Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-963

Recommendation of Standardized Health Learning Contents using Archetypes and Semantic Web Technologies

María del Carmen LEGAZ-GARCÍA[a], Catalina MARTÍNEZ-COSTA[a], Marcos MENÁRGUEZ-TORTOSA[a], Jesualdo Tomás FERNÁNDEZ-BREIS[a,1]
[a] *Departamento de Informática y Sistemas, Facultad de Informática, Universidad de Murcia, CP 30100, Murcia, Spain*

Abstract. Linking Electronic Healthcare Records (EHR) content to educational materials has been considered a key international recommendation to enable clinical engagement and to promote patient safety. This would suggest citizens to access reliable information available on the web and to guide them properly. In this paper, we describe an approach in that direction, based on the use of dual model EHR standards and standardized educational contents. The recommendation method will be based on the semantic coverage of the learning content repository for a particular archetype, which will be calculated by applying semantic web technologies like ontologies and semantic annotations.

Keywords. Electronic Healthcare Records, Archetypes, eLearning, SCORM, Semantic Web Technologies

Introduction

Studies about the quality of health-related contents report that 70-80% of Internet users search for health-related contents [1]. This constitutes a situation of potential risk since the reliability of the information is not guaranteed. Such situation can generate contrary effects in terms of patient safety, inefficiency and overrun to the healthcare system. For this purpose, in recent years, the need for finding quality criteria of healthcare web contents has been identified.

The SemanticHealth project [2] identified the need for linking the content of the electronic health records with educative contents to support the training of both professionals and citizens. Such project also considered that EHR standards should play a crucial role in the achievement of semantic interoperability, so any attempt for developing solutions in this domain should use them. In this work, we are interested in dual model architecture standards [3], which are based on the definition of information and knowledge separately. Such standards definine two conceptual levels by using the reference and archetype models. Archetypes represent the minimal information unit that clinical information systems can exchange, and archetypes also include the context for interpreting clinical information [4]. Archetypes include bindings to terminologies,

[1] Corresponding Author. María del Carmen Legaz-Garcia. Departamento de Informática y Sistemas, Facultad de Informática, Universidad de Murcia, CP 30100, Murcia, Spain

which provide the clinical meaning of the terms of the archetype. Moreover, they are used as templates for data capture what means that EHR extracts are interpreted in the context of its archetype. Consequently, it is sensible to think that archetypes should play a fundamental role in the process of finding appropriate contents according to the EHR of a citizen.

Thus, in this work, we will present an approach that provides recommendation of contents at archetype level, that is, given a particular archetype, a set of related contents will be retrieved from a repository of learning contents. Such repository may include any kind of document, which will be classified from the clinical perspective. Clinical guidelines are an example of such resource of interest, especially for professionals. The clinical condition of a patient is part of the EHR but also may be associated with a particular guideline. Therefore, they have to be considered as learning resource. Finally, our approach will provide the recommendations based on the semantic similarity of the annotations of the archetype and the annotations associated with the learning contents.

1. Methods

1.1 The Learning Contents Repository

A learning object can be defined as a digitized entity which can be used, reused or referenced during technology supported learning [5]. SCORM (Sharable Content Object Reference Model) (http://www.adlnet.gov/capabilities/scorm) is a set of technical norms that allows online learning systems to import and reuse standardized learning contents. SCORM is the current standard for exchanging learning contents, so our learning repository uses it. In fact, our solution uses SICARA [6], which is our semantic repository of SCORM learning contents.

The semantic representation of SCORM objects in SICARA is enabled by a semantic extension done to the SCORM standard by our research group. The semantic annotations of a learning object are defined over the set of ontologies managed by SICARA. An annotation in SICARA associates a learning object with an ontology class. The repository of such semantic annotations will play a very important role in this approach, since those annotated learning objects will be the ones used for the recommendations. For this purpose, SICARA makes use of domain ontologies in OWL format. Despite SICARA has a user interface, in this work we will use it only as repository.

1.2 The Archetype Management System

ArchMS (Archetype Management System http://sele.inf.um.es/archms) [7,8] is a system for managing clinical archetypes that works with both ISO 13606 (http://iso.org/iso) and openEHR (http://www.openehr.org) archetypes and facilitates activities like importing, exporting, validating, annotating and searching archetypes as well as generating web applications from archetypes.

