

Applying Ontologies for Semantic Information Integration on Electronic Medical Records (EMRs)

Suarez Barón M. J.¹, Ospina Becerra V.E.², Salinas Valencia K.E.³

¹Researcher, Research Center at Unitec University, Bogotá, Colombia.

²Researcher Professor, Colombian School of Engineering, Bogotá, Colombia.

³Teacher, Department Education of Boyacá, Tunja, Colombia.

Abstract

This work describes the development of ontology in the domain of Electronic Medical Records (EMRs). The purpose is to apply the semantic web as an integrating tool for information and knowledge management in clinical histories, supported by the design of an ontology of the medical domain in a telemedicine environment. The final result is the design of an ontology in the medical domain that allows the management of patient records in Spanish language. In addition, the ontology is complemented by the implementation of a semantic web service that facilitates the interrogation, maintenance and consultation of the ontology.

Keywords: e-health, Ontologies, Semantic web, information retrieval, information indexing, Electronic Clinical Records.

INTRODUCTION

The Semantic Web is a novel system of web considered for permitting human and informatics tools to development and shares the same sources of information. The Semantic Web relies on a set of standards which deliver syntactic consistency and semantic value to all of its resources and contents. E.g., the other site, description Logic is applied as the theoretic base for the description of web items, and the languages RDF and OWL for their syntactic representation. Description Logic defines a family of knowledge representation languages which can be used to represent, in a well-understood formal way, the knowledge of an application domain.

Now, within the Semantic Web there are tools or applications that solve specific problems, although they apply common techniques, the developments are fragmented, generating a problem of disarticulation of these topics under the same thematic axis; causing a low production of key documentation of the essential paradigms of the Semantic Web specifically in relation to information retrieval, integration of information and indexing of information as fundamental tools in the management of information and knowledge.

On the other hand, the application of the semantic Web in real scenarios is another lack that can be evidenced in the environment and this leads to a lack of knowledge in the specialty, to the application of this type of projects, to marked scientific gaps and technological and to a limited application

in the management of information and knowledge because there is not a documentary base or prototypes sufficiently articulated to technically support the projects from the curricular to the investigative.

The final result obtained is OGHC ontology building (*its Spanish acronym Ontología para Gestión de Historias Clínicas*). The development is based on the methodology "Methontology"[1]. This methodology is one the most known and applied in the development of semantic web applications.

BACKGROUND

Much of the current literature on ontologies for health (ehealth) is focused to the application of information extraction (IE), applications running in different domains, from treatment, diagnosis and medical records to manage patient information hospitals and clinics. In environments real-time tracking ontologies emerge as an initiative to provide semantics to unstructured data supported medically specific vocabularies, Snomed or libraries give a concrete application of these vocabularies as medline.

One of the most coherently structured ontologies in biomedical and specifically in telemedicine is the Fundamental Model of Anatomy (FMA), it is an ontology of domain of the classes and the relationships that belong to the structural organization of the body[2]. Its developers have extensively described the disciplined approach they used, which was based on a set of stated principles, high-level schemes, Aristotelian definitions, and a formalism based on frames also known as knowledge representation schemes.

SNOMED-CT terminology covers diseases, clinical findings and procedures, and helps with indexing, storing, retrieving and aggregating clinical data consistently. This allows computer-structure and manages medical records, reducing the variability in the way in which you can acquire, use and code the data necessary for the clinical care of patients. Its basic elements are:

- Concepts: Represent a minimum unit of meaning.
- Hierarchies: Composed of top-level categories and their respective subcategories.

- Relationships: The type "IS_A" link concepts allow hierarchies; attribute relationships between hierarchies connected concepts.
- Descriptions: Terms or names associated with a concept

The other hand, the Taxonomies is utilized to represent the knowledge on the ontological base. They are used to classify the terms contained in the semantic vocabulary for the sake of performing a simpler searching for the system.

“The taxonomy groups several tags around a set of concepts, therefore by mapping its tags to the taxonomy, they can infer the concepts the user is interested in, even if these concepts are not explicit in the user's tags” [2].

The Semantic Web has provided advances in several computer science fields, being information systems for medicine an example. It is important to remark that the admission operation of a patient is a critical point for the electronic medical recording (EMR) procedure. The use of information technologies has offered opportunities, particularly in hospitals, for the support and improvement of the EMR process [3].

