Assessment of the Most Relevant Learning Object Metadata

Relieving the Learner-User from Information Overload

Alessandro da Silveira Dias and Leandro Krug Wives
Informatics Institute, Universidade Federal do Rio Grande do Sul,
Av. Bento Gonçalves, 9500, Porto Alegre, Brazil

Keywords: Metadata, Learning Object, Information Overload, End User, IEEE LOM, Learner-driven Learning.

Abstract: E-learning systems created new learning spaces and enabled users to participate more actively in the construction of their own knowledge. In these, users can learn in a self-directed way, make choices regarding their learning depending on the possibilities provided by the system. One of the most important choices is "how to learn", which in this work corresponds to which learning object the user will choose. For this, the user, considering of a list of relevant learning objects, uses their metadata to make a decision. The problem is that current metadata standards have many types of information, so, the user suffers from the metadata information overload. For relieving the user, this work assesses the most relevant metadata from a set of learning objects and ranks them based on this assessment. A case study was conducted to show the application of this ranking on the AdaptWeb® e-learning system and indicated that the vast majority of subjects did not suffer from the metadata overload.

1 INTRODUCTION

With the advancement of technology and telecommunications, computers began to be used in the context of Education, which created new learning spaces and enabled users to participate more actively in the construction of their knowledge. For instance, there are several learning resources available in open learning repositories on the Internet. In these repositories, users can learn in a self-directed way and make choices (decisions) regarding their learning. For example, users must decide "what to learn", "how to learn", "where to learn", "in which learning pathway to learn", "how to perform self-assessment", among others. From the point of view of Pedagogy, these choices belong to the learner-driven learning paradigm (Alexander et al., 2004; Watkins et al., 2007).

On the other hand, the use of computers in the context of Education has brought numerous pedagogical and technological challenges. Among the technological ones, one of them is the implementation of techniques that allow ways of designing, developing and distributing educational material, which gave rise to *Learning Objects* (LO).

LOs can be defined as any entity that can be used, reused or referenced during computer-supported learning. They can contain a variety of features, from the simplest ones, such as text, to some more sophisticated ones like hypertext or animation with interactive features (IEEE Learning Technology Standards Committee, 2016).

Over time, LOs became available quickly, cheaply and widely disseminated. Therefore, to facilitate the search, evaluation, acquisition, sharing, and use of LOs, different metadata standards have emerged, like IEEE LOM (IEEE Learning Objects Metadata), SCORM (Shareable Content Object Reference Mode) and IMS-Metadata.

As previously mentioned, users can make different choices or decisions during their learning in e-learning systems, depending on the possibilities provided by the system. One of the most important decisions is "how to learn", which in this work corresponds to which LO users will use to learn – a simple text, a video lesson, a multimedia presentation, a simulation application, etc. For this, the user, in front of a list of relevant LOs (provided by a recommender system or an information retrieval system) uses the LO's metadata to make the final decision, about what LO to use. The problem is that the current metadata standards have many types of information (general information metadata, technical metadata, educational metadata,

da Silveira Dias, A. and Krug Wives, L.
Assessment of the Most Relevant Learning Object Metadata.
DOI: 10.5220/0006660601750182
In *Proceedings of the 10th International Conference on Computer Supported Education (CSEDU 2018)*, pages 175-182
ISBN: 978-989-758-291-2

administrative metadata, etc.) and, so, the user suffers a kind of information overload, known as *metadata overload* (Beeson, 2006).

In this work, it was performed a research about the relevance of different LOM-based metadata for university students. The goal was to decrease the amount of metadata available and, thus, relieve users from being overloaded with information during their LO selection process. Based on the relevance indicated by the students, LOM metadata was ranked.

IEEE LOM was chosen for this work because it is the most widely used in e-learning systems and it served as a basis for the development of other metadata standards, as SCORM (Advanced Distributed Learning Network, 2004) and the Agent-Based Learning Objects (OBAA) (Vaccari et al., 2010).

A case study is presented to show the application of the developed ranking of the most relevant IEEE LOM metadata for university students on an e-learning system, namely AdaptWeb® (Gasparini et al., 2013). This ranking was used to assemble the screen where LOs are listed to users in this system. A group of 30 users attended an online course in this system. At the end of the course, an online satisfaction survey was conducted about the choices of LOs performed during the system usage, i.e., during the learning activity, and about the set of metadata displayed by the system to the user. This survey showed that the vast majority of subjects did not suffer from the metadata information overload.

