Combining *Folksonomies* and Automatic Information Techniques for *LO* Semantic Indexing

DOINA ANA CERNEA¹, ESTHER DEL MORAL², EMILIO LABRA³

¹Mathematics Department, ²Education Science, ³Computer Science Engineering Department
University of Oviedo
SPAIN

Abstract: - The actual changes produced in the e-Learning field are being driven by two primary facts: the first one is the steady increase of information creation and the second is the new collaborative framework of the Web 2.0. While *LO* (Learning Objects) repositories are more and more extensive, the automatic tools for search and location of proper learning content in these repositories, based on semantic tags are not yet effective.

This article proposes a system architecture that combines automatic techniques of information retrieval with the user’s collaborative annotation of the documents, for the semantic indexing of *LO* in a repository. Collaborative tagging provides real meaning derived from the learning experiences in real user’s communities and improves *LO* reusability in new learning contexts.

Key-Words: - Semantic indexing, semantic gap, Learning Objects, social learning, social tags, folksonomy, e-Learning 2.0

1 Introduction

The opportunities offered by the Web 2.0 and its collaborative tools applied to the on-line learning/teaching lead us to reformulate the e-Learning process, called now the e-Learning 2.0. In this new framework, the search and location of adequate learning assets through this huge amount of information is more and more difficult.

We need to design new information management systems based on semantic descriptions that could guarantee *LO* reusability in different learning contexts.

However, the existing technologies for concept based description of the Web resources are still not very advanced. At the same time, the multiple and distinct media types used in the design of *LO* requires specific techniques for each of these types.

While the analysis of textual information is more affordable and the assigning of keywords for the syntactic searches is made possible by breaking up the text into separate paragraphs and/or words, in the case of multimedia assets the minimum unit is much more difficult to divide and there is still a great distance between the high-level descriptions (or concept based descriptions) according to the human perceptions and low-level visual features that can be automatically extracted from data, like color histograms, texture indexing, shape or contours, etc. The main challenge in this field has been quoted as automatically generated concept based description for multimedia information and this problem is known as “bridging the semantic gap”. Currently, there is a lack of tools that can automatically manipulate the high-level concepts of an image or video.

There have been many attempts to solve the problem of the semantic description of the multimedia assets [17]. One of the first – using the MPEG-7 format to store multimedia metadata – was made by Hunter in 2001 [11], who first developed a specific RDF ontology enriched by DAML-OIL and latter by OWL. As part of the aceMedia Project, an ontology was created for the visual descriptors of multimedia content which have been incorporated in the MPEG-7 visual metadata.

Most of the proposed solutions for the automatic semantic description are dividing the image into regions with similar visual features and assigning them semantic labels using statistical methods, EM algorithms [7], or recent probabilistic models like the Cross Relevance Model [12].

Recently, more and more websites allow their users to use collaborative tagging of different types of resources, for their social classification or social indexing. In contrast with the traditional method of adding semantics, metadata is not created by experts but it is spontaneously generated by consumers [4], [14].

Among the actual web sites which base their strategy on the collaborative tagging, there are
services like del.icio.us, Tecnocrati and Flickr [14] that allow their users to annotate the web resources like a web page, a blog post or an image, normally using free sets of tags and permitting their sharing and reutilization.

The system we propose combines users’ collaborative annotations, based on their real learning experiences, with information retrieval techniques in order to improve emergent semantic descriptions for LO and learning assets.

2 e-Learning 2.0: Collaborative Learning in the Web 2.0 Framework

Social software arises in the new context created by the Web 2.0 and reformulates the e-Learning paradigm, based now on collaboration and connections between people, organizations and web resources. This new concept of on-line learning/teaching is known as e-Learning 2.0 [6].

Web 2.0 collaborative tools enable the ease of content creation on Internet and the social interactions as a result creating a new framework for e-Learning. Learning environments and virtual communities designed from this new collaborative perspective introduce new alternative ways of working and bring a social dimension of knowledge construction.

e-Learning 2.0 recognize the ability of synthesize and to see connections between fields, ideas and concepts, as a core skill. Learning is now based on information pattern recognition. Moreover, learning and knowledge can also reside in others’ experiences since the process is supported by social connections creation [6].

From this collaborative framework, users’ tags are themselves learning resources, generating a social learning context. Students are motivated and involved in the construction of knowledge process through the creation of new personal meanings for the shared learning resources and are developing a new learning context in this way.

