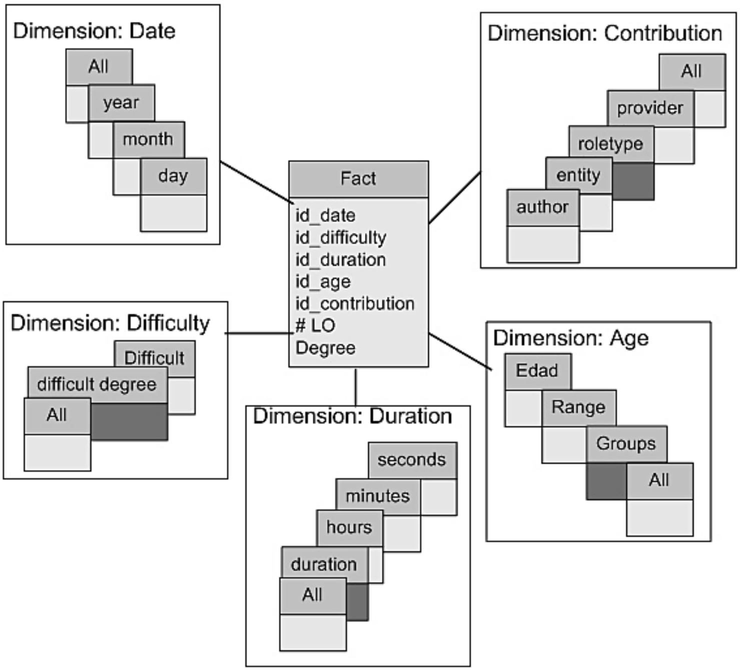

Fig. 7. Datacube Scheme for the Analysis of LO contributions

Table 2. Facts

#LO	Degree	Difficulty Degree	Contribution	Duration	Age Range
2	0,6	5 (very difficult)	Provider 1 (large community)	75 min (verylarge)	23 (Young)
3	0,7	4 (difficult)	Provider 2 (medium community)	20 min (short)	18 (Teen)
1	0,8	4 (difficult)	Provider 2 (medium community)	36 min (medium)	15 (Teen)
3	0,3	1 (very easy)	Provider 4 (short community)	10 min (very short)	7 (Child)
3	0,5	5 (very difficult)	Provider 5 (large community)	30 min (medium)	48 (Adult) t

Some example facts are shown in the table 2.

The operations to be performed to respond to the query are:

- Dice on Age dimension with the condition ”adult” at the group level.
- Dice on Contribution role type on the condition ”large community”
- Dice on the condition Difficulty Degree dimension ”very difficult”
- Dice on Duration on the condition ”medium”.
- Roll-up on the scale Creation Date to define level Year 2010.

5 Conclusions and Future Work

In this paper we describe and apply concepts of Learning Objects, Repositories, Precise and Fuzzy Data Warehouses. We show how to use concepts and tools used in business intelligence to the area of management of learning object repositories widely used in educational communities and e-learning. It raised issues related to the standardization and management of learning object repositories and proposes a solution through the use of fuzzy multidimensional modelling that can ease the management at the e-learning. The main scientific contribution is the design of buckets for the flexible management of Learning Objects Repositories.

Currently we are working in the proposal of new multidimensional models to analyze content demand, quality of content, user profiles in the various communities and in the development of data mining techniques to define models containing grouping rules and regulations prediction. These prediction models could be used to automatically perform sophisticated analysis of data to identify trends that will help to identify new opportunities and choose the best for learning object repositories.