Archetype terms usually contain terminological bindings that can be used as annotations. Besides, healthcare institutions may need additional annotations at archetype level, which can be based on criteria like application domains or related diseases. To define such annotations, ArchMS permits users to upload annotation

resources in OWL format and to use the ontologies included in Bioportal through the NCBO web services (http://rest.bioontology.org/). ArchMS is able to suggest annotations by using the purpose, keyword and term definitions of the archetypes. For such purpose, the Lucene API is used.

As a result of the present work, the functionality of ArchMS will be extended. First, the user will be able of annotating learning resources from ArchMS and the annotations will be stored in SICARA. Second, the user will have the possibility of retrieving the relevant learning contents for a particular archetype.

1.3 The Recommendation Method

The recommendation approach is based on calculating the semantic similarity between the semantic profile of an archetype and the annotations of the learning contents. The semantic profile of an archetype is obtained by retrieving the terminological bindings of the archetype terms and the annotations of the archetype defined in ArchMS.

The recommendation algorithm returns the set of learning objects that provides the best coverage for the archetype. Thus, the recommendation method consists of searching for those objects annotated with the annotations of the archetype. For each concept included in the profile, the most similar concept from the annotations of the learning object is obtained. Our semantic similarity function is based on the one defined in [9] and combines the following three factors:

- Linguistic similarity: The similarity between the terms associated with the annotation concepts using a string-based calculation.
- Taxonomic similarity: The distance between the annotation concepts in the taxonomic structure of the annotation resource.
- Properties similarity: The similarity between the set of properties associated with the annotation concepts. This means that not only the taxonomic structure of the ontology is used but also the axioms defined in the annotation resources.

The weighted average of these factors returns a score between 0 and 1. If such score is greater than the similarity threshold, then the concept is considered similar to the annotation and that learning object is included in the recommendation list.

2. Results

In this section, an example of how ArchMS is used to retrieve annotated learning content related to a given archetype is described. Let us consider that we want to look for the learning resources associated with the openEHR archetype for non-TNM staging scores for the colorectal cancer, namely, *openEHR-EHR-CLUSTER.tumour_colorectal_staging_non_tnm.v1*. This archetype does not contain terminological bindings, so we can only use the archetype annotations defined using ArchMS. Now the recommendation method can be invoked. In this example, the threshold value is 0.7 and the weights for the linguistic similarity, taxonomic similarity and properties similarity are, respectively, 0.3, 0.5 and 0.2. Therefore, the taxonomic similarity factor is given more importance. In this case, the learning resource recommended by the algorithm is *clinical guideline for colorectal cancers*. The explanation for such recommendation comes from Table 1, where the annotations of the archetype are shown in the left column, and the clinical guideline ones are shown in

the right one. Both the archetype and the learning resource were annotated through independent processes. The archetype was annotated using the suggestions from ArchMS, whereas the clinical guideline was annotated using the information provided in one of its appendices. This resource has been selected because the similarity of the archetype annotation *"Multiple malignancy"* with the resource ones *"Adenosquamous carcinoma"* and *"Carcinoid tumor"* is greater than the threshold value. Such similarity can be observed in Figure 1, which shows an excerpt of how such concepts are located in the SNOMED-CT disorder taxonomy, because such concepts are taxonomically close. This means that the recommendation method is able to use such conceptual proximity to recommend relevant, related contents.

If we look at both sets of annotations, we could think that the algorithm should have considered the archetype annotation "*colon and rectum (combined site)"* similar enough to the guideline annotations "*colon structure*" or "*rectum part*" because of the similarity of the labels and the intuitive closeness of the concepts. However, the location of such concepts in SNOMED-CT avoids them to be found similar. Despite, being subclasses of the body structure concept "*physical anatomical entity*", "*Colon and rectum*" is a subclass of "*group of anatomical entities*" at depth 2, but both "*colon structure*" and "*rectum part*" concepts are subtypes of "*anatomical structure*", but their minimum depth is 6. Consequently, they are not considered taxonomically close enough.

The recommended content is displayed, and the different files can be opened and explored by the user. Finally, the user can use the results of the recommendation to create a SCORM course.