2 THE IEEE LOM STANDARD

The IEEE LOM standard is based on a conceptual data schema that defines the structured metadata instance of a LO. This instance describes the relevant characteristics of the resource which applies and is composed of data elements (IEEE Learning Technology Standards Committee, 2016). These characteristics are stored in a structure composed of 56 data elements, organized into nine categories:

- The **General** category groups the general information that describes the learning object as a whole;
- The **Lifecycle** category groups the features related to the history and current state of this learning object and those who have affected this learning object during its evolution;
- The **Meta-Metadata** category groups information about the metadata instance itself (rather than the learning object that the metadata instance describes);
- The **Technical** category groups the technical requirements and technical characteristics of the learning object;
- The **Educational** category groups the educational and pedagogic characteristics of the learning object;
- The **Rights** category groups the intellectual property rights and conditions of use for the learning object;
- The **Relation** category groups features that define the relationship between the learning object and other related learning objects;
- The **Annotation** category provides comments on the educational use of the learning object and provides information on when and by whom the comments were created;
- The **Classification** category describes this learning object in relation to a particular classification system.

In the IEEE LOM standard, metadata is organized into XML documents. As mentioned, the standard has a structure composed of 56 elements, organized into nine categories (detailed in Figure 1). Many of the elements can be repeated, for example, *general.keyword* can have up to 10 values. Moreover, many categories can be repeated, such as the *annotation* group, that can have up to 30 set of values. With this amount of data, it is common to find a LO containing hundreds of metadata. In addition, some of them are focused on structural, referential and organizational aspects, which may not be relevant for end users when selection materials to use for learning.

3 ASSESSING THE MOST RELEVANT LOM METADATA

In this work a quantitative research was carried out; the data collection was done through a questionnaire developed as an online Web form that could be accessed through an invitation e-mail that was sent to the participants of the experiment. Through this form, subjects were asked to indicate what were the ten most relevant metadata to help them choose a LO to learn a specific learning topic of a course. With these data, the ranking of the most relevant IEEE LOM metadata for university students was set up. This process is detailed in the following sections.

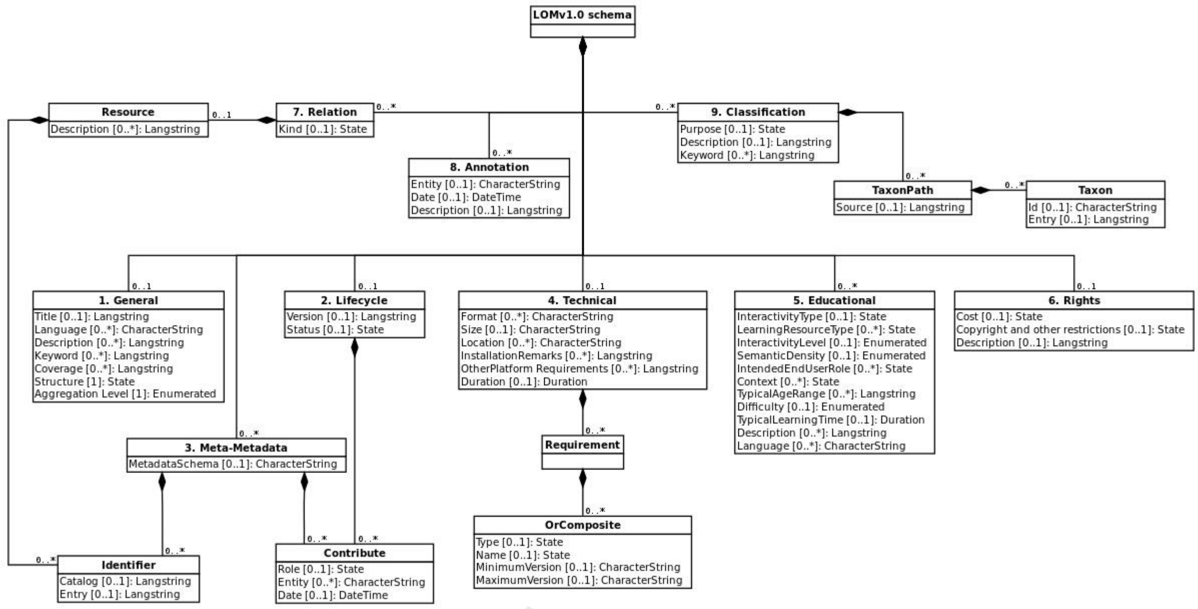

Figure 1: Hierarchical representation of metadata in the IEEE LOM standard (Ciloglugil and Inceoglu, 2016).

