Utilizing Semantic Web Technologies in Healthcare

Vassileios D. Kolias, John Stoitsis, Spyretta Golemati and Konstantina S. Nikita

Abstract The technological breakthrough in biomedical engineering and health informatics has produced several Health Information Systems (HIS) and medical devices that are used on a daily basis in hospitals producing a vast amount of data. The data that are produced come from different sources and are not stored in a unified storage repository or database even in a single hospital. As a result of that the interoperability of HIS is limited, the retrieval of information is difficult and there is hidden knowledge that remains unexploited in vast and diverse pools of medical data. In order to overcome the above, scientific community suggests the use of the semantic web technologies. The semantic web technologies provide the tools that allow to process data in a more effective and accurate way, create the framework for interoperability between HIS and also integrate data from various sources with their semantic meaning

Keywords Semantic web · Ontologies · Biomedical ontologies · Healthcare services · Health information systems

1 Introduction

The technological breakthrough in biomedical engineering and health informatics has produced several Health Information Systems (HIS) and medical devices that are used on a daily basis in hospitals producing a vast amount of data. Although the systems and the devices have improved the healthcare services, the main issue that has emerged is the utilization of the produced data. The data that are produced come from different sources and are not stored in a unified storage repository or database

V. D. Kolias (✉) · J. Stoitsis · K. S. Nikita
Biosim Laboratory, School of Electrical and Computer Engineering,
National Technical University of Athens, Athens, Greece
e-mail: vaskolias@biosim.ntua.gr

S. Golemati
Medical School, National and Kapodistrian University of Athens,
Athens, Greece
e-mail: sgolemati@med.uoa.gr

D.-D. Koutsouris, A. A. Lazakidou (eds.), *Concepts and Trends in Healthcare Information Systems,* Annals of Information Systems 16,
DOI 10.1007/978-3-319-06844-2_2,

even in a single hospital. In many cases one hospital has many Radiological Information System (RIS), Laboratory Information Management System (LIMS) and HISs which are not interconnected. In addition to this, the medical data have not a structured and unified form and, as a result, data are not utilized in an efficient and is very difficult to be retrieved. In order to overcome these issues several standards such as Electronic Health Records (EHR) and define abbreviation Health Level 7 (HL7) have been suggested and used by the HIS. However, their major limitation is that they do not contain the semantic information of the medical data in a form that can be easily processed by computers. Due to this, knowledge is hidden in vast and diverse medical data pools. The scientific community has suggested the use of the semantic web technologies to overcome the problems that are mentioned above. The semantic web technologies provide the tools that allow to process data in a more effective and accurate way, create the framework for interoperability between HIS and also integrate data from various sources with their semantic meaning. In the past years several ontologies and terminologies have been introduced, which are the core element of the semantic web, in healthcare for describing and integrating medical data such as Unified Medical Lexicon System (UMLS) (Bodenreider 2004), Foundational Model of Anatomy (FMA) (Rosse and Mejino Jr. 2003), Radiological Lexicon (Langlotz 2006) and International Classification of Diseases (ICD)-11 (Tudorache et al. 2013). Also new standards are evolving such as OpenEHR (Kalra et al. 2005) which have integrated the main biomedical ontologies in order to provide to the HIS systems the capabilities of the semantic web that are mentioned above.

The mission of this chapter is to analyze how semantic web technologies can be integrated to the HIS systems in order to solve the integration of the medical data, to provide enhanced capabilities for retrieving medical data, to provide personalized healthcare services to patients and also how the research in medical and biomedical fields can be enhanced.

2 Semantic Web and Ontologies

According to Tim Bernee Lee the semantic web is a web of data that can be processed directly and indirectly from the machines (Berners-Lee et al. 2001). As a result of that many tasks will be automated, data processing will be more accurate, faster and can lead to greater exploitation of data by sharing and reusing. In health sciences and health informatics systems that leads to better information retrieval, discovery of new knowledge from unexploited data, enhanced interoperability between institutions and better health services for the patients. Core element of the semantic web is the ontology. Ontology has several definitions but in computer science ontology is a formal, explicit specification of a shared conceptualization (Gruber 1993). This definition implies that ontology provides a vocabulary which can be used to model a domain, that is, the types of concepts and objects that exist and their relationships between them. There are two types of ontologies: the reference or domain ontologies and the application ontologies. Reference ontologies represent knowledge about a particular part of the world in a way that is independent from specific objectives, through a theory of the domain (Burgun 2006). In the contrary application ontologies are designed to perform specific tasks and are narrower than the reference ontologies.