References

1. Ballard, C., Herreman, D., Schau, D., Bell, R., Kim, E., Valencic, A.: Data modeling techniques for data warehousing. Tech. Rep. SG24-2238-00, IBM - International Technical Support Organization (February 1998), http://www.redbooks.ibm.com
2. Baruque, C.B., Melo, R.N.: Using data warehousing and data mining techniques for the development of digital libraries for lms. In: Isaías, P.T., Karmakar, N., Rodrigues, L., Barbosa, P. (eds.) ICWI, pp. 227–234. IADIS (2004)
3. Bohl, O., Schellhase, J., Sengler, R., Winand, U.: The sharable content object reference model (scorm). a critical review. In: Proceedings of the International Conference on Computers in Education, ICCE 2002, p. 950. IEEE Computer Society Press, Los Alamitos (2002)
4. Castillo-Ortega, R., Marín, N., Molina, C., Sánchez, D.: Building linguistic summaries with f-cube factory. In: Actas del XVI Congreso Español sobre Tecnologás y Lógica Fuzzy, pp. 620–625 (2012)
5. Connolly, T.M., Begg, C.E.: Database Systems: A Practical Approach to Design, Implementation and Management, 5/E. Addison-Wesley (2010)
6. Currier, S.: Metadata for learning resources: An update on standards activity for 2008. Tech. Rep. 55, ARIADNE (April 2008), http://www.ariadne.ac.uk/issue55/currier/
7. Delgado, M., Molina, C., Rodríguez Ariza, L., Sánchez, D., Vila Miranda, M.A.: F-cube factory: a fuzzy olap system for supporting imprecision. International Journal of Uncertainty. Fuzziness and Knowledge-Based Systems 15(suppl.-1), 59–81 (2007)
8. Delgado, M., Molina, C., Sánchez, D., Vila, M.A., Rodriguez-Ariza, L.: A fuzzy multidimensional model for supporting imprecision in olap. In: Proceedings of the 2004 IEEE International Conference on Fuzzy Systems, vol. 3, pp. 1331–1336 (2004)
9. Hodgin, W.: Draft standard for learning object metadata. Tech. Rep. IEEE 1484.12.1-2002, Learning Technology Standards Committee of the IEEE (July 2002)
10. Inmon, W.: Building the Data Warehouse. Wiley (2005)
11. Kraan, W.: Learning object metadata use survey: sticking the short and wide in the long and thin. Tech. rep., CETIS - Centre for Educational Technology and Interoperability Standards (2004), http://zope.cetis.ac.uk/content2/20041015015543
12. Laurent, A.: Querying fuzzy multidimensional databases: unary operators and their properties. Int. J. Uncertain. Fuzziness Knowl. -Based Syst. 11(suppl.), 31–45 (2003), http://dx.doi.org/10.1142/S0218488503002259

13. López Guzmán, C.: Los Repositorios de Objetos de Aprendizaje como soporte a un entorno e-learning. Master's thesis, University of Salamanca (2005)
14. Małysiak-Mrozek, B.z., Mrozek, D., Kozielski, S.: Processing of crisp and fuzzy measures in the fuzzy data warehouse for global natural resources. In: García-Pedrajas, N., Herrera, F., Fyfe, C., Benítez, J.M., Ali, M. (eds.) IEA/AIE 2010, Part III. LNCS, vol. 6098, pp. 616–625. Springer, Heidelberg (2010)
15. Molina, C., Rodríguez Ariza, L., Sánchez, D., Vila, M.A.: A new fuzzy multidimensional model. IEEE T. Fuzzy Systems 14(6), 897–912 (2006)
16. Poltrack, J., Hruska, N., Johnson, A., Haag, J.: The next generation of scorm: Innovation for the global force. In: Proceedings of Interservice-Industry Training, Simulation, and Education Conference (I-ITSEC), pp. 1–9 (2012)
17. Sapir, L., Shmilovici, A., Rokach, L.: A methodology for the design of a fuzzy data warehouse. In: Proceedings of the 4th International IEEE Conference on Intelligent Systems (IS 2008), vol. 1, pp. 2.14–2.21. IEEE (2008)
18. Soto-Carrión, J., Garcia, E., Sánchez-Alonso, S.: Problemas de almacenamiento e inferencia sobre grandes conjuntos de metadatos en un repositorio semántico de objetos de aprendizaje. In: Actas del III Simposio Plurisciplinar sobre objetos y diseños de aprendizaje, Oviedo, Spain (2006)