

Semantic Digital Libraries

(MAP-I thesis proposal – 18th January 2008)

Proponent:

Álvaro Reis Figueira
DCC – FCUP
Research Unit: CRACS

Context and Motivation

One way to sustain and foster development of elearning is commonly agreed to be the reuse of existing learning material. For a long time we are used to search the Web for text files, images and generic material that can be used in a particular subject in a course. Afterwards, the procedure is basically to copy and paste, creating new documents. This process is facilitated by search engines that treat the Web as a vast repository (or library), or even through sites that provide precisely this functionality (“Merlot”, “Cameo” ...).

However, despite this technique being most common used, it is many times hampered by a difficulty in finding the exact piece that is needed to fit in the new created document.

A proposed solution for this problem is the creation of Semantic Digital Libraries as a fundamental tool to deal with information recovering and retrieval where it is not important the presence of the “exact word”, but of the “intended meaning”.

Moreover, we are proposing to move one step further by allowing the search of, not only “static” documents, but also of quizzes, forums, sites, audio and videos.

Our proposal involves the creation of a system that will be able to provide accurate search results by using text mining for a process of automatic classification and categorization of files, for later insertion into an ontology-based storage, thus providing a semantic layer on the top of the digital files stored in repositories, eventually distributed among different locations.

Although our main motivation is to facilitate the access and reuse of digital educational assets in our University, this kind of strategy is also especially useful for areas that require gathering of huge amounts of sample data to be classified, and many comparisons of related documents (describing similar subjects) which often is the case of Life Sciences, particularly Medicine, Biology and Molecular biology.

Research Plan and Goals

To create a Repository integrated in a Federation of several Higher-Education institutions, capable of storing, searching and delivering learning objects, or simple assets, existent in digital form. Specifically, the research will involve:

- A short period of readings for acquiring basic knowledge in Learning Objects, normalization and in e-learning standards

- Creation of a pre-processing of files using text mining techniques for classification and categorization
- Creation of a repository based on semantic segments maintained by an ontology
- Development of an interface for social classification of stored files (folksonomy)
- Development of a protocol for communicating with other repositories
- Evaluation of the work

Contacts:

Prof. Álvaro Reis Figueira

arf@dcc.fc.up.pt

Room: 1.32

Computer Science Department

Rua do Campo Alegre 1021/1055

4169-007 Porto

Faculty of Sciences – University of Porto