Current systems make use of patients' data and the reported symptoms in order to perform a posterior semantic search to infer those most probable diseases affecting the patient. This helps to significantly improve the speed of attention in admission process in hospitals.

Regarding clinical data analysis, the principal problem is related to the ability to recognize information embedded in heterogeneous data and the description of the relations between them; therefore, important actions should be performed towards information and knowledge management.

In the Information management, the ontologies give a set of procedures that allow controlling the life cycle of clinical data. The main goal of these processes is to minimize the interaction between the human beings and the machine such that the machine independence is increased. The idea is to establish a query and information management system highly autonomous such that it minimizes dependence on user experience and associated errors [4]. The main function of the semantic web in the information management system is the indexing and retrieving of information for the later use. Ontologies provide facilities for the understanding of data to the same machine without requiring human interaction [5].

For the authors in [6], knowledge management (KM) has been one of the success factors for competitive advantage in organizations. On the other hand, new developments in information technology (IT) have provided innovations in the field of knowledge management. That is, IT innovation has become a crucial factor in the survival and success of organizations in the information era.

Semantic annotation tools allow fully retrieving of the knowledge required by the manager; and, research in this area is constantly growing and evolving. In addition, new advances in information technology have provided innovations in the field of KM.

The motivation for working on knowledge management supported by IT comes from the need of availability of information accumulated from previous experiences, such that once the information is stored and organized, this can be useful in case of facing similar challenges, where the storage and organizing information are provided by using ontologies.

The Semantic Web is designed to allow humans and software applications to process and share the same sources of information [7]. The Semantic Web is based on a set of standards that provide consistency and syntactic and semantic value to all its contents. For example, the logical detail is used as the theoretical basis for the description of the articles in the web; and, the RDF (Resource Description Framework) and OWL (Onto Web Language) languages for syntactic representation.

The success of the Semantic Web in an organization ultimately depends on its adoption by an important mass of users; that is, it should be applied to realistic scenarios outside laboratories, theoretical models and research centers. In fact, this situation has not occurred, as reported in some studies [8]. It is probably caused by the fact that users find SW difficult for using, complex to develop and only offers trivial solutions. It saves time and money, and provides access to specialized medical services in remote geographical areas [9]. In the telemedicine systems, the ontologies can provide several information management services, where the storage of medical recordings plays an important role (e.g. big data). However, if we want to get access to relevant information including medical vocabulary, the new telemedicine systems should include alternatives that provide semantic information management solutions.

Among other properties involving telemedicine we mention: efficiency, effectiveness, robustness, quality and coverage to a large number of patients. Summarizing, telemedicine is a widespread conception that involves all those fields of medicine supported by IT.

METHODOLOGY

The scheme applied in the development of the ontology is given from the methodology known as "Methontology"[1]. It uses this methodology because an electronic medical record system is a type of system based on data, information and knowledge; therefore in the development arises problems in the acquisition, conceptualization, formalization, implementation and maintenance of the ontology. The methodology implemented applies the following stages:

Design

The aim of this phase in Ontology (OGHC) is to represent the concepts that refer to the individual patient data and information entities required in the automation of information processing. In particular, OGHC indicates the health data relationship, data concerning the patient's medical records and information entities that are used in its management.

On the other side, from the point of view of knowledge management, the ontology allows that other users re-uses domain knowledge such as: diagnostics, patient monitoring data, and medical treatment of patients.

The knowledge base also allows knowledge engineers declare valid other terminology used in knowledge representation. The ontology allows adding, deleting and querying data on patients and determining the types of clinical diagnosis that might involve these actors; in addition, it can allow defining the collection and analysis of data for which three data sources were analyzed: practical experiences, review literature and medical records of patients.

Concepts Integration

Concept integration phase creates the organization and the structure of knowledge acquired during acquisition and formalization activities; using schemas that are independent respect to paradigms of knowledge representation (e.g. semantic net) and implementation in which the ontology will be formalized and implemented. This ontology applies a conceptual model of natural language for semantic annotations given knowledge domain (see Figures 1(a) and 1(b); here it can be seen three semantic annotations given for "Antecedente1", equivalent to "Antecedent" in English language.

From the point of view of ontology: Cancer, Parkinson, Diabetes are semantically representations of the "family antecedent" type; although the ontology could take that as a diagnosis, this would be considered as cases of equivalence or synonym. A similar behavior occurs with semantic annotations and relationships between ID and record personal.