The emergent social interactions through the collaborative annotations of the shared resources in dynamic folksonomies force users to constantly reflect on the connections between items and tags. This fact directly implicates the embodied social conceptualization of the learning assets.

In this collaborative learning environments the reusability of LO relies on the user’s annotations that will identify the learning resources in e repository. This is the philosophy behind the social tagging systems that we use in the design of our system architecture that we describe below.

2.1 Collaborative tagging systems: folksonomies

Folksonomy systems enable users to freely add numerous descriptive keywords to content such as web pages, images, videos, etc, and also to learning assets and LO. The social tags assigned to the LO or web resources serve to classify and locate this content. At the same time, social tagging generates a navigation system based on the labels assigned to different web resources and on the existing links between the users that participate in a folksonomy.

From this point of view, the use of collaboratively generated tags improves the management of resources such as search, spam detection, reputation systems and personal organization through the introduction of new forms of communication and new opportunities for data mining, due to the emergent social structure that underlies folksonomy systems.

Social tagging systems allow their users to share their tags of particular resources. Each tag serves as a link to additional resources tagged in the same way by other users [14]. Certain resources may be linked to each other; at the same time, there may be relationships between users according to their own social interests, so the shared tags of a folksonomy come to interconnect the three groups of protagonists in social labelling systems: Users, Resources and Tags [14].

Figure 1 shows a conceptual model of a social tagging system where users and resources are connected through the tags they assign.

Social tags, introduced by a specific person, are very different from expert tags which we assume are objective and consistent. In this way, social tags are mostly associative and subjective. They become social tags when they are shared over a users’ community that creates an implicit feedback for both users and tags. It can be observed that over time, the relative frequency of the tags used to label a specific web resource tends to approach a constant value.

Fig. 1. Connections network of a collaborative tagging system
This shows that collaborative tagging is able to coordinate Web users’ actions to create coherent and consistent semantic annotations for the shared content. For a given tag chosen from one of the four popular tags on del.icio.us, a power-law distribution between the number of tags and co-occurring tags has been shown by Cattuto, who predicted that "the most used tags are more likely to be used by other users since they are more likely to be seen".

There are three hypotheses regarding the behaviour of the tag-sets over time [9]:

Tag convergence: after a certain amount of time, tags given to a resource (tag-set) stabilize as the relevant categorizations are made and the most common words for those categorizations become the majority.

Tag divergence would then be that tag-sets never converge to a smaller group of more stable tags, and the tag distribution continually changes.

Tag periodicity, where after one group of users tag to some local optimal tag-set, another group uses a divergent set and so perturbs the original optimal set, and after a period of time the new group's set becomes the new local optimal tag-set. This process may repeat and so lead to convergence after a period of instability, or it may repeat ad infinitum and so act like a chaotic attractor.

At the same time, from a pedagogical point of view, the use of social tagging systems in e-Learning proves to be a way of constructing personal context and a technology with a great potential to support self-steered learning. From a Constructionist perspective, social tags become meaningful learning assets, where the learner himself is consciously involved in constructing a public entity.

So the use of folksonomies, besides being a semantic mark-up collaborative method, is also an opportunity for learning because it generates social learning context. Social tags respond to Vygotsky constructivist definition that evolution involves transformation from an interpersonal process into an intrapersonal one, and that learning takes place through communication among people in learning communities.

3 LO Semantic Indexing

The Semantic Web approach breaks the concept of a web page as an information unit, enabling the creation of resource RDF descriptions with finer granularity. For example, from the semantic description of the Learning Object “Introduction to the Fractal Geometry”, we could extract the definition of the fractal structure concept or the recursivity property. So, in the Semantic Web context, LO and learning resources have to be accompanied by the semantic descriptions of their content [13].

Example: Semantic description in RDF code of a Learning Object about Fractal Geometry:

```
<rdf:RDF xmlns:rdf="..." ...

<rdf:Description rdf:about="http://.../geofractal/esStruFractal">

<dc:title>Introducción a la Estructura Fractal</dc:title>
<dc:creator>
</rdf:Description>
</rdf:RDF>
```

At the present time, Web resources mark-up process still requires manual annotations, especially multimedia resources such as images, videos, sound assets, etc., which need subjective interpretations. Managing large data bases of images, for instance, is the well known problem of Image Retrieval (IR). The main challenge in this field is the automatic generation of concept-based descriptions that reflect the real meaning of the images.

