Application of ontologies and semantic web technologies in the field of medicine

Petrika Manika
Department of Informatics
University of Tirana
petrika.manika@fshn.edu.al

Elda Xhumari
Department of Informatics
University of Tirana,
elda.xhumari@fshn.edu.al

Ana Ktona
Department of Informatics
University of Tirana
ana.ktona@fshn.edu.al.com

Aurela Demiri
Master of Science in Informatics
University of Tirana
aurela.demiri2@fshnstudent.info

Abstract

The complexity and diversity of knowledge and terminology of medical science is one of the main obstacles to successful interdisciplinary studies. Important data are difficult to find and to be selected mainly due to different formats, schemes and semantics they have. Extracting automatic knowledge from previous practices and past medical history is also very difficult due to the diversity of medical systems. This paper aims to present a solution to these problems by using ontologies and semantic web technologies. In this paper we intend to emphasize that the use of ontologies and semantic web technologies like RDF, OWL and SPARQL can provide the necessary semantics for a variety of medical domains and, moreover, can serve as tools for building innovative solutions technology to existing problems.

1. Introduction

One of the most challenging problems in the field of health care is to ensure interaction between health care systems "(Bicer, Laleci, Dogac, & Kabak, 2005). Ontology improves interaction in healthcare systems. In this area, interaction between systems is more delayed than in other areas such as finance where ontology is better. Ontology can help build robust and more interoperable information systems. They can support the need for a healthcare process for re-use, transmitting and sharing personal data of the patient. The use of ontology in medicine is mainly focused on the representation and organization or reorganization of medical terminologies. If it is understandable, ontology is an essential part of any solution to the problems of medical terminology. In Albania the knowledge about ontology are very little compared to developed countries, or there is almost no knowledge in this area. This paper will present the role of semantic network and ontology in the field of medicine. It will be determined the role of information technology in this area and the features that need to be taken into account by the technological solutions. Knowledge in this material can serve as a beginning for anyone wishing to learn and practice the ontology and technologies of Web Semantics. Specifically, in this study were taken the patient's medical records, about who is created an ontology scheme and some simple query. Ontology built here is just a small part of what can be done for medical records. The field of study for the application of ontology is supposed to be applied precisely in medicine to the importance of maintaining data retention in this field.

2. Characteristics of Semantic Web

The web has become an inseparable object of our daily lives and the amount of online information is growing. In addition to simple texts which occupy most of the multimedia information like graphics, audio or video have become a prevalent part. The question is how to find useful information in this vast of information? It means to find what you need without much research while there is infinite information about what you are looking for. Traditional search engines will come to the limits of their powers when it comes to understanding the content of the information. Here comes Semantic Web Help [Bern98]. Tim Berners-Lee identifies two main objectives the Web has to fulfill [Bern97].
a) The first objective is to enable people to work together by allowing them to share knowledge between them.
b) The second objective is to include tools that can help people analyze and manage the information they share

in a meaningful way. This vision has become known as the Semantic Web (SW). So, initially, the term "Semantic Web" has been invented by Sir WWW Sir Tim Berners-Lee and has also inspired many researches in this field. Ever since its inception, the development of Semantic Web technologies has been closely linked to WWW. The semantic web is an extension of the traditional web, in the sense that the information contained in the natural text is supplemented by clear semantics based on the formal representation of knowledge. It has also been conceived as an extension of the WWW that allows computers to search, combine, and intelligently process web content, based on the understanding that this content has on people. In the absence of artificial intelligence at the human level, this can only be achieved if the intended meaning of web resources is clearly defined in a format that can be processed by computers. In order to achieve this, it is not enough to store data in a machine-readable syntax, but it is required that these data are equipped with a semantics that clearly specifies the conclusions that should be drawn from the available information on our disposal. Often it is difficult for people to agree on the validity of the content of a document, and it is therefore much more difficult to formalize that document in order to make sense even for the computer. So this is an impossible attempt. The purpose of the Semantic Web is to enable cars to have more information that so far has required human time and attention. As a consequence, the Web Semantics does not refer to a specific extension of WWW, but is rather an ideal against which the network evolves over time. Meanwhile any progress in this area can also be useful in applications that are not closely related to the web. The Semantic Web Vision is to add a syntax web so that resources are more easily interpreted by programs (or 'intelligent agents'). As mentioned above, the Semantic Web is a vision for the future of the web in which information is given clear meaning, making it easier for machines to process and automatically integrate available information into the Web. In order to better understand the Semantic Web, the architecture built by Tim Berners-Lee, called "Semantic web stack" [Alq16] or "Semantic Web Layer Cake" (Figure 1), comes in handy.

