Semantic Annotation and Linking of Medical Educational Resources

N. Dovrolis[1], T. Stefanut[2], S. Dietze[3], H.Q. Yu[3], C. Valentine[3], and E. Kaldoudi[1]

[1] Democritus University of Thrace/School of Medicine, Alexandroupoli, Greece
[2] Technical University of Cluj-Napoca/Computer Science Department, Cluj-Napoca, Romania
[3] The Open University/The Knowledge Media Institute, Milton Keynes, UK

Abstract—**Educational content is often shared among different educators and is enriched, adapted and in general repurposed so that it can be re-used in different contexts. This paper presents the MetaMorphosis+ environment for publishing, sharing and repurposing educational content in medical education. The environment meshes the paradigms of social Web and semantic Web to publish richly annotated educational resources that are further semantically enriched and exposed in the Linked Open Data cloud. The goal is to enable more relevant searching and retrieval of medical educational resources, as well as linking to other related resources in the medical domain, including scientific publications and clinical data.**

Keywords— **medical education, semantic Web, Linked Open Data, ontologies, social Web.**

I. Introduction

Continuous advances in medicine and life sciences lead to an ever expanding core knowledge relevant to the medical practice. Thus, medical academic institutions are increasingly required to invest in order to enrich their curricula by developing overspecialized courses and corresponding educational content. Educational content in medicine includes a broad range of learning object types that address both the theoretical as well as the clinical aspects of medical education. Its unique nature lies along with the fact that it is produced by both academics and clinical teachers, in a variety of places like hospital wards, healthcare practice units, laboratories, classrooms/lecture theaters, and recently the collaborative web and virtual reality spaces. In contemporary education, educational resources can be of a variety of different types. These include [1]: (a) conventional educational resources, such as lecture notes, books, lecture presentations; (b) educational content types unique in medical education, e.g. teaching files, virtual patients, evidence based medicine forms, anatomical atlases, etc; (c) alternative educational content types, either reflecting active learning techniques and/or stemming from newly introduced web 2.0 technologies, such as problem/case based learning sessions, serious games (2D/3D), web traces, wikis, blogs/discussion forums, etc.; and (d) user generated content.

Considering the state-of-the-art nature, the complexity and, consecutively, the cost of state-of-the-art educational content, it is imperative that such content can be repurposed, enriched, and embedded effectively into respective curricula and continuing education, as well as public dissemination and awareness.

This need for sharing, re-using and repurposing educational resources actually makes them a natural candidate for social objects in professional educational social networks. The social Web, or Web 2.0 [2], has become an important trend during the last few years. Recently the term object-centered sociality was introduced [3] to describe the fact that strong social relationships are built mainly when individuals are grouped together around a shared object that mediates the ties between them. Therefore, we can assume that each content item on a social network site can be a source of social connectivity, catalysing social networking in virtual spaces. At the same time, Semantic Web technologies [4] are specifically designed to address the challenge of data and knowledge management in a world with highly distributed resources. The semantic Web promises an infrastructure that comprises machine understandable content and, therefore, a worldwide Web made of semantically linked data instead of a mere collection of HTML documents.

These two paradigms of the social Web and the semantic Web are merged in the MetaMorphosis+ [5] semantic social network which aims to provide an environment for resource publishing, sharing and repurposing in medical education. The MetaMorphosis+ semantic social network can be viewed as two distinctive and interacting networks. The first one is a network of persons, including authors, potential authors and final users of learning objects (students, teachers or others, e.g. educational managers, etc). The second is a network of educational resources. The network of persons is functioning in a way similar to other social networks. Persons can interact with each other via their personal blogs, declare friends and create their own interest groups. At a different level, educational resources themselves create an equivalent social network. Educational resources in MetaMorphosis+ can be resources residing in a Learning Management System (LMS), in another educational repository, or merely available on the Web. Re-

Á. Jobbágy (Ed.): 5th European IFMBE Conference, IFMBE Proceedings 37, pp. 1400–1403, 2011.
www.springerlink.com

sources (like humans) are represented in MetaMorphosis+ by their profile. Educational resources as social objects can exhibit different aspects of 'object sociality' [6]: (a) the obvious connections via common tags; (b) connections based on collective usage and other related interaction of human users; (c) social connections based on the inheritance as realized via educational content repurposing; (d) semantic connections realized via semantic annotations and linking of educational resources.