Our solution defines a specific system that allows automating the LO semantic indexing process, combining the emergent semantic with the users’ collaborative annotations.

LO are defined as minimal units with self meaning, built by information assets interactive, represented in various digital formats and based on a single learning objective. The most important property relating to LO is their reusability in new and different learning contexts, a consequence of their flexible and granular structure [5]:

First level: the lesson items (Information Objects)
Second level of granularity: the learning assets that compose the Information Objects (text assets, images, videos, audio assets, …)

While the automatic semantic description could be simpler in the case of text resources, and is based on the keywords extracted from the break up of text into small units, in the case of images, videos and audio assets, the information retrieval needs specific techniques for each type of multimedia resource.

Generally, e-learning repositories contain a high density of multimedia resources. But in the case of multimedia assets, the automatic methods for semantic indexing are still inefficient, and for that reason, we propose a system that integrates
folksonomy techniques to complete and refine the emergent semantics for LO in a repository.

3.1 The Semantic Gap

There is a large distance between multimedia automatic descriptions and their real content, known as the semantic gap. The semantic gap is defined as the lack of coincidence between the information extracted automatically by computers and the human perception of the real content of multimedia resources, based on high-level concepts [10]. The state of art of the automatic systems for multimedia analysis, particularly for image analysis, are still limited to the low-level feature extraction, such as colour attributes, texture parameters, borders, etc.

We will represent all LO and learning assets through a vector space of visual features that are processed using the Latent Semantic Indexing algorithm in order to extract the semantic definition for the media resources.

The results in bridging the semantic gap are mainly referred to information retrieval in text data bases, leaving the automatic semantic indexing for multimedia resources as an open problem. Our solution is based on the integration of users’ collaborative annotations to improve the automatic methods.

4 Integrating Collaborative Annotations with Emergent Semantics

The proposed architecture intends to classify LO in a repository by combining the emergent semantics with the users’ collaborative tags. A folksonomy-based technique is used for the capture of users’ annotations, testimonies of real learning experiences with a concrete learning resource. These tags are processed to obtain a set of terms connected with specific field ontology or an ontology network. This way, the semantic indexing obtained by automatic methods is completed with the incorporation of users’ collaborative contributions within a learning community.

From this point of view, social tags constitute an optimal solution for completing and refining the emergent semantic descriptions.

4.1 Automatic semantic indexing

Each type of media resource is represented through its low-level features in a vector space and processes these vectors using matrix analysis algorithms in order to extract the concept based description.

The representation of an image through a vector containing the visual characteristics is made using a set of low level features that might be relevant for image identification such as color analysis, textures and boundaries. For color process we chose the normalized histograms and color coherence vector that contribute to describing the distribution of color in the image. Color histograms are a popular method to compare images. Color histograms are computationally efficient, invariant to geometric transformations such as rotations or scaling, and sensitive to illumination changes and noise. But color histograms do not provide any information about the spatial distribution of the colors in the image, they indicate just what colors are present in the image and in what quantities [Fig. 3]

In order to overcome this problem and to take into account the spatial distributions of each color present in the image, we consider the Color Coherence Vector (CCV) method proposed by Greg Pass [15]. The coherence measure defined classifies pixels as either coherent or incoherent. A coherent pixel is part of a large group of pixels of the same color, while an incoherent one is not. The CCV represents this classification for each color in the image. Java libraries provided by the Discovir project are implemented by our system.

Another set of features computed in the case of images are the texture’s descriptors, which are represented through the grey scale co-occurrence matrix. The final part of the low-level features vector that represents an image contains the invariant moments for shape and boundary

Audio assets are represented through the Mel frequency cepstral coefficients (MFCCs) derived from the energy spectrum and are adequate for speech and non speech audio resources. MFCCs are obtained through a frame- based analysis of the signal. A Discrete Fourier Transform (DFT) is performed using a hamming window overlapping each frame to obtain an amplitude spectrum. This spectrum is converted to a Mel scale spectrum using
triangular filters emphasizing frequencies according to their perceptual importance on this scale [3].