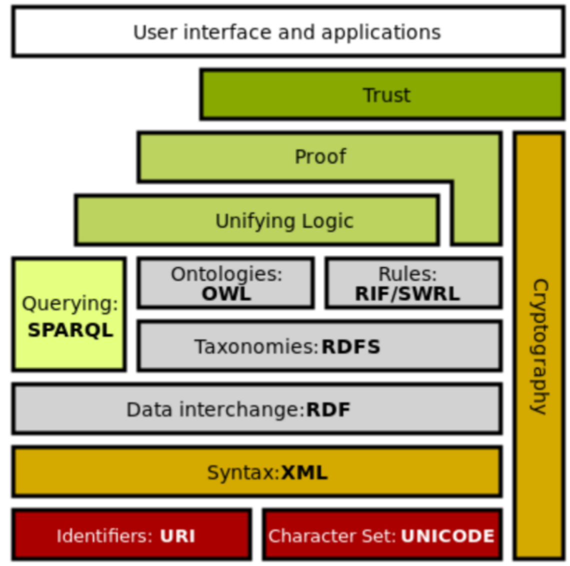

Figure 1: Semantic Web Stack

2.1. Why is semantic needed?

Natural language is amazing; it is hard to imagine a better API for knowledge [Seg,Eva,Tay12]. Without much effort, we can ask a foreigner how to find a place where we want to go; we can share our knowledge of different things in our community of friends; we can go to the library, take a book, and learn from an author who lived hundreds of years ago. Semantics is the process of communicating enough meaning to result in an action. The sequence of symbols can be used to communicate meaning, and this communication can affect behavior. For example, when we read this material, what we are doing is integrating the ideas expressed in these sentences with everything we knew before. If the writing semantic in the material is clear, it should help readers to create an idea about ontology and web semantics to serve as a starting point for other larger jobs. If the semantics is not reached, then the reader will not get much benefit from the study. So semantics is very important and is the main thing to be considered during any kind of work. The foundations of Semantic Web technologies are data formats that can be used to encode knowledge for computer-based processing, although the focus is on different forms of knowledge. We summarize three key issues that provide conceptual support for the Semantic Web: Model Building, search to describe the world in an abstract way, to allow for easier understanding of a complex reality. Knowledge-based computing, attempting to build reasoning machines that can make meaningful conclusions from encrypted knowledge. Exchange information, transmit complex source information between computers that allow us to disseminate and link knowledge at a global scale.

2.2. The scope of the semantic web

Semantic Web technologies and semantics provide us with a new approach to information management and processes, the basic principle of which is the creation and use of semantic metadata. Metadata is a data set that provides information about other data [Dav,Stu,War06]. Metadata can exist on two levels. On one hand, they can describe a document, for example a website, or part of a document, for example a paragraph. On the other hand, they can describe subjects within the document, for example a person, country, or company. In both cases the most important thing is that metadata is semantic, so they tell us about the content of a document and the link to other documents or even about a subject within the document. This contrast with metadata on today's Web site simply describes the format in which information should appear: Using HTML, it can be specified that a given string should be displayed in bold, red color, but can not specify that the string indicates a product price, a copyright name, and so on. There are a number of additional services that metadata can do (Davies et al., 2003). Initially, based on the meaning of information we can arrange it and find it, not just the text. Using semantics the systems can understand where words or expressions are equivalent. When we search for 'George W Bush' we can get a document that is equally valid referring to the "United States President". Inversely systems may differ when the same word is used with different meanings. If we search for references to the word 'Jaguar' in the context of the motor industry, the system may ignore references to large cats or the operating system of the same name. The way information is presented can be improved by using semantics. Instead of a search that provides a linear list of results, results can be summed up by meaning. So a search for 'Jaguar' can provide documents assembled on whether they are related to cars, large cats, operating systems, or different themes all together. However, we can go further by using semantics to merge information from all related documents, removing surpluses and summing up where appropriate. Links between the main entities in the documents can be represented visually. The reason behind all this is the ability to justify and draw conclusions from existing knowledge in order to create new knowledge. The use of semantic metadata is also crucial for integrating information from heterogeneous sources, whether within an organization or across organizations. Various schemes are used to describe and classify information, and different terminologies are used together with information. By creating different schemes, it is possible to create a unified view and achieve interaction between processes that use the information. Semantic descriptions can also be applied to processes presented as web services. When the function of an Internet service can be described semantically, then the web service can be more easily detected. When existing web services are provided with metadata describing their function and context, then new services can be automatically compiled from a combination of existing services.

