An Application Ontology to Support the Access to Data of Medical Doctors and Health Facilities in Brazilian Municipalities

Aline da Cruz R. Souza, Adriana P. de Medeiros, Carlos Bazilio Martins

Department of Computing - Science and Technology Institute, Fluminense Federal University (UFF), Rio das Ostras – RJ – Brazil

acrsouza@id.uff.br, adriana@vm.uff.br, carlosbazilio@id.uff.br

Abstract. *The Web of Data is a global data space based on open standards. However, it is still far from reality found in websites: unstructured and disconnected data focused on human understanding. This work aims to mitigate this problem for a portion of data in the health area, data about medical doctors and health facilities. Thus, this paper presents an application ontology designed to accurately represent such data and some examples of instances and queries, which can be used on the development of applications in order to provide precise information for Brazilian citizens.*

1. Introduction

According to the Google Trends tool, the search volume for the keyword "medical doctor" from the Google search engine had index 60, on a 0 to 100 scale, in the period from 2008 to 2013. The index shows the number of times a keyword was searched on Google in relation to the total number of searches performed in the period. It suggests a considerable use of services like Google to search for data about medical doctors and related terms in the health area. Commonly, Brazilian websites returned from such queries contains unstructured and often incomplete data, mixed with other types of content, such as advertising, hindering the efficient use of these data by citizens. Moreover, published data are, usually, for human processing, which makes hard the reuse of such data in applications.

The Semantic Web (SW) is a Web of Data - dates, names, and any other data that could be conceived. Its technologies (RDF, OWL, SPARQL, etc.) provide an environment in which an application can query this data, make inferences using vocabularies, etc. The set of interrelated data on the Web of Data is called Linked Data (LD) [World Wide Web Consortium (W3C), s.d.]. It allows interconnections to be defined between items in different data sources, aiming a unique global information space [Heath & Bizer, 2011]. The LD principles introduced by Tim Berners Lee [Bizer, et al., 2009] are the following: use URIs as names for things; use HTTP URIs, so that people can look up those names; provide useful information, using standards (RDF, SPARQL); and include links to other URIs for discover more things.

This work presents an application ontology designed to accurately represent data of medical doctors and health facilities of Brazilian municipalities. Instances of the ontology classes were created following LD principles and will become available through a public repository. Some examples of queries that a SW application could perform to aid citizens to access these semantically structured data are presented. Then, final considerations are provided about the use of the SW technologies in the development of applications to allow access to such data.

2. The Proposed Ontology

Application ontologies describe concepts of a domain and specific tasks for implementing systems, the practical part [Guarino, 1997]. The proposed ontology was created following the "Ontology Development 101" [Noy & McGuiness, 2001] guide. Domain was defined as medical doctors, requiring data about medical doctors and health care facilities of Brazilian municipalities. The scope of the ontology was determined by drafting the following list of Competency Questions (CQ) for which the repository should provide answers: CQ1: *What is the specialty of a particular doctor?*; CQ2: *What are the workplaces of a doctor?*; CQ3: *Does a doctor have more than one CRM?;* CQ4: *Do medical doctors have more than one specialty?*; CQ5: *What are the characteristics considered by a citizen when choosing a doctor?*; CQ6: *Is there any hospital in my neighborhood with a particular specialty?;* and *CQ7: What are the available medical specialties in a given clinic?*.

Thereafter, searches were performed in the DAML (http://www.daml.org/) and Schemapedia (http://schemapedia.com/) repositories, in order to locate validated ontologies that could be integrated to this work by reusing their terms. None of them completely met the work needs, possibly because it is a very specific theme. In this ontology version some terms and resources provided by FOAF (http://xmlns.com/foaf/spec), Geonames (http://www.geonames.org/) and DBpedia (http://dbpedia.org/) regarding to cities were used. Thus, the ontology contains internal terms, data, and references to resources from other repositories.

2.1. Classes and Properties

The ontology was constructed using the OWL 2 DL language, with 98 classes and 656 axioms identified by the prefix *med*. Figure 1 depicts the main classes of the ontology. Vertices are classes and edges are relationships between classes. Dashed arrows represent object properties while continuous arrows represent subclasses.