In this paper we present our approach towards semantic annotation and linking of medical educational resources within the MetaMorphosis+ environment. The aim is to semantically annotate educational resources, then enrich their metadata and semantically link them with other related resources. The goal is to enable more relevant searching and retrieval, as well as linking to other related resources in the medical domain, such as scientific literature, medical data and healthcare information systems, medical and biological databases, etc., and at the same time expose medical educational resources as part of the emerging semantic web.

II. Semantic exposing and linking of educational resources

In the conventional Web, a resource can be described via an XHTML/XML document, where various tags are used to annotate the document, mainly regarding its presentation, not conveying any semantics about the resource itself. In order to describe a resource the W3C Resource Description Framework (RDF) [7] is commonly used to represent metadata about a resource in the form of triples: subject, predicate, object (Fig.1). Generally, the subject can be the resource itself while the predicate can be any relationship as defined in any XML namespace published on the Web. The object can be an explicit value but also a dereferencable URI. This way, an RDF triple can link the description of a resource with other sources of information on the Web, thus creating a worldwide graph-like linking of resources, what is currently termed Linked Open Data (LOD) cloud [8], [9], a community project of the World Wide Web Consortium's Semantic Web Education and Outreach Group (W3C SWEO).

Following this approach, educational resources in MetaMorphosis+ are primarily described by the mEducator RDF metadata scheme [10] for describing medical educational resources. This includes a number of fields addressing different aspects of the educational resource: (a) general fields: resource title, unique identifier, URL, URN, intellectual property rights clearance/license, quality stamp (if any); (b) fields related to a general resource description: resource authors, creation date, citation (i.e. how the resource should be formally cited), keywords, content description, technical description (including any technical requirements to access and use the resource); (c) fields related to the educational aspect of the resource: educational context (for which the resource is intended), teaching/using instructions, educational objectives, expected learning outcomes, suggested assessment methods, educational prerequisites; (d) fields related to classification/taxonomy information: resource language, type, discipline, discipline subspecialty, educational level; and (e) fields addressing repurposing: resource parents, repurposing context, repurposing description.

In the mEducator RDF scheme a number of predicates are part of other popular namespaces, e.g. the Dublin Core (ISO 15836:2009) Elements and Terms namespaces [11] for a generic resource description, FOAF [12] for describing people, SIOC [13] for integrating online communities and SKOS [14] for knowledge organization, while a number of novel domain-specific properties are also included. Most of the triples defined by the mEducator RDF scheme however allow for explicit string values for objects, to be as free text by human users.

In our work we extend this basic semantic annotation schema via enrichments by means of links to, established data sets. This is of particular importance to extend rather unstructured metadata, such as keywords or free text subject and discipline descriptions with structured data based on well-known vocabularies. This is achieved by exploiting a variety of medical domain ontologies and the expanding LOD cloud to semantically annotate the existing RDF description of a resource and then expose its metadata back to the LOD cloud for further exploitation by third parties which make use of the web of LOD.

Biomedical ontologies provide essential domain knowledge to drive data integration, information retrieval, data annotation, natural-language processing and decision support. BioPortal (http:// bioportal.bioontology.org) is an open repository of biomedical ontologies that provides access via Web services and Web browsers to ontologies developed in various formats including OWL, RDF, OBO format and Protégé frames [15]. In MetaMorphosis+ we have utilized the NCBO BioPortal's RESTful Web services programming interface to access and incorporated terms and concepts from the more than 260 ontologies provided to this day, corresponding to more than 4.5 million medical and life sciences terms. This way the MetaMorphosis+ user can annotate an educational resource with suggested standardized terms and concepts from a variety of ontologies, enriching the RDF output with dereferencable standardized terms as values for the various fields, e.g. keywords, discipline, specialty, etc. The ontologies used include amongst else for prominent medical ontologies such as SNOMED-CT (Systematized Nomenclature of Medicine – Clinical Terms), ICD9/10 (International Statistical Classification

Diseases and Related Health Problems), Body System (body system terms used in ICD11), MeSH (Medical Subject Headings), NCI (Meta)Thesaurus, Galen (the high level ontology for the medical domain), HL7 (the Normative RIM model v2), Biomedical Resource Ontology (BRO, a controlled terminology of resources to improve sensitivity and specificity of Web searches).

