OntoPHC: AN ONTOLOGY APPLIED FOR PRIMARY HEALTH CARE

Eduardo C. Moraes 1,2, Kellyton Brito 2, Silvio Meira 2
1Instituto Federal de Alagoas – IFAL – Maceió-AL – Brazil,
2Centro de Informática(Cin) – Universidade Federal de Pernambuco (UFPE)

{ecm2,kcb,silvio}@cin.ufpe.br

Abstract

Today new researches create a considerable amount of knowledge and scientific expertise related to health problem. There is a need to constantly update their professional knowledge base that enable them to face the diversity and “adversity” of the health process in the daily routine. Within this wealth of information is almost impossible to less experienced health professionals store all specific knowledge of their areas, and there is still the problem of difficulty of fast and easy access to knowledge in health, because this knowledge is mainly internalized in people minds and are not explicitly organized and documented. The use of ontologies is a proposal to represent knowledge in a more rich and detailed way and is a very interesting way to share and reuse formally represented knowledge. Ontology is a representation that can be understood by humans and processed by computers; it is useful to define the common and mathematically based vocabulary in which shared knowledge is represented. This paper describes the development of an ontology to represent part of the Brazilian knowledge in the Primary Health Care and creates a semantic web system to serve as a basis for health professionals better control their routine activities and serves as an information source for health professionals.

Keywords: Semantic Web, Ontology, Primary Health Care, Knowledge Management, Knowledge Engineering

1 Introduction

Today new researches create a considerable amount of knowledge and scientific expertise related to health problem. This knowledge is essential for analysis, validating and efficient detection, storage and share for new researches. It thus became clear that the adequate use of this knowledge is quite a major task. This paper describes the development of a semantic-based system and an ontology representing knowledge for Brazilian Primary Health Care Program.

We observed the ontology can be an important source of relevant knowledge for health professionals experienced in computers. They have few difficult after training, to use the Protégé ontology editor(http://protege.stanford.edu/). New users were trained using the ontology, and with this tool could describe the information flow and the main processes in terms of a conceptual framework. Beyond a web system developed with python/Django was created to support health professionals routines based on Brazilian primary health care program.

The Brazilian Government is investing in primary health care to provide higher quality, lower costs and meet the people where traditional programs cannot achieve. The government has created the Family Health Strategy, where Family Health Program(PSF) is the main program and is part of the so-called Unified Health Care System (SUS).
According to Rosa[24]. It is one of most important programs of the Brazilian national public health, which implements a national policy for primary care settings with the aim of reformulate part of the traditional model of assistance to a preventive model. As its name says, its main focus is on families instead of individuals, and it is organized around multidisciplinary Family Health Teams, formed by a core of professionals such as physicians, nurses, dentists and social workers, as well as community health agents.

With the grown of PSF, a large quantity of knowledge about primary health care have been produced and not represented/managed in a correct way. The evolution of computer technology has enormous potential to improve primary care in the areas like communication between physicians and patients, information sharing among health care providers, training and rapid access to reliable medical information for both physicians and patients. And inside of this context this work proposes a solution. Represent primary health care content throw ontology and create a system based on this ontology.

This paper discusses the development of the ontology OntoPHC and proposes a websystem called SemanticPHC. This paper is composed by: Section 2 introduces the Brazilian Health Care Program, section 3 presents the how the ontology was created and methodologies used, section 4 presents the SemanticPHC System and section 6 discuss the contributions of this work and future works.

2 The Brazilian Health Care Program

Since 1994, the Family Health Program (PSF) has been an important pillar in the reorganization of the Unified National Health System, whose organizational principles include universality and equity. Since 2004, when the program had been implemented in 82% of Brazil's 5561 municipalities, covering some 40% of the total national population. The PSF is centred on a family and community approach in with multiprofessional teams (including physicians, nurses, community health agents, and oral health professionals) work under the principles of comprehensive care[3]. Our main goal is improve the follow-up care by family physicians. They usually work in faraway areas with few or no hospitals nearby and have to support a vast number of cases in different medical areas. They are a link to more complete medical centers. With the grown of PSF(Brazilian family health program), a large quantity of knowledge about primary health care have been produced and not represented/managed in a correct way. The evolution of computer technology has enormous potential to improve primary care in the areas like communication between physicians and patients, information sharing among health care providers, training and rapid access to reliable medical information for both physicians and patients.