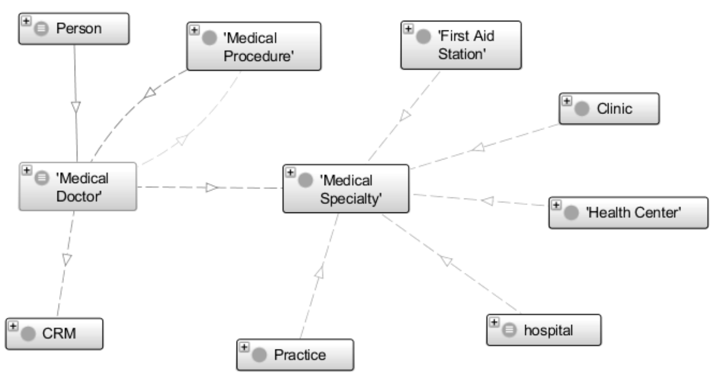

Figure 1. Graph visualization of the ontology.

The class *med:MedicalDoctor* describes a medical doctor, defined as a subclass of *foaf:Person* and as an equivalent class to *dbo:Medician* class. The class *med:CRM* represents a CRM (registry at the Regional Council of Medicine), defined as a subclass of the class *foaf:Document*. *med:MedicalSpecialty* represents a medical specialty and has medical specialties as subclasses, such as *med:Dermatology*. *med:SurgicalSpecialty* describes a surgical specialty, defined as subclass of *med:MedicalSpecialty*, and has surgical specialties as subclasses, such as *med:SpineSurgery*. The subclasses of *med:MedicalSpecialty* and *med:SurgicalSpecialty* allow answers for the CQ1. The classes *med:Clinic*, *med:Practice*, *dbo:Hospital*, *med:HealthCenter* and *med:FirstAidStation* describe workplaces of a doctor, answering the CQ2. Finally, the class *med:MedicalProcedure* represents a medical procedure that a doctor can perform.

The datatype property *foaf:name* describes the name of something, *dbo:address* represents the address of a place, *dbo:date* corresponds to the date of some event, *dbo:status* is used to represent the status of a CRM (active, inactive, etc.), *med:UF* represents the Unity of Federation in which it was issued, and *dbo:number* describes the number. The object property *med:HasCRM* relates a medical doctor to a CRM and is characterized as inverse functional, which guarantees that an instance of the class CRM relates with a single instance of the class *med:MedicalDoctor*. Regarding the CQ3, the answer is "Yes, it is possible.", but a CRM must be associated with only one doctor. The property *med:WorkplaceOf* relates a person (*foaf:Person*) to his/her workplace (*dbo:Place*) and has an inverse property, the *med:WorksAt*, where the domain and range are reversed. Likewise, *med:PerformedBy* also has an inverse property called *med:Perfoms*. This property represents a medical procedure (*med*:*MedicalProcedure*) performed by a medical doctor (*med:MedicalDoctor*). *The property med:HasExpertise* indicates that a doctor has an expertise (*med:MedicalSpecialty*) and *med:HasSpecialty* relates a health unit to a medical specialty.

Restrictions work as basis for the inferences made by the reasoner, defining which features an instance must have to belong to a certain class. For example, an individual is associated to *med:MedicalDoctor* class when he/she has at least one CRM and has expertise on at least one medical specialty. Regarding CQ4, the answer is "Yes, it's possible", so no restrictions were made to constrain the number of specialties associated to a doctor. Another example of restriction specified for the classes *med:Clinic*, *med:Practice*, *dbo:Hospital*, *med:HealthCenter* and *med:FirstAidStation* is that these health facilities shall have at least one medical specialty.

2.2. Instances

The instances were created according to information extracted manually from websites of Brazilian private health plans and data sources of the Brazilian government. The main data sources used were: Unimed Medical Guide (http://www.unimed.coop.br/) - where were extracted the medical doctors names, number of CRM and specializations; Consulta CRM (http://www.consultacrm.com.br/) - where can be collected the remaining CRM data through an API; DATASUS (http://cnes.datasus.gov.br/) - where were extracted URIs of pages that describe each health facility; Website of Rio das Ostras prefecture (http://www.riodasostras.rj.gov.br) - where were collected the medical specialties provided by the health facilities; Geonames and DBPedia - where were collected URIs of resources that represent the cities of Rio das Ostras and Macaé focus of this work. For privacy reasons fictitious data were used in this section examples. Listing 1 shows an instance of *med:MedicalDoctor*, specifically the doctor "Sara de Sa".

