Learning Object Metadata and Automatic Processes: Issues and Perspectives

Olivier Motelet, Nelson Baloian, José A. Pino

{omotelet, nbaloian, jpino}@dcc.uchile.cl
DCC - Universidad de Chile

Learning Objectives

Generation of learning object metadata

- Understanding the issues related with the generation of learning object metadata.
- Identifying the opportunities and drawbacks of using automatic techniques for generating learning object metadata.

Validation of learning object metadata

- Understanding the validation of learning object metadata.
- Identifying the opportunities and drawbacks of automatic techniques for validating learning object metadata.

Learning object retrieval

- Understanding the retrieval of learning objects using learning object metadata.
- Identifying the opportunities and drawbacks of automatic query generation for enabling semantically rich retrieval of learning objects.

Use of learning object metadata

- Visualizing the impact of automatic processes on the practical use of learning object metadata.

Executive Summary

Reuse of learning material has recently become a leitmotiv for research on computer-aided education. The most obvious motivation is the economic interest of reusing learning material instead of repeatedly authoring it. Other motivations can be found in the pedagogical area since learner-centric teaching theories invite instructors to use a wide variety of didactic material. Since sharing and retrieving learning material is a basic requirement to ease learning material reuse, it is not surprising to see the research community specially focusing on these topics. Learning material retrieval should not only stand on common document characteristics - like DublinCore Metadata Intitiative (DCMI, 2005) - in order to be pedagogically relevant, but also on specific educational data. The Learning Object Metadata (LOM) standard includes such data. Consequently, Learning Object Repositories typically use this metadata for storage and retrieval of learning objects. However, creating a LOM document means to instantiate the almost 60 metadata attributes of the IEEE LTSC LOM specification (LOM, 2005). Such a fastidious task is not compatible with making learning material sharing a customary activity for regular teachers. Therefore, several researchers seriously focus on the metadata generation issue (Downes, 2004; Duval et al., 2004; Simon et al., 2004). The main difficulty for generating LOM documents stands in the educational part of the metadata. Educational information is generally implicit in the learning objects. Therefore, existing metadata extraction methods based on content analysis cannot totally serve LOM generation. The same situation occurs with learning material retrieval. Retrieval effectiveness depends on LOM-based query precision. Consequently, generating effective queries could rapidly be as complex as generating LOM documents. In such a context, the future of learning object repositories will definitely depend on the ability of current systems to facilitate the generation of metadata as well as queries on these metadata. This chapter deals with the existing methods for facilitating the generation of LOM documents and LOM-based queries. First, LOM generation methods are discussed. The efficiency of metadata generation methods based on content analysis will be bounded with the specific characteristics of LOM. Then, alternative methods are presented. In particular, context-based methods for supporting LOM generation are discussed. Second, LOM-based query generation is reviewed. Visual approaches and recommender systems are discussed. A special focus is placed on context-based generation methods. Finally, the impact of these technologies on the practical use of LOM is discussed.

1. Introduction

One of the main motivations for having Learning Objects and Learning Object Repositories is to allow educational resources to be reused by as many people as possible (Wiley, 2000). In order to make this possible, the characteristics of the learning objects should be exposed, so that others than their authors could locate and retrieve them. A very critical issue in this process is how to describe the educational resource and how to search for it in order to find those who really match the needs of certain potential users.

Document retrieval is largely dominated today by search engines like Google (Google, 2005). Basically, these systems automatically index documents by analyzing their content. The algorithms used for analyzing and indexing documents are so efficient that they make possible the retrieval of documents among the billions pages of the Web in a few milliseconds. This fact seems to make indexing engines the ideal tool for supporting learning objects retrieval. However, there are many factors that make these search engines inapplicable while retrieving learning material. First, learning material is not only text-based; it may be a video, a picture, or an audio file, a simulation program or a file with any other structure which is difficult to efficiently mine with content analysis methods. Second, educational resources could be copyrighted and their access restricted. In this case, documents cannot be indexed since their content is not available. Third, the efficiency of the retrieval process for educational resources mainly deals with information about its usage like its educational context, its pedagogical methodology, or/and its interactivity. Nevertheless, this kind of data is generally not contained in the document itself. Google-like tools cannot consider such implicit information. An alternative method to content analysis and indexing engines is necessary.

The usage of metadata for describing the educational contribution of a learning object has been a widely accepted approach for classifying educational material. Metadata is data about data. Typically, the metadata of a document describes its relationships with other documents, its content, its life cycle, its technical characteristics, its potential usage, and its expected position inside higher-level processes. Kim (2005) defines *legitimate metadata* as data which "helps the user to understand the semantics and lineage of stored data and to properly run the applications in support of the business needs". Metadata has the advantage to remain independent from the format and accessibility of the learning material to which it relates. Moreover, it is able to store information not present in the described document. With such features, the problems, which arise when trying to use indexing engines to find educational material, may be solved with metadata.

Metadata for educational resources aims at effectively facilitating the retrieval and reuse of learning material. However, some standard metadata specification schemas like basic Dublin Core (DC) (DCMI, 2005) may not completely satisfy this requirement. Although DC attributes holding content data such as authors, title, or granularity are definitely useful for describing educational resource content, DC does not contain any field for describing the pedagogical perspective of a document. In order to cope with educational concerns, various metadata sets were defined such as DC Educational extension (Edu, 2005), CANCORE (Cancore, 2005), IMS Metadata (IMS, 2005), SCORM Metadata (SCORM, 2005), and IEEE Learning Object Metadata (LOM) (LOM, 2005). While LOM is considered as a standard reference, sometimes it does not meet all the requirements of communities of practice requiring local extensions or modifications. Thus, the term of Application Profile for LOM (or LOM Profile) has emerged as a means of describing this practical reality. Application profile (the way the standard is applied to specific requirements) may therefore include local elements or even elements taken from other metadata standards.

Most metadata specification schemas for educational use can be seen as LOM Profiles in the sense that they just customize the LOM standard for certain purposes. For instance, the metadata definitions used by the SCORM specification (SCORM, 2005) or the CANCORE repository (Cancore, 2005) may be considered as LOM Profiles since they have few differences between them and it is always possible to transform a set of SCORM or CANCORE metadata to the format specified by IEEE LOM. That is why we will refer to the IEEE LOM metadata specification when mentioning metadata for educational resources without losing generality in this chapter.

Officially, the LTSC workgroup defined the LOM standard for educational resource metadata in order to enable the retrieval of learning objects, define an interoperable, standard, extensible and secure language for their metadata, enhance the modularity, computability and ability to be composed of learning objects, permit the sharing of educational methods, and enable reuse economics of learning contents. Practice shows that these ambitious goals are not completely achieved yet. Nevertheless, this metadata specification remains a fundamental work used in most present efforts for enabling relevant retrieval of learning material and promoting reuse.

The LOM specification contains almost 60 attributes. These elements cover various aspects of the learning material including data about general content, life cycle, meta-metadata, technical characteristics, educational usage, rights, relations, annotation, and classification. Instantiating the metadata may be a straightforward task for attributes like author, title, and language. They can even be automatically generated. However, other elements are less obvious to instantiate and often may require a serious evaluation effort from the user defining their values (Hodgins, 2001; Kabel et al., 2003). This is the case with most educational attributes. A recent survey about metadata usage (Friesen, 2004) confirms that metadata instantiation is a difficult task which is generally avoided. This conclusion may indicate present metadata creators have a technical bias, which make them prefer the technical aspects of the learning objects and also to use the corresponding metadata vocabulary in an effective way. As a conclusion, we can say that correct instantiation of metadata for learning objects requires too much time and combined educational and technical skills. This may be too costly to be practical for all educational institutions. Therefore, metadata creation needs to be simplified: technical aspects should be interfaced and vocabulary should suit the user terminology. In this regard, many works are focusing on automatically generating the learning object metadata (Downes, 2004; Duval et al., 2004; Simon et al., 2004).

Automatic generation has an obvious relevant advantage: the economy of work for not having to “manually” create the metadata, apart from standardizing the value of the metadata. Nevertheless, it has also a serious weakness related to the quality of the generated metadata. Metadata is useful not only when it is complete and accessible but also when it is correct. However, the correctness of some metadata values, especially those related to educational characteristics is very difficult to verify, because such information deals with the intended usage of the material, which remains implicit in the material itself.

The required level of validity of the learning object metadata has to suit the actual usage of this metadata. Nonetheless, using metadata for retrieving learning material involves the specification of value range for each element of the specification. Such task may be still more tedious than creating metadata for existing material. Recent usability studies (Najjar et al., 2005) shows that searching interfaces available for some learning repositories are not satisfying. In particular, interfaces using the metadata fields in order to generate queries are poorly used: most people are lost in the complex structure of the metadata and the specific vocabulary required in the fields. These conclusions are similar to Friesen’s conclusions concerning LOM creation: the metadata creation issue leading to focus on automatic generation is replicated in the area of retrieval. Consequently, similar efforts are done in order to facilitate the generation of queries for retrieving learning material (Pinkwart et al., 2004; Motelet and Baloian, 2005; Najjar et al., 2005).