3.1 The Online Questionnaire

The questionnaire was developed as a dynamic web page; it was composed of a header, with static content, and a body, with dynamic content. In the header, we have presented general information about the research, i.e., researchers and their institution, the period in which it would be open and the objectives of the research. The body contained a script to assign one learning topic to each individual randomly. For the topic chosen (i.e., to be learned), the participant was asked to indicate the metadata (10 at most) that she finds most relevant to understand and select a LO.

Figure 2 shows the main part of the body of the questionnaire, that is a form. Before displaying the form, the following text is presented to the subject: "Imagine that you are a student of a distance education course and that, using the e-learning system of the course, you should learn the topic *X*. There are 2 digital learning materials available to learn this topic: LO_1 and LO_2. In the table below, these 2 digital learning materials are listed. For each one there is a set of information fields that describe it. Each field of information is accompanied by its meaning. Read all of this information about each digital learning material and, after that, indicate in the *Your Opinion* column the 10 information fields that you think are the most important when choosing a digital learning material to learn".

X, LO_1, and LO_2 were variables defined dynamically by the script, when the subject entered the online questionnaire. It could be: the learning topic *X* = "*Cardiovascular Human System"* (from the Biological Science area), with the LOs LO_1 = "text with figures" and LO_2 = "interactive map", as presented in Figure 2; or it could be: *X* = "*Geoprocessing"* (from the Exact Science area), with LOs LO_1 = "package (a text with graphics and a small statistical dataset)" and LO_2 = "simulation application"; or it could be: *X* = "*Civil Procedural Law* (from the Human Science area), with Los LO_1 = "video lesson" and LO_2 = "document (text only)". This variability of topics (from different areas of knowledge) and learning objects (with different formats and granularities) ensures that the research results (the measurement of metadata relevance) are not biased, for a particular learning topic or a particular set of LOs.

Moreover, for this script was used a logic of assignment of topics to the subjects in a balanced way: this logic ensures that 1/3 of subjects was assigned to each learning topic, that is, 1/3 of subjects was assigned to *Cardiovascular Human System*, 1/3 to *Geoprocessing* and 1/3 to *Civil Procedural Law*. In this way, no learning topic received more research evaluations than others, that is, all different learning topics were evaluated equivalently by the subjects.

Some metadata categories were not presented (*LifeCycle*, *Classification*, and *Meta-metadata*) because they contain information that is not so relevant to the learners when they choose which LO to use in their learning. For instance, the *LifeCycle*

Your Opinion	Field + Meaning	Text with Figures	Interactive Map
	Title name given to this learning material	Cardiovascular System	Cardiovascular System - Interactive Map
	Aggregation Level the functional granularity of this learning material: • **level 1**: the smallest level of aggregation, e.g., one file. • **level 2**: a collection of level 1 learning materials, e.g., a lesson. • **level 3**: a collection of level 2 learning materials, e.g., a course. • **level 4**: the largest level of granularity, e.g., a set of courses that lead to a certificate.	level 1	level 1
	Description a textual description of the content of this learning material	This material shows the Human Cardiovascular System. It is divided into 2 parts: In the Anatomy part, it describes in detail the components of the cardiovascular system: blood, heart and blood vessels. In the part of Physiology, it describes the functioning of the system and interaction with other systems of the human body. The text consists of 5 pages and has beautiful color illustrations and photos in high definition.	This material presents the cardiovascular system in an interactive way. The student can choose 2 options: By clicking on the "Images" option, when moving with the mouse pointer on certain parts of the model figure of the cardiovascular system, texts and photos of the indicated place are shown. By clicking the "Videos" option, when you move with the mouse pointer over certain part of the model figure, 3D video of the indicated location is displayed.
	Keywords a keyword or phrase describing the topic of this learning material	cardiovascular system, anatomy, physiology	circulatory system, heart, blood vessels

Figure 2: The main part of the body of the questionnaire.

category contains metadata about the LO's lifecycle (e.g., "*version*", "*status*", whose values are *draft*, *final*, *revised*, *unavailable*). This kind of metadata is more relevant to other types of users, such as instructional designers.