The object properties *#HasCRM* (*med:HasCRM*) and *#HasExpertise* (*med:HasExpertise*) were defined as required conditions for association with the *med:MedicalDoctor*. The property *#WorksAt* (*med:WorksAt*) relates this instance to instances of health facilities in which the medical doctor provides services.

```
<ClassAssertion>
    <Class IRI="#MedicalDoctor"/>
    <NamedIndividual IRI="#Sara_de_Sa"/>
</ClassAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="#HasCRM"/>
    <NamedIndividual IRI="#Sara_de_Sa"/>
    <NamedIndividual IRI="#CRM_841631"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="#HasExpertise"/>
    <NamedIndividual IRI="#Sara_de_Sa"/>
    <NamedIndividual IRI="#REC_Pediatria"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="#WorksAt"/>
    <NamedIndividual IRI="#Sara_de_Sa"/>
    <NamedIndividual IRI="#PS_Familia_Rocha"/>
</ObjectPropertyAssertion>
```

Listing 1. Example of instance of *med:MedicalDoctor*.

Listing 2 shows an instance of *med:HealthCenter* representing the Família Rocha Health Center. Note that the property *med:WorkplaceOf*) is inferred by the reasoner from its inverse property *med:WorksAt*. Also worth highlighting relationships with resources located in the external repository Geonames via *foaf:based_near*.

```
<ClassAssertion>
    <Class IRI="#HealthCenter"/>
    <NamedIndividual IRI="#PS_Familia_Rocha"/>
</ClassAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="#HasSpecialty"/>
    <NamedIndividual IRI="#PS_Familia_Rocha"/>
    <NamedIndividual IRI="#REC_Pediatria"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="#WorkplaceOf"/>
    <NamedIndividual IRI="#PS_Familia_Rocha"/>
    <NamedIndividual IRI="#Sara_de_Sa"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty abbreviatedIRI="foaf:based_near"/>
    <NamedIndividual IRI="#PS_Familia_Rocha"/>
    <NamedIndividual IRI="http://sws.geonames.org/3451205/"/>
</ObjectPropertyAssertion>
```

Listing 2. Example of instance of *med:HealthCenter*.

3. Examples of Queries

The following examples show some queries that a SW application could perform from the repository to present useful information for its users. For instance, consider that a woman wants to search for a female gynecologist. Queries like these help to answer the CQ5. The query in SPARQL and its result are shown in Figure 2. This query searches the name, the address and the phone of the medical doctor's workplace. It searches an individual of the type “medical doctor”, whose sex is female, i.e., whose sex is related to the string “FEMININO” by *dbo:sex* and whose medical specialization is related, by the property *med:HasExpertise,* to the instance *med:REC_Ginecologia_e_Obstetricia*. The workplace, represented by *?health_unit*, is associated to name, address and phone by the properties *foaf:name*, *dbo:address* and *foaf:phone*, respectively. The FILTER clause attends an application specification of returning only health facilities of the type *med:Clinic* or *med:HealthCenter*.

```
SELECT  ?name_doctor ?health_unit_name ?health_unit_address ?health_unit_phone
   WHERE
   {
        ?medico         foaf:name ?name_doctor; a med:MedicalDoctor ;
                        med:WorksAt ?health_unit ;
                        med:HasExpertise med:REC_Ginecologia_e_Obstetricia ;
                        dbo:sex "FEMININO"^^xsd:string.
        ?health_unit    a ?tipo ; foaf:name ?health_unit_name ;
                        dbo:address ?health_unit_address ;
                        foaf:phone ?health_unit_phone .
        FILTER(?tipo IN (med:HealthCenter, med:Clinic))
   }
```

name_doctor	health_unit_name
"KARIN DA PENHA BARROS"	"POSTO DE SAUDE DA CIDADE"

health_unit_address	health_unit_phone
"RUA SANTA CATARINA, S/N	tel:+55-22-2768-2008

Figure 2. Query workplace information of a female gynecologist.