This chapter attempts to clarify the present situation about generation, validation and retrieval of metadata for educational resources. First metadata generation is explored. In particular, this review focuses on methods for automatically generating metadata. Then, metadata validation is introduced as a necessary process to guarantee the completeness, correctness and accessibility of educational resource metadata. Next, methods using metadata in order to retrieve learning objects are reviewed. Finally, perspectives for taking advantage of the metadata in order to share and reuse learning material are discussed.

2. Generation of metadata for educational resources

2.1. *Creation of metadata: people and tools*

Greenberg (2003) distinguishes four people roles, which may be involved with learning instantiation: professional metadata creator, technical metadata creators, content creator, and community enthusiasts. Applied to the educational field, these roles can be described as follows:

- Professional metadata creators are third party metadata creators because they produce metadata content created by individuals. Typically, they are specialized catalogers, indexers or pedagogical engineers. They have sufficient pedagogical, domain and technical skills “to make sophisticated interpretative metadata-relative decisions” (Greenberg 2003). They work with classification systems such as vocabulary and ontology, LOM Profile schemas and LOM standard.

- Technical metadata creators include paraprofessionals (e.g., library assistants, or planning assistants in educational institutions) who have had metadata training but their hiring costs are lower than those of metadata professionals. They are generally responsible for instantiating simple metadata such as authors, title, format, or creation date.
- Content creators are typically instructors developing the intellectual content of the educational resource. Nevertheless, they generally have no special training on LOM Profile, vocabulary and ontology. Facilitating the creation of LOM by content creators may be less expensive than hiring metadata professionals.
- Community enthusiasts are typically instructors with special knowledge and interest in the shared educational resources. They typically have no formal LOM creation training, but they may participate in the completion of the metadata. Collaborative creation has of course many economical advantages but the coordination required is often difficult to achieve.

Greenberg also distinguishes three types of agents for supporting creation of metadata:

- Human beings,
- Standards and documentation – for instance, the LOM specification, a LOM Profile including recommendations for the instantiation, a LOM vocabulary.
- Tools – technical support for capturing and storing LOM in its conformant XML format. Tools for supporting LOM creation are templates, editors and generators.

Templates are sometimes used for creating LOM. They consist of a void instance of XML code including all the LOM attributes. However, they are not very frequently used. In fact, LOM users are rather familiar with higher-level tools such as editors. Editors are typical examples of those forms available in most learning object repositories. The Reload editor (Reload, 2005) is a very sophisticated standalone platform for creating metadata. This editor has the advantage of being compatible with various versions of LOM and it offers access to vocabulary terms with combo-lists available for each attribute needed for a vocabulary instance. The last type of tools supporting creation of metadata deals with generators for automatic production of metadata. This category is further explored below.

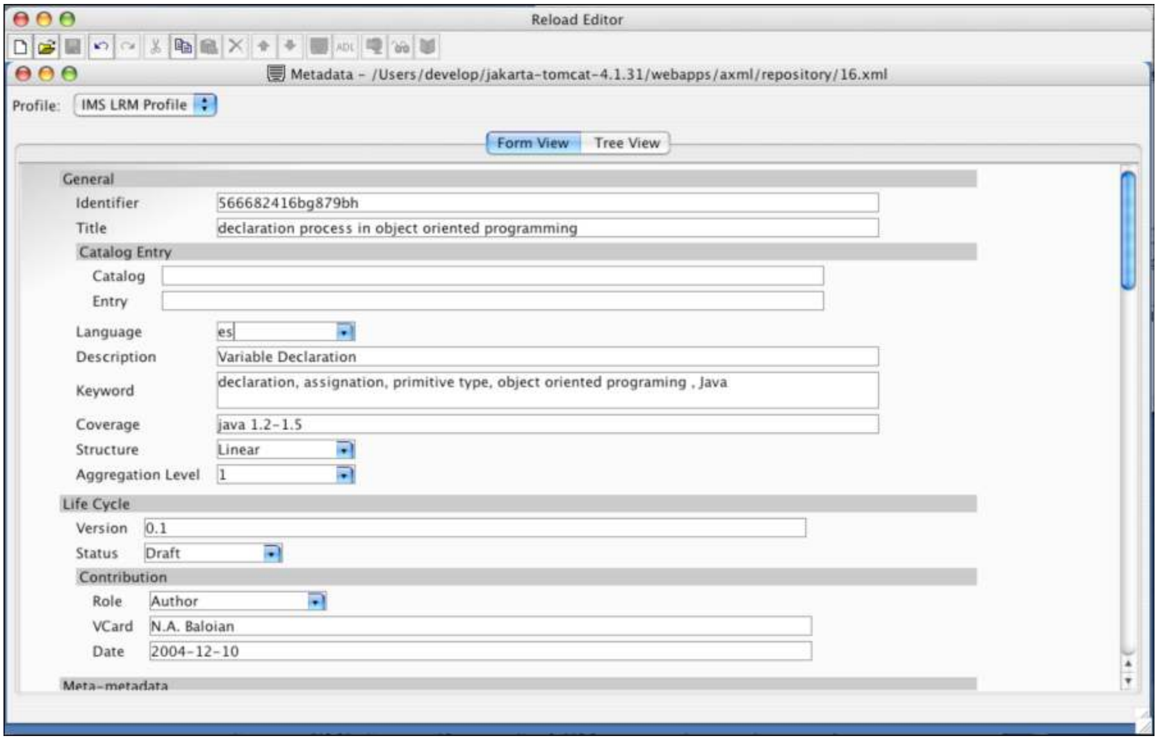

Figure 1. Interface of the Reload, multi-format editor of learning object metadata.

2.2. Information sources for generating metadata

A generator for automatic production of educational metadata needs information sources in order to deduce metadata values. We distinguish two typical sources of information:

- *Document contents,* i.e. the educational resource itself and its technical characteristics,
- *Document context,* i.e. the learning environment, the resource usage, the environment in which it is used (e.g., is it part of a course? and if yes, which are the characteristics of this course), and the related material.

Documents content may deal with structured documents, free-text documents, images, sound, video, animations, simulations, or hybrid media. The difficulty of using the content as a source for generating the metadata stands mainly in the fact that not all media can be easily analyzed. Technical characteristics of a document are also part of this source. They are the format, the size, the modification date and the access rights.

Document context presents other opportunities for generating metadata. Typically, the usage of a certain learning resource reveals the learning time, the level of interactivity, or the density of the learning material. The learning environment also gives information on the profile of the learner who uses it, the used language, or the necessary material. The lesson or learning unit in which a learning object is being (re)used and the related material may also give some interesting information about the pedagogical context of the learning material. Nevertheless, the document context is not always computable. Electronic learning environment such as most Learning Management Systems attempt to rationalize and trace this information as much as possible. Standards like SCORM (2005) and IMS (2005) are initiatives that contribute by giving information about the context in which the learning object is intended to be used. Moreover, they are especially intended to define uniform interfaces for the electronic learning environments.

2.3. Generation techniques based on content analysis

In this subsection, various metadata generation techniques based on document content analysis are reviewed. Methods using LOM as well as methods using other types of metadata definition schema are presented mentioning their advantages and drawbacks. The generic approaches and the semantic based approaches are presented separately.

2.3.1. Generic approaches

Search engines like Google are based on document content analysis. For that purpose, they usually extract a so-called standard logical view from the documents. The most used logical view for documents in search engines is the "bag of words" model, in which each document is seen only as an unordered set of words. In modern Web search engines, this view is extended with extra information concerning word frequencies and text formatting attributes, as well as meta-information about Web pages including embedded descriptions and explicit keywords in the HTML markup. There are several text normalization operations that are executed for extracting keywords. The most used ones are: *tokenization* (dividing stream of text into words), *stopword* removal (removing functional words without semantic information) and *stemming* (extracting the morphological root of every word). Those three methods are more efficient working with languages similar to English. For example, tokenization does not work with languages like Chinese in which word decomposition is not obvious, and stemming is very difficult with Arabic languages.

After text normalization, keyword ranking is done. The most used methods for text ranking are: *Vector Space Models* (representing natural language documents in a formal manner by the use of vectors in a multi-dimensional space), and *Naïve Bayes Classifiers* (based on probability models incorporating strong independence assumptions which often have no bearing in reality). Some libraries implementing such methods are, e.g., Classifier4J (C4J, 2005), which enables keyword extraction and summarization from a text-based document, or KEA (2005) for automatic key phrase extraction. Generic content analysis may be useful in order to generate keywords or description. More specific characteristics like title, author, size, creation date, etc. are generally retrieved from already existing metadata embedded in the educational resource itself. This method is used by DC.dot (2005), a web application generating DublinCore (DCMI, 2005) metadata for a web resource. This system mainly works on retrieving text embedded by tags targeting DC-related information.