3 The OntoPHC ontology

Due to the volume of information generated by projects in the health area, which is clearly an interdisciplinary area with health professionals from various fields such as nursing, physiotherapy, nutrition, psychology, among others. When there is a need to retrieve relevant information about health from "cyberspace", we face the limitation that until now the data of "cyberspace" in his most content is meant to be understood by people, but incomprehensible to programs. Computers can expertly analyze web pages in structural terms, how to check the reference of the links connecting the pages to each other, for example.

In the primary health field, due its interdisciplinary it is quite impossible to formalize all existing concepts. We have observed the rise of the scientific studies about ontologies, but none of the selected studies found englobes the Brazilian primary health experience (PSF). Before the development of the ontology a systematic literature review was made. It founds that there were important proposals in the domain, we outstand: the proposal to create a tool to training primary health care to medical students [2], the clinical practice guideline based ontology driven clinical decision [4] and an ontology change management system [1]. The others studies don’t study deeply the subject primary health care with a semantic perspective/application or only cover this field very superficially. This field still needs more related works and the Brazilian experience can help a lot. An important point is that the search did not find any study in Portuguese language

This work has these specific goals:

1-To represent formally Primary Health Knowledge based on Brazilian Health Program (PSF).
2-To analyze different formalisms to represent knowledge and Upper Ontologies.
3-To extent the BFO Upper ontology.
4- To reuse software components and frameworks
5- To develop a semantic-based system to evaluate the ontology created.

3.1 Methodology of development of OntoPHC

During the development of OntoPHC we have followed Methontology Ontology engineering second Corcho (2005)[16]. This section describes the phases and important activities in the development of OnToPHC according to the steps defined by Methontology methodology.

Following Methontology methodology we will detail the conceptualisation phase. The objective of conceptualization according to Fernandez[21] is precisely this: to organize and structure the acquired knowledge using external representations that are independent of the implementation languages and environments. In others words, this phase organizes and converts an informally perceived view of a domain into a semiformal specification. Our main sources of knowledge were the manuals from the Health’s Ministry, ordinance number 1886/97, ordinance number 3.295/98, ordinance number 648/2006, ordinance number 649/2006, Operational Standard for SUS (NOB 96), website of Brazilian Primary Health Care (http://dab.saude.gov.br/). Other important source of knowledge was interviews with PSF Health Professionals from two Brazilian cities from the State of Alagoas, Penedo and Junqueiro were essential partner of this work.

At Penedo the PSF team were composed by a Physician, an Odontologist and the manager of the medical center. At Junqueiro by a nurse.

Preliminary interviews and meetings with the primary health care experts were done to structure the knowledge captured and to analyse and validate with the conceptual model. They need to decide what would be the granularity and limits of the ontology. It was a barrier to overcome, to teach to health professional why/how/what is an ontology and how to concepts should interconnected. They have to see primary health program knowledge represented into concepts, their properties and their relationships, to evaluate the conceptual model. Then, a need was identified, some specific rules needed to be more detailed for applying the knowledge base. These rules shall be initially described in natural language, to simplify their understanding by health professionals for validation. During the creation of these rules, may have detected the need for new concepts or relationships to simplify or facilitate their definitions such cases, it is important validation of the ontology. Besides procedures related to Primary Health care we added into OnToPHC management data about PSF, more specifically which knowledge is necessary for answer forms solicited by Brazilian health ministry, which monthly requires reports about the ongoing local work, and this knowledge has no structured data and no system to support the process. The main goal is to define relevant knowledge, how to quickly retrieve intelligent and relevant information, the possibility of do queries allowing him to know what to do, how to manage, and which knowledge is faced by Brazilian Health Program daily.