3 Main Biomedical Ontologies

In the past years several biomedical ontologies have been created. Most of them have been created to describe thoroughly a domain in medicine and biology such as FMA which describes the human anatomy terms and their relations. Other ontologies such as SNOMED-CT are used to model the clinical terms and processes in order to provide better communication and interoperability between HIS. The most significant biomedical ontologies will be described in terms of domain and use. Most of the ontologies and terminologies are available through web services via the BioPortal.

3.1 Foundational Model of Anatomy (FMA)

FMA ontology describes the domain of human anatomy. FMA is a reference ontology which contains over 75,000 distinct anatomic types which cover the human anatomy from sub-cellular components to major body parts and the whole organism itself. The anatomic types are connected with more than 130,000 terms either as preferred names, synonyms or their non-English equivalents. FMA describes and defines the relationships of the types. There are over 2.1 million relationships which are grouped in more than 200 types of spatial structural and non-structural relationships. FMA is the basis of the human anatomy for many projects and also subparts of FMA are used for representing the anatomy for specific anatomy or application ontologies such as NeuroFMA, MEDICO (Möller et al. 2009), SEMIA (Kyriazos et al. 2011) and others.

FMA was developed and maintained by the Structural Informatics Group from the University of Washington and is currently at version 3.1 and its format is OWL. FMA is available either with the use of the web services and widgets from BioPortal or it can been downloaded as a file from the Structural Informatics Group[1] site.

3.2 International Classification of Diseases (ICD-10)

ICD-10 was endorsed by the Forty-third World Health Assembly in May 1990 and came into use in World Health Organization (WHO) Member States as from 1994.

[1] http://sig.biostr.washington.edu/projects/fm/AboutFM.html.

The classification is the latest in a series, which has its origins in the 1850s. The first edition, which is known as the International List of Causes of Death, was adopted by the International Statistical Institute in 1893. WHO took over the responsibility for the ICD at its creation in 1948 when the Sixth Revision, which included causes of morbidity for the first time, was published. The World Health Assembly adopted in 1967 the WHO Nomenclature Regulations that stipulate use of ICD in its most current revision for mortality and morbidity statistics by all Member States.

The ICD is the international standard diagnostic classification for all general epidemiological, many health management purposes and clinical use. These include the analysis of the general health situation of population groups and monitoring of the incidence and prevalence of diseases and other health problems in relation to other variables such as the characteristics and circumstances of the individuals affected, reimbursement, resource allocation, quality and guidelines.

It is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records. In addition to enabling the storage and retrieval of diagnostic information for clinical, epidemiological and quality purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by WHO Member States.

ICD-10 is usually kept as a thesaurus but there is also an ontology of ICD-10 in OWL (Möller et al. 2010). The ICD-10 in OWL format can describe in a better and more formal way except from the terms and the relationships between the ontologies entities. As a result of that the query capabilities will be enhanced compared with a simple thesaurus. The latest version of ICD is ICD-11, which is under construction, and it is estimated to be available for use at 2015. The main difference from older versions of ICD is that there will be mappings and relationships to other ontologies, terminologies and classifications such as SNOMED-CT to provide semantic interoperability (Tudorache et al. 2013). Also the development will be based on an open collaborative tool (Tudorache et al. 2010) which will be used for authoring by medical experts from all over the world. ICD-10 is available through BioPortal at UMLS format via web services.

3.3 Radiological Lexicon (RadLex)

RadLex is created by the Radiological Society of North America (RSNA). The goal of RadLex development was to unify the variety of terminologies that radiologists use into one unified lexicon to serve all their needs. Additional to that is that the use of standardized terminologies is now vital in medical practice. The majority of the benefits that the HIS systems can provide in healthcare services cannot be exploited if the data are not stored using standardized terms in a structured format. Unfortunately, almost all radiology reports are stored in free text rather than in a structured format and due to this radiologists struggle to follow the changes in health care systems which are based on informatics innovations. RadLex provides a uniform standard for all radiology related information.

RadLex development started in 2005 and today the 3rd version of Radlex is available. RadLex includes many complex domains that are necessary for radiologists. These domains range from basic sciences to imaging technologies and acquisition.

The lexicon is organized into a subsumption hierarchy with RadLex term as the root. RADLEX contains over 7,400 terms which are organized in 9 main categories or types such as anatomic location, treatment, uncertainty and image quality.

Radlex provides a unified lexicon and it can evolve to an ontology but it is not an ontological framework. From an ontological prospective RADLEX has 3 limitations

- It is term-oriented and as a result of that it ignores the entities to which its terms project
- There is a lack of taxonomy grounded on biomedical reality
- The ambiguity and mixing of relations such as "is_a", "part_of", "contained_in"

For these reasons the application ontology FMA-RADLEX is proposed in (Mejino Jr et al. 2008).