2.3.2. Approaches based on specific domain semantics

Metadata contains semantic information about the documented resources. Consequently, using the domain semantic for generating metadata may offer interesting opportunities.

Concerning document classification, Jenkins et al. (Jenkins et al., 1999) describe a system generating DC-like metadata by matching correspondences with a classification hierarchy (in this case the Dewey Decimal Classification System). Stuckenschmidt and Harmelen (2001) also propose semantic-based classification by referring to an ontology. Some available commercial tools enable semantic classification of documents as well. The Klarity API (Klarity, 2005) and Metatagger (Interwoven, 2005), e.g., enable such a classification process based on a user-defined ontology.

Yilmazel et al. (2004) introduce natural language processing methods in order to extract pedagogical information from an educational resource. They succeed in efficiently generating DC metadata and GEM (Gateway to Educational Material) Metadata, which are quite similar to the educational category of LOM. However, their system focuses on a specific type of educational resource since it deals with lesson outlines.

2.4. Generation based on context analysis

Metadata generation techniques using document context analysis are discussed. As above, we present generic approaches first and then the semantic based approaches.

2.4.1. Generic approaches

There are only few generic approaches for generating metadata based on the analysis of the context, being the work of Marchiori (1998) the most interesting one. He suggests a propagation system based on simple fuzzy logic. Basically, metadata values are propagated from one document to another one, where a hyperlink between the two documents exists. A weight variable is introduced to each metadata value. At each diffusion step, this weight is decreased. Since a document can be related to various elements of the graph, the propagation system may generate multiple values for the same metadata. In these cases, a composition is made in which the value of the metadata with the maximum weight is kept. This approach does not consider the semantics of the elements, but it obtains exceptionally good performance. The propagation method drastically increases the mass of documents with instantiated metadata whereas the produced metadata values remain acceptably correct. However, these results are valid in systems using ontologies based on one category, since fuzzy logic is not efficient with various, possibly ambiguously overlapping categories. If these ambiguities are carefully identified for the case of metadata for educational resources, it may be very interesting to investigate the application of this method to the elements of learning object repositories.

2.4.2. Approaches based on domain semantics

Several research works focus on using domain semantics for context-analysis-based generation of metadata. In order to enable metadata generation, these tools process sets of rules concerning the semantics of (1) the other metadata of the resource, (2) the associated learning environment, and (3) the related material.

Lattner and Gehrke (2004) propose to use Inductive Logic Programming in order to create sets of rules relating the semantics of different elements of the same metadata. Then, they suggest to apply them in order to generate missing values of metadata using the values of the instantiated elements. As learned rules might not be completely correct or they might miss values for attributes of objects, they provide a rule relaxation algorithm while applying learned rules. This algorithm permits to systematically deduce metadata. However, this weakens the overall quality of the generated metadata. Thus, the authors argue that this feature has to be used in the scope of post-verification of metadata. There are no available results of experiments using this methodology on metadata for educational resources. The commercial tool Metatagger (Interwoven, 2005) enables the processing of similar rules in order to generate metadata values. Nevertheless, the used rules are not induced from system usage but user-defined.

Learning Management Systems are also semantically rich source of information for generating metadata. Ochoa et al. (Ochoa et al., 2005) suggest deducing metadata values from the information available in the LMS. Metadata like author-related information and educational context (course level, area, prerequisites, student level, etc.) may be generated. They realized two implementations for specific LMS (BlackBoard (2005) and SIDWeb (2005)). Hatala and Richards (2003) propose a similar approach but they base their implementation on the standard SCORM and IMS Packaging. Since this system uses standards instead of specific implementations, it could be easily applied on most learning management systems.

Another method for generating educational metadata is based on using semantics of the metadata of the related material. Hatala and Richards (Hatala and Richards, 2003) also develop this approach in their system. In particular, a set of specific rules concerning the inheritance between educational resource (from parent to children), the accumulation (from children to parent), and the content similarity between educational resources are defined. These inference rules may not be valid in all potential uses. This work does not aim at creating metadata values for automatic instantiation, but it provides suggestions for metadata values during the metadata creation process. Motelet and Baloian (Motelet and Baloian, 2005; Motelet, 2005) also suggest using inference rules based on the relations between a certain educational resource and other material. This work is based on the principle that a semantic relation between two educational resources may impose a certain mutual influence between their metadata values. For instance, a relation of type *explainedBy* between two learning objects may imply that the *keywords* of the first are quite similar to the *keywords* of the second. Such a rule enables the deduction of suggestions for creating metadata values. Furthermore, the principle is extended so that rules could also formulate restrictions on metadata values. If two learning objects, e.g., are related with the semantic relation *introducesTo*, then the *semanticDensity* of the first resource is certainly lower to the *semanticDensity* of the second. This assumption may not be valid for all potential users, so they provide a simple domain specific language for defining rules. With this language, each community may be able to define their own rules according to their needs.

2.5. Conclusion

Most metadata providing technical information (e.g., size, format, creation date, duration, etc.) about an educational resource can be generated automatically by the applications used to create them. Approaches for generating metadata using document contents are aimed at extracting information identifying the subject of the educational resource (title, keywords, description, classification). The efficiency of the method may be improved by using domain ontology. However, these techniques are based on text mining but not all educational resources are text-based. Moreover, most educational-related metadata remain implicit in the learning material. Document context offers other opportunities for generating such data. In particular, techniques using the semantics of this context seem to give interesting results also for generating educational metadata. Nonetheless, the correctness of the metadata generated with these techniques remains approximate. Consequently, part of the literature claims that such values should be used for suggestions or validations only.

In a recent study about metadata generation methods, Greenberg (Greenberg, 2004) claims "best metadata generation option is to integrate both human and automatic processes". Using generation methods for suggestion and validation of metadata suits this conclusion. According to this trend, metadata for educational resources should remain instantiated by content authors or professional metadata creators, but this instantiation process should be supported by the suggestions and restrictions of various automatic generation methods. Tools supporting such a process are hybrid engines between editors and generators.

Other research in the field is grounded on the fact that metadata cannot be a human concern anymore. Duval et al. (2004) claim that human-based metadata generation "does not work in real life". Metadata creation is too constraining for content authors that do no see their immediate benefit. Professional metadata creators are considered as too expensive to be employed in most educational institutions. They agree with the fact that automatic generation of metadata does not generate perfect metadata, but they advocate that it could "be good enough for sharing material". This assumption deals with the expected correctness of the metadata for educational resources. In the next section, metadata validation including correctness is further discussed.

3. Validation of metadata for educational resources

Techniques for automatic generation of LOM were presented in the previous chapter. Some possibilities were shown for generating those metadata describing the educational characteristics of the learning resources. This section introduces the validity of the metadata for educational resources issue. The validity of some metadata depends on their completeness, their correctness, and their accessibility. These topics are described in three subsections and some conclusions are presented.

3.1. Completeness of metadata

According to Stuckenschmidt and Harmelen (2004), "in order to provide full access to an educational resource, it has to be ensured that all the information is annotated with the metadata. Otherwise, important or useful parts of an information source may be missed or cannot be indexed correctly". For that reason, LOM completeness is an important topic. This characteristic should be assessed in order to decide the validity of metadata. Basically, it means to check if each metadata element is effectively instantiated.

LOM Profiles provide useful information about the relevance of metadata elements. For instance, in the application profile of CANCORE, the metadata *general/coverage* is not considered whereas it is part of the LOM specification. In this specific context, evaluation of completeness should not take into account this particular metadata element. On the other hand, if the CANCORE resources are retrieved from another context, completeness evaluation may negatively consider this missing metadata.

3.2. Correctness of metadata

Metadata about an educational resource is only useful if it correctly describes its contents and pedagogical contexts. In fact, inconsistent metadata may be a more difficult problem than missing metadata, "because mechanisms relying on metadata will produce wrong results without warning" (Stuckenschmidt and Harmelen, 2004). Therefore, validity of metadata should depend on the evaluation of the correctness of this metadata.

Correctness evaluation is more complex than completeness assessment because it deals with the semantics of the metadata values. A first method consists of checking the validity of the data type. Such test can be done at low level using the XML Schema validation mechanism. Most tools for editing educational metadata (e.g., Reload editor (Reload, 2005)) and most metadata creation forms available in current learning object repositories provide this kind of type checking. Moreover, these devices permit to check whether the given value belongs to a set of vocabulary terms. The Reload editor, e.g., enables to check the vocabulary corresponding to a different version of the IMS Metadata definition. Nevertheless, none of the available editors provides a deep checking of the meaning validity, because this topic has to do with both the semantics of the metadata and the semantics of the educational resource.

