Using Semantic Technologies in E-Learning Platforms: a Case Study

Paolo Bouquet¹, Andrea Molinari²,³

¹ Dept. of Information Engineering and Computer Science - University of Trento, Italy
² Dept. of Industrial Engineering - University of Trento, Italy
³ School of Industrial Engineering and Management, Lappeenranta University of Technology, Finland

Abstract—The research presented in this paper aims at taking advantage of emerging semantic technologies in a specific part of Learning Management Systems, specifically where all the contents are stored. Semantic technologies have been studied and used in different areas of computer science. E-learning has been one of this, but the most frequent use of semantic technologies has been in the extraction and indexing of contents, like forums, blogs, learning objects etc. Our research follows a different approach, i.e., using semantics technologies as a support, if not an entire replacement, of the backend and persistence mechanisms of Learning Management Systems (LMS).

Index Terms—Learning Management Systems, Semantic Technologies, Learning Objects

I. INTRODUCTION

In this paper we will present a line of investigation regarding the application of semantic technologies to “Online Communities” (OC), a virtual communities platform that we are using in several projects with public and private institutions to support educational processes. E-learning and virtual communities have been an interesting research and application field for semantic technologies, but most of the attention has been devoted to tools and techniques able to enrich, categorize and retrieve contents inside e-learning platforms. Minor attention has been devoted to semantics technologies as a replacement, of the backend and persistence mechanisms of Learning Management Systems (LMS).

In this paper we present the design and early results of the application of semantic technologies to the persistence layer of a Virtual Communities System, where these technologies enrich the platform to address two fundamental issues: a) entities disambiguation and identification inside the persisted objects; b) adding new features to the platform without refactoring it, by using graph-based representation techniques to add knowledge to the existing datasets. In order to address the first item, our implementation takes advantage of the entity-centric tools developed in the Okkam EU-funded project. These tools provide a solution to uniquely and permanently identify entities (people, locations, organizations etc.) inside contents, specifically using the Entity Name System (ENS). The ENS supplies a persistent identifier, called OKKAMid, to any entity included in the knowledge base, together with advanced entity matching methods for detecting the occurrence of the same entity in different contexts and data sources. Once the entity in the LMS has been profiled, it is possible to connect any content where the different occurrences of the same entity have been mentioned inside the LMS, and also to connect any other content outside the learning platform where that entity has been named, for example web pages or social network contents.

Furthermore, we extended a domain ontology (SIOC) for the conceptual representation needs of the application, and added an RDF graph mapped onto the database, in order to add new functionalities to our virtual community platform. This allows us to take advantage of the inference processes available through a reasoner, and to substitute some parts of the business logic of the application.

To improve the semantic representation of concepts and to allow the mechanisms to interface with this knowledge base, RDF data stores have been implemented. Changing the traditional database-oriented representation of data towards a richer format (i.e. RDF statements), other than a richer representation model, gives a high level of flexibility in the definition of persistence and data representation models. RDF has been selected for data representation, while tools like Hadoop, Flink and NoSQL databases have been used and tested as a replacement of “traditional” relational databases, and a bridge towards big data scenarios. These tools and techniques have been experimented in other application fields, specifically in data integration inside the fiscal evasion domain, where different data sources are reconciled through the use of the ENS for entity disambiguation.

The paper is divided as follows. The next section will be devoted to a quick overview of how semantic technologies have been used in e-learning. The third section will briefly present the virtual communities platform where we are experimenting the entity-centric approach to e-learning platforms. The fourth section will present a summary of the entity-centric approach promoted by the OKKAM project, and discuss how this approach can be applied for our refactoring and extension objectives. The last section will briefly present the achievements obtained with the application of semantic technologies and an entity-centric approach to our virtual communities platform.
II. E-LEARNING AND SEMANTIC TECHNOLOGIES: STATE OF ART

Most of educational institutions have recognized e-learning resources as fundamental assets for their training processes, mainly for the capabilities of delivering educational contents to participants over the Internet anytime and anywhere at competitive costs[1]. In these institutions, we can find many different implementations and customizations of available approaches (from blended to full online e-learning) and tools (platforms like LMS, technologies like videoconference, standards for learning objects (Los) metadata like LOM or SCORM [2][3]).

The maturity level of e-learning is visible also in the increasing amount of educational material available under various forms, and in the availability of many Massive Open Online Courses (MOOCs) involving people around the world. Considering the complexity of learning tasks, any new technology that can help to accelerate and improve these processes, catalyzes the attention of specialists. Semantic (web) technologies made no exception, providing a landscape where web-based information and services are understood, processed, interlinked and reused not only by humans but also by machines [4]. Consequently, content classification, retrieval and enrichment through the use of knowledge representation instruments, like vocabularies, taxonomies and ontologies, have been the widest application fields of semantic technologies inside e-learning contexts.

Another perspective sees the semantic web as the possible implementation of a reliable, large-scale environment of machine-understandable and interoperable services that intelligent agents can discover, execute, and compose automatically [5]. Other researchers used semantic technologies to build a brand new generation of learning applications from scratch, or to enrich existing software platforms that deal with educational settings [6]. In any case, despite the excellent opportunities of combining e-Learning platforms with semantic technologies, there is no magic solution to exploit this integration: spreading the “magic power” of semantic technologies over a LMS does not guarantee measurable improvements.

Currently, different educational standards for describing contents in learning resources exist, and a number of organizations have been involved in producing metadata standards specifically for learning technology: SCORM, IEEE LOM and IMS Learning Resource Meta-data Specification can be identified as the commonest and most robust ones[7]. Standard metadata are used by IEEE-LOM, mainly for interoperability between different LMSs but, unlike RDF based metadata, the standard only allows for a hierarchical structure. Semi-semantic metadata extend the IEEE-LOM standard with some semantic component, for example extending the relational field in the standard with a semantic net to interconnect different LOs [8], or adding term associated to some pedagogical or domain ontologies [9]. Semantic metadata can be defined as “…the process of attaching semantic descriptions to Web resources by linking them to a number of classes and properties defined in Ontologies” [10]. Applications using semantic metadata rely on domain ontologies to define their metadata using RDF to express the semantics of a learning resource.

There are several advantages of using RDF over the standard metadata approach [11]: a) an RDF data model is based on the assumption of selecting metadata potentially from heterogeneous ontologies, while standard metadata are taken from a LOM-based, closed-world approach confining metadata to the particular LMS implementation; b) with RDF, complex statement can be created, thus expressing logical networks, while LOM can express simple composition of statement possibly extended through taxonomic classification; c) simple forms of inference (e.g. class inheritance, consistency check, transitivity) can be applied, this way reducing the cost of developing ad hoc solutions to implement the functions which requires them.

Annotating LOs is therefore a fundamental task to guarantee and facilitate access, sharing and reuse of the learning resource. Annotation is also a keyword for the semantic world; in fact, annotated contents transform a full text to be scanned by keyword into a structured, semantically-enabled content. However, there are some obstacles to use structured learning material as a perfect knowledge base for learning activities.

Firstly, most LOs have not been enriched with metadata, or have been enriched with automatic, title-related or filename-related attributes that are semantically poor and sometimes counterproductive. Second, learning objects are not the only source of knowledge inside a LMS, and LMS platforms are not built just of learning objects. Web 2.0 tools and services, like blogs, wikis, forums, FAQs, glossaries, questionnaires etc. are most of the time very useful for the conduction of the learning process [12], especially in educational paths with a high degree of interaction among participants and instructors.

As a further element, organizations can replace their tools and platforms along time, but the investments on LOs should be preserved. This means that a great attention should be paid to content and data interoperability and migration. For those materials that have been created using some standard, the problem should not exist, but for other contents like those created with Web 2.0 tools, the availability of an RDF-based representation simplifies the mapping process between data schemas of different e-Learning platforms [13][14] thus facilitating contents migration among different LMSs. As LOs could be very complex multimedia artifacts, these problems could be frustrating for any interchange of educational material.

Another aspect where semantic technologies could play a fundamental role in learning settings is the addition of search capabilities to a LMS. The integration of semantic technologies is mainly devoted to get meaningful results from user queries about the knowledge base managed by the LMS itself. Parts of this Knowledge Base that could be affected by semantic categorization could be contents, course materials, students’ profiles, etc. [15].

A field where semantic technologies frequently intercept...
approximately 15,000 students in our University.

The core of the application is in abstract entities called “Virtual Communities”, i.e., an aggregation of people to which some collaboration and communication services are available in order to obtain certain objectives [25]. (Virtual) Communities can be aggregated into larger ones, with hierarchical relationships and unlimited nesting levels, thus allowing to represent hierarchies between different types of communities (such as Faculties, Didactic Paths, Master Degrees, Courses, Departments, Organization charts etc.). Each Community has a certain number of services, applications that enable users to communicate in synchronous and asynchronous ways, publish contents, exchange files, coordinate events, etc. The community is a container ready for training processes, but not only: research teams, recreation groups, groups of friends / colleagues / classmates, secretariats, board of directors, sport teams, anything representable as an aggregation of people around a scope could take advantage of OC services.

From the viewpoint of the application’s persistence layer, semantic representation standards like RDF and OWL could be put aside relational representation in order to extend and empower it, allowing the implementation of new features that would be otherwise very expensive in terms of software refactoring. Apart from this economic consideration, other aspects have stimulated the development of this integration:

- the relationship among virtual communities, both hierarchical and transversal, can be more expressive than those implemented in relational databases;
- the implementation of services that use “relationship” among communities in order to inherit contents and services from related communities (files, wiki, FAQ, forums etc.);
- a graph-based navigation interface for the end-users, based on an RDF graph and interpreted through an ontology starting from the SIOC ontology [26];
- the categorization of communities and contents through a tagging mechanism that allows aggregation of any object inside the platform;
- the implementation of an inferential process to access to objects/services available from parent communities, allowing the creation, for example, of “transversal wikis”, “inherited blogs”, “parent’s files”, “similar communities” and so on;
- last but not least, the availability of a development team with a deep knowledge of the source code of the application, completely developed from scratch;

For any of the above needs, traditionally engineered software application can provide solutions. In OC, most of these mechanisms are “hard coded” in the business logic and in the persistence layer of the application, thus creating relevant issues:

- query performance with dramatically poor response time when traversing hierarchies for cross-community learning objects;
- complexity in managing graph-based representations;
- user interface for rendering inheritance and graph
relationships among objects.

Building an RDF graph representing relevant (labeled) connections among objects can open new scenarios, especially when applying reasoning thus allowing the inference of logical consequences in the knowledge base. Another example regards users and their management of contact lists. This is different respect to managing community members as a list of “friends”, or respect to connecting people inside the platform and outside the platform through the FOAF ontology [27]. As a final example, users can have a lot of communities' enrolments over time. These communities have contents that could be related to what the user needs, but with no possibility to be inter-related with other contents in other communities. Thanks to classification and semantic representation of relationships among communities, we can create now different views and aggregation of communities (fig. 1), and in the future any other content that has been semantically tagged.

![Fig. 1: the tile view of tagged communities](image)

IV ENABLING SEMANTIC TECHNOLOGIES IN LEARNING MANAGEMENT SYSTEM

In this section, we describe how we are modifying OC with semantic technologies to achieve the results discussed above. One of the most evident advantage for the learning ecosystem was to start integrating semantic technologies with the persistence layer, in order to achieve the following results:

- linking internal contents with contents that are external to the platform;
- the creation of new “semantic-enabled” services;
- the replacement of existing services with new semantically-enriched services;
- the improvement of existing services through the use of semantic technologies;
- the provision of new graph-based navigation in the platform’s entities;
- inference on the facts represented in the semantic knowledge base.

The work is inspired by the results produced by the OKKAM project, a Large Scale Integrating Project co-funded by the European Commission between 2008 and 2010, where the core technology is the Entity Name System (ENS) [28]. The main purpose of the ENS is to provide unique and uniform names for entities for the use in information collections, so that the same name (identifier) is used for an entity, even when it is referenced in different contexts, inside and outside our platform. By reconciling entities inside the OC database, it is possible to tag “entities” inside the contents with a globally unique and persistent identifier (technically, an HTTP URI), this allowing the integration of OC with linked data available on the web.

Recognizing that information from different sources refers to the same (real world) entity is a crucial challenge in instance-level information integration, as it is a pre-requisite for combining the information about one entity from different sources. The first conceptual block is the adoption of what we called “entity-centric” view at the level of data and content in the persistence layer. The hypothesis is the following: if all the contents inside the eLearning platform, inside the organization's information system and inside the external resources were “entity-centric” (i.e., use the notion of entity for annotating and classifying data and content), the linking process among instances of the same entity through the different data sources would be straightforward. In a nutshell, this was the vision behind the Okkam vision.

The ENS is a global service (centralized or distributed, application- or private- or public-oriented) that acts as an authority for recognizing and disambiguating every single entity, storing some specific attributes of the entity together with a unique identifier assigned to that entity. The ENS should “cut to the root the proliferation of unnecessary new identifiers for naming the entities which already have a public identifier” (inspired by the well-known Ockham's razor from the XIV century).

The first step towards the creation of a new semantic layer for OC is coupling primary keys in our database with OKKAM identifiers as provided by the ENS. We identify any relevant entity (person, place, organization, object, event, etc.) in different columns of the database tables using advanced machine learning techniques, customized for each type of entity. This is in order to be sure that (i) the entity is recognized by the system (the “things, not strings” concept) and (ii) the same entity is always recognized as being the same entity in any type of data and context (entity resolution).

Identifying and annotating entities with OKKAM identifiers allows us to overcome several limitations of primary key usage in entity identification. While a primary key in a database represents an identifier for that record, many problems could arise from the usage of primary keys as entity identifiers:

- if the entity “XYZ” is mentioned for any reason in any content of the LMS, this will be simply taken as descriptive data, and cannot easily be connected to the other occurrences by simply annotating the text with the primary key of our table;
- identifiers generated in this way must be kept aligned with other applications of the information system where that entity is referred;
- identifiers will be invalid/useless outside the scope of the application (e.g. a LMS) where they have been created.
• identifiers should be forced by the application as a foreign key every time a relationship between our tables and other tables is established, not forgetting that this referential integrity must be reinforced by the DBMS.

• Finally, if our entity “XYZ” is present in other information systems, a) the two applications should share and preserve a common identifier, or b) the different development/management teams need to share/exchange/align the different unique IDs, or c) the two systems will identify “XYZ” with different identifiers, thus preventing the idea of interlinked data.

This last point is particularly relevant for our argumentation, as it is one of our main objectives in extending OC towards the semantic web. Indeed, the metadata about an entity in the ENS contains also a list of pre-existing web URIs for data about an entity on the web (mappings between an OKKAMid and other linked data URIs). So an entity in our system can be easily linked not only to any occurrence in the system itself, but also with external data which can be fetched and integrated with a simple HTTP call to other datasets about it. This is exactly paving the way towards the “web of entities” envisioned by the OKKAM project that we are embracing and presenting in this paper.

Entity identification and mapping is not enough to enable a full-fledged semantic application. Ontologies can be very helpful in content interpretation and integration. Also the different level of knowledge and lexicon between teacher and learner can complicate the relationship and the learning processes, also here an ontology can help a lot in sharing and transmitting understanding and knowledge. Another interesting application of ontologies in “traditional” software is to map the columns of DB tables onto concepts represented in the ontology, i.e., resolving differences among heterogeneous databases from different domains using different concepts to represent the same entity.

Using ontologies to associate unambiguous content to columns is a common approach in semantic data integration [29]. In this specific field, the ontology is a middle layer mainly used to map references to the same concepts among multiple data sources. Several techniques of schema matching have been presented, aimed at mapping elements of different database schemas that are semantically corresponding to each other in order to enable co-processing of data collected against different models.