3.2 Research Subjects

After implementing the online questionnaire, a set of e-mails of university students were obtained with professors, from diversified courses of different universities. The invitation e-mail to participate in the research was then sent directly to the students.

The e-mail was used not only to invite people to participate in the research but also to ensure they were university students and to make an automated check so that one person did not participate than once.

3.3 Ranking the Most Relevant LOM Metadata

The questionnaire was available for seven days for people to respond. In the end, 87 students voluntarily participated out of 900 invited students. From the resulting data, the ranking of the most relevant metadata was created (see Figure 3).

This ranking indicates that “description” is the most important information. Among the ten most relevant metadata fields we can see that users are interested in the price of the object (i.e., if it is paid or free), on technical information (usage and installation requirements), and on educational information such as typical learning time. Among the ten least relevant metadata fields we can perceive the interactivity level and the aggregation level. The description of each element is available on the IEEE LOM standard (IEEE Learning Technology Standards Committee, 2002).

4 CASE STUDY

The ranking of the most relevant metadata from the IEEE LOM standard for university students was used to assemble the screen where LOs are listed to users in the AdaptWeb® e-learning environment. In Adaptweb®, each course is divided into topics. Each topic can have dozens of LOs that the user can select and use to learn the topic. LOs consist of video lessons, multimedia presentations, simulators, tools for cooperative learning, for self-assessment, etc. These LOs come from a repository integrated into the system.

Also, Adaptweb® has a LO recommender system that provides the student with a personalized list of recommended LOs. Over this list, the user can select "how to learn", i.e., which LO she will use to learn the current topic. Therefore, on the list of recommended LOs, the user makes a finer filtering on which LO will use, using metadata.

Figure 4 shows this screen where the metadata is presented to the user. On it, we can check that the user is attending an online web course of UML diagrams, and she is currently learning the Time Diagram. On the left side of the screen is the list of LOs available to her to learn this topic - with 17 objects (only the first five appear in the figure). This listing is personalized to each user; it is generated using a LO recommender system. When the user marks a LO in this listing, through the checkbox, the metadata is displayed on the right side of the screen. As a matter of screen space, only the top 14 most relevant LO metadata from the ranking are displayed. If all metadata from the IEEE LOM standard were displayed, the user would suffer from the issue of metadata overload.

Up to three LOs can be marked at a time in the LO’s list to compare LOs through metadata. In this comparison, metadata from different LOs are available, side by side, which facilitates comparison. In this way, the user makes a finer filtering of which LOs to use over the set of LOs defined by the recommender system. This selection process performed by the user has to do with the “how to learn” dimension and takes into account the user knowledge about the future and about probabilistic situations, which are usually not taken into consideration by recommender and information retrieval systems.

A class containing 30 students attended this online course of UML interaction diagrams over the AdaptWeb® e-learning system at the end of 2016. After the course, an online satisfaction survey was conducted among these users. They were university students (undergraduate level) from two courses, Computer Science and Computer Engineering, at Federal University of Rio Grande do Sul, with ages between 18 and 29 years old. This survey has two open-ended questions (openly ask the opinion). The advantage of this type of survey questions, over closed-ended questions, is that subjects can respond to the questions exactly as how they would like to answer them, it is, they do not only choose among generic response alternatives (Reja et al., 2003).

The first question was technical: “Give us your opinion about the set of LO metadata displayed, i.e., about the set of information shown concerning each digital learning material”. In brief, users reported that they find it useful to access different types of metadata beyond general metadata (usually *title*, *description*, and *file format* only). Some students commented that they could better plan their learning activity with information from metadata, for instance, the field *educational.typical_learning_time* that presents the typical time it takes to work with or through the LO. Moreover, students commented they use metadata to make a finer filtering over the set of recommended LOs. One user commented that “in one topic the system chose good LOs for me, but I chose those LOs that taught the content from a general point of view and then it went into detailing the parts, not the inverse”. Finally, from the 30 subjects only three complained that there was too much information about LOs, that is, the vast majority of subjects did not suffer from the metadata information overload.