3.4 Systemized Nomeclature of Medicine-Clinical Terms (SNOMED-CT)

SNOMED-CT (Stearns et al. 2001) is the most commonly used multilingual clinical healthcare terminology in the world and it provides the core terminology for the Electronic Health Records (EHR). The goal of SNOMED-CT is to encode the meanings that are used in health information in order to improve the healthcare services. SNOMED-CT covers most of the areas that are used in medical practice such as clinical findings, symptoms, diagnoses, procedures, body structures, organisms and other etiologies, substances, pharmaceuticals, devices and specimen.

The use of SNOMED-CT in Health Informatics Systems can lead to an interoperable EHR and due to that can enhance the interoperability between different HIS. Additional to that it provides a consistent way for indexing, storing, retrieving and aggregating clinical data across specialties and sites of care. SNOMED was created by the College of American Pathologists (CAP) and since April of 2007 is owned, maintained and distributed by the International Health Terminology Standards Development Organization (IHTSDO)[2].

3.5 Unified Medical Lexicon System (UMLS)

UMLS repository is consisted from biomedical vocabularies and ontologies that are developed by the US National Library of Medicine. UMLS covers most of the

[2] http://www.ihtsdo.org/.

biomedical terminology and it consists of over 60 vocabularies with 900,000 concepts and over 12 million relationships among these concepts. The most notable vocabularies that are integrated in UMLS are SNOMED-CT, ICD-10, Medical Subject Headings (MeSH), Gene Ontology and others. UMLS covers from clinical terms to genetic information.

UMLS consists of 3 components. The first component is the Metathesaurus that is a repository of inter-related biomedical concepts. The relation of the concepts can be inherited from the structure or are manually created be the editors using the hierarchical or associative relationships. MeSH is used to derive the statistical relations between the concepts from the MeSH indexing terms in MEDLINE[3] citations. The knowledge of the repository is organized by concept. In order to achieve this the similar and synonymous terms are clustered together to a concept and then are linked to other concepts with relationships types that were described above in this section. Additional to that the concepts are categorized in terms of semantic types,which are assigned by the Metathesaurus editors. The structure of the Metathesaurus allows users to collect the various terms that name a concept, to extract the relationships of the concepts and to collect the concepts based on their semantic meaning.

The other two components of UMLS are the Semantic Network and lexical resources. Semantic Network provides the semantic types to semantically categorize the concepts of the Metathesaurus. Lexical resources such as the SPECIALIST lexicon[4] and programs are used to generate variants of the biomedical terms.

3.6 Open Biomedical Ontologies (OBO)

Unlike the ontologies and terminologies that are described previous in this section OBO is not an ontology itself but and ontology library and a framework for ontology developers. The purpose of OBO is to create an evolving set of shared principles for the ontology development at the biomedical field. Most of the ontologies that are created in the biomedical field do not follow the same principles in design and as a result of that the integration of them meets obstacles.

An ontology that is developed with the OBO principles must be open, orthogonal, syntactically well-specified and to share a common space of identifiers. Open ontology means that the use of the ontology and the data that are described by its terms must be free without the need of license. Additional to that is that the evolution of the ontology is open to community debate. Ontologies that are created must be orthogonal in order to add new annotations without the need of restructuring the ontology and also to exploit the benefits of the modular development. The well-specified syntax enhances the use of ontologies because it allows processing from algorithms. Crucial principle of OBO is the use of commonly shared identifiers. The goal of the commonly shared identifiers is to provide backward compatibility with legacy annotations as the ontology is evolving.

[3] http://www.nlm.nih.gov/bsd/pmresources.html.

[4] http://www.nlm.nih.gov/pubs/factsheets/umlslex.html.

OBO is comprised of over 60 ontologies and many are submitted as candidates. OBO is supported by the NIH Roadmap National Center for Biomedical Ontology (NCBO) through BioPortal[5](Rubin et al. 2006b). Except from the support by the NCBO the developers for a subset of OBO ontologies have created the OBO foundry[6] (Smith et al. 2007) which is a trial of the use of evolving principles in a voluntary basis by the participants. Significant biomedical ontologies such as FMA and Gene Ontology have been reformed with the use of OBO foundry basis.