3. Ontologies

Semantic technology has the potential to provide solutions for many limitations, providing enhanced access to knowledge based on the use of metadata processed by the machines. By using semantics we can improve the way information is presented. Basically, in all semantic Web applications is the use of ontology. They facilitate the sharing of knowledge and reuse between agents, whether human or artificial. The word "Ontology" comes from onto-logos and is the science of being. In computer science, ontology is a formal representation of knowledge from a variety of concepts within a domain and the relationship between concepts. The most widespread in computing is the definition of Gruber [Gru05]. Gruber initially defined ontology as an "explicit specification of a conceptualization". From the beginning there were some discussions about this definition and is thought to be incomplete. Therefore, other authors over the years have improved this interpretation by making it more expressive and comprehensible. In 1997, Borst defined ontology as a "formal specification of a common conceptualization" [Bor97]. Several years later, Studer et al. expanded this definition by introducing a new and more complete definition: Ontology is a formal and explicit specification of separate conceptualization" [Stu, Ben, Fen98]. The formal term has to do with what the ontology should be readable by the machine. Explicit means that the type of concepts used and the limitations during their use are explicitly defined, so clearly stated and in details that leave no room for confusion. The meaning of the concepts should be clearly defined. While conceptualization is an abstract, simplified picture of the world we want to present [Gen, Nil87]. Conceptualization describes knowledge on the domain (domain). The divided term emphasizes that ontology should include knowledge received from different groups rather than individual ones.

3.1. The scope for the use of ontologies

Ontology is developed and defined to share knowledge among researchers working in the same field. Some of the main purposes of using ontology are: sharing knowledge on the same domain, reuse of previously used ontology, sharing domain knowledge from operational knowledge. Some of the design criteria for ontology are: clarity, coherence and extension. Ontology consists of a number of different components. The names of these components vary depending on the ontological language used and the authors. But the core components are common among different ontology. These components can be divided into two types: those that describe domain entities - called concepts, individuals, and relationships; and those that either enable the use of ontology or describe ontology itself. The common components of ontology include: Classes, Can be real-world concrete objects or abstract concepts. Individuals, instances or objects, represent the basic or atomic level of ontology. Attributes, aspects, properties, features, or parameters that objects might have. Links or Relationships: Define the ways in which classes and individuals can connect to one another.

4. Ontology in the field of medicine

The use of ontology in medicine mainly focuses on the representation and organization of medical terminologies. Doctors have developed their specialized languages in order to help them maintain and communicate effectively general medical knowledge and patient information. Such terminologies, which are optimized for human processing, are characterized by a considerable amount of knowledge not clearly expressed. Meanwhile medical information systems should be able to communicate complex and detailed medical concepts. This is a difficult task and requires a profound analysis of the structure and concepts of medical terminology. It can be achieved by building ontology of the medical field in order to present medical terminology systems. The benefits of ontology in this area:

• Ontology can help build stronger systems and higher interaction of information in health care. Interaction in health care is the ability of various technological information systems and software applications to exchange data and use this information that is exchanged.

• Ontology may support the need for a healthcare process to transmit and reuse patient records.

• Ontology has the ability to support the integration of knowledge and data, which can be considered as the most important benefit they can bring to healthcare systems.