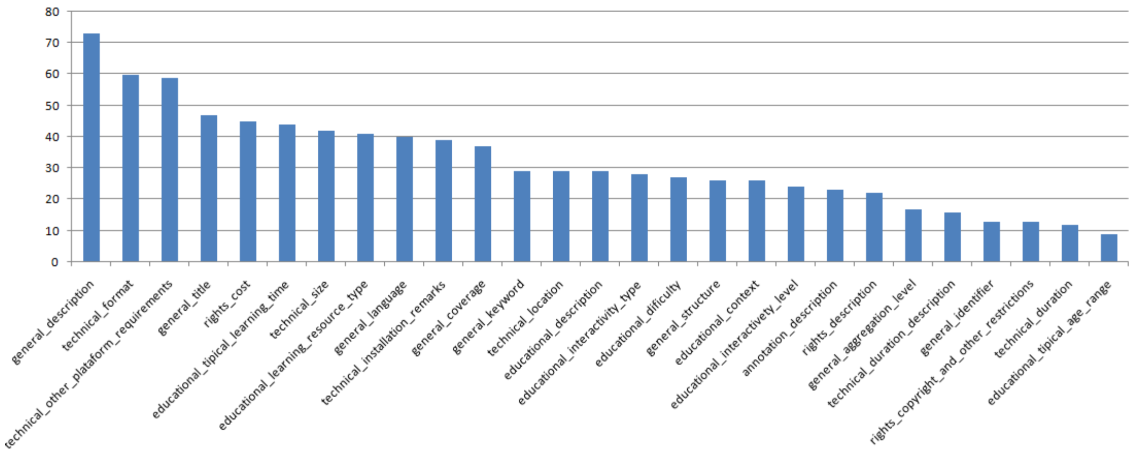

Figure 3: The ranking of the most relevant metadata from the IEEE LOM standard for university students.

The second question was pedagogical: "Give us your opinion on the possibility of choosing learning objects during the learning activity". Summarizing, users reported that they like to choose how to learn each topic, and with that, they felt more motivated to learn. One user even commented that "enabling students to choose learning objects empowers them to conduct a personalized study, and they can progress in class at their own pace and in the way they prefer." Once users can better choose items (LOs, in this case) according to their preferences, the recommender system can better learn their preferences; consequently, the recommender system can improve its items/rating prediction accuracy (Zhao and Shen, 2016).

5 RELATED WORK

Beeson (2006) presented the term "metadata overload" as a challenge in the Information Age. It presents the three major causes for the explosive spread of metadata: ease of publishing documents, dissolution of documents in small pieces and the drive to machine processing of documents. This work focuses on metadata of digital documents mainly on Web, but it is also related to enterprise information management.

Over time this problem began to be perceived in other areas. For example, Kelby and Nelson (2006) describe metadata overload on images, Happel (2008) describes metadata overload on social media systems and Yang, Huang and Hsu (2010) describe metadata overload related to data replication on grid environments. Moreover, in the last years, in the field of data science, metadata is gaining more attention when viewed as *big metadata* (Zhao et al., 2014; Smith et al., 2014; Greenberg and Kroeger, 2017).

For the best of our knowledge, metadata overload about LOs and relevance of LO metadata for end-users in e-learning has never been studied in the literature so far.

In terms of decision-making process, it has been addressed in works of different fields, such as Psychology, Administration, and Economics. In the Computer Science field, Jameson et al. (2015) present a work on Human Decision Making and Recommender Systems. It addresses recommender systems as tools for helping people to make better choices — not large, complex choices, such as where to build a new airport, but the small — to medium-sized choices that people make every day: what products to buy, what documents to read, which people to contact. In this context, a recommender system can keep the chooser (the user) in the loop: arriving at a choice is, in general, best seen as involving collaboration between the chooser and the recommender system. One of the ways in which a recommender system can keep the chooser in this loop takes over only a part of the processing that is required to make a choice, leaving the rest to the chooser. For example, many recommenders use their algorithms to reduce a very large number of options to a smaller subset but then leave it to the chooser to select an option from the subset.

