Ontology-Based Automatic Annotation of Learning Content

Jelena Jovanović, University of Belgrade, Serbia and Montenegro

Dragan Gašević, Simon Fraser University Surrey, Canada

Vladan Devedžić, University of Belgrade, Serbia and Montenegro

ABSTRACT

This paper presents an ontology-based approach to automatic annotation of learning objects’ (LOs) content units that we tested in TANGRAM, an integrated learning environment for the domain of Intelligent Information Systems. The approach does not primarily focus on automatic annotation of entire LOs, as other relevant solutions do. Instead, it provides a solution for automatic metadata generation for LOs’ components (i.e., smaller, potentially reusable, content units). Here we mainly report on the content-mining algorithms and heuristics applied for determining values of certain metadata elements used to annotate content units. Specifically, the focus is on the following elements: title, description, unique identifier, subject (based on a domain ontology), and pedagogical role (based on an ontology of pedagogical roles). Additionally, as TANGRAM is grounded on an LO content structure ontology that drives the process of an LO decomposition into its constituent content units, each thus generated content unit is implicitly semantically annotated with its role/position in the LO’s structure. Employing such semantic annotations, TANGRAM allows assembling content units into new LOs personalized to the users’ goals, preferences, and learning styles. In order to provide the evaluation of the proposed solution, we describe our experiences with automatic annotation of slide presentations, one of the most common LO types.

Keywords: learning content; learning objects; ontology-based approach

INTRODUCTION

Over the past few years we have witnessed a tremendous amount of activity taking place in the development of Web-based e-learning systems (Mohan & Greer, 2003). A substantial percentage of those activities have been related to learning content authoring. As authoring of high quality learning materials proved to be a highly expensive task in terms of both time and money, reuse of once created learning content soon become one of the hottest research issues. Learning content represented in the form of reusable learning objects (LOs) promised to significantly reduce the time...
and cost of authoring high-quality learning materials, making them more affordable and readily available. The principal objective is to enable faster, cheaper, and better learning (Duval & Hodgins, 2003).

Current research efforts are almost exclusively oriented toward reusability of LOs in their entirety. Annotations of LOs with the standard-compliant metadata sets (e.g., IEEE Learning Object Metadata [LOM, 2002] [LOM] and Dublin Core) aim at enabling search and retrieval of existing LOs stored in LO repositories. Accordingly, metadata is seen as the primary mean for fostering LOs reusability. However, very often a content author needs to reuse just some specific parts of an LO, rather than the entire LO, for example, just a couple of slides out of a slide presentation, or an image or a table out of a text document. Faced with such a need, the content author typically turns to what we call the search-read-copy-paste approach. Specifically, the process of authoring new learning materials typically proceeds in the following steps: an author first searches both LO repositories and the Web to find potentially useful learning content. Then (s)he reads the retrieved materials to determine whether they really contain content relevant for the course under development. Having recognized relevant parts of the retrieved materials, the author copies/pastes them in the new materials (s)he is authoring. The process finishes by fine-tuning content units collected from different sources and optionally adding some new, original contents. Obviously, the content authoring process demands an LOt of time and effort. Additionally, it is not scalable in terms of maintenance (Verbert, Jovanović, Gašević, & Duval, 2005). (Semi-)automating reuse of individual components of LOs’ can improve the current practice by reducing the effort that content authors put in preparation of learning materials. However, an approach to such a kind of automation is still an open question.

To enable reusability of content units of varying granularity levels, an explicit definition of the LO’s structure is needed. Additionally, if the process of reusing content units has to be (semi-)automatic, the definition of the LO’s structure must be formally specified and expressed in a machine understandable language. Furthermore, to facilitate search and retrieval of content units based on the semantics of their content, those content units must be semantically annotated, that is, semantic metadata must be attached to them. Ontologies and Semantic Web languages provide means to achieve both things.

In this paper we present our approach to automatic annotation of LOs’ components based on a number of ontologies. The approach is tested in TANGRAM — an integrated learning environment for the domain of Intelligent Information Systems (IIS).

PROBLEM STATEMENT

The objectives of this paper are:

- To present the rational for using Semantic Web technologies, ontologies in particular, to annotate LOs and their components, and thus facilitate LOs reusability at the component (content unit) level;
- To present how automatic semantic annotation is implemented in a practical learning environment — TANGRAM — developed applying Semantic Web technologies to enable reusability at the level of LOs’ components;
- To discuss our experiences with automatic annotation of individual content units.

The principles we discuss are implementation-independent. On the other hand, their implementation in TANGRAM helped us reveal important practical details and problems we were not aware of initially.

The rest of the paper is structured to follow the order of the objectives stated above.

THE RATIONAL FOR SEMANTIC ANNOTATION OF LOS

The starting point in our approach to ontology-based LO annotation is the classifi-
Related Content

Semantics Discovery via Human Computation Games
www.igi-global.com/article/semantics-discovery-via-human-computation/62561?camid=4v1a

Semantics for the Semantic Web: The Implicit, the Formal and the Powerful
www.igi-global.com/article/semantics-semantic-web/2802?camid=4v1a
Enhancing Folksonomy-Based Content Retrieval with Semantic Web Technology
www.igi-global.com/chapter/enhancing-folksonomy-based-content-retrieval/64022?camid=4v1a

Application of Web 2.0 Technology for Clinical Training
www.igi-global.com/chapter/application-web-technology-clinical-training/29293?camid=4v1a