4 Semantic Web Technologies and Ontologies in Healthcare

Ontologies in healthcare and semantic web technologies in general are used to share, integrate and reuse biomedical data to enhance diagnostic procedures, limit costs and to enhance research in biomedical field. In this section ontology based integration in the biomedical field will be discussed, how ontologies are integrated in HIS and how interoperability is promoted in healthcare with the use of ontologies.

4.1 Ontology Based Integration of Heterogeneous Biomedical Data

Data that are produced from healthcare providers are huge, unstructured and diversed. There are in many formats (images, free text, structured text) and stored in different systems that in many cases cannot communicate. Due to that the management and the retrieval of that data is very difficult and it is an extra burden for the quality of the healthcare service. A first to bring order to the chaos is to create distributed or centralized repositories of medical data. Although repositories are a first solution if the data are not integrated at a semantic level the access to that data remains difficult and also there is no capabilities for semantic search. Due to that the exploitation of the data remains low and hidden knowledge is at the exploited data.

The semantic web technologies can provide data integration in semantic level with the use of ontologies. This approach has been used in ONTOFUSION (Pérez-Rey et al. 2006) and Bio2RDF (Belleau et al. 2008). Core element of ontology-based integration is the domain ontologies. The domain ontologies are mapped with the data sources, which can be relational databases of RDF triplestores, or views that are derived from the data sources. The domain ontologies are used to provide a common vocabulary and also the relationships between the different concepts. Different elements from the data must be described only with concept names from the ontologies. As a result of that the semantically equivalent elements that are

[5] http://bioportal.bioontology.org/.

[6] http://www.obofoundry.org/.

stored in difference databases are now described by the same concept. The creation of such a semantic repository allows users to make semantic queries to a unified biomedical pool of data from one user interface and also allows the semantic integration of new databases.

4.2 Ontologies and HIS

HIS are informatics systems for managing the data that are produced in hospitals and healthcare providers in general. The core element of the HIS is the Electronic Health Record (EHR). EHR contains the personal info of a patient, patient history and also medical data that are derived from the subsystems of HIS. Additional to EHR there is DICOM[7] and HL7[8]. DICOM is used for the storing and transmitting of the medical images and HL7 is for exchanging, integrating, sharing and retrieving medical electronic information. The main issue of the HIS and standard that is used for storage, communication and information retrieval is that the data are not structured and unified. Even though the standards provide mechanisms for storing metadata these metadata are not based on standardized terminologies but they are in free text. As a result of that interoperability between institutions and even hospital departments is difficult, information retrieval is limited to text queries results and huge amounts of medical data remain unexploited.

Due to the above the integration of standardized terminologies and ontologies to standards is suggested as a solution. The main terminology which was selected was SNOMED-CT because it covers most of the areas that are used in medical practice such as clinical findings, symptoms, diagnoses, procedures, body structures, organisms and other etiologies, substances, pharmaceuticals, devices and specimen. Also SNOMED is used for structuring the reports at the EHR. The mapping of the current version of HL7 (HL7 RIM[9]) with SNOMED-CT is suggested (Ryan 2006; Ryan et al. 2007) for enhancing interoperability between HIS.

Additional to the above openEHR (Kalra et al. 2005) is created for enhancing interoperability between HIS and healthcare organizations. The community of openEHR focuses its work at interoperability between EHRs and HIS. OpenEHR inserts a reference model, which is called archetype, which models the clinical procedures (Chen et al. 2009) and the query language. That model in order to provide interoperability is designed to make use of external terminologies and ontologies such as SNOMED, ICDx, LOINC[10] and others. Also there are several works that are proposing the mapping of the archetypes with the Semantic Web Rule Language (SWRL[11]) (Lezcano et al. 2008; Viklund and Karlsson 2009) for better reasoning and information retrieval (Lezcano et al. 2011).

[7] http://medical.nema.org/.

[8] http://www.hl7.org/index.cfm.

[9] http://www.hl7.org/implement/standards/product_brief.cfm?product_id=77.

[10] http://loinc.org/.

[11] http://www.w3.org/Submission/SWRL/.

5 Use of Ontologies in Biomedical Research

Biomedical research has a broad spectrum of fields from Biology to Medicine, from Bioinformatics to Biomechanics. Due to the chaotic form of the data the exploitation of them and the production of information is very difficult. Many techniques are used from Artificial Intelligence such as neural networks, pattern recognition and data mining but the main drawback of them is that they cannot integrate the semantic meaning of the data they process. An example of that is the Semantic Gap (Hare et al. 2006) when low level features cannot be connected with high level features of the image. The semantic gap limits the power of the Content Based Image retrieval (CBIR) (Akgül et al. 2011) systems in retrieval and also hides important knowledge from the researchers. The major advantages that ontologies and semantic web technologies bring in biomedical research can be easily shown in the research field of epidemiology (Ferreira et al. 2013). The heterogeneous resources of epidemiology with the use of commonly shared concepts have a harmonized description. The terms of the resources are mapped to concepts that have defined vocabulary and relations among them. On the other hand the process of the data is done with the semantic relations of the concepts and the retrieval is more accurate and relevant. This is shown in several research works that use ontologies for annotating medical data (Kyriazos et al. 2011; Möller and Sintek 2007; Rubin et al. 2009; Möller et al. 2009).