To understand the choice process, it presents an overview of the ASPECT and the ARCADE model. The former distinguishes six human choice patterns (and its combinations). The latter provides a high-level overview of strategies for helping people make better choices. They discuss how recommender systems can make use of these patterns and strategies to support aspects of human choice. One

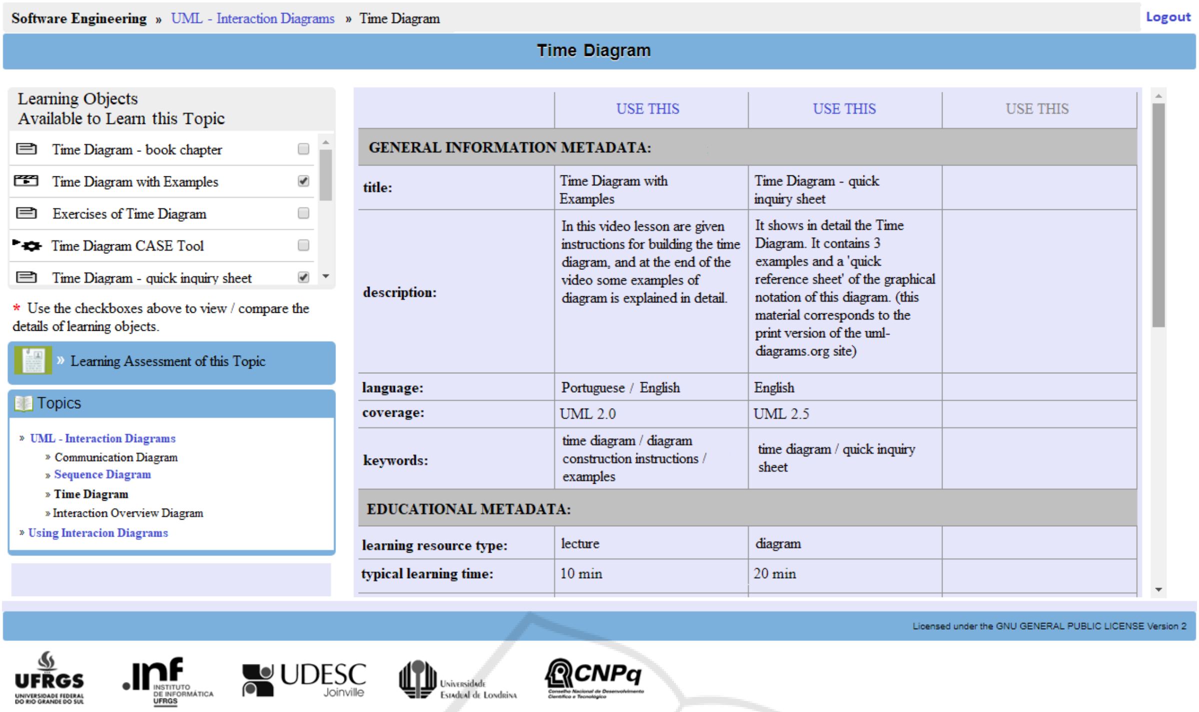

Figure 4: Screen of the AdaptWeb® e-learning system with the list of LOs to learn the topic Time Diagram showing metadata of 2 LOs.

of these patterns is the Attribute-based Choice. According to this, the options can be viewed meaningfully as items that can be described in terms of attributes and levels (item's metadata). And the (relative) desirability of an item can be estimated in terms of evaluations of its levels of various attributes. Then, the typical procedure is: the chooser reduces the total set of options to a smaller consideration set on the basis of attribute information, then he/she chooses from a manageable set of options.

6 CONCLUSIONS

E-learning systems enable students to participate more actively in the construction of their knowledge. Users can learn in a self-directed way, making decisions regarding their learning depending on the possibilities provided by the system. One of the most important choices is "how to learn", which in this work corresponds to which LO someone will use to learn. For this, the user, considering a list of relevant LOs uses metadata to make the final decision. The problem is that the current metadata standards have many types of information and, so, the user suffers from the metadata information overload.

In this work, a study was performed to rank the most relevant metadata from the IEEE LOM standard. The goal was to decrease the amount of metadata available and, thus, prevent students from being overloaded with metadata information. This process takes into account user information that are not usually taken into consideration by recommender and information retrieval systems.

A case study was presented to show the application of the developed ranking of the most relevant metadata of IEEE LOM for university students on the AdaptWeb® e-learning system. After a course in this system, an online satisfaction survey was conducted among their participants. There were 30 subjects. This survey was based on open-ended questions and showed that only three subjects complained that there was too much information about LOs, that is, the vast majority of subjects did not suffer from the metadata information overload.

Regarding limitations, it is essential to state that the study does not cover users from non-formal learning environments, students with disabilities or e-learning systems with open-corpus. However, this is the first research about the assessment of LO' metadata relevance. It can be used as a baseline to evaluate future approaches.

ACKNOWLEDGEMENTS

This work is partially supported by CNPq (Brazilian Council for Scientific and Technological Development), FAPERGS, and CAPES.