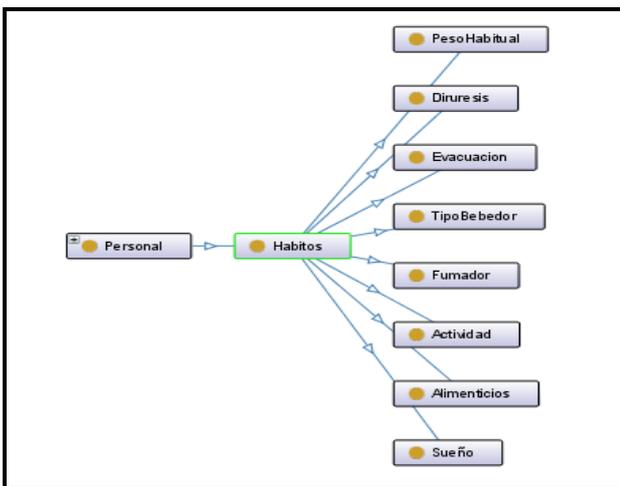


Figure 1(a). Conceptual model

Annotations: Datos1	
Raza	"Mulato"^^string
Sexo	"Masculino"^^string

Figure 1(b). Semantic annotations

The ontology domain (OGHC) involves the concepts and relationships of a history between a health care provider (medical personnel) and their patients; these terms and concepts often are distributed like heterogeneous but through the semantics provided by the ontology the desired information can be naturally integrated. It is important to note that these concepts are considered as actions of an administrative and clinical order given between staff and patients.

Validation

Based on the principle of usability of the information, it is assumed that the OGHC ontology will be applied both to structure a knowledge base (KB) and to support its reused and interoperability to be shared by different levels of users, using initially a structured and precise language natural representation. In the design and construction of ontology it is used the ontologies programming language OWL executed under the platform protégé. As seen in Figure 2, the design of frame of classes is performed taking into account four main entities: "FrecuenciaRespiratoria, ReflejoPupilar, TensionArterial....". These classes are related to generate the complete structure to the ontology.

```

RelacionDiagnostico1
UR: http://www.owl-ontologies.com/OntoHipertension.owl#RelacionDiagnostico1
Annotations:
  FrecuenciaRespiratoria ""^^string
  ReflejoPupilar ""^^string
  TensionArterial ""^^string
  Diagnostico "Asma"^^string
  AntecedentePersonal "Asma"^^string
  Edad "23"^^string
  FrecuenciaCardiaca "62"^^string
  Temperatura "37"^^string
  AntecedenteFamiliar "Asma"^^string
    
```

Figure 2. Frame of Classes for ontology OGHC

RESULTS

The Ontology Implementation

In this phase of ontology development is applied Top Down method starting with the definition of global concepts to reach the minimum level of specificity; for example, the super classes defined are associated in medical record as "Antecedentes, Datos Personales y examen fisico", the classes associated to "Antecedentes" are subordinate in the "familiar" type, and antecedents in the "personal" type.

In Knowledge management, some classes of ontology as race "EstadoAlerta" and "EstadoConciencia" (see figure 3) are key criteria for determining possible of preliminary diagnoses if exist case of some very particular symptomatology with characterization that may occur in some populations and cultures.



Figure 3. Levels OGHC Ontology

Table 1. Notation OWL in terms RDF for the OGHC Ontology

```
<owl:NamedIndividual rdf:about="http://www.owl-ontologies.com/OnoHipertension.owl#Datos2">
  <rdf:type rdf:resource="http://www.owl-ontologies.com/OnoHipertension.owl#DatosPersonales"/>
  <Nombre>Marcela</Nombre>
  <Identificacion>1026873748</Identificacion>
  <Sexo>Femenino</Sexo>
  <Edad>30</Edad>
  <Raza>Caucasico</Raza>
</owl:NamedIndividual>
```

The evaluation and maintenance

The evaluation of the solution was performed by knowledge engineers who will provide theoretical and practical judgment to incorporate new knowledge derived from medical experience that may lead to optimal performance of managed medical record. This support will allow a constant refinement and product quality strengthening.

To information retrieve from ontology, it is used a specific syntax that allows the return and integration of data from relational databases. Currently, there are several languages that allow this work; nevertheless, the W3C consortium was suggested as recommendation to the SPARQL language. SPARQL uses syntax based on Transact-SQL [11] given that multiple statements can be compared. An example of this syntax can be seen in Table 1 which uses a series of commands to return information from the medical record where metadata are stored in terms RDF / OWL and SPARQL recovered through flexible queries.

Jena [12] is the base component in building the semantic web service. This service was designed to allow the end user to view the ontology. Jena provides high operability of information management function and the metadata management functionality is complemented with the characteristic to be system "reasoning" since this system is integrated with the Protégé OWL of inference [13].

The Jena architecture consists of several layers which allow deciphering the content semantics files such as RDF, to subsequently execute the search. The framework built for the launch of OGHC ontology and knowledge management can be seen in Figure 4. In addition to this, you can note the interaction between components and levels at the time to present and display the contents of the ontology in the designed semantic service.

The subclass habits ("Edad") of class "DatosPersonales" contains instances: regular weight "Nombre" and Sexo ("Gender"); it is evaluated through instances "type" if a patient is drinker and/or smoker; it details the feeding practices and the time of their sleep. Each level of detail and specialization of OGHC ontology allows experts and support staff updating the Knowledge Base, in this case by means of type inference rules SWRL (Semantic Web Rule Language). The generation, validation and application of knowledge by the experts [10] are reflected in the continuous updating and feeding of ontology and their semantic relationships between terms that lead to transform tacit knowledge to explicit knowledge.

OWL is the development language used in the construction of OGHC ontology. OWL allows the information management of medical records contained in the ontology and further facilitates the interpretation of the vocabulary explicitly, its significance in terms of domain knowledge and relationships between these terms. Being oriented the ontology as a solution to the interoperability of clinical information on telemedicine platforms, the ontology operates according to semantic specifications, given by W3C consortium.

In the web context through OWL, the ontology works as a graph of nodes relations, properties and attributes, called RDF.

Additionally, RDF has as native language XML (eXtensible Markup Language) and thus any OWL annotation may be interpreted in terms RDF / XML. The above analysis can be seen in Table 1; we used XML syntax which allows you to specify the relationships between the various entities. It also allows the specification of the properties of each entity and its type.

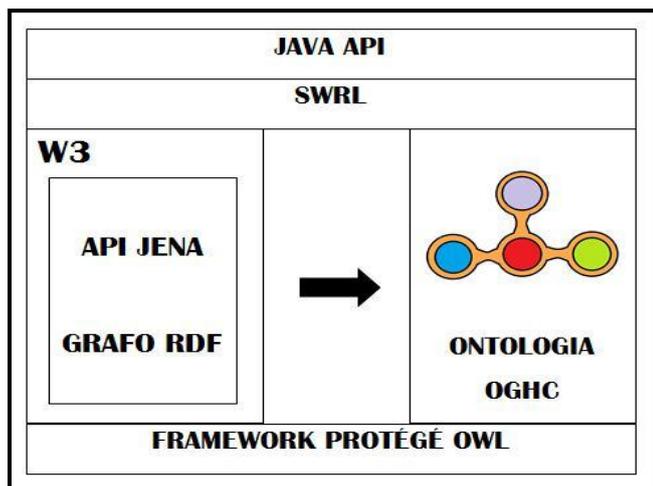


Figure 4. Model of Knowledge and its integration to the service web

Figure 5. Screenshot of the semantic web service for displaying Ontology OGCH

Data acquisition of ontology is semiautomatic through XML documents. The result of this stage allows establishing that the ontology scope will handle basic medical knowledge for patients with mild pathologies. Supported on the conceptualization phase of ontologies where one of the tasks is to define logical rules expressed in natural language, we proceed to implement the rules with the use of the reasoners and specific purpose languages focused to writing rules under SWRL syntax.

To populate the ontology, it starts a load in the ETL processes (Extracting, Transformation and Load). This operation is done by manually incorporating semantic information from the ontology so that processing and extracting tasks are executed from multiple XML documents containing information from Electronic Medical Record of patients; these data XML allow retrieving information from relational databases stored in health centers.

CONCLUSIONS AND DISCUSSION

The management of information and knowledge through OGCH Ontology is a novel proposal that in the near future will allow sectors as health care to have access to multiple resources and global technologies. Indexing information retrieval and multiple data links are some of the alternatives that may have them to implement semantic solutions.

To display the actual functionality and consumption of the OGCH ontology prototype it is implemented a semantic Web service that provides customers the ability to perform queries about the attributes of the ontology. A part of the semantic web service is shown in figure 5; each change, adjust or maintenance on the ontological knowledge model will be visualized to the customer through semantic service.

In the future, one of the tasks that must be done is the indexing to global level of the medical nomenclatures such as SNOMED. Bilingual access to medical libraries as MEDLINE, information retrieval of pharmaceutical terminologies and finally integration with other medical domain ontologies are currently under development and implementation, models in which generic relations are expressed in the medical domain and can be useful to specialists. As a future work, we propose to incorporate a service that facilitates the performing of the ETL tasks.

This service would work automatically for multiple XML documents containing information from relational databases found in hospitals. The DTD design (Document Type Definition) is considered as the source file that will populate the ontology OGCH in an automatically way. The knowledge model will be updated in real time by the ontology OGCH, together with implementing HL7 standards.

REFERENCES

- [1] Kuziemy, C. and Lau, F. 2010. *A four stage approach for ontology-based health information system design*. Elsevier Science Publishers Ltd. 50, 3 (Nov.2010), 133-148. DOI= <http://dx.doi.org/10.1016/j.artmed.2010.04.012>.
- [2] Ilaria, T. 2009. Adaptive systems in the era of the semantic and social web, a survey. *User Modeling and User-Adapted Interaction*. 19, 5 (Dec. 2009), 433-486. <http://dx.doi.org/10.1007/s11257-009-9067-3>
- [3] Paolo, B., Marc, C., Oussama, Z., Annabel, B., and Anita B. 2010. Using semantic web technologies for clinical trial recruitment. In *Proceedings of the 9th international semantic web conference on The semantic web - Volume Part II (ISWC'10)*, Peter F. Patel-Schneider, Yue Pan, Pascal Hitzler, Peter Mika, and Lei

Zhang (Eds.), Vol. Part II. Springer-Verlag, Berlin, Heidelberg, 34-49.

- [4] Gebhardt, S., Wehrmann, T., Klinger, V., Schettler, I., Huth, J., Künzer, C. 2012. *Improving data management and dissemination in web based information systems by semantic enrichment of descriptive data aspects*. Computers & Geosciences. 36,10. (Oct. 2012), 1362-1373 <http://dx.doi.org/10.1016/j.cageo.2010.03.010>
- [5] Colomo, P., Ángel, G., Pedro, S., Marcos, R., and Diego, J. 2010. Case study: A case analysis of semantic technologies for R&D intermediation information management. *Int. J. Inf. Manag.* 30, 5. (Oct. 2010), 465-469. DOI=10.1016/j.ijinfomgt.2010.05.012 <http://dx.doi.org/10.1016/j.ijinfomgt.2010.05.012>
- [6] Jaehun, J. and Sang M, L. 2009. Adoption of the Semantic Web for overcoming technical limitations of knowledge management systems. *Expert Syst. Appl.* 36, 3 (April 2009), 7318-7327. <http://dx.doi.org/10.1016/j.eswa.2008.09.005>
- [7] Dange, R., and Berlanga, R. 2009. A Semantic Web Approach for Ontological Instances. *Software and Data Technologies.* 22. (Nov.2009). 269-282. http://dx.doi.org/10.1007/978-3-540-88655-6_20
- [8] Roberto G., Juan G., Ferran P., Rosa G., Marta O., Juan L., Afra P. and Montserrat S. 2010. Building a Usable and Accessible Semantic Web Interaction Platform. *World Wide Web* 13, 1-2 (March. 2010), 143-167. <http://dx.doi.org/10.1007/s11280-009-0076-2>
- [9] Casado, M., and Santervas, A. Available from <http://www.mecg.es/archivos/DSSI1%20-%20Bloque3.pdf>. (2012); accessed 25 August 2012.
- [10] Nonaka I., Takeuchi H. *The knowledge-creating company*. Oxford: Oxford University Press; 1995.
- [11] SPARQL-SPARQL Protocol and RDF Query Language. Available from <http://www.cambridgesemantics.com/semantic-university/sparql-vs-sql-intro>. (2009); ccessed 20 December 2012.
- [12] Jena, Available from <http://jena.apache.org/>.(2005); accessed 28 December 2012.
- [13] Protégé. Available from <http://protege.stanford.edu/>. (2010); accessed 28 December 2012