Analyzing the semantics of an educational resource means to be able to mine this document. This issue is related with the automatic generation of metadata, which was further discussed in the previous section. Ochoa et al. (Ochoa et al., 2005) suggest a framework for using various automatic metadata generation methods for cross-validating metadata. The idea is to compare the value of a metadata element with the ones provided by several generators. They also propose developing a system based on Bayesian networks in order to fairly attribute weights to the wide range of produced values.

Rule-Processing-based systems may also be relevant for checking the correctness of the metadata. For instance, the system developed by Lattner and Gehrke (2004) applies rules by processing some semantic relationships between the various elements of the metadata of a same resource. This approach may provide interesting information on the consistency of the metadata elements between them. Other systems such as the proposals by Marchiori (Marchiori, 1998) and Motelet and Baloian (Motelet and Baloian, 2005; Motelet, 2005) use semantics of resource graphs in order to deduce suggestions and restrictions for element values. Such techniques may provide relevant data about metadata coherence for a set of educational resources.

None of the available generation methods for metadata values provide sufficiently accurate result to safely validate the correctness of the metadata. However, combining various generation techniques may be an efficient solution in order to build a validation system having a minimum fairness. No research has currently seriously studied that issue and humans (e.g., professional metadata creators/evaluators) remain the most reliable means to verify metadata correctness.

3.3. Metadata Accessibility

Metadata for educational resources has to be accessible for people and applications wanting to use them in order to be useful. Accessible metadata has the following features: (1) it is defined within an interoperable format, (2) it uses an accessible vocabulary, and (3) it is possible to localize. Evaluating the accessibility of metadata for educational resources means to assess the completeness of these three characteristics.

Most XML-based standards such as LOM or IMS Metadata attempt to offer a concrete means to achieve interoperability for educational resource metadata. For being interoperable, metadata should be defined in a format transformable to an existing standard or format, which an end-user system can understand. Consider some metadata M defined in a certain format F1. M is interoperable with another format F2 if it exists a direct transformation function F1>F2 between the two languages *or* a standard format F and two transformation functions F1>F and F>F2 (and their inverse function, of course) in order to use F as a pivot. This simple definition may be used as a proof of the interoperability of a metadata. XML (XML, 2005) is a powerful tool for implementing interoperability of formats. In particular, it permits the use of XSLT, a language for defining transformation sheets between XML documents. Since LOM is based on XML, it is recommended for metadata to be defined in a format based on XML or extensions (e.g., RDF).

Accessibility of metadata also depends on the accessibility of the used vocabulary. A vocabulary is accessible when (1) it is a standard, or (2) a translation table from this vocabulary to a standard vocabulary is available, or (3) a translation table from this vocabulary to the end-user vocabulary is available. Current definitions of LOM Vocabulary are generally based on RDF (RDF, 2005). RDF is an XML-based language enabling the definition of semantic relations between elements. In particular, it permits the definition of ontology, classes or instances of vocabulary. RDF may also be used in order to define a translation table between vocabularies.

Finally, metadata is said to be accessible if it is localizable. Basically, metadata for educational resource is localizable when it is distributed by a medium (e.g., internet or cd-rom). Localizable metadata should also succeed in reaching the intended public (*pull*, e.g., mailing or RSS) *or* be easily available to be found by them (*push*, e.g., electronic libraries, learning object repositories). Accessibility may be improved with advanced visualization techniques of specialized databases (Klerkx et al., 2005).

3.4. Conclusion

Validating metadata for educational resources is a difficult task. Completeness and accessibility are well-defined concepts and their validity is easily computable. However, correctness deals with the semantics of the educational resource. Thus, metadata correctness confronts with the same issues and problems as metadata generation. In particular, both processes face the difficulty of deriving educational information from the educational resource contents. Nonetheless, composition of existing metadata generation techniques seems to be a promising field to be explored in order to increase the possibilities of developing accurate automatic methods for metadata validation.

Metadata correctness evaluation may depend on the subjective expectation of their users. Some authors (Duval et al., 2004; Ochoa et al., 2005) advocate for automatic systems ala Google in which approximation is sufficient to offer great results. However, this assumption is not valid anymore when retrieval expectations concern services not available with Google. Such expectations may deal with collaborative filtering or seamless integration of the retrieval process during authoring of the lesson. Metadata correctness evaluation should be related to the characteristics of the needed retrieval processes. Retrieval strategies for educational resource metadata are reviewed in the next section.

4. Learning Object Retrieval based on Metadata

The challenge of finding appropriate learning objects is a key issue for end-users in learning object repositories. Recent usability studies (Najjar et al., 2005) show that the user search interfaces for the available repositories are not satisfying. In particular, those using the metadata fields in order to generate queries are poorly used: most people are lost in the complex structure and the specific vocabulary required. As a consequence, research about educational retrieval has now three main directions. First, end-users need simple interfaces like the Google one in order to make queries on learning objects. Such a retrieval method may radically simplify the query process for end-users but it might also lack the advantage of metadata being a semantically rich information source. Thus, the second research direction deals with the generation of semantically rich queries. Finally, a third direction deals with the design of comfortable navigation methods in query results.

4.1. Keyword Queries for Learning Object Retrieval

Indexing engines - like Google - index the WWW using text normalization operations like those described in the section about content-based techniques for metadata generation. Elaborated methods of indexes are used when finding query results in very short time. Query interfaces for such systems are based on simple keywords; users enter a sequence of expressions in order to retrieve documents containing them. Then, the resulting documents are ranked considering the order of the keywords and their occurrence in the retrieved contents. Other document characteristics like keywords frequency and related web documents also play an important role in the ranking process.

Most people are now familiar using the *keyword queries* of search engines. Consequently, interfaces implementing such queries have appeared in most learning object repositories. In these systems, the sequence of keywords is processed on one specific element of metadata (e.g., title) or on a limited set of elements (e.g., title, keywords, description). Some research (Najjar et al., 2005) indicates the convenience of searching the occurrences of the expressions given by the users in all the elements of metadata.

In most learning object repositories, keyword queries are limited to the metadata of the resources available in this repository. Therefore, they strongly depend on the validity of these metadata. Sometimes, the incompleteness of available LOM documents is a serious drawback for processing queries. Thus, in some situations, it is still faster to look for an educational resource with an indexing engine like Google than with repositories for educational resources. Metadata and queries on this metadata should not be restricted to data describing the contents, because educational context is an important characteristic to consider when reusing learning material. However, in order to query this kind of data, the semantics of LOM must be considered.

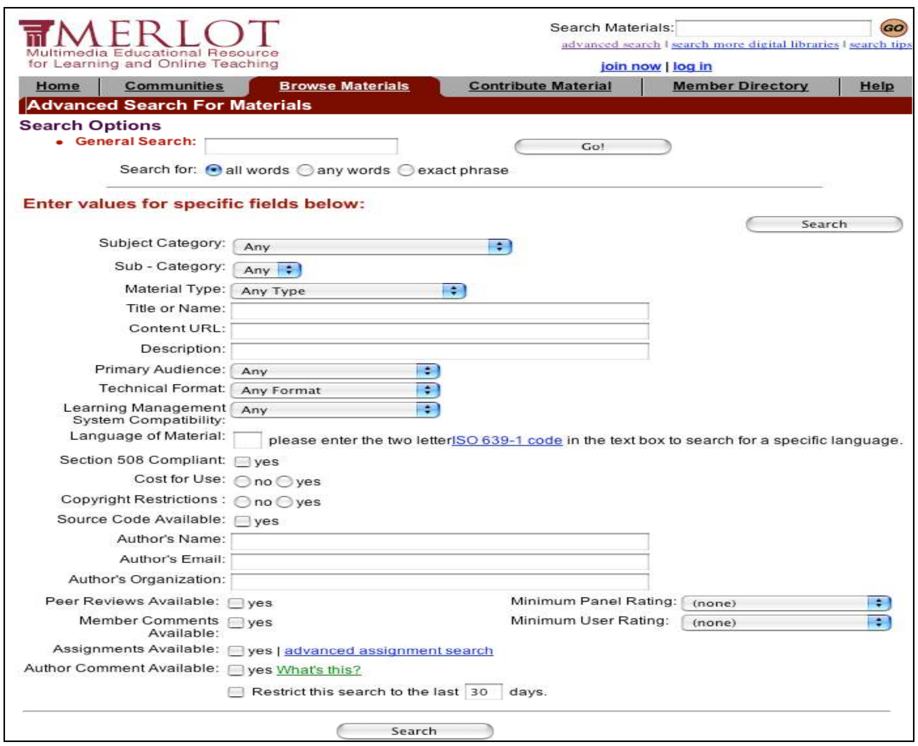

Figure 2. Search interface for the Merlot learning object repository

4.2. *Semantically Rich Queries*

Metadata for educational resources usually contains more information than content description. Such information (e.g., educational usage) is generally implicit in the document and search engines are not able to grasp it. Educational metadata was especially designed for improving the retrieval of learning material. However, users are not comfortable with querying such data. First, most people are used to query document content and not document context. Second, it implies the understanding of the semantics characterizing the context and the knowledge of the associated vocabulary. In order to support users in formulating semantically rich queries, most learning object repositories provide forms. They are restricted version of those used for metadata edition except that they accept the input of restrictions instead of values. Lists of vocabulary along with each metadata element are intended to guide users in formulating their queries (Figure 2). Nevertheless, Najjar et al. (2005) claim that the complex structure of the forms and the wide range and non-familiarity of vocabulary terms hinder the generation of semantically rich queries. This section reviews the available methods supporting users in generating semantically rich queries without falling in the previous problems. First, keyword queries targeting all semantics of metadata for educational resources are discussed. Then, recommender systems are introduced since their purpose is the generation of semantically rich queries. Finally, approaches supporting the user in generating such queries are presented.