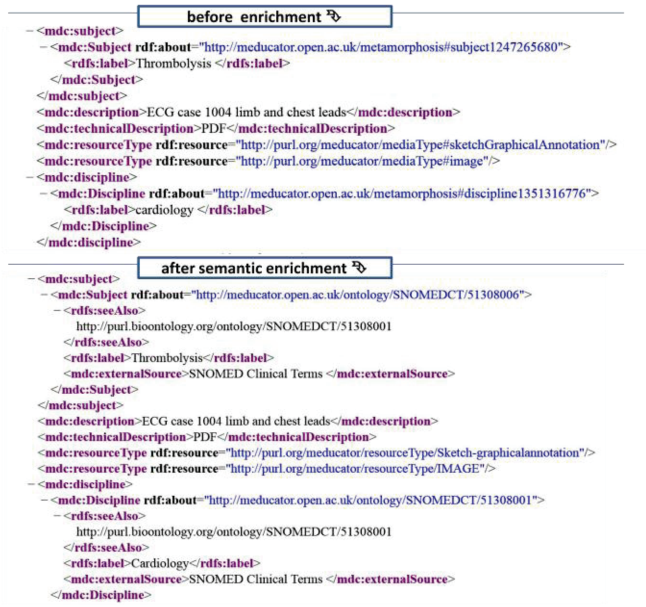

Fig. 1 The above depictions of RDF/XML snippets show the differences of the RDF output before (above) and after (below) the semantic enrichment via the BioPortal. Before enrichment the values for the 'Keyword' and the 'Discipline' fields appear as strings ('Thrombolysis' and 'Cardiology' respectively) with no linking to an ontology, while after enrichment these values are substitutes with dereferencable URIs corresponding to standardized ontology terms.

As an example, suppose a user intends to describe an educational resource by using the term/concept *Telemedicine*, in the list of 'Keywords' or in the 'Discipline' and 'Specialty' fields of the metadata description of the educational resource. Semantic annotation in MetaMorphosis can suggest a number of related standardized terms from the available ontologies – for example the equivalent term from the NCI Thesaurus which is represented by http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#Telemedicine, a dereferencable URI, thus enabling rich linking and reasoning within the Semantic Web. Figure 1 shows an example of RDF triples in the mEducator scheme before and after semantic annotation in MetaMophosis+ .

The semantic annotation of the educational resources as described above can then be the basis for expanding search queries via semantically related terms and even retrieve additional but relevant information from external sources. An example is linking educational resources with related published work in PubMed [16] based on semantic linking of MeSH terms. Another example is linking to teaching files and related data collections via semantic meshing of e.g. SNOMED-CT or ICD9/10/11 terms.

A different equally interesting aspect is that of exposing the semantic descriptions of educational resources in MetaMorphosis+ as RDF triples in the LOD cloud. This enriches the semantic Web and allows for bi-directional linking, querying and retrieving of relevant educational resources from other Web environments, thus fully integrating MetaMorphosis+ and the educational resources there in with the semantic Web.

In order to achieve this integration three different technological approaches can be exploited. The first one involves using technologies such as RDFa or MicroFormats, approaches that add 'tags' within conventional XHTML documents in order to semantically annotate data. Such semantic annotations can then be retrieved by special Web programs including semantic search engines, semantic browsers or plug-ins for conventional browsers. This is a simple but rather limited approach, not fully realizing the semantic Web paradigm. A more semantically oriented approach involves using special technology that transforms the entries of a relational database to RDF format exposing them as triples in the semantic Web. A prominent such technology is Triplify [17] which was originally used with MetaMorphosis+ to expose triples and link to the LOD cloud. However, the fully semantic approach involves implementing a triplestore, a specialized database for the storage and retrieval of RDF metadata via a query language like SPARQL. Examples of commonly used triplestores include ARC [18] and Sesame [19]. The most important attribute of this approach is that data are being stored as RDF triples thus creating a graph.

MetaMorphosis+ is implemented using the Elgg open source social engine [20] while it employs a SESAME triplestore instance for storing annotated and enriched metadata as RDF triples. Further details about the MetaMorphosis+ architecture can be found in [21] while the underlying RDF schema and a first implementation is described in [22]. This specific implementation has been achieved partly within the mEducator project, an EU funded best practice network (under the eContentPlus2008 programme, Contract Nr: ECP 2008 EDU 418006) with the aim to implement and critically evaluate existing standards and reference models in the field of e-learning in order to enable specialized state-of-the-art medical educational content to be discovered, retrieved, shared and re-used across European higher academic institutions. During the first few months of deployment the environment exhibits more than 100 registered users and more than 350 educational resources, including 80 repurposed resources. Although about half of the resources are in English language, there is a representation of more than 15 other European languages. The resources included in the environment are distributed among the various educational levels, 33% intended for undergraduate medical education, 23% intended for postgraduate/resident studies and 21% for continuing life-long education, while 22% are intended for educating the public. Currently, there is a mean of around 50 triples exported per educational resource, which results in a total of more than 15,000 triples exposed via the MetaMorphosis+ triplestore.