REFERENCES

Alexander, S., Kernohan, G. & McCullagh, P. (2004) Self Directed and Lifelong Learning. *Global Health Informatics Education - Studies in Health Technology and Informatics*, (109), 152-166.

Advanced Distributed Learning Network (2004) *SCORM 4th Edition, Version 1.1, Overview*. [Online] http://adlnet.gov/adl-research/scorm/scorm-2004-4th-edition [Accessed 27th November 2016].

Beeson, I. (2006) *Metadata Overload*. Higher Education Academy. [Online] http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.210.7290 [Accessed 7th September 2017].

Ciloglugil, B. & Inceoglu, M. (2016) Ontology Usage in E-Learning Systems Focusing on Metadata Modeling of Learning Objects. *Proceedings of the 3rd International Conference on New Trends in Education* (ICNTE 2016), pp. 80-95, At Izmir, Turkey, 2016.

IEEE Learning Technology Standards Committee (2002) *Draft Standard for Learning Object Metadata*. [Online] Available from: http://grouper.ieee.org/groups/ltsc/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf [Accessed 14th June 2016].

IEEE Learning Technology Standards Committee (2016) *Systems Interoperability in Education and Training*, [Online] Available from: https://ieee-sa.imeetcentral.com/ltsc/ [Accessed 16th December 2016].

Gasparini, I., Pimenta, M. S. & Oliveira, J. P. M. (2013) How to Apply Context-Awareness in an Adaptive e-Learning Environment to Improve Personalization Capabilities? *Proceedings of 30th International Conference of the Chilean Computer Science Society*.

Greenberg, J. & Kroeger, A. B. (2017) Big Metadata, Smart Metadata, and Metadata Capital: Toward Greater Synergy Between Data Science and Metadata. Expert Review. *Journal of Data and Information Science*. 2. 2017-2036. 10.1515/jdis-2017-0012.

Happel, H. (2008) Growing the Semantic Web with Inverse Semantic Search. *Proceedings of the 1st Workshop on Incentives for the Semantic Web* – INSEMTIVE'08. Karlsruhe, Germany, 2008.

Jameson, A., Willemsen, M., Felfernig, A., Gemmis, M., Lops, P., Semeraro, G. & Chen, L. (2015) Human Decision Making and Recommender Systems. In F. Ricci, L. Rokach, B. Shapira (Eds.), *Recommender systems handbook (2nd edition)*. Springer, pp.611-648.

Kelby, S. & Nelson, F. (2006) *Photoshop CS2 Killer Tips* (first edition). pp.196. USA, New Riders Publishing.

Reja, U., Manfreda, K. L., Hlebec, V. & Vehovar, V. (2003) Open-ended vs. Close-ended Questions in Web Questionnaires, *Developments in Applied Statistics*, (19), 159-77.

Smith, K., Seligman, L., Rosenthal, A., Kurcz, C., Greer, M., Macheret, C. & Eckstein, A. (2014) Big metadata: The need for principled metadata management in big data ecosystems. In *Proceedings of Workshop on Data Analytics in the Cloud* (pp. 1–4). New York: ACM.

Vaccari, R., Gluz, J., Passerino, L. M., Santos, E., Primo, T., Rossi, L., Bordignon, A., Behar, P., Filho, R. & Roesler, V. (2010) The OBAA Proposal for Learning Objects Supported by Agents. *Proceedings of MASEIE Workshop* – AAMAS 2010. Toronto, Canada, 2010.

Watkins, C., Carnell, E. & Lodge, C. (2007) *Effective Learning in Classrooms*. London, Paul C. Publishing.

Yang, C., Huang, C., Chen, T. & Hsu, C. (2010) A Dynamic File Maintenance Scheme with Bayesian Network for Data Grids. *Proceedings of the Second Russia-Taiwan Symposium Methods and Tools of Parallel Programming Multicomputers (MTTP), 2010.*

Zhao Y. & Shen B. (1999) Empirical Study of User Preferences Based on Rating Data of Movies. Braunstein LA, ed. PLoS ONE. 2016;11(1).

Zhao, X., Ma, H., Zhang, H., Tang, Y. & Fu, G. (2014) Metadata extraction and correction for large-scale traffic surveillance videos. 2014 *IEEE International Conference on Big Data* (pp. 412–420). Washington, DC: IEEE Computer Society Press.