For video content semantic indexing, we use the approach of Adams and all [1] that begin by assuming the a priori definition of a set of atomic semantic-concepts (objects, scenes and events) which is broad enough to cover the semantic query space of interest and starting with a training set of videos. The low-level feature representation is made by a key-frames basis using colour and texture descriptors such as in image processing.

The proposed architecture is based on Latent Semantic Indexing using the representation of the resources in the $\mathbb{R}^n$ vector space through their visual features. This method uses linear algebra matrix computation to process the multimedia resources in a large data base. Truncated Singular Value Decomposition method is applied to the features by documents matrix to estimate the structure of the visual features across resources within an observation set. The similarity of the analyzed resource with other resources that already possess a semantic description is calculated using a Euclidean measure. The output of the algorithm is a terms vector containing the weights of each term of preset vocabulary which reflects the importance of that concept for the analyzed multimedia asset. The extracted terms represent the automatic semantic index of the multimedia resource.

The Information Retrieval process in the case of text resources is associated with the complexity of the natural language analyze. For the semantic indexing of this type of e-learning resources, we use Natural Language Processing (NLP) techniques, very popular in the world of web searchers. Semantic indexing of text assets implies that the documents are classified using interpretable concepts described in a field modeled by ontology. The main steps in the text analysis are text segmentation, word sense disambiguation and normalization of the concepts extracted [18].

Despite all, automatic methods for semantic indexing don’t guarantee the reliability of the extracted concepts, because often, these algorithms introduce errors in the final information. That’s why other complementary methods are needed to complete and correct the resulted semantic descriptions, especially in the case of the multimedia resources.

Three types of meta information will be considered to describe LO or learning resources:

- Automatic semantic indexing generated from the low-level features
- High-level descriptors provided by authors (normally referred to the title, date of creation, names of the authors, learning objectives…)
- Users’ semantics through the collaborative annotations.

### 4.2 Collaborative tags semantics

Without any doubt, the collaborative tags of users constitute a powerful source of semantic information for LO in a repository. Social tags bring new knowledge based on users’ the real learning experiences. We will integrate these collaborative annotations in order to complete and refine automatic semantic indexing based on low-level features extraction.

Collaborative tags are a constant feedback for the whole system: the initial vocabulary used by the LSI algorithm is actualized and completed with the new terms introduced by users and the whole algorithm is recalculated. After processing the tags we obtain a normalized set of new terms that we intent to integrate in ontology. But the philosophy behind collaborative tagging is so different to ontology based approaches, where a group of experts develops an ontology and the resources are linked to that ontology. In the social tagging case, the approach is bottom-up, because the users develop their own tags without any coordination.

A Spreading Activation algorithm is applied to the normalized set of terms in order to relate them with ontology.

### 4.3 Social trust filter for collaborative annotations

Finally, users’ tags set can be filtered applying a social trust test based on the similarity of users’ profiles to guarantee their reliability and relevancy [8], [16]. We will apply a Collaborative Filtering (CF) algorithm using the “Word of Mouth” principle used by the Ringo music recommendation system based on the similarity between users’ profiles. A measure of similarity proposed by Sardanand and Maes is the PCC (Pearson Correlation Coefficients) that aims to measure the degree of agreement between two users $a$ and $b$. PCC are computed from the ratings $r_{ai}$ and $r_{bi}$ that the users $a$ and $b$ rate for the resource $i$ and their user mean $r_a$ and $r_b$:

\[
\text{PCC}_{ab}(i) = \frac{\sum (r_{ai} - \bar{r}_a) (r_{bi} - \bar{r}_b)}{\sqrt{\sum (r_{ai} - \bar{r}_a)^2 \sum (r_{bi} - \bar{r}_b)^2}}
\]
Through the social filtering, we pretend to assign a greater probability to a tag taking into account if the users have similar profiles in order to assign the most relevant semantic information.

5 Conclusion
The automatic generation of semantic description for learning resources, is still an open field. This article proposes a system’s architecture for the semantic indexing of the LO in a repository, combining automatic techniques for information retrieval based on LSI algorithm with social tagging technologies. The tags set assigned by the users in a learning community are filtered using a trust test based on the similarity of users’ profiles and, later on is processed and linked to the ontologies.

Through the technology presented in this article, the meta-data of the LO is itself a learning resource incorporating the meanings derived from the real learning experiences. Users’ annotations contribute to a better identification of the learning resources and improve their reusability in new learning contexts.

References: