Semantic Technologies and e-Learning: Towards an entity-centric approach for Learning Management Systems

Paolo Bouquet[1,3], Andrea Molinari[2,3]

[1] Dept. of Information Engineering and Computer Science
[2] Dept. of Computer and Management Sciences
University of Trento, Italy
[3] OKKAM s.r.l. – Trento, Italy
{paolo.bouquet, andrea.molinari}@unitn.it

In the past, the application of semantic technologies to educational settings attracted a lot of attention, in particular regarding those approaches and software tools able to enrich, categorize and retrieve learning objects. In this paper we present a different perspective, specifically the evolution of a traditionally engineered learning platform into a semantic-enabled application. The semantic enrichment addresses two issues: a) the unique identification of entities inside the contents and b) adding new features to the platform without refactoring it through the usage of semantic representation of information. We believe that these two modifications are missing elements for semantic learning, and in order to apply these changes, we integrated our Virtual communities platform with the entity-centric tools developed in the Okkam EU-funded project. Okkam tools provide a solution to uniquely identifying entities inside data and contents, specifically using the Entity Name System (ENS), which supplies a persistent identifier, called OKKAMid,

to any entity included in the dataset and advanced entity matching methods for detecting the occurrence of the same entity in different contexts. This allows a deeper connection between entities' information available inside content, inside the platform and outside the platform. 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 used, and also to connect any other content outside the learning platform where that entity has been identified, for example web pages or social networks. Furthermore, we extended a domain ontology (SIOC) for the conceptual representation needs of the application, and created an RDF graph mapped onto the database in order to add new functionalities to our virtual community platform, taking advantage of the inference processes available through a reasoner and substituting some parts of the business logic of the application.

1 Introduction

Currently, educational institutions have recognized e-learning and web-based learning resources as fundamental elements of their training processes, mainly for the capabilities of delivering educational contents to participants over the Internet anytime and anywhere at competitive costs (Leghris *et al.*, 2006). Almost any educational institution has at least considered using one of the different available approaches, blended or full online, and tools that the e-learning world can offer today: platforms like Learning Management Systems-LMS, technologies like videoconference, standards for learning metadata or objects like SCORM (SCORM, 2004) or LOM (IEEE, 2002; IMS, 2012; Friesen, 2004; LTSC, 2002). Another effect of the maturity of e-learning regards the large amount of educational material (Learning Objects - LO) that has been produced and that is now available under various forms.

Learning is a complex task, so any new technology that could help to improve these processes, from mobile computing to edutainment devices, has been considered in educational field. The advent of semantic (web) technologies makes no exception to this rule, as they have a special impact on how we categorize, search and retrieve information online (Sampson *et al.*, 2004). The World Wide Web was conceived for linking documents to be read and interpreted by humans, not for machine consumption. So even if the Web is machine-readable, it is not machine-understandable (Lassila, 1998). The fascinating idea of an appropriate distributed infrastructure with intelligent agents going around the Web while performing complex tasks for the learner is a scenario that has been analysed quite deeply in research studies (Primo *et al.*, 2011; Celik *et al.*, 2011). Another perspective, not relevant for our argumentation but surely interesting for e-learning, sees the semantic web as the possible implementation of a reliable, large-scale interoperation of machine-understandable and interoperable services that intelligent agents can discover, execute, and compose automatically (McIIraith, 2001; Moreale *et al.*, 2004). Other research directions have been followed as possible perspectives for semantic technologies in e-

learning, specifically for web based learning-related applications (Khamnayev, 2007; Anderson *et al.*, 2004; Ghaleb *et al.*, 2006). Semantic technologies could be used to build from scratch a brand new generation of learning applications, or possibly to enrich existing software platforms that deal with educational settings, like LMS, LCMS, Virtual communities, Document Management Systems, Workflow engines, and all those software used, at different levels, to support learning tasks (Pandit, 2010).

In this paper we will present a line of investigation regarding the application of semantic technologies to "Online Communities", a virtual communities platform that we are using in several projects with public and private institutions to mainly support educational processes (Colazzo *et al.*, 2009a). The implementation of semantic tools have been facilitated by some intrinsic features of the platform:

- the relationship among virtual communities, both hierarchical and trasversal, needed more computable and more expressive relationships;
- services that inherit contents and permissions from related communities (Colazzo *et al.*, 2009b): these would be features otherwise very expensive in terms of maintenance costs and management;
- being the platform created from scratch, the development team has the total knowledge of the source code of the application.

These assets have been used for a unique approach to extending the software, thus simplifying the management and integration of entities inside the content, i.e., any structured (coming from the relational backend of the platform) or unstructured information available in the platform. This entity-centric approach and a set of tools coming from the OKKAM project (Bouquet *et al.*, 2008) will be briefly presented, and will be applied on different components to achieve these main results:

- a graph-based navigation on parts of the database exported in RDF and interpreted through an ontology that extends SIOC ontology (SIOC, 2010);
- services that extend the visibility of files in different communities respect to the owner's one, respecting parent-child relationships. This allows, for example, child communities' members to see parent community's files without duplicating or moving the files;
- implementation of inference processes to modify the visibility of web 2.0 services for parent communities, thus allowing the creation, for example, of "trasversal wikis", "inherited blogs" and so on (Colazzo *et al.*, *op. cit.*).

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 present the virtual communities platform where we applied the entity-centric approach. 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.

2 Semantic web and E-learning

"Semantic technologies" is one of the many buzzwords that passed from the research arena to industry, with some interesting application fields and other twisting that could damage the positive aspects. The web is a perfect example of the potential of introducing "semantics" into computer-based applications. Today we are able to extract information from heterogeneously structured Web documents using keyword-based search engines that use statistical or reputation-based approaches. The traditional objection to this is that the information contained in web pages is human-readable, but not easily computer-processable, specifically for knowledge discovery and inference processes (Berners Lee, 1998). To overcome these difficulties, new languages, protocols and standards have been created to enable the World Wide Web to become the Semantic Web, with the objective of enabling machines to access more information automatically. E-learning, on the other side, is a well-established, long experienced research and application field, with many years of research threads, successful stories and embarrassing pitfalls. Today the discipline can take off thanks to a market that is overtaking the traditional educational institutions, a set of technologies that provide rich media contents to users, and an adequate capillarity of network connection combined with acceptable bandwidth (at least in developed countries).

Content management always represented one of the critical issues for e-learning, particularly contents' standardization and interoperability. 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 (Robson, 2000). The representation of metadata in e-learning applications (Al Khalifa *et al.*, 2006) has been classified into three categories: a) Standard Metadata, b) Semi-Semantic Metadata c) Semantic Metadata. Standard metadata are used by the IEEE-LOM standard, mainly for interoperability between different LMSs, while semi-semantic metadata

extend the IEEE-LOM standard with some semantic component (Engelhardt *et al.*, 2006), for example adding term associated to some pedagogical or domain ontologies (Sancho, 2005). 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" (Scerri *et al.*, 2005). Applications using semantic metadata rely on domain ontologies to define their metadata, using RDF to express the semantics of a learning resource (Brooks *et al.*, 2006), thus allowing simple forms of inference (e.g. class inheritance, consistency check, transitivity) and reducing the costs of developing ad hoc solutions. Annotating LOs is therefore a fundamental task in order to guarantee and facilitate access, sharing and reuse of the learning resource. Annotation is also a keyword for the semantic world; 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.

As a first element, most LO have not been enriched with metadata, or have been enriched with automatic, title-related or filename-related attributes that are semantically poor and sometimes even counterproductive. Second, learning objects are not the only source of knowledge inside a LMS. Web 2.0 tools and services, like blogs, wikis, forums, FAQs, glossaries, questionnaires etc. are very useful for the conduction of the learning process (Khamnayev, 2007) especially in educational paths with a high degree of interaction among participants and instructors. Furthermore, the connection of information contained inside a platform with the external information space would be today a very powerful added value for learning processes. Connecting a LO with materials available from the web requires a manual effort by the instructor/learner with search engines, and then this relationship should be "hardwired" inside the educational material. Last but not least, organizations can replace their tools and platforms along time, but the investments on LOs should be preserved. For those materials that have been created under some standard's umbrella, the problem should not exist, but for other contents (like Web 2.0 contents), the availability of an RDF representation simplifies the mapping process between schemas of different e-Learning platforms (Abels *et al.*, 2009; Garcia *et al.*, 2009) thus facilitating contents migration.

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 (KB) managed by the LMS itself. Parts of such a KB that could be affected by semantic categorization could be contents, course materials, students' profiles, etc (Çelik *et al.*, 2011). For example, many different semantic representation using ontological descriptions

could affect materials, users' profile, Courses and Exams etc.

Another direction where semantic technologies intercept the e-learning field is the connection with the WWW. The Semantic Web extends the categorization of existing WWW resources, allowing "computers to intelligently search, combine, and process Web content based on the meaning that this content has to humans" (Hitzler, 2010). There have been several projects and researches that combined these three factors into e-learning systems, focusing on determining the standard architecture and format for learning environments. IMS and SCORM are XML-based interoperable specifications for exchanging and sequencing LOs, and this of course helps the integration with what has been famously illustrated in Tim Berners-Lee's "Semantic Web Stack" representation (Berners Lee, 2003). These standards, however, are trying to model the interoperability of educational information that are relevant to the educational process (Adelsberger, 2003), but not the educational contents knowledge in educational activities. Other authors (Quemanda *et al.*, 2003) used the taxonomy of learning resources and stereotypes of teaching models for educational contents and sequences, but these aspects are heavily platform-dependent and lack standardization and reusability.

Regarding contents and semantic integration, including ontology-based properties and hierarchical semantic associations, e-learning systems enriched with semantic tools could provide adaptable and personalized LO to participants. Furthermore, most of relationships across educational contents ('equivalent', 'inverse', 'similar', 'aggregate', 'classified') can provide useful information for a semantic LMS. The crucial role is played by ontologies, here intended (Gómez-Pérez, 2002) as conceptualizations of a specific domain in terms of concepts, attributes, and relationships. Ontologies enable the representation, processing, sharing and reuse of knowledge among applications. In e-learning settings, they play a crucial role in a number of ontology-centered researches where technology standards, such as XML, RDF(S) and OWL, allow to share and reuse any web-based knowledge system (Devedzic, 2001; Brewster *et al.*,2004).

To sum up, the impact of the merge between e-Learning and semantic technologies is, in our opinion, deep and permanent, for reasons that we shortly presented but that have been largely discussed (Pandit, 2010). Nevertheless, much of the current research seems to limit the discussions on how the semantic web will enable knowledge engineers, instructors or instructional designers to construct elegant ontology-based annotations for web-based resources and Los, and to further expand metadata schemes (Sicilia *et al.*, 2004; Anderson *et al.*, 2004; Wolf *et al.*, 2001). Our approach is less focused on contents and more oriented to a different usage of semantic technologies, for managing unique identification of entities and for the improvement of provided services.

3 Integrating semantic technologies in existing e-learning platforms: a different perspective

The above aspects of semantic e-learning emerged very vividly in our virtual communities platform called "OnLine Communities" (OC) when we started to enrich the platform with tools that connected external resources (like social networks web pages) with platform internal contents.

Since 2003, OC is a collaborative environment totally designed and developed by the Laboratory of Maieutics – University of Trento (Italy) (Colazzo *et al.*, 2011) which aims at supporting cooperative processes, and teaching/ learning activities in particular. The platform is up-and-running since 1998, and has been often modified and updated due to the addition of new services and/ or due to customers' requests for personalization and customization. In these conditions, even if fascinated by new semantic technologies, a rewrite process, but even a refactoring process is not sustainable in terms of costs and time of development. Furthermore, any request of update constantly fights against the need of keeping the platform consistent (forking) on one side, and on the other side with some functionalities that have not been updated at the same level or at the same time all over the years. This is a typical problem of any large scale software application, that on one side tries to be modern and up-to-date, and that on the other side has to come to terms with the past. The core of the application is composed by some abstract entities, called Virtual Communities, viewed as an aggregation of people to which some collaboration and communication services are available in order to obtain certain objectives (Boyd, 2006). In detail, a virtual community (Rheingold, 2000), is a space on the web dedicated to a collaboration objective, populated by people who communicate with each other's, using a collection of communication systems. With this approach, it is possible to represent all the hierarchical relationships between different types of communities (such as an organization chart, or the dependencies between Faculties, Didactic Paths, Master Degrees, Courses, etc.).

The architecture of Online Communities is based on five pillars: *Person*, *Community*, *Service*, *Role* and *Permission*. The combination of roles and permissions defines the Profile for each user. Communities can be aggregated into larger communities with hierarchical mechanisms and unlimited nesting levels. Each Community avails itself of a certain number of services, i.e., applications that enable users to communicate in synchronous and asynchronous ways, publish contents, exchange files, coordinate events, etc. Services of a community are activated by a community manager on demand, and users of a community have different rights on them. OC is not an anonymous platform: all users are profiled and recognized inside the system. Over the years, the system evolved into a platform for professional training oriented to lifelong

learning outside academia, specialized in creating relations with the information system of educational institution. E-Learning platforms seem to act in a restricted circle made up only of teachers, tutors and students. In OC, instead, the community is a container ready for teaching processes, but not only: research teams, recreation groups, friends, secretariats, board of directors, colleagues, anything that could represent an aggregation of people around a scope could be implemented in OC.

Currently, Online Communities is mainly used outside the university campus, serving approximately 40.000 users from different public and private customers against approximately 15.000 students in our University. The platform provides a wide range of articulated functionalities:

- "*traditional*" services: synchronous and asynchronous services (like forum, LO management, newsgroup, chat, streaming audio/video) and "Personalized" Services, closer to life-long learning and "training on the job" needs (tutorship, training on demand, FAQ etc.).
- *Integration services with external information systems* (for example, the Personnel information system of the organization)
- *Services for the fruition of "off-line" courses*, i.e., courses already held and recorded, digitalized and available to controlled communities (synchronize video with slides, podcast, webcast, SCORM modules, etc.).
- Services for the creation of evaluation test, quizzes, polls etc.
- Statistics about the users behaviour (using an internal data warehouse enriched by activity logs)
- *mobile Services*. There are some innovative services which meet the mobility needs of the subject who wants to learn "on the move", performing learning/collaboration activities directly through his/her mobile device.

The platform is constantly extended with new services, coming from research projects, users requests and our intuitions. Among these functionalities, one is particularly frequent in users' requests and in our "future development" discussions, i.e., propagating the visibility of files along hierarchical paths inside the communities, and this extension is particularly suitable for semantic technologies usage. Imagine a situation of communities' relationship like the one presented in figure 1. The "traditional" file management service presents several issues and limitations when requested to provide the following:

1. (A member of) The Community "Project Management" wants to share "File 2" with its sub-communities;
2. (A member of) The Community "Group A" wants to share "File 3" with community "Master Degree in Business Administration";
3. The Dean of the Faculty of Economics delivers the document "File 1"

that should be distributed to all the communities of the Faculty.

The three examples above are very frequent in everyday activities of OC users, and represent the need of the following mechanisms:
- Propagation, i.e., allowing a document to be spread in sub-communities
- Inheritance, i.e., sub communities that want to share documents with super-communities, typically when a student produces a content that could be shared with parent communities.
- "Trasversality", in the sense of linking communities located in different branches of the communities' hierarchy.
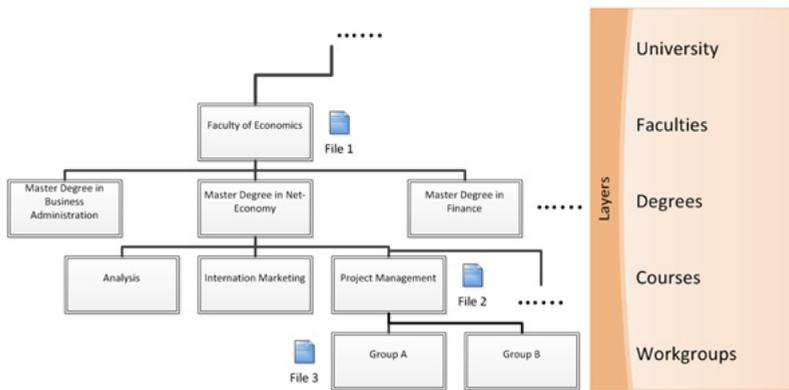

Fig. 1 - example of files' inheritance in Online Communities

The problem is even more compelling, considering that in our platform, the inheritance mechanism is applicable not only to communities or documents, but also to contents (in the sense of posts of a forum, entries in a FAQ, comments in a wiki etc.) and, most of all, to services (the user could inherit permissions on a service in sub- or super-communities, like writing permissions in a forum). In all these cases, we have two technical solutions in traditional software:
- Replicating the documents where requested, with obvious drawbacks (disk usage, alignment, complexity, permissions management etc.).
- Create a (soft) link to the document, with a certain number of problems.

All over these mechanisms, permissions and rights on services and documents must be guaranteed for every user. Having these problems, it has been clear that semantic technologies and representation languages provide a very

interesting path to explore, specifically for the set of relations that are created to implement the above features. This perfectly fits into a language of knowledge representation that allows a method of automatic reasoning to provide responses and take decisions. In whatever of the above conditions, traditionally engineered software application can provide solutions. In OC, most of these mechanisms are implemented and "hard coded" in the business logic and in the persistence layer of the application, but we have to face many issues:

- Query performance: we have dramatically poor response time in presence of large datasets;
- The interface for rendering the inheritance relationships: for example, when a document is displayed in a list of the documents of the community, how do we represent documents shared or coming from other communities?
- difficulties in navigating results for the user, due to complexity of the information to be presented.

More than this, the fascinating perspectives provided by semantic representation regard the idea of graph, naturally related with our idea of communities network: building an RDF graph that represents all the possible (labeled) connections among objects can open new scenarios when a reasoner and inference mechanisms are applied to the set of triples. An example regards users and their management of contact lists: it is clearly different to manage community members as a list of "friends", or to connect people enrolled in the platform with the FOAF vocabulary in order to link them inside and outside the platform through the FOAF ontology (FOAF, 2010; Ounnas *et al.*, 2006; Ounnas *et al.*, 2009).

Different possibilities were available for integration semantic technologies into existing e-learning platforms: in the next sections, we will describe a preliminary implementation in OC. In particular, we focused on entity identification and interlinking (step 1) and on the description of structural elements (posts, blogs, authors, etc.) through an ontology (step 2). As a side effect, we also present a general method for migrating legacy data and applications into the semantic world, a very important step for avoiding the costs of a reimplementation of a system to deploy new semantic-enabled services.

4 An entity-centric semantic view for a LMS

The first step was to adopt the *entity-centric* view on OC. This is a key element of the Semantic Web (the centrality of the notion of resource in its models and languages). However, it was reinterpreted and strongly promoted by the EU co-funded project "OKKAM – Enabling the Web of Entities"[1],

---

[1] See http://fp7.okkam.org/. The name of the project was inspired by the objective of cutting to the root the proliferation of

whose goals were:

- to provide an open and scalable service (called Entity Name System-ENS) for creating and managing persistent identifiers (called OKKAM ids) for any entity named on the web (technically, OKKAM ids are HTTP URIs);
- to support the largest possible reuse of these identifiers across any format, application and context by enabling a service for entity matching (i.e. detecting when an entity description corresponds to an entity whose OKKAM id is already stored in the ENS);
- to provide a first interoperability service by mapping the OKKAM id of each entity to any known pre-existing identifier for the same entity.

All these services are offered through a Web interface[2] or through a collection of documented APIs[3].The adoption of the OKKAM approach and the corresponding technologies for rrecognizing that information from different sources refers to the same (real world) entity is a fundamental step in instance-level information integration, as it is a pre-requisite for combining the information about one entity from different sources. The systematic use of the ENS reduces the entity matching problem to spotting the same entity identifier in different information collections. Enabling this entity-centric approach in OC required some work on the persistence layer of the application, in order to achieve a more powerful inter-linking of contents (documents, posts, events, etc.) inside the platform and with additional contents inside or even outside an organization's information system, thus allowing a controlled integration of the information system with external resources. The underlying hypothesis was that, if all the contents inside the e-learning platform, inside the organization's information system and inside the external resources were "entity-centric", then the linking process among instances of the same entity through the different data sources would be straightforward (as the recent launch of the "Knowledge Graph: things, not strings" by Google proves perfectly well).

The first step towards this entity-centric approach was replacing primary keys in our database with OKKAM ids as provided by the ENS. We needed to recognize any relevant entity (person, place, organization, object, event, etc.) in our data and make sure that (i) it is recognized as an entity by the system (the "things, not strings" concept) and (ii) the same entity is always recognized

---

unnecessary new identifiers for naming the entities which already have a public identifier (the OKKAM's razor, of course inspired by the well-known Ockham's razor from the XIV century philosopher). OKKAM was co-funded between 2008 and 2010.

[2] See http://api.okkam.org/

[3] See http://api.okkam.org/okkam-core/WebServices?wsdl

as being the same entity in any type of data and context (entity matching)[4]. Identifying entities and annotating contents with OKKAM ids inside records allows us to overtake several limitations of primary key usage in entity identification. Indeed, a primary key in a database represents an identifier for that record, namely for a partial representation for the entity itself. Consider, for example, a database containing a table for "Teachers", where each "teacher" is assigned a primary key as a unique identifier by the system itself. So the teacher "Alice Bob" will be assigned a unique identifier (or recently, a UUID/GUID, which can solve the local problem of uniquely identify the record). There are several reasons why this is not an adequate solution for identifying the real-world entity "Alice Bob":

1. unique identifiers managed this way must be aligned across the information system, e.g. when in one of the components this unique identifier is modified.
2. they are meaningless outside the application where they have been created.
3. they should be forced as foreign keys every time a relationship between table "Teachers" and other tables (for example "Courses") must be established, not forgetting that this referential integrity must be enforced by the DBMS.
4. Most important, if the entity "Alice Bob" is mentioned in any unstructured content of the LMS, this will be simply taken as descriptive data, and cannot be connected to other occurrences by simply annotating the text with the primary key, as primary keys are unknown to most users and are meant to be used only by the DBMS for record identification.
5. Finally, if our teacher "Alice Bob" is present in other information systems, for example in the ERP records: a) in the best case scenario, the two applications should share and preserve a common identifier; b) in an intermediate scenario, the different management teams need to talk and align the different unique IDs; c) in the worst scenario, the two systems will identify "Alice Bob" with different identifiers, thus preventing the idea of interlinked data.

This last point is particularly relevant for our argumentation, as it is one of the main objectives in extending our learning platform 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 OKKAM id and other linked data URIs). So an entity in our system can

---

[4] The detailed description of this process from a technical point of view is beyond the scope of this paper, so we refer the interested reader to a general description available at https://docs.google.com/document/preview?id=1RVQKuqz_xgeyHcyM GSu1e31sTcbZX8cVQ2F3alEkVUQ&pli=1

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 of course not enough to enable a full-fledged semantic application. We need one or more ontologies that describe the interpretation of statements made about entities and enable reasoning about them. In OC, ontologies can be very helpful in content mashups, for example when you have different authors discussing the same discipline and you want to integrate their knowledge. 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, transmitting and understanding. Another interesting usage 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 (Wolf *et al.*, 2001). Using ontologies to unambiguously assign a meaning to columns and map elements of different database schemas that are semantically corresponding to each other (Yi *et al.*, 2005) is a common approach in semantic data integration (Guarino, 1998; Gardner, 2005). Semantic integration of data models (Saake *et al.*,2005) should also provide a way to interpret relationships between entities (Li *et al.*, 2000; Knappl, 1998).

For the purposes of the first implementation we extended a well-known ontology, SIOC (Semantically-Interlinked Online Communities) to fit 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. In fig.2 an example of the extensions of SIOC, limited to the FAQ service provided by the platform.
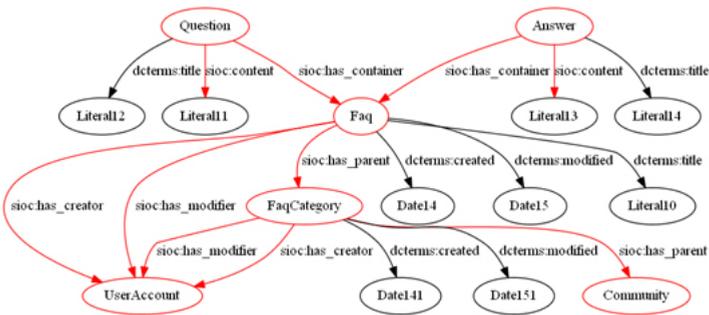
Fig.2 A partial view of SIOC ontology and the extensions added

As a final result, 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. The resulting knowledge base can now be navigated with a browser and queried via SPARQL.

Regarding files and the complexity of allowing a web of relationships among the them, at the moment we are implementing the first prototype of a "visibility manager" for files, whose general functioning schema is presented in figure 3.
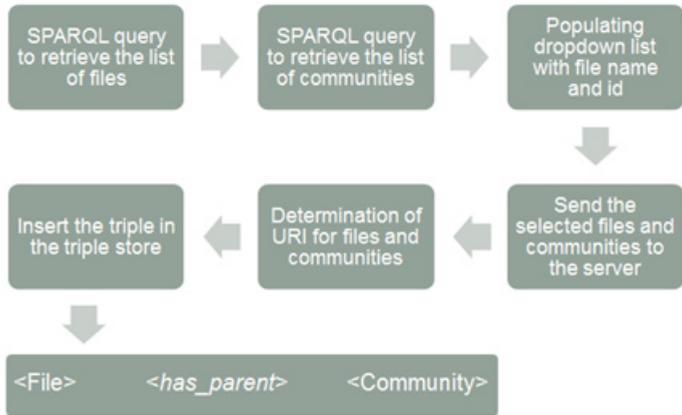

Fig. 3 - General functioning schema for file visibility manager

In a near future, the extensions and improvements obtained using Okkam's entity-centric approach will be implemented relatively quickly inside OC, due

to the existence of a populated ENS that, by growing with new entities, will be the reference point for interlinked data across various types of contents. For example, adding a semantic search engine where we can search by Entity (not by keyword) inside many different contents and services of the platform will be a very powerful and interesting perspective for e-learning environment, for example for educational content aggregation and customized educational paths. As another example of implemented services that, using Okkam technologies, enriched the platform with previously unavailable features, we briefly cite the "trasversal wiki". With this term, we intend the possibility of inheriting wiki's discussions and comments from communities that are hierarchically related, but also to extend this relationship to communities that are in a different branch, thus creating a network effect that bypasses the hierarchy's rigid relationships. We used this concept in the past to magnify the capabilities of our hierarchical structure of communities and extend the learning settings. The implementation of this feature has always been confined in the test versions of the platform, due to feasibility, performance and interface issues. However, the most relevant element here is the complexity that should be implemented inside the business logic of the application, and inside those parts where the aggregation of every object of the wiki is created every time the user interacts with the parent -wiki or any of its sub-components. What we did in the semantic extension was simply to transfer the structure of communities and wikis in the triple store, insert 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.

Concluding, if we look at the market, most of semantic applications available today are built from scratch, thus not having the robustness of a users' long-tested platform. Building a new e-learning management systems with semantic foundations is complex to imagine in terms of investments, due to the availability of valid and robust traditional applications like Moodle™. When dealing with the perspective of increasing the features of the platform with semantic technologies, we believe that another approach should be used, i.e., enriching the existing architecture with semantic extensions where they could perform better. Our platform, with the propagation, inheritance and trasversality mechanism that are native inside data structures and business logic, is well-equipped to be a successful semantically-enabled LMS with a strong orientation towards collaboration. The addition of an entity-centric approach allows 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. So, the final architecture of the semantic application is presented in fig.4:
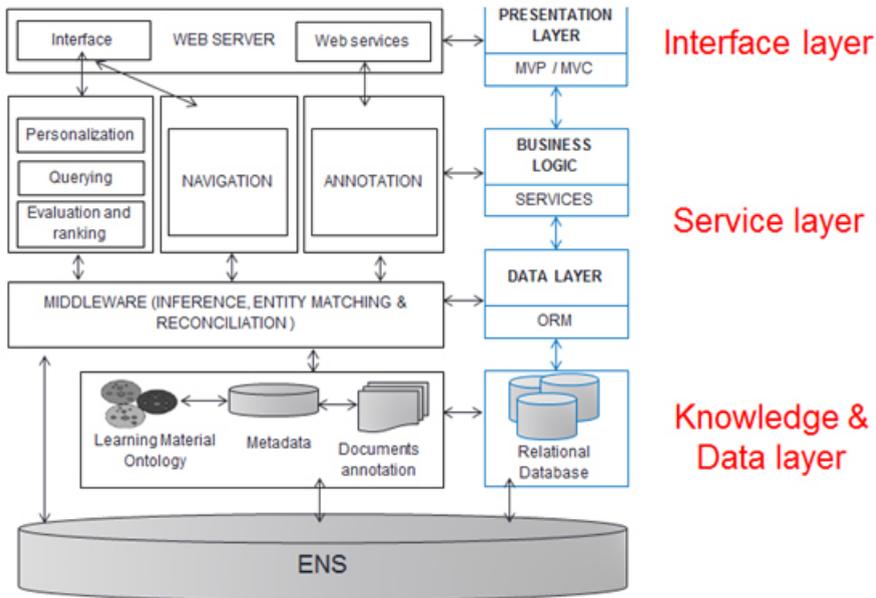
Fig. 4 - The OC semantically-enabled architecture

In this architecture, the "old" part of the software is maintained thus guaranteeing continuity and preserving the investments, without excluding refactoring activities

Conclusion and future developments

In this paper, a new perspective of integrating semantic technologies into the e-learning field is presented, specifically using semantic technologies to extend functionalities provided by software platforms that support educational tasks. We presented a few developments of our e-learning platform "Online Communities" and a set of preliminary implementations which have already provided support to our main thesis, namely the ease and speed of implementation, and the possibility to add features otherwise not feasible or too expensive to be implemented. These improvements have been created using and extending available ontologies related with virtual communities (mainly the SIOC ontology), and adopting the entity-centric approach promoted in the Okkam EU project. In this way, the creation of a graph of content nodes referring to the same entity is straightforward, thus allowing the aggregation of navigable contents for that entity, the creation of more specific search and retrieval tools thanks to a non-ambiguous identification of entities, and the possibility of connecting

information inside the repository of the platform with the information available outside the e-learning environment. The RDF graph thus created contains new relationships among entities without having to further modify the persistence layer and the corresponding application logic, and inference mechanisms can be applied in order to provide new, previously unstated connections among entities available inside the knowledge base. On this growing knowledge base, new software functionalities can be created, otherwise complex or expensive to be introduced. In this direction, what we created so far in a prototypical way is a rough interface of services previously unavailable in Online Communities: wiki and FAQ trasversality, file visibility across parent-child communities, and inference processes on the triple store derived from the persistence layer that provide new information to the end-user. The usage of Okkam entity-centric approach provides the backbone for this linked data-aware environment, and the new services have been built relatively quickly without modifying anything in the original software.

This last point deserves further investigations, as the idea is particularly interesting for software engineers that are facing software extension or refactoring. Some further elements should be investigated, in order to have a clear vision about pros and cons of this approach, but the early functionalities that have been implemented in the "Online Communities" platform seem to be encouraging. Other issues are currently under investigation, first of all performances of semantic persistence layer that should provide these extensions.

REFERENCES

Abels S., Chepegin V., Campbell S. (2009), *Semantic Interoperability for Technology-Enhanced Learning Platforms*. 2009 Ninth IEEE International Conference on Advanced Learning Technologies, 564-568.

Adelsberger H. (2003), *The Essen model: a step towards a standard learning process* http://citeseer.ist.psu.edu/515384.html, 2003.

Al-Khalifa H., Davis H. (2006), *The Evolution of Metadata from Standards to Semantics* in E-Learning Applications HT'06, August 22–25, 2006, Odense, Denmark

Anderson T., Whitelock D. (2004), *The educational semantic web: Visioning and practicing the future of education (special issue)*. Journal of interactive Media in Education,1, 2004. URL http://jime.open.ac.uk/article/2004-1/181.

Berners-Lee. T (1998), *A roadmap to the semantic web*. http://www.w3.org/DesignIssues/Semantic.html.

Berners-Lee T (2003), *WWW past & future, 2003*. Available at http://www.w3.org/2003/Talks/0922-rsoc-tbl/.

Bouquet P., Stoermer H., Niederee C., Mana A. (2008), *Entity Name System: The Backbone of an Open and Scalable Web of Data*. In Proceedings of the IEEE

International Conference on Semantic Computing, ICSC 2008, number CSS-ICSC 2008-4-28-25, pages 554-561. IEEE Computer Society, August 2008.

Boyd D. (2006), *Friends, friendsters and mySpace Top 8: writing community into being on social network sites*. First Monday 11(2), December.

Brewster C. *et al.*(2004), *Knowledge representation with ontologies: the present and future*, IEEE Intelligent Systems, Vol. 19, pp. 72-81, 2004.

Brooks, C. McCalla G. (2006), *Towards flexible learning object metadata*. International Journal of Continuing Engineering Education and Life-Long Learning, 2006. 16(No.1/2): pp. 50 - 63.

Çelik D., Elçi A., Elverici E. (2011), *Finding Suitable Course Material through a Semantic Search Agent for Learning Management Systems of Distance Education*, Computer Software and Applications Conference Workshops (COMPSACW), 2011 IEEE 35th Annual, vol., no., pp.386-391, 18-22 July 2011.

Colazzo L., Molinari A., Maresca P., Stanganelli L. (2009), *Mashup Learning and Learning Communities*. International Workshop on Distance Education Technologies (DET 2009), San Francisco (USA), 10-12/09/2009, p.122-128.

Colazzo L., Molinari A., Villa N. (2011). *Formal and informal lifelong learning in a virtual communities platform*. Lecture Notes in Computer Science including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics.(2011), Volume: 6537 LNCS, Pages: 291-300.

Colazzo L. Molinari A. Villa N.(2009), *Collaboration vs. Participation: The Role of Virtual Communities in a Web 2.0 World*, Education Technology and Computer, 2009. ICETC '09. International Conference on, vol., no., pp.321-325, 17-20 April 2009.

Devedzic V. (2001), *Knowledge Modeling - State of the Art*. Integrated Computer-Aided Engineering 8, 257-281, 2001.

Engelhardt M., Hildebrand A., Lange D., Schmidt T. C. (2006), *Semantic Overlays in Educational Content Networks*, in TERENA Networking Conference. 2006. Catania, Italy.

FOAF (2010), *FOAF Vocabulary Specification 0.98*, Namespace Document 9 August 2010 - Marco Polo Edition at http://xmlns.com/foaf/spec/

Friesen N.(2004), *International LOM survey: Report. Technical report*, Digital Library of Information Science and Technology, July 2004.

Garcia R., Pariente T. (2009), *Interoperability of Learning Objects Copyright in the LUISA Semantic Learning Management System*. Information Systems Management, 26(3), 252–261. Taylor & Francis.

Gardner S. (2005), *Ontologies and semantic data integration*, Drug Discovery Today, Volume 10, Issue 14, 15 July 2005, Pages 1001-1007.

Ghaleb F., Daoud S., Hasna A., ALJa'am J., El-Seoud S., El-Sofany H.(2006), *E-Learning Model Based On Semantic Web Technology*, International Journal of Computing & Information Sciences Vol. 4, No. 2, August 2006, pp. 63 – 71.

Gómez-Pérez A., Corcho O. (2002), *Ontology Languages for the Semantic Web*. IEEE Intelligent Systems 17(1), 54-60, 2002.

Guarino N. (ed.) (1998), *Formal Ontology in Information Systems*. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.

Hitzler P., Krotzsch M., Rudolph S. (2010), *Foundations of Semantic Web Technologies*. CRC Press, 2010.

IMS Learning Resource Meta-data Specification http://www.imsglobal.org/metadata/index.html

Khamnayev T. (2007), *Applying Some Principles of Semantic Web to Alleviate Lack of Collaboration and Integration in eLearning.* In proceedings PICMET 2007, 5-9 August, Portland, Oregon - USA

Knapp JL (1998), *ER isomorphisms and uniqueness conditions*. Data Knowl Eng 1998;26:271–90.

Lassila, O.(1998), *Web Metadata: A Matter of Semantics*. IEEE Internet Computing 2(4), 30-37, 1998.

Leghris C.; Mrabet R. (2006), *Cost Comparison of E-Learning Solutions*", Information Technology Based Higher Education and Training, 2006. ITHET '06. 7th International Conference on, vol., no., pp.817-824, 10-13 July 2006

Li W-S, Clifton C. (2000), *SEMINT: A tool for identifying attribute correspondences in heterogeneous databases using neural networks*. Data Knowl Eng 2000;33(1):49–84.

LOM: IEEE Learning Object Metadata standard http://ltsc.ieee.org/wg12/

LTSC (2011), *IEEE Standard for Learning Object Metadata*, 1484.12.1-2002, IEEE LTSC, 2002.

McIlraith S.A., Son T.C., Zeng H. (2001), *Semantic Web Services*. IEEE Intelligent Systems 16(2), 46-53.

Moreale E., Vargas-Vera M. (2004), *Semantic Services in e-Learning: an Argumentation Case Study*. Educational Technology & Society, 7 (4), 112-128.

Ounnas A., Liccardi I., Davis H. C., Millard D. E., White S. A. (2006), *Towards a Semantic Modeling of Learners for Social Networks*. International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL) at the AH2006 Conference, 20-23 June, Dublin, Ireland.

Pandit V.R (2010), *E-Learning System Based on Semantic Web*, Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on, vol., no., pp.559-564, 19-21 Nov. 2010.

Primo T.T., Vicari R.M. (2011), *A Recommender System that Allows Reasoning and Interoperability over Educational Content Metadata*, Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on, vol., no., pp.598-599, 6-8 July 2011.

Quemanda J., Simon B. (2003), *A use-case based model for learning resources in educational mediators*, Educational Technology and Society, Vol. 6, pp. 149-163.

Rheingold H.(2000), *The Virtual Community: Homesteading on the Electronic Frontier*, The MIT Press; revised edition edition (November 1, 2000)

Robson R. (2000), *Report on Learning Technology Standards*. in The World Conference on Educational Multimedia, Hypermedia and Telecommunications, 2000.

Saake G, Sattler KU, Conrad S. (2005), *Rule-based schema matching for ontology-based mediators*. Journal of Applied Logic 2005;3(2):253–70.

Sampson D. G., Lytras M. D., Wagner G., Diaz P. (2004), *Ontologies and the Semantic Web for E-learning*. Educational Technology & Society, 7 (4), 26-28.

Sancho P., Martínez I., Fernández-Manjón B. (2005), *Semantic Web Technologies Applied to e-learning Personalization in <e-aula>*. Journal of Universal Computer Science, 2005. 11(9): p. 1470-1481.

Scerri S., Abela C., Montebello M. (2005), *semantExplorer: A Semantic Web Browser*. In IADIS International Conference WWW/Internet 2005. 2005. Lisbon, Portugal.

SCORM 2004 3rd Edition Documentation Suite, http://www.adlnet.gov/scorm/20043ED/Documentation.aspx, 2006.

Sicilia M.A., García E. (2004), *On the convergence of formal ontologies and standardized e-learning*. Journal of Distance Education Technologies, 2(4), 2004.

SIOC Core Ontology Specification, 25 march 2010, http://sioc-project.org/ontology

Wolf B., Dhraief H., Wolpers M., Nejdl W. (2001), *Open Learning Repositories and Metadata Modeling,* SWWS, 2001.

Yi S, Huang B, Tat Chan W. (2005), *XML application schema matching using similarity measure and relaxation labeling*. Information Science 2005;169(1–2):27–46.

Ounnas A., Davis H. C., Millard D. E. (2009), *A Framework for Semantic Group Formation in Education*. Educational Technology & Society, 12 (4), 43–55.