Semantic integration of data models should also provide a way to interpret relationships between entities. The point presented in this paper sees the integration of ontologies with the Okkam entity-centric approach, useful to connect data from multiple structured and unstructured data sources referring to the same entity. What has been done so far in “okkamizing” the persistence layer of “Online Communities” platform is the extraction of primary entities from database tables, and some experiments in the analysis of forum and communities. For the purposes of the prototype we extended a very famous ontology, i.e., SIOC (Semantically-Interlinked Online Communities) to our needs. SIOC provides a Semantic Web ontology for representing rich data from the Social Web in RDF, and is commonly used together with FOAF vocabulary in order to conceptualize and present personal profiles and social networking information. We therefore used SIOC to ontologically describe some services existing in the platform, and we extended SIOC with time, events and other specific concepts available in services present in OC and not provided by SIOC or by social networks.

As a final results, we applied our idea of semantically transforming a “legacy” application into a semantic application starting from some contents of the database, mapping this part onto the ENS, adding OkkamID to entities found in this part of contents, and then creating an RDF graph with the mapped portion of the DB. This knowledge base is navigable with a browser and queryable via SPARQL.

The addition of an entity-centric approach will allow to identify entities inside contents of the platform and connect these contents with the URI referring to the same entity, thus creating a true linked data environment for e-learning. In a near future, the improvements obtained using Okkam’s entity-centric approach will be quickly usable inside OC. The following are some examples of an entity-based enrichment of the knowledge base, where we can search for an Entity (not for a text) inside many different contents and services of the platform. What we did was to transfer the structure of communities and of wikis in the triple store, insert of a rule by which the reasoner infers new triples, and then prepare a web page where the prototype shows the list of Wikis and their communities before and after inference.

V. CONCLUSION AND FUTURE DEVELOPMENTS

In this paper we presented a short description of experiments regarding how semantic technologies could be coupled with e-learning software, with some final considerations about the application of semantic technologies to our collaboration platform that, thanks to the metaphor of virtual communities, facilitated this kind of integration and evolution.

The early implementation revealed two of the most interesting aspects we wanted to test, i.e., the ease of implementing and/or new functionalities to a software platform when semantic technologies are wisely integrated with them. A partial objective that has been investigated regards the identification of those issues that could be generalized to any initiative aimed at extending a software platform. The idea is to put semantic technologies aside with traditional development technologies, identifying entities in contents, using persistent identifiers for storing their unique ID taking from an entities’ central repository called ENS.

Some further elements should be investigated, in order to have a clear vision about pros and cons of this approach, specifically regarding sustainability and investments in re-factoring e-learning applications, and in
general software applications. A second item that emerged from our experiments regards performances of semantic persistence layers. Intensive tests have been conducted using the same big data-oriented technologies (Hadoop, Hbase, Flink etc.) but in different contexts, where big-data range of operation is required. As results are very encouraging, we have conducted some preliminary tests with the knowledge base available in the virtual community platform. Another items regards the exact moment where to align the database and the semantic store in case some management operations are available to the public. Finally, the potential that semantic technologies could provide to e-learning and collaboration in general is very vast, but the problem here is the usability of contents and services. Enriched by this semantic functionalities and contents, the risk is that they could become complex to be understood and therefore unusable. A great help will come from semantic tools for data visualization and navigation.

REFERENCES


AUTHORS

Paolo Bouquet graduated in Philosophy in 1991 (University of Milan) and then got his PhD in Epistemology in 1997 (University of Genoa). In 2004 he became Associate Professor in Computer Science at the University of Trento (email: paolo.bouquet@unitn.it)

Andrea Molinari, graduated in Economics (University of Trento), is contract professor at University of Trento, Bolzano and Abo Akademi (Finland) (email: andrea.molinari@unitn.it)

Manuscript received 15 March 2016

Published as submitted by the author(s)