4.2.1. Keyword queries targeting semantics of metadata

According to Najjar et al. (2005), keyword queries should target all the elements of learning object metadata. In particular, they may focus on semantically different attributes of the metadata (e.g., keywords, title, and also interactivity type and learning resource type). In keyword queries, targeted attributes are not specified. For instance, a user looking for an educational resource in order to make an interactive simulation in Java about the life of beavers could formulate her request as “active java simulation beaver life”. “Active” stands for the interactivity type, “java simulation” for the learning resource type and “beaver life” for keyword or title. This approach has the advantage that users do not have to design long queries but they just need to summarize their demand. However, it is obvious that for generating such queries users should have sufficient knowledge about the semantics of learning object metadata, because some specific vocabulary (e.g., “active”) is not especially intuitive. Moreover, not specifying the semantic of the query terms can lead to confusions due to possibly overlapping categories. Indeed, the example query “high java simulation beaver life” may match with a simulation, intended to life-long learning, on using ActiveXML (a Java library based on another Java library called Beaver) for pushing information in XML documents. Considering such cases, it seems preferable to limit the use of keyword queries to basic requests. Natural language processing is an interesting tradeoff between keyword queries and semantically rich queries (Woodley and Geva, 2004). However, this approach has not yet been explored in the field of learning object metadata retrieval.

4.2.2. Recommender systems

Recommender systems are basically programs, which try to match a certain document with a user. Thus the main task they accomplish is trying to answer the question: “given a certain document, how interesting or useful may it be for a certain user?”. Recommender systems may be based on content analysis or collaborative filtering. In the first case, the content of a document is usually automatically analyzed to extract its relevant characteristics by the application of different metrics. The characteristics are then compared with those desired by the user. In the second case, these systems base their decisions on opinions previously given by other users. Both types of systems apply user-modeling methodologies. Some techniques, which have been successfully used to accomplish this task are Markov chains (Wong, 2000) and similarity metrics (Savia, 1998).

A variety of collaborative filters or recommender systems have been designed and deployed. The Tapestry system relied on each user to manually identify like-minded users (Goldberg 1992) and it is one of the earliest implementation of collaborative filtering-based recommender systems. However, since this system depends on each person knowing the others, it is not suitable for large user communities. Thereafter, several rating-based automated recommender systems were developed. The work of Breese et al. (Breese, 1998) identifies a general class of Collaborative Filtering algorithms called model-based algorithms. The authors describe and evaluate two probabilistic models, which they term the Bayesian clustering and Bayesian network models. In the first model, like-minded users are clustered together into classes. Given her class membership, a user's ratings are assumed to be independent (i.e., the model structure is that of a Naive Bayes network). The second model also employs Bayesian networks, but of a different form. Variables in the network are titles and their values are the allowable ratings. Bayesian networks create a model based on a training set with a decision tree at each node and edges representing user information. The model can be built off-line on a matter of hours or days. The resulting model is very small, very fast, and essentially as accurate as nearest neighbor methods (Breese, 1998). Other technology that has been used is Horting, that is a graph-based technique in which nodes are users, and edges between nodes indicate degree of similarity between the users (Aggarwal, 99). Walking the graph nearby nodes and combining the opinions of the nearby users produces predictions.

As with any document, recommender systems can be used with learning objects. They would make a difference in locating appropriate learning material especially if they are based on the opinion of other users, which have similar profiles and/or are "culturally" near, thus reducing the subjectivity of interpreting the metadata describing the educational contribution of the object. Baloian et al. (2004) propose a recommender system for learning material based on the characteristics of the metadata as well as the recommendations of other users with similar profiles. The system uses the LOM metadata schema for characterizing a learning object.

4.2.3. Supporting the generation of semantically rich queries

Generation of semantically rich queries may also be supported at user level. Pinkwart et al. (2004) define a system complying with this proposition. Their work is implemented over the CoolMode platform enabling collaborative learning. A user working within this environment has the option of asking "is there something in the repository similar to the object I am working with?" without providing additional information. Then, the metadata values of the current document are used to formulate the query to the resource repository. Users may also specify the required level of similarity by qualifying metadata elements "free", "required", or "not required". To summarize, the current material defines a specific working context, which is used by the system in order to automatically generate a semantically rich query tuned to users expectations. Nonetheless, this system is implemented for supporting a specific context of collaborative learning.

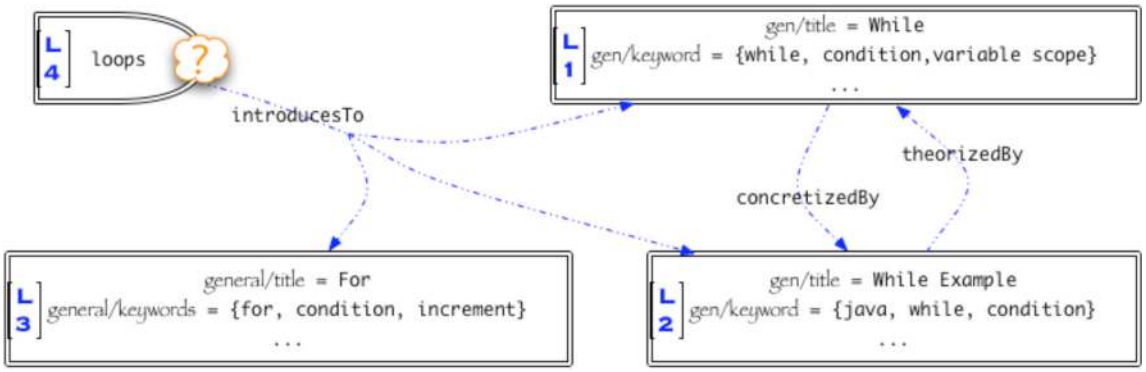

Figure 3. L1, L2, and L3 are learning objects with LOM description. The instructor is looking for a learning object L4 in order to introduce L1, L2, and L3. (From (Motelet and Baloian, 2005)**)**

Motelet and Baloian (2005) suggest to use the semantics of a lesson graph in order to provide semantically rich automatic contextualization of simple keyword queries defined by the user. This system is based on the fact that a syllabus of a certain learning unit might be represented with a graph, in which nodes represent a concept to be learnt and contain the corresponding learning material while the edges represent the inter-relations. A lesson syllabus graph could refer to already existing material and also material that has still to be provided or retrieved. The idea is to take advantage of the semantics of the graph, i.e. the nature of the edges relating the nodes, in order to identify some characteristics of the missing material. In Figure 3, the author of the syllabus of a lesson about programming languages is looking for some learning material in order to introduce the concept of “loops”. First, s/he creates a new node L4 characterized with the key-sentence “loops”. Then, s/he specifies that L4 introduce L1, L2, and L3 by creating some links of type “introducesTo”. The key-sentence is used as a keyword query. This query is then automatically contextualized by the system with semantically rich information being deduced from the graph semantics. This approach is based on the rule-processing method presented in the previous sections in order to deduce sets of restrictions for the metadata values of the required learning resource.

4.3. Query Result Processing

Result processing basically consists of organizing and presenting the elements retrieved by a query in order to facilitate the analysis. As a result of this analysis, one or more learning resources is chosen and (re)used and/or the user reformulates the query in order to better match his/her expectations.

4.3.1. Results ranking

Results ranking is the most common task of information retrieval. In web search engines, web pages are indexed via web crawling. When a user issues a query, term indexes are used to locate the documents where the query terms appear. The documents are then ordered in decreasing order of relevance with different criteria: page authoring, frequency and importance of query terms in the page, and so on. It should be noted that a document could be relevant even if not all keywords appear in it (Baeza-Yates and Ribeiro-Neto, 1999).

In databases and learning object repositories, the answers to a query basically consist of the set of documents exactly matching the query. Some work has been done in order to use relevance ranking in XML databases (INEX, 2003, 2004). However, such techniques are yet not used in learning object repositories. Learning object repositories will definitely benefit from being part of the Initiative for the Evaluation of XML Retrieval (INEX).