Our main goal will be create a formal and shared knowledge based on the Brazilian experience in primary health care.
Some competency questions were created. Competency questions by Fernandes[21] are the basis for a rigorous characterization of the knowledge that the ontology has to cover. They specify the problem and what constitutes a good solution. Some Competence Questions:

1- How is composed the PSF Health Team?
2- What are the main subprograms of Family Health Program?
3- How is the procedure and treatment for a Diabetic person?
4- What are the main routine activities for the Woman Health Program?
5- How many families a Health Team should cover?

3.2 Implementation

After the conceptualisation step, the implementation step starts, and is into this step the ontology is implemented. The OnToPHC was created using OWL-DL (Description Logic). The idea of OnToPHC was to define the roles, legal terminology, procedures and relations that influence the Primary health Care domain, based in Brazilian experience in PSF.

The OnToPHC ontology was developed following the bottom-up strategy proposed by Van der Vet (2002)[22]. This strategy results in well-structured ontologies. This methodology is defined over a set of terms specific to identify more general concepts. This type of approach allows us to have a broader vision of the initial terms of ontology. The ontology was created in Portuguese language.

The first step in conceptualization was build the glossary of terms: The terms were obtained from meetings with Health experts, from Governmental documents(laws, manuals, and ordinances) and from Family Health Portal http://dab.saude.gov.br/), no other ontology about this domain was reused because no one was found with relevant knowledge about the domain. The next step was integrate the concepts and define the binary relations. The OnToPHC has 131 Classes, 130 subclasses, 58 Equivalent Classes, 44 Disjoint axioms, 20 Object properties and 8 data properties.

Let’s see now part of the code of OntoPHC. First,in figure above we will see the header of the ontology.

```
<Ontology xmlns=”http://www.cin.ufpe.br/ecm2/OntoPHC.owl#”
xml:base=http://www.cin.ufpe.br/ecm2/OntoPHC.owl
ontologyIRI=”http://www.cin.ufpe.br/ecm2/OntoPHC.owl”
xmns:rdfs=http://www.w3.org/2000/01/rdf-schema#
xmlns:xsd=http://www.w3.org/2001/XMLSchema#
xmlns:owlrdfs=http://www.w3.org/1999/02/22-rdf-syntax-ns#”/>
</Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#” />
<Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#” />
<Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#” />
<Import href="http://www.ifmis.org/bfo/1.1/Import”>

Fig. 2 OntoPHC Header
```

Lines 1 to 3 of the figure above represent the namespace for the ontology OntoPHC, and lines 4 to 10 represent the identification of the namespace for the language RDF, RDF Schema, XML Schema and OWL. Namespace are alias to other resources, it usually links to the W3C(World wide Web Consortium) reference. Line 11 represents the BFO upper ontology import. The OntoPHC attributes are of type owl: OBJECTPROPERTY that defines the relationships between the concepts defined in ontology and owl: DatatypeProperty that allows restrictions to be established as cardinality and instances that an attribute can have.

```

Figure 3. Example of Object Property
```
The example above shows the object property is diagnosed by, and analysing the code above we realize the defined relationship (line 4), the relationship of each domain class (line 5) and the class range (line 6). The first three lines are comments, they are important for a better understanding but do not mean anything. The Family Health Program is composed by the union of many subprograms. Some of them are temporary, others no. In this research we represent formally knowledge about these subprograms.

Let’s see an example of a rule represented on OntoPHC.

See the following example: “IF Patient 1 has excessive thirst and fatigue and excessive urinate is recommended to do the Glycemia Exam.” The figure below describe how it can be represented in OWL.

Here we established a relationship between the classes Patient, Symptom and Glycemia Exam as follow. Patient-1 is an individual of class Patient who has some symptoms, (they were represented as Excessive Thirst, Fatigue, Excessive Urinate). Excessive Thirst, Fatigue, Excessive Urinate are individuals of the Symptom Class and have an object type property called Has symptom connecting to Patient 1. Is recommended is defined as an object type property and has value Glycemia exam. This patient may be suffering of Diabetes.

The OnToPHC was created using OWL-DL (Description Logic) and modelled with Protégé Framework (protege.stanford.edu) version 4.0.1 that allows to automatically generate the code OWL ontology.

Evaluation is an important phase of the development process. The evaluation activity judges the developed ontologies, software and documentation against a frame of reference. Ontologies should be evaluated before they are used or reused. There are two kinds of evaluation, the technical one, which is carried out by developers, and users evaluation.

To evaluate the OnToPHC we have used two mechanisms. Metric-based techniques and user evaluation.

The metric-base technique has used a new inference machine (reasoner), Protégé version 4.1 innovate and now it comes with a new reasoner called HermiT. Hermit reasoner defined OntoPHC as ALCHI(D) as figure 4 shows.

The user evaluation will be described below together with the system evaluation.

4 SemanticPHC : a semantic-based system

The need of improve the primary health care attendance and the possibility of evaluate by more users the created ontology, motivate the development of a semantic application, aiming allow through a web browser, access the modelled ontology and transform it into a concrete solution for Brazilian health professionals, providing an interesting way of retrieve information and test the knowledge formalized in OntoPHC ontology.
To test the applicability of OntoPHC we have use an innovative platform. We developed an application called SemanticPHC using the DJANGO framework, and python language for translate the OWL ontology into python classes. The application has features such as: add new families covered by the program, add citizens, add vaccine, schedule physician, register a visit, and extract reports, retrieve data by a semantic search, navigate through classes and subclasses, properties visualization.

We have chosen DJANGO because it is an open source web application framework, written in Python, which follows the model-template-view architectural pattern (similar to Model-View-Controller pattern). It provides a clear separation between what is “domain logic” (the application logic for the user) from the user interface (input and presentation), permitting independent development, testing and maintenance of each one. We included our concepts defined in our ontology OntoPHC, now will became classes of the business model.

The system is in test phase and the primary health center of Penedo is participating.

The objective of the system were two: 1- To be a source of knowledge for new users, were they can search for concepts, routines of primary health care, roles attributions, relations between roles as modelled by OntoPHC. 2- A system to support the daily routine based on our abstraction and modelled ontology.

The system has many features as described before, now we will detail the semantic retriever component. This component provides the functionality responsible for enhancing the user query with related terms in order to match relevant rules or related concepts. The objective is to contextualize the keyword into the primary health care domain. Also it should avoid ambiguity in search return.

The semantic retriever will use the RDFLib\textsuperscript{a} library and minimal python toolkit for viewing and querying RDFS/OWL ontologies called OntoInspector\textsuperscript{b}. RDFLib is the most prominent library for "Phytonic"(pure python language) programming with ontologies.

\textsuperscript{a} Website: http://www.rdflib.net/
\textsuperscript{b} Website: http://www.michelepasin.org/artifacts/software/ontoinspector/
Internally the webserver will receive the keyword and redirect it to the semantic retriever. The semantic retrieve will search for the object properties, rules and data properties related to the keyword. If the search returns not null, all related data will be displayed for the user.

The system has two levels of access. The administrator access can view, edit and manipulate the others users profiles. The administrator is responsible for feed an intranet blog created to disseminate important news related to family health program, calendar for future trainings and related news. This blog is showed in the index page of the system. Other activity from administrator is to receive the critics and the problems by email and provide solutions.

4.1 Software Evaluation

The plan of this evaluation follows the model proposed according to Wohlin et al. (2000)[29], the experimental process can be divided into the following main activities: the definition is the first step, where the experiment is defined in terms of problem, objective and goals. The planning comes next, where the design of the experiment is determined, the instrumentation is considered and the threats to the experiment are evaluated. The operation of the experiment follows from the design. In the operational phase, measurements are collected, analysed and evaluated in the analysis and interpretation. Finally, the results are presented and packaged in the presentation and package.

In order to define the experiment, the GQM paradigm defined by Basili et al., (1994)[30] was used. The GQM is based upon three pillars: GOALS, Questions and Metrics.

Now we will present our experiment definition using the GQM paradigm.

Goal:
According to the paradigm, the main objective of this study is:
To analyse the system developed for the purpose of evaluating it with respect to the relevancy of the tool from the point of view of primary health professionals in the context of Brazilian Primary Health Care.

Questions:
In addition, the questions to be answered are:
Q1. Does the ontology can be a source of knowledge and is useful to the primary health care professionals?
Q2. Does the tool is useful to primary health care professionals in their daily routine?
Q3. Do the subjects have difficulties to use the system?

Metrics:
M1. Usefulness of OntoPHC: P1 - The amount of knowledge that is acquired from ontology in three minutes; P2 - The number of procedures that was retrieved by subjects in one hour;
M2. Usefulness of SemanticPHC: The relation between the increment of use the system and not use the paper forms;

M3. Difficulty to use the system: In order to evaluate the difficulty of the subjects is necessary to identify and analyze the difficulties found by users, it could be misunderstanding or weakness of the system model and to define improvements.

process. Thus, two metrics related to difficulty will measure this issue:

* DM: %Subjects that had difficulties of system understanding;

* DU: %Subjects that had difficulties during the use of SemanticPHC;

These metrics have no well-known values for them. So arbitrary values were chosen, based on practical experience and previous work. Thus, values above 30% for DM and 20% for DU were considered as an indicative that the process, in this particular step, is too difficult and should be improved.

4.2 The Planning

Following Basili work [30] after the definition of the step, the planning is started. The definition determines the foundations for the experiment, the reason for it, while the planning prepares for how the experiment is conducted. We planned the case study as follows:

Context. The objective of this study is to evaluate the viability of using the SemanticPHC tool in primary health care. The analysis will be conducted in a primary health center located in Penedo.

Subjects. The subjects of the study will be the primary health staff. The staff is composed by physician, nurse and health agents.

Instrumentation. All subjects received a questionnaire (Q1) on his/her education and experience, in addition to questions about strong and weak points of the tool.

Criteria. The quality focus of the study demands criteria that evaluate the benefits obtained by the use of the Ontology and the tool and the difficulties of the users. The criteria will be evaluated quantitatively through the amount of effort to understand the system, and the time to retrieve knowledge from the ontology.

Null Hypothesis. This is the hypothesis that we want to reject with a high significance as possible. In this study, the null hypothesis determines that the use of SemanticPHC tool in primary health care does not produce benefits that justify its use and that the subjects have difficulties to use the tool.

Independent Variables. In a study, all variables in a process that are manipulated and controlled are called independent variables. The independent variables are the tool, the experience of the subjects, the technology and the staff size.

Qualitative Analysis. The qualitative analysis aims to evaluate the usefulness of the tool and the quality of the material used in the study. This analysis will be performed through questionnaire Q1.

Internal Validity. Considers whether the experimental design is able to support conclusions on causality or correlations[29]. The internal validity of the study is dependent of the number of subjects. This study is supposed to have at least five subjects to guarantee a good internal validity.

External Validity. The external validity of the study measures its capability to be affected by the generalization, i.e., the capability to repeat the same study in other research groups[29]. In this study, a possible problem with external validity is the subjects’ experience, since the experience of subjects can interfere in the productivity results.

Construct Validity. The validation of the construction of the study refers to the relation between the theory that is to be proved and the instruments and subjects of the study. This validity is concerned with the relationship between the treatment and the outcome, and determines the capability of the study to generate conclusions[29]. This conclusion will be drawn by the use of qualitative analysis due our main concern is the relevancy of the ontology and usefulness of the tool.

The Instrumentation
Selection of Subjects For the execution of the study, five primary health professionals of the Penedo’s primary health care were selected.

Instrumentation. Before the experiment can be executed, all experiment instruments must be ready. In this study, we used the questionnaire Q1.

The Operation
Experimental Environment. The case study was conducted during September-November 2011, at Penedo primary health care center. The case study was conducted directly by one nurse, and by four health agents.

Training. The training was performed according to plan: three high level meetings, divided in two hours among two weeks; 6 hours of subject specific training; and 16 hours of tool usage by the subjects.

**The Analysis and Interpretation**

Training Analysis. The training was divided in two steps. Step1- Training for the use of Protégé Tool, Step 2- Training for the use of SemanticPHC Toll. The subjects and all people involved considered the training essential to the usage of the tool. They related too about the three initial high level meetings were very important, to gain knowledge avoid misunderstandings.

Four subjects considered the use of the Protégé tool very difficult mainly due three things: 1- Its knowledge in English language. 2-The knowledge in computer usage. 3- The ontology paradigm, which is different from what the health professional are used to work. Two subjects who directly used the tool considered the training time were not enough to an appropriate tool understanding. Besides, both considered themselves not dedicated enough to learn.

Qualitative Analysis. The qualitative analysis is based on the answers defined for the Questionnaire.

Usefulness of ONTOPHC. The subjects reported that the ontology was useful and cover relevant concepts. Based on Questionnaire Q1. 60% of the subjects classified the ontology as useful and 20% classified as Very Useful. And 20% classified as Very few useful.

Usefulness of SemanticPHC. The subjects reported that the tool was useful to electronically register the process of primary health care. Beyond, the tool can improve answers to the reports solicited by Brazilian Health Ministry, they ask for a concept and the system can find not trivial rules retrieved by a semantic search on OntoPHC. Based on Questionnaire Q1. 80% of the subjects classified the tool as very useful and 20% classified as Useful.

Difficulty to use the system. In order, the subjects reported difficult of the infrastructure:
1. The Internet connection is very low and difficult the use if the system is not locally installed. Most of the medical primary health center from far away areas has no Internet connection or it has limitation
2. As described above some primary health center has no computers too.

Another difficult reported was the difficulty for using the Protégé Tool, and the English language. One difficulty reported is the afraid from some health professionals of using computer systems. Another difficulty reported was locomotion of the health agents.

They suggest the possibility of using the tool in mobile phones or smartphones or PDA to provide orientation to health agents into the field, not only in the medical center. And based on our previous defined metric for difficulty of the system, (DU) due the results of our, metric our goal was achieved and the usage of tool was considered not difficult.

5  Conclusion and future works

Nowadays there is a real need for knowledge management in healthcare due the explosion of information. Primary Health care is a rich domain in semantic relationships between entities and it often needs expressivity in modelling knowledge. Ontology-based approach is a new and interesting trend to support the sharing and reuse of formally represented knowledge among systems.

To develop an ontology is no simple task and it is necessary to use methodologies, methods, processes, tools and development languages in order to the development become more reliable and accurate. And an initiative from a country like Brazil with a huge program to cover the primary health problem, with no represented knowledge, and still with no work process, destitute of knowledge representation. We can conclude that OnToPHC is consistent and expressive due its formal development according to Description Logic metrics and users evaluations.

The Brazilian Family Health Program is achieving important results, raising health indicators, rising coverage. This project englobes variables like procedures, roles, professionals and actions that influences the PHC domain. Other goal was achieved, the objective of apply the formal knowledge in the system SemanticPHC for semantic clarification of Family Health's domain, which still is lacking of terminology standardization and involves a lot of concepts with subtle differences.

Future works of OnToPHC may be the reuse of the ontology for others ontologies or applications associated with semantic knowledge.

A future work is the inclusion of more detailed medical formal terminology, the reuse of biomedical ontologies like Snomed. Other interesting future work may be the creation of medical guidelines. These clinical guidelines can be
used by any health professional, they search for short description of a word and the system shows related terms, orientation and a graphical report with the relations between the words.

6 References