4.1. Current work on medical ontologies

There are some ontology and systems that have been created nowadays in the field of medicine. Some of them are: CO-ODE: Cooperative Open Environment Development Project, Medical Informatics Group at the University of Manchester. LinKBase: is a knowledge base of over one million language independent medical concepts. Contains an ontology with a formal conceptual description of the medical field. MedO - a bio-medical ontology developed at the Institute of Formal Ontology and Medical Information Systems, Germany. The Basic Model of Anatomy - an ontology which represents a coherent body of explicit declarative knowledge about human anatomy. The Consortium of Ontology of Geneva which aims to produce a controlled dictionary that can be applied to all organisms as knowledge even if the role of genes and proteins in the cell is accumulating or changing.

4.1.1. GALEN

GALEN and "Galen-Core" is top level ontology for medicine [Noy, McG]. GALEN is the abbreviation for Generalized Architecture for Languages, Encyclopedia and Nomenclature in Medicine. This is a European project (1992-1999) developed for the reuse of terminology in clinical systems. GALEN was developed from the terminology based on the knowledge of the electronic records system. GALEN unlike most traditional terminologies provides terminology in order to construct terminology description blocks. GALEN developed appropriate technology for [Gua]:

• Allowing clinical information to be captured, displayed, modified and appear more powerful.

• Support the reuse of information to integrate medical records and other clinical systems.

4.1.2. Gene Ontology

Gene Ontology (GO) is a controlled biological terminology that is created by a bioinformatics society. It is a relatively new ontology compared to other ontology but has a greater impact on the bioinformatics community. Initially, GO addressed terminologies from three databases: Flybase, Saccharomyces Genome database, and Mouse Genome database. The gene database is the central repository for genomic mapping data that results from the Human Genome. Later then developed three hierarchies of terms to describe biological processes, cellular components, and molecular functions. Authors have found it reasonable to put definitions and comments about genes on the ontology notes.

5. An ontology example for medical records

The example that will be shown below is the proposal of an ontology for medical records in the republic of Albania. For this purpose, Kartele.owl has been built using Protégé. The classes that are part of this ontology are listed and described below:

• **Person:** a person can be a patient or a clinically qualified person.

• **Mjek:** a person qualified to treat people who are ill.

• **Pacient:** a person with a health problem.

• **Diagnoze:** Determining or identifying a diseased condition. Diagnosis has two subclasses that are *Diagnoze_shtrimi* and *Diagnoze_e_daljes.*

• **Sëmundje:** a disorder of structure or function in a person. The disease has subclasses: *Ekzaminimi_i_pergjithshem, Histori_e_semundjes dhe Rezultatet_e_Ekzaminimit, respectively Overall Examination, Injury History and Examination Results*

• **Trajtimi:** medical care given to a sick patient. Treatment has subclasses: *Barnat_e_marra, Dhenia_e_mjekimit,Dozat_e_marra dhe Kohezgjatja,* respectively: *Barn, Housekeeping, Dummies and Duration*

• **Spitali:** the institution providing medical treatment and health care for people with health problems. A hospital has doctors, nurses, and other staff that we do not need for our ontology.

In addition to class-subclass links that are easily distinguished from the above figure, Object Properties that are created through Protégé are

• ka_pacientë

- domain: Mjek
- range: Pacient

• ka_mjek

- domain: Spital
- range: Mjek

• punon_në

- domain: Mjek
- range: Spital

• ka_sëmundje

- domain: Sëmundje
- range: Pacient

• ka_trajtim

- domain: Mjekimi
- range: Sëmundja

• varet_nga

- domain: Sëmundja
- range: Trajtim

For each of the Object Properties, define the domain and range that are the class or subclass. Then there are Data Properties that describe the connection between instances and data veils. Some built Data_properties are:

• Id

- domain: Person
- range: int

• emri

- domain: Person
- range: String

• gjinia

- domain: Person
- range: String

• adresa

- domain: Person
- range: Double

• specializim

- domain: Mjek
- range: String

Also for Data Properties we define the domain and range, but in this case the domain is a class whereas the range is the type of value.

These were some Data Properties created for the Person class, and are automatically applied to the subclasses of this class: **Mjeku** and **Pacient**.