Suppose now a search for hospitals that perform general surgery near Rio das Ostras city. Figure 3 shows this query that could answer the CQ6.

```
SELECT ?nome_unidade ?endereco_unid_saude ?tel_unid_saude
   WHERE
   {
        ?unidade        foaf:name ?nome_unidade;    a ?type ;
                        med:HasSpecialty med:REC_Cirurgia_Geral ;
                        dbo:address ?endereco_unid_saude ; foaf:phone ?tel_unid_saude ;
                        foaf:based_near <http://dbpedia.org/resource/Rio_das_Ostras> .
                        FILTER(?type IN (dbo:Hospital, med:FirstAidStation))
   }
```

nome_unidade
"HOSPITAL MUNICIPAL "@pt

endereco_unid_saude	tel_unid_saude
"RUA EUDON MUSTOSA - S/N - PARQUE"	tel:+55-22-2779-6329

Figure 3. Query hospitals near Rio das Ostras, which perform General Surgery.

Such query returns an individual related to the individual *med:REC_Cirurgia_Geral* by the property *med:HasSpecialty*. It shows a link between a resources of the DBpedia and of the local repository by the property *foaf:based_near*. According to the LD recommendations, links with other repositories allow applications to obtain useful information following these links [Heath & Bizer, 2011]. For instance, from this link with DBpedia it is possible obtain other information, e.g., a place description. An application could get this data via HTTP requests sent to a SPARQL endpoint [Sequeda, 2012]. This is a fundamental difference between SPARQL and other query languages such as SQL, which assume that all data being queried are local and conform to a single model. To answer the CQ7, a query returning the medical specialties associated with the clinic would be enough.

4. Conclusions and Future Works

This paper presented an application ontology for describing data of medical doctors and health facilities in a semantic way, in order to facilitate the development of applications for providing access to these data by Brazilian citizens. The data were represented as axioms structured in RDF and expressed from links with Geonames and DBpedia. Finally, some examples of queries that an application could perform were presented.

Many ontologies and vocabularies are available for the health area, such as OMRSE [Brochhausen, et al., s.d.] and those stored in the OBO-Foundry repository [Ashburner, et al., s.d.]. However, most of them describe information that is not provided in this work, like diseases and human body anatomy. The ontology SNOMED CT [IHTSDO, s.d] and the upper ontology UMBEL [Bergman, M. K. & Giasson, F., s.d.] specify terms relating to medical specialties and health facilities. Specifying the relationship between the terms of the proposed ontology and the terms of these ontologies is an ongoing work. Future works include: the development of an application to provide access to the semantically structured data about medical doctors and health facilities and the creation of a repository with data obtained from municipalities, in order to allow the interoperability of information between medical and government institutions and the data management to support the decision-making.

References

Ashburner, M. et al. (s.d.), Open Biological and Biomedical Ontologies (OBO) Foundry, http://www.obofoundry.org, February 2015.

Bergman, M. K. & Giasson, F. (s.d.), UMBEL (UMBEL Vocabulary and Reference Concept Ontology), http://umbel.org, February 2015.

Bizer, C., Heath, T. & Berners-Lee, T. (2009), “Linked data-the story so far”. International Journal on Semantic Web and, 5(3), pp. 1-22.

Brochhausen, M. et al. (s.d.), Ontology for Medically Relevant Social Entities, https://github.com/ufbmi/omrse, April 2015.

Guarino, N. (1997), “Understanding, building and using ontologies”, International Journal of Human-Computer Studies, v. 46, p. 293–310.

Heath, T. & Bizer, C. (2011), Linked Data: Evolving the Web Into a Global Data Space, 1st ed. San Rafael, California: Morgan & Claypool.

Hitzler, P. et al. (2009), “OWL 2 Web Ontology Language Primer”, http://www.w3.org/TR/2009/REC-owl2-primer-20091027, August 2014.

IHTSDO (s.d.), SNOMED-CT (SNOMED Clinical Terms), http://www.ihtsdo.org/snomed-ct, August 2014.

Noy, N. F. & McGuiness, D. L. (2001), “Ontology Development 101: A Guide to Creating Your First Ontology”, http://goo.gl/NvcVTS, July 2014.

Sequeda, J. (2012), “SPARQL 101”, http://goo.gl/zZzubj, August 2014.

World Wide Web Consortium (W3C) (s.d.), “Linked Data”, http://www.w3.org/standards/semanticweb/data, December 2014.