III. DISCUSSION

Recent technological advances have induced a concept and paradigm shift for the local, opinion driven physician approach towards the global expert notion, including aspects of evidence-based medicine and information aggregation towards individualized patient care. This modifies quickly and essentially not only the way healthcare professionals work, but also the way medical educational processes are designed and hopefully practiced, placing requirements for adaptive and ubiquitous online expert educational content sharing. Technology-supported educational interventions are usually successful when specific training requirements are aligned with the learning potential and the educational use of technology. Thus, requirements for flexible, adaptive and ubiquitous online content sharing should evoke notions, practices and technologies from respective state-of-the-art evolutions of the Web technologies, namely the emerging paradigm of participative, collaborative Web 2.0 and the semantic Web suite of technologies. The specific implementation of MetaMorphosis+ presented in this paper is only an example of the various different ways one can combine object sociality and semantic annotation and linking to create powerful personalized application. Another application domain can be that of scientific knowledge management as well as that of personalized patient empowerment services.

ACKNOWLEDGMENT

This work is funded in part by the mEducator project (Contract Nr: ECP 2008 EDU 418006 mEducator), a project funded under the eContentplus programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

REFERENCES

1. Kaldoudi E, Bamidis P D, Pattichis C (2009) Multi-type content repurposing and sharing in medical education, INTED2009 Proc. Valencia, 2009, pp 5129-5139
2. O'Reilly T (2005) What is Web 2.0: Design patterns and business models for the next generation of software at http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html
3. Engeström J (2005) Why some social network services work and others don't. Or: the case for object-centered sociality at http://www.zengestrom.com/blog/2005/04/why_some_social.html
4. Berners-Lee T, Hendler J, Lassila O (2001) The semantic Web – a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Sci Am 284:34–43
5. MetaMorphosis+ at http://metamorphosis.med.duth.gr
6. Kaldoudi E, Dovrolis N, Giordano D, Dietze S (2011) Educational resources as social objects in semantic social networks, Linked Learning 2011 Proc, 1st Int. Workshop on eLearning Approaches for the Linked Data Age, Herakleio, Greece, 2011
7. RDF at http://www.w3.org/TR/rdf-syntax/
8. Bizer C, Heath T, Berners-Lee T (2009): Linked Data – The story so far. Int J on Sem Web Inf Systems (IJSWIS) 5:1-22
9. LOD Cloud Diagram at http://richard.cyganiak.de/2007/10/lod/
10. mEducator RDF scheme at http://purl.org/meducator/ns/
11. Dublin Core at http://dublincore.org/documents/dces/
12. FOAF at http://www.foaf-project.org/
13. SIOC at http://sioc-project.org/
14. SKOS at http://www.w3.org/2004/02/skos/
15. Noy N F, Shah N H, Whetzel P L, et al (2009) BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res 1:37
16. PubMed at http://www.ncbi.nlm.nih.gov/PubMed
17. Auer S, Dietzold S, Lehmann J, Hellmann S, Aumueller D (2009) "Triplify: light-weight linked data publication from relational databases", Proc. of the 18th international conference on World wide web, ACM, NY, 2009, pp. 621-630
18. ARC at https://github.com/semsol/arc2/wiki
19. Sesame triplestore at http://www.openrdf.org/
20. The ELGG social network engine at http://elgg.com
21. Yu H Q, Dietze S, Li N, Pedrinaci C, Taibi D, Dovrolis N, Stefanut T, Kaldoudi E, Domingue J (2011) A Linked data-driven & service-oriented architecture for sharing educational resources, Linked Learning 2011 Proc, 1st Int. Workshop on eLearning Approaches for the Linked Data Age, Herakleio, Greece, 2011
22. Mitsopoulou E, Taibi D, Giordano D, Dietze S, Yu H Q, Bamidis P, Bratsas C, Woodham L. (2011).Connecting medical educational resources to the Linked Data cloud: the mEducator RDF Schema, store and API, Linked Learning 2011 Proc, 1st Int. Workshop on eLearning Approaches for the Linked Data Age, Herakleio, Greece, 2011

Corresponding author:

Author: Eleni Kaldoudi
Institute: School of Medicine, Democritus University of Thrace
Street: University Campus, Dragana
City: Alexandroupolis
Country: Greece
Email: kaldoudi@med.duth.gr