Only *specializim (specialization)* will be special for the **Mjeku** subclass and will not apply to the **Pacient.**

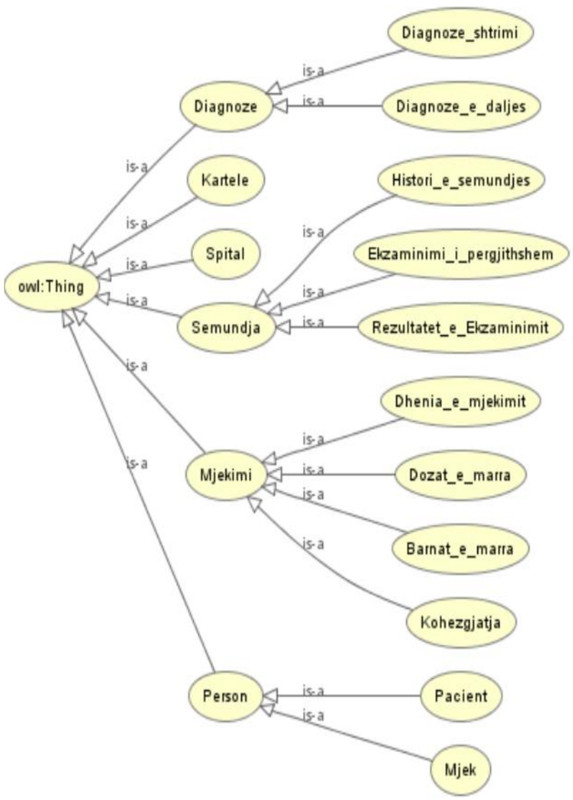

Figure 1. Ontology graph

The figure shows a graph built with OWLViz where it is clearly seen that the subclasses are linked to their superclasses according to the IS-A connection. OWLViz is one of the important functionalities of Protégé, explained in the Technology Section used.

6. CONCLUSIONS

The origin of the challenge of extracting or storing information dates back to the ancient times of humanity.

Using Web Semantic information can be exchanged more easily through different systems and they can communicate more efficiently with each other. Ontology is the best way to structure and model information.

In recent years, the development of ontologies has taken a tremendous shift from the product of artificial intelligence labs to experts from various fields.

Ontology defines a dictionary for researchers who want to share information in a field, it includes definitions that can be interpreted by computers, definitions of field-based concepts, and links between them.

The main reasons and advantages why one might want to develop an ontology are:

• Sharing knowledge on the same domain.

• To reuse previously built ontologies

• To share domain knowledge from operational knowledge

7. REFERENCES

[Bern98] T. Berners-lee. Semantic Web Road Map, Oct 1998

[Bern97] Tim Berners-Lee. Realising the Full Potential of the Web, Dec 1997

[Alq16] A. Alqazzaz. State of the Art of Semantic Web. International Conference on Industrial Engineering and Operations Management Detroit, Michigan, USA, September 23-25, 2016

[Seg,Eva,Tay12] Tobby Segaran, Colin Evans, Jamie Taylor. Programming the semantic web, 2012

[Dav,Stu,War06] John Davies, Rudi Studer, Paul Warren. Semantic Web Technologies - Trends and Research in Ontology-based Systems, 2006

[Gru05] Gruber, Thomas R. "A translation approach to portable ontology specifications". Knowledge Acquisition, 2005

[Bor97] W. Borst. Construction of Engineering Ontologies. PhD thesis, Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands, 1997

[Stu, Ben, Fen98] R. Studer, R. Benjamins, and D. Fensel. Knowledge engineering: Principles and methods. Data & Knowledge Engineering, 1998

[Gen, Nil87] M. R. Genesereth and N. J. Nilsson. Logical Foundations of Artificial Intelligence. Morgan Kaufmann, Los Altos, CA, 1987

[Noy, McG] Ontology Development 101: A Guide to Creating Your First Ontology, Natalya F. Noy and Deborah L. McGuinness, Stanford University, Stanford, CA, 94305

[Gua] Formal Ontology and Information Systems, Nicola Guarino, National Research Council, LADSEB-CNR, Corso Stati Uniti 4, I-35127 Padova, Italy