Table 1. Extract of the SNOMED-CT concepts used to annotate the archetype and the clinical guideline

Archetype annotations	Clinical guideline annotations
▪ Body structure o Colon and rectum (combined site) ▪ Disorder o **Multiple malignancy**	▪ Body structure o Appendix structure o Colon structure o Rectum part ▪ Disorder o Adenoma of rectum o **Adenosquamous carcinoma** o **Carcinoid tumor** o Carcinoma in situ of colon o Carcinoma in situ of rectum o Squamous cell carcinoma

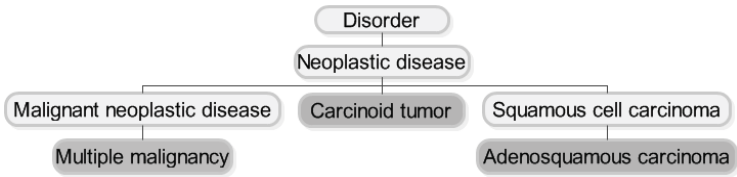

Figure 1. SNOMED-CT taxonomy of concepts with similarity>0.7

3. Conclusions

Recent international reports and recommendations suggest linking EHR content with learning resources. In this paper we have presented the first step of an approach that goes in that direction. We have described how archetypes can be used as an instrument for retrieving relevant learning contents. In our solution, standards have been used for the representation of both clinical and learning content.

The current extension to ArchMS allows users to retrieve contents by archetype. We are currently working on improving the efficiency of the recommendation method when working with large resources such as SNOMED-CT. We aim at extending our approach to perform the analysis at data level by using archetyped EHR extracts. This would permit to obtain a more specific set of educational materials. It would be also interesting to add classifications of resources according to the target audience, that is, professionals or citizens.

In line with this objective, we are working on including external resources such as PubMed articles as part of the recommended contents for professionals. For such purpose, we plan to exploit the annotations defined by NCBI for such papers.

Acknowledgements

This work has been possible thanks to the Spanish Ministry of Science and Innovation through grant TIN2010-21388-C02-02 and to the Fundación Séneca through grant 15295/PI/10. María del Carmen Legaz García is supported by the Fundación Séneca through fellowship 15555/FPI/2010.

References

[1] Mayer MA, Karampiperis P, Kukurikos A, Karkaletsis V, Stamatakis K, Villarroel D, Leis A. Applying Semantic Web technologies to improve the retrieval, credibility and use of health-related web resources. Health Informatics Journal June 2011; 17(2): 95-115

[2] European Commission, Semantic interoperability for better health and safer healthcare deployment and research roadmap for Europe. 2009; ISBN-13: 978-92-79-11139-6

[3] Beale T. Archetypes: Constraint-based Domain Models for future-proof Information Systems. In Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer, 2002.

[4] Kalra D, Archana T. The EHR and Clinical Archetypes: time for clinical engagement!, in: eHealth Planning and Management Symposium, Copenhagen, Denmark, 2008.

[5] Rehak DR, Mason R, "Engaging with the Learning Object Economy", in Littlejohn, Allison, Reusing Online Resources: A Sustainable Approach to E-Learning, London: Kogan Page, 2003

[6] Esteban-Gil A, Fernandez-Breis JT, Castellanos-Nieves D, Valencia-Garcia R, Garcia-Sanchez F. Semantic enrichment of SCORM metadata for efficient management of educative contents Procedia - Social and Behavioral Sciences 2009; 1(1): 927-932.

[7] Fernández-Breis JT, Menárguez-Tortosa M, Martinez-Costa C, Fernández Breis E, Herrero J, Sánchez Cuadrado J, Moner D, Valencia-García D, Robles M. Semantic Web-based System for Managing Clinical Archetypes.30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2008; 1482-1485

[8] Legaz-García, M.C., Martínez-Costa, C., Menárguez-Tortosa, M., Fernández-Breis, J.T. Towards semantic platforms for archetype management: ArchMS, OpenHealth 2011, Barcelona, Spain.

[9] Castellanos-Nieves D, Fernández-Breis JT, Valencia-García R, Martínez-Béjar R, Iniesta. Semantic Web Technologies for Supporting Learning Assessment. Information Sciences 2011;181:1517-37