Result ranking may also be related to evaluation techniques of recommender systems. In such approaches evaluation of resources is based on profile analysis and/or pattern matching.

Motelet and Baloian (2005) propose to use the suggested values provided by metadata generation systems in order to rank the results set of a user query. Such ranking should be based on generation techniques taking advantage of the query context (e.g., the lesson unit). In that work, semantics of lesson syllabus graphs are suggested as an additional source of information. Profiles of teachers, learners, institutions and learning environment may also be used for that purpose.

4.3.2. Results visualization

Since metadata of educational resources has explicit semantics, the use of information visualization methods can help the user to analyze the result set. For instance, metadata elements referring to a hierarchy may be used to visually organize sets of documents. Klerkx et al. (2005) suggest using advanced hierarchy visualization systems like a tree-map. A tree-map is a visualization technique for hierarchical structure attempting to use the complete display area. With this method, they made a portal for the Ariadne (2005) repository in which the results of a query are visualized within the tree-map reflecting domain taxonomy of the whole repository. Hence, the results of a query are contextualized. Another interesting approach may focus on facilitating result overview. In particular, the various semantics of metadata should be explored and compared. For example, users may benefit from comparing the characteristics of various instances of metadata. In most web sites for e-commerce, this feature consists in displaying the characteristics as lists. Using more refined techniques may certainly enhance results comparison. The Periscope 3D Search System (Wiza et al., 2004), enables a flexible visualization of various characteristics of large sets of documents. However, this complex technique requires a non-negligible training period.

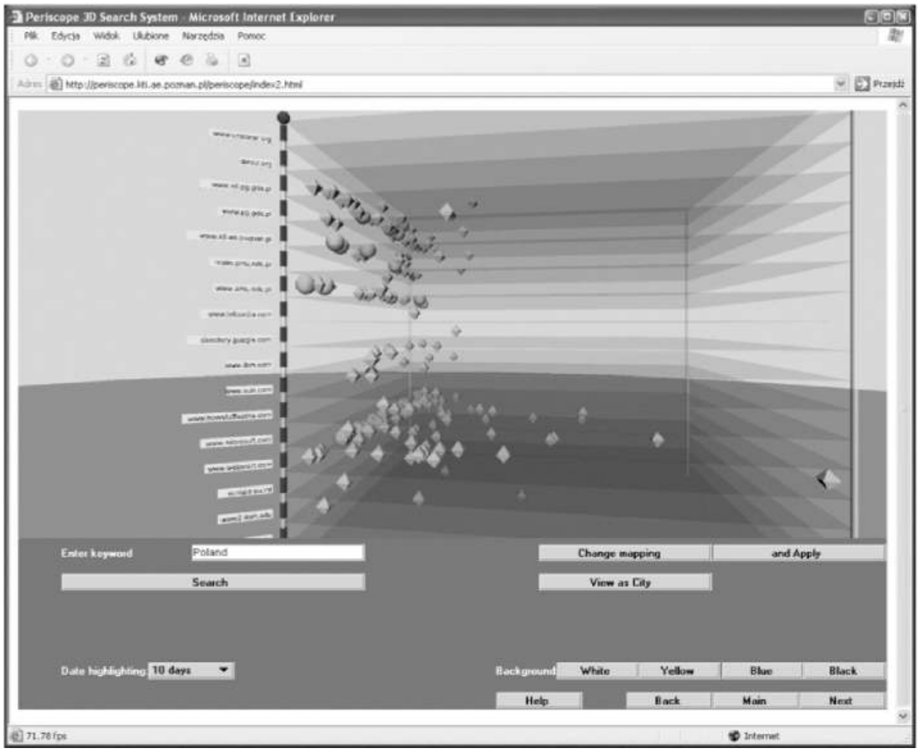

Figure 4. Example of search result visualization in Periscope 3D: the metaphor of multilevel store (From (Wiza et al., 2004))

4.3.3. Query reformulation

When the user is not satisfied with her/his results, s/he could modify her/his query. As a result of frequent web search engine use, users tend to reformulate their query more often. Query reformulation can also be an automatic process, or semi-automatic though user interaction (Baeza-Yates and Ribeiro-Neto, 1999). Similarly, most people are able to reformulate a keyword query when looking for learning objects. However, when using semantically rich queries, the reformulation process may be a tedious task, since the user has to consider for each element of the metadata. An alternative process is the relaxation. Relaxing a query means to weaken some of its constraints in order to widen the scope of results. This process is commonly used when querying the semantic web (Stuckenschmidt, 2003). An interesting point is that the relaxation of semantically rich queries for learning object metadata may follow some predefined patterns. For instance, the elements of the query constraining the categories *general*, *lifeCycle*, *technical*, and *classification* may be relaxed in order to enlarge the search to educational resources matching a certain pedagogical context but not limited to a specific discipline or format. The learning objects resulting from this relaxation process may offer interesting hints for defining methods supporting the particular educational objective of the authored lesson. The development of a set of pedagogically sound relaxation strategies may be a powerful tool for using the semantic aspect of LOM.

4.4. Conclusion

Creating semantically rich queries may be a process even more fastidious than creating metadata for sharing learning material. Therefore, it is necessary to offer simple means to query learning object repositories like keyword query interfaces. However, such approaches may also tend to disregard the semantic aspect of metadata for educational resources. Discarding this characteristic results in under-using the LOM possibilities.

Other approaches like recommender systems and support for semantic based queries can offer new perspectives to the teacher/learners wanting to retrieve learning material. Restricting users to keyword queries may improve the usability of learning object repositories, but semantics support is still needed in order to enhance the retrieval process. An interesting approach may consist in combining full-text search engine with XML databases offering semantic precision. We could thus combine the performance and efficiency of search engines like Google while enhancing the precision though the use of metadata semantics. Such hybrid systems have already been developed (Pehcevski et al., 2005) for domains not related to learning. Adapting such a proposal may be more relevant for the future of learning object repositories than trying to reinvent content-based search engines.

5. Perspectives for the Use of Learning Object Metadata

Learning object metadata is without doubt a keystone for promoting reuse of learning objects. Paradoxically, they are also the bottleneck of this process, since many authors are not comfortable with the work of producing meaningful metadata for their created learning objects. Automatic generation of metadata seems to be – if not a definitive solution – important help to simplify this work (Section 2). It is in fact very easy for some metadata to be automatically generated by the authoring tool used to create the associated learning objects. For example, digital format, weight, author's name and even language can be derived from the environment in which the object is being created. However, there is a set of metadata, which is difficult to generate automatically: it includes information related to the educational characteristics of the learning object such as difficulty, interactivity type and required learning time. Nevertheless, there is some possibilities of inferring the values for this metadata from the context in which the learning object has been, is, or is intended to be used (Section 2). This information can be supplied by people who have already used it in a so-called recommender system. Another interesting approach for deriving the values for the metadata of one learning object is to gather information from the metadata of other learning objects that are used in conjunction with this one. This approach is reasonable because learning objects are seldom used in isolation: a learning unit typically contains several learning objects. Furthermore, learning units are generally structured, thus their learning objects are inter-related.

The automatic generation of metadata cannot only be used for instantiating the actual values for a certain existing learning object but also for generating the queries for retrieving objects from repositories. This process can also be used for validating the metadata of a certain object (Section 3). This is important because invalid metadata will lead to errors when trying to retrieve suitable learning objects from a repository for a certain learning purpose. Moreover, these errors are difficult to discover by the user.

In order retrieving learning material, data mining methods typically used in information retrieval are not sufficient. Indeed, these methods cannot extract implicit information, e.g. most pedagogical data, from the material contents. In contrast, metadata present the advantage to hold such information. Moreover, learning object metadata are semantically rich data. This characteristic offers interesting perspectives for learning object retrieval (Section 4). Nevertheless, creating semantically rich queries may be a process even more fastidious than creating metadata for sharing learning material. Therefore, it is necessary to offer simple means to query learning object repositories like keyword query interfaces. Restricting users to keyword queries may improve the usability of learning object repositories, but semantics support is still needed in order to enhance the retrieval process. Consequently, complementary approaches like recommender systems and automatic systems for generating semantically rich queries may emerge as fundamental tools in order to support users in retrieving learning material.

Since metadata creation is a necessary step for posting learning material in learning object repositories, there is a trend to associate the negative reputation of the metadata business to the learning object repositories. Most creators of educational resources wish to see metadata disappear. Some years ago, certainly the same people were also reluctant to use multimedia applications like MS Powerpoint or Macromedia Flash in order to build educational material. Nevertheless, they finally are using them because these tools represent a powerful support for their work. However, while the efforts done for using multimedia applications may rapidly result in significant improvement of the comprehensibility and interactivity of the learning objects, the efforts done for creating valid metadata are generally unsuccessful.