A major field in biomedical research is the simulation models of Physiology. Simulation models of physiology and pathology as modeled and represented as a set of equations and graphical schematics. In current form of the models there is no link between the data, the equations and the graphic schema. Due to that the update of a equation or an alteration at the graphical scheme is a very time consuming task and also error prone. With the use of ontologies it is feasible to create a framework that in the same time can represent the graphical and the mathematical part of the model. Also changes that can occur in the mathematical model can be propagated to the graphical model and vice versa. As a result of that the maintenance and the extension is easier and is done in an explicit framework, which is given from the ontology. Several research works propose such frameworks in cardiovascular dynamics (Rubin et al. 2006a) and the VPH[12] European Commission projects create such frameworks for physiology and pathology.

6 Conclusions

Healthcare sector produces heterogeneous, diversed and huge data. Current information systems and health standards are not able to exploit the data in an efficient way. The main reason is that these systems cannot interpret the semantic content of the data and also the structure of the data and the metadata that describe them are not structured in a formal and unified way. As a result of that there is hidden

[12] http://www.vph-noe.eu/.

information that lies in the pools of data, retrieval of information is limited and also there is limited interoperability in between HIS and healthcare institutions. The drawbacks that are mentioned above can slow down diagnostic procedures, make integration of new data difficult and limit the research capabilities. The use of semantic web technologies and ontologies can create the infrastructure for homogeneous information access through the unification of data pools in a semantic basis, structure the data form and provide enhance systems interoperability. There are several ontology engineering tools such as Protégé[13] that allow the development of new ontologies and create ontology based mappings with data pools. Additional to that there are active communities such as OBO foundry and openEHR that through collaborative work set principles in creation of biomedical ontologies and clinical models respectively in order to promote semantic interoperability of data and systems. Moreover NIH has created NCBO BioPortal, which is a repository of ontologies, and are available through REST web services.

However there are some certain drawbacks in the use of semantic web technologies and ontologies in healthcare. One major drawback is that currently the mappings that are used in data integration with the use ontologies is made semi-automatically. Research community must focus in the creation of tools that can integrate with an automatically way the new data pools. Also the development of ontologies is not made with the same principles and due to this the alignment between them is a very difficult task. Finally the maintenance of the ontologies has to be made in a way so that every HIS has the most updated version. A solution of this is creation of web services that provide the ontologies and the terminologies as the BioPortal provides or the use of local sub-ontologies as it is suggested in (Sari et al. 2013)

References

Akgül CB, Rubin DL, Napel S, Beaulieu CF, Greenspan H, Acar B (2011) Content-based image retrieval in radiology: current status and future directions. J Digit Imaging 24:208–222

Belleau F, Nolin M-A, Tourigny N, Rigault P, Morissette J (2008) Bio2RDF: towards a mashup to build bioinformatics knowledge systems. J Biomed Inform 41:706–716

Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. Sci Am 284:28–37

Bodenreider O (2004) The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res 32:D267–D270

Burgun A (2006) Desiderata for domain reference ontologies in biomedicine. J Biomed Inf 39:307–313

Chen R, Georgii-Hemming P, Åhlfeldt H (2009) Representing a chemotherapy guideline using openEHR and rules. In: MIE. pp 653–657

Ferreira JD, Paolotti D, Couto FM, Silva MJ (2013) On the usefulness of ontologies in epidemiology research and practice. J Epidemiol Community Health 67:385–388

Gruber TR (1993) A translation approach to portable ontology specifications. Knowl Acquis 5:199–220

Hare JS, Lewis PH, Enser PGB, Sandom CJ (2006) Mind the gap: another look at the problem of the semantic gap in image retrieval. 607309-607309-12

[13] http://protege.stanford.edu/.

http://www.springer.com/978-3-319-06843-5

Concepts and Trends in Healthcare Information Systems
Koutsouris, D.-D.; Lazakidou, A.A. (Eds.)
2014, XI, 235 p. 79 illus., 43 illus. in color., Softcover
ISBN: 978-3-319-06843-5