Currently, the activities related to learning object metadata processing seem to stay in a vicious circle (Figure 5). The process of generating metadata is tedious and complex (1). For that reason, produced metadata has poor quality in general (2). Therefore, it cannot be correctly processed when trying to retrieve a suitable learning object using the objects metadata (3). Thus, metadata may not provide interesting benefits for the user (4). Consequently, generating metadata is considered a tedious process (1). As we can see, unfortunately, the vicious circle is closed.

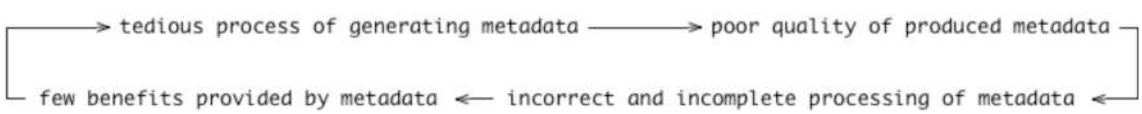

Figure 5. The vicious circle of activities related to leaning object metadata processing.

In order to change this situation for a virtuous circle as the one shown in Figure 6, all steps of the circle must be modified according to a *shifting strategy*. In particular, each step of the circle should be translated from its present state to the following position: (1) generation of metadata should be supported, (2) the quality of the metadata should be improved, (3) metadata processing should be complete and relevant, and finally (4) the benefits of this process should be noticeable.

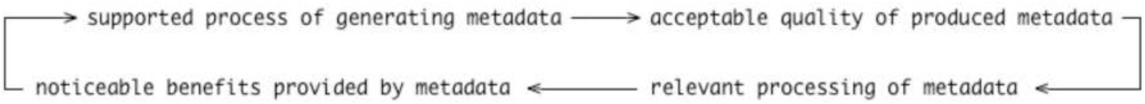

Figure 6. The virtuous circle of activities related to leaning object metadata processing.

As this chapter has shown, automatic processes may facilitate the generation and validation of the metadata for educational resources and also the generation of relevant queries to retrieve the learning material. However, it was also shown that these systems could not completely replace the human participation. A shifting strategy may benefit from the relevant support of automatic processes, but this strategy should be centered on the users. We advocate for automatic processes helping humans to accomplish their tasks. No wonder, this has been a recurring shift of paradigm in the artificial intelligence area and especially in the development of intelligent tutoring systems.

Lesson authoring and metadata generation are now two separated processes. Lesson authors generally take care of metadata only when sharing their production. Most of them create the metadata at publishing time when the learning object repository portal asks for this information. Other authors delegate this work to third-party metadata creators. However, this solution is costly and few authors are practically involved in this latter category. We suggest using the lesson authoring process in order to define a user-centric shifting strategy. For this, the authoring process should encourage and support the retrieval and reuse of existing learning material created by other people and stored in repositories in addition to creating the own learning material from scratch. This strategy combines the learning material retrieval with the lesson authoring process. This approach motivates the merging of lesson authoring with metadata generation.

Basically, our strategy consists of three steps. First, the benefits of using learning object metadata must be immediately noticeable by the creators of learning material, i.e., during lesson authoring. Second, this motivation should invite them to achieve a minimum level of validity for the metadata of the various elements of the lesson being created. Third, creation of valid metadata should be highly supported. An instance scenario of this strategy is described below.

Motelet and Baloian (2005) describe a lesson-authoring interface merged with the interface for querying learning object repository (Section 4). With this system, while the instructor organizes the activities and material for the lesson, an element of the authored lesson graph may be seamlessly used as input for automatically generating a query on learning object repositories. It is possible to significantly help the instructor in finding missing learning material by combining efficient retrieval systems based on full-text search engines with automatic generation of semantically rich information about the context of the requested material. If such a technique is integrated in the system supporting the authoring process, the instructor may efficiently reach relevant educational resources without having to open another third-party application (e.g., a web browser).

The lesson author should be invited to understand that the efficiency of her query increases with the size and precision of the context description in which the learning object is going to be used (e.g., the lesson graph of a learning unit). Moreover, she may notice that the quality of the metadata of the rest of the material contained in the lesson should also improve the retrieval process. Hence, the author will be continuously interested in checking the level of validity of the metadata of the elements of the lesson. Again, automatic processes may have an important role.

Completeness and accessibility of metadata are easily computable (Section 3). Smart composition of generation techniques enables validity checking of most elements of the metadata. When trying to retrieve some learning material for a certain lesson, the author may notice that the results are numerous and not always adequate for the lesson context. She may also be confronted to the fact that the validity of the metadata of most learning objects already included in the lesson is quite low. Therefore, instead of specifying her query, the author may try to increase the validity of the material since automatic processes may benefit from this information to deduce the required context of the missing material.

Metadata edition should be as easy as possible (Section 2). For example, we may say that users should assist the work of automatic generators and not that automatic processes should assist the work of the users. Suggested values have to be comprehensible, i.e., referring to vocabulary used by the lesson author (e.g., community ontology). Access to the definition of the presented terminology should be ubiquitous. An ontology may be used and presented to the users in a human-readable manner using for instance the extension of the Unified Modeling Language suggested by Naeve (2005). The validation process should support the navigation in the complex structure of learning object metadata. For example, users may be invited to reconsider the values of some specific parts of the metadata for more than one learning object at the same time. This process may help him to grasp the meaning of the metadata by comparing the instantiation of various educational resources. Access to thumbnails presenting the learning material should also be handy. Educational resources authoring should be seamlessly integrated to the lesson syllabus authoring system. Finally, edition of learning object contents should work in parallel with learning object metadata instantiation. Systematic reevaluation of the metadata may occur when material contents are modified.

Even today learning object metadata still appears for the beginner user as a self-referent concept since she may not see much concrete benefit of creating them. On the opposite, the usage scenario we are proposing deals with a tangible and immediate justification for creating valid metadata. This justification is founded on the efficient retrieval of learning material during the authoring process. Obviously, many parts of the proposed scenario still have to be investigated. However, this example explains the advantage of automatic processes in order to shift from a complex and meaningless use to a user-centric and practical use of learning object metadata.

Acknowledgments

The authors are sincerely grateful for the important collaboration of Benjamin Piwowarski. This work was partially supported by grant No. 1040952 from Fondecyt (Chile) and by Chile-Corea IT Cooperation Center.

Biographies

Olivier Motelet obtained a degree of Engineer from the Ecole des Mines de Nantes (France) in 2000. The same year, he also acquired a degree of European Master of Science in Computer Science from the Vridje Universiteit Brussel (Belgium). His research interests include Learning Management Systems, Human Computer Interaction, Software Usability and Domain Specific Languages. He is presently PhD student in the Universidad de Chile.

Nelson Baloian obtained his Degree of Engineer in Computer Science in 1989 from the Universidad de Chile. In 1998 obtained his PhD from the University of Duisburg (Germany). He has been working on the fields of Computer Supported Learning and Distributed Computing since then. He is author of about 30 journal and conference articles of these subjects. He is currently Assistant Professor at the Department of Computwer Science of the Universidad de Chile. He is also Visiting Professor of the University of Waseda (Japan) and the University of Duisburg-Essen (Germany).

José A. Pino is Associate Professor of Computer Science and Director of the PhD Program in Computer Science at the Universidad de Chile. His research interests include Computer-Supported Collaborative Work, Human-Computer Interaction and Software Industry Studies. He has served as President of the Chilean Computer Science Society (SCCC) and President of CLEI (the Latin American Association of Universities concerning Information Technology). He has co-authored six books and published research papers in international conferences and journals, including Journal of the ACM, Communications of the ACM, Decision Support Systems, Interacting with Computers and Information Technology and People.

References

Aggarwal, C., Wolf, J., Wu, K., & Yu, P., (1999). Horting Hatches an Egg: A New Graph-theoretic Approach to Collaborative Filtering. Proc. of the KDD-99: The Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, pp.201-212. ACM Press.

Ariadne. (2005) ARIADNE - Foundation for the European Knowledge Pool. http://www.ariadne-eu.org. Last accessed on July 5, 2005.

Baloian, N, Galdamez, P, Collazos, C, Guerrero, J.C.(2004). A model for a Collaborative Recommender system for Multimedia Learning Material. In: Groupware Design, Implementation, and Use. CRIWG 04 Conferece in Costa Rica, September 2004 Lecture

Notes in Computer Science 3198 pp. 281-288.

Baeza-Yates, Ricardo and Ribeiro-Neto, Berthier (1999) Modern Information Retrieval. New York: Addison Wesley, 1999.

Blackboard. (2005). BlackBoard Learning Management System. http://www.blackboard.com. Last accessed on July 5, 2005.

Breese, J., Heckerman, D., & Carl, K. (1998). Empirical analysis of predictive algorithms for collaborative filtering. Proc. of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, pages 43–52, July 1998.

C4J. (2005). Sourceforge Classifier4J. http://classifier4j.sourceforge.net. Last accessed on July 5, 2005.

Cancore. (2005). CANCORE. http://www.cancore.ca. Last accessed on July 5, 2005.

DC.dot. (2005). DC.dot. http://www.ukoln.ac.uk/metadata/dcdot. Last accessed on July 5, 2005.

DCMI. (2005). Dublin Core Metadata Initiative. http://dublincore.org. Last accessed on July 5, 2005.

Downes, S. (2004). Resource Profiles. *Journal of Interactive Media in Education*, Special Issue on the Educational Semantic Web, 2004 (5).

Duval, E., Hodgins, W. (2004). Metadata matters. Proc. of International Conference on metadata and Dublin Core specifications – DC-2004, Shanghai, China.

Edu, D. C. (2005). DCMI Education Working Group. http://dublincore.org/groups/education. Last accessed on July 5, 2005.

Friesen, N. (2004). International LOM Survey: Report. *ISO/IEC JTC1/SC36 sub-committee*

Goldberg, D., Nichols, D., Oki, B., & Terry, D. (1992) Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12): 61–70.

Google. (2005). Google. http://www.google.com. Last accessed on July 5, 2005.

Greenberg, J. (2003). Metadata Generation: Processes, People and Tools. *Bulletin of American Society for Information Science and Technology*, 29(2)

Greenberg, J. (2004). Metadata Extraction and Harvesting: A Comparison of Two Automatic Metadata Generation Applications. *Journal of Internet Cataloging: The International Quarterly of Digital Organization, Classification, and Access*, 6(4), 58-82.

Hatala, M., & Richards, G. (2003). Value-Added Metatagging: Ontology and Rule Based Methods for Smarter Metadata. Proc. of International Conference on Rules and Rule Markup Languages for the Semantic Web - RuleML, Lecture Notes in Computer Science, 2876, pp. 65-80.

Hodgins, W. (2001). The Future of Learning Objects. In D. Wiley (Ed.), *The Instructional Use of Learning Objects* Association for Instructional Technology.

INEX (2003) Fuhr, N., Lalmas, M. and Malik, S. (Ed.) INEX 2003 Proceedings.

INEX (2004) Fuhr, N., Lalmas, M. and Malik, S. (Ed.) INEX 2004 Proceedings.

IMS. (2005). IMS Global Learning Consortium. http://www.imsglobal.org. Last accessed on July 5, 2005.

Interwoven. (2005). Metatagger. http://www.interwoven.com. Last accessed on July 5, 2005.

Jenkins, C., Jackson, M., Burden, P., & Wallis, J. (1999). Automatic RDF metadata generation for resource discovery. Proc. of the eighth international conference on World Wide Web. Toronto, Canada, pp. 1305–1320.

Kabel, S., de Hoog, R., & Wielinga, B. J. Consistency in indexing learning objects: an empirical investigation. Proc. of Learning Objects 2003 Symposium: Lessons Learned, Questions Asked – World Conference on Educational Multimedia, Hypermedia & Telecommunications ED-MEDIA 2003, Honolulu, Hawaii.

KEA. (2005). Kea - Automatic Keyphrase Extraction. http://www.nzdl.org/Kea. Last accessed on July 5, 2005.

Kim, W. (2005). On Metadata Management Technology: Status and Issues. *Journal of Object Technology*, 4(2), 41-47.

Klarity. (2005). Karity API. http://www.intology.com.au. Last accessed on July 5, 2005.

Klerkx, J., Meire, M., Ternier, S., Verbert, K., & Duval, E. (2005). Information Visualisation: Towards an Extensible Framework for Accessing Learning Object Repositories. Proc. of World Conference on Educational Multimedia, Hypermedia & Telecommunications ED-MEDIA 2005, Montreal, Canada, pp. 4281-4287.

Lattner, A. D., & Gehrke, J. D. (2004). Applying Inductive Logic Programming and Rule Relaxation for the Generation of Metadata. Proc. of FGWM'04 Workshop, Annual Meeting of the Special Interest Group on Knowledge Management of the German Society for Computer Science (GI), Berlin, Germany, pp.267-273.

LOM. (2005). IEEE LOM specification. http://ltsc.ieee.org/wg12. Last accessed on July 5, 2005.

Marchiori, M. (1998). The limits of Web metadata, and beyond. Proc. of the seventh international conference on World Wide Web, Brisbane, Australia, pp. 1-9.

Motelet, O. (2005). Relation-based heuristic diffusion framework for LOM generation. Proc. of 12th International Conference on Artificial Intelligence in Education AIED 2005 - Young Researcher Track, Amsterdam, Holland.

Motelet, O., & Baloian, N. A. (2005). Taking advantage of LOM semantics for supporting lesson authoring. *Research Report - DCC - Universidad de Chile*.

Naeve, A. (2005). The Human Semantic Web Shifting from Knowledge Push to Knowledge Pull. *International Journal of Semantic Web and Information Systems*, 1(3), 1-30.

Najjar, J., Klerkx, J., Vuorikari, R., & Duval, E. (2005). Finding Appropriate Learning Objects: An Empirical Evaluation. Proc. of ECDL 2005 9th European Conference on Research and Advanced Technology for Digital Libraries, to be published.

Ochoa, X., Cardinaels, K., Meire, M., & Duval, E. (2005). Frameworks for the Automatic Indexation of Learning Management Systems Content into Learning Object Repositories. World Conference on Educational Multimedia, Hypermedia & Telecommunications ED-MEDIA 2005, Montreal, Canada, pp. 1407-1414.

Pehcevski, J., Thom, J. A., & Vercoustre, A.-M. (2005). Hybrid XML Retrieval: Combining Information Retrieval and a Native XML Database. *Information Retrieval*, 8(4), 571-600.

Pinkwart, N., Jansen, M., Oelinger, M., Korchounova, L., & Hoppe, U. (2004). Partial Generation of Contextualized Metadata in a Collaborative Modeling Environment. Proc. of 2nd International Workshop on Applications of Semantic Web Technologies for E-Learning - International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems AH 2004, Eindhoven, The Netherlands.

RDF. (2005). W3C Specification of Ressource Description Framework. http://www.w3.org/RDF. Last accessed on July 5, 2005.

Reload. (2005). Reload editor. http://www.reload.ac.uk/editor.html. Last accessed on July 5, 2005.

Savia, E., Kurki, T., & Jokela, S. (1998). Metadata Based Matching of Documents and User Profiles, in Human and Artificial Information Processing. Proc. of the 8th Finnish Artificial Intelligence Conference, Jyväskylä, Finland, pp. 61-70.

SCORM. (2005). SCORM. http://www.adlnet.org. Last accessed on July 5, 2005.

SIDWeb. (2005). SIDWeb. http://www.cti.espol.edu.ec/sidweb. Last accessed on July 5, 2005.

Simon, B., Dolog, P., Miklos, Z., Olmedilla, D., & Michael, S. (2004). Conceptualizing Smart Spaces for Learning. *Journal of Interactive Media in Education*, Special Issue on the Educational Semantic Web, 2004(5).

Stuckenschmidt, H. (2003). Query Processing on the Semantic Web. *Kunstliche Intelligenz: Special Issue on the Semantic Web*, 3(03), 22-26.

Stuckenschmidt, H., & Harmelen, F. v. (2001). Ontology-based metadata generation from semi-structured information. Proc. of the first international conference on Knowledge Capture K-CAP, Victoria, Canada, pp. 163-70.

Stuckenschmidt, H., & Harmelen, F. v. (2004). Generating and managing Metadata for Web-based Information Systems. *Knowledge-based Systems*, 17(5-6), 201-06.

Wiley, D. (2000). Learning object design and sequencing theory. *Doctoral dissertation*, Brigham Young University.

Wiza, W., Walczak, K., & Cellary, W. (2004). Periscope: a system for adaptive 3D visualization of search results. Proc. of the ninth international conference on 3D Web technology – Web3D 2004, Monterey, California, USA, 29-40. ACM Press.

Woodley, A. and Geva, S. (2004) NLPX - An XML-IR System with a Natural Language Interface. Proc. of ADCS 2004, Ninth Australasian Document Computing Symposium, Melbourne Australia, pp. 71-74.

Wong, S., & Butz, C. (2000). A Bayesian Approach to User Profiling in Information Retrieval. *Technology Letters*, 4(1), 50-56, 2000.

XML. (2005). Extensible Markup Language. http://www.w3.org/xml. Last accessed on July 5, 2005.

Yilmazel, O., Finneran, C. M., & Liddy, E. D. (2004). Metaextract: an NLP system to automatically assign metadata. Proc of the 4th ACM/IEEE-CS joint conference on Digital libraries - JCDL '04, Tucson, Arizona, USA, pp. 241–42. ACM Press.