A Framework to describe problems and algorithms in medical informatics via ontologies

Ch. Bratsas*, N Maglaveras** and P Quaresma***

* Aristotle University /Lab of Medical Informatics, PhD student, Thessaloniki, Greece.
**Aristotle University /Lab of Medical Informatics, Professor, Thessaloniki, Greece.
***University of Evora/Informatics Department, Professor, Evora, Portugal.

E-Mail Address
mpampis@med.auth.gr

Abstract: In this paper the need to create ontologies in order to describe, classify and correlate the problems of medical informatics and the algorithms that solve them will be presented and analyzed. The final objective is to build ontologies that will standardize the process of describing problems and algorithms from expert users –doctors and computer scientists respectively -. These ontologies along with a set of instances-descriptions will constitute a medical informatics problems and algorithms knowledge base. Firstly, the ontologies will be defined and the extent of their use in medicine will be mentioned. Secondly, the definition of problems and algorithms in medical informatics will be specified. Finally, a specific ontology and the RDF-Schema will be presented. It should be noted that the tool for the creation of this ontology and RDF-Schema is Protégé –2000

Introduction

Although there is a considerably large variety of medical informatics algorithms' referring to health care, the majority of practitioners use only a small subset routinely. The reasons that lead to this situation are numerous: first of all practitioners are not usually aware of the existence of all the algorithms that could help them; secondly, even when they are aware of their existence it is rather difficult to find them; thirdly, practitioners are not confident that these algorithms can be useful in solving their problem and finally, these algorithms may not be straightway available [1]. The particular feature that characterizes these algorithms is the fact that they are created and implemented by informatics scientists even though they are used in order to solve medical problems.

Medical algorithms can help to reduce errors by ensuring proper selection and application of an algorithm. Some errors may be introduced as well, although this can be minimized through proper design and use of the system. Other authors have noted the value of medical algorithms in health care. In particular, McGinn et al. note that validated clinical decision rules have "the potential to inform clinical judgment, to change clinical behavior, and to reduce unnecessary costs, while maintaining quality of care and patient satisfaction"[2].

In order to accomplish an efficient search of appropriate algorithms in medical informatics a framework able to describe problems and algorithms in medical informatics via ontologies will be suggested.

The remainder of this article is structured as follows: firstly, definitions of ontologies, Medical Informatics Problems and algorithms as also the state of the problem are given; secondly, a short description of MedProAlg ontology is presented; thirdly, a test case of the ontology is illustrated and finally the conclusions and possible future work are discussed.

Definitions and stating the problem

Although ontologies have become a popular research topic since the early 1990s, researchers can find more than one definitions of them in the literature. Generally, ontology can be defined as an explicit and specific description of a domain knowledge [3].

Ontologies provide a structure framework for modeling the concepts and the relationships of some domain of expertise. Ontologies support the creation of repositories of domain-specific reference knowledge, domain knowledge, for communication and sharing of this knowledge among people and application [4].

As fas as medical ontologies are concerned, the schemas that have already been categorised are mostly repositories of concepts and controlled dictionaries, which provide definitions and semantic relationships among the concepts [5.] The main functionality of these schemas is to facilitate the interoperability between systems that store, elaborate and query biologic and medical data, and to establish the basis for the communication between individuals by offering officially registered definitions of concepts.

A medical problem shall be called Medical Informatics Problem (MIP), if its solution is an algorithm that can be implemented via informatics.

In general, an algorithm is defined as "a step-by-step procedure for solving a problem or accomplishing some end especially by a computer"[6].

Generally, a problem in medical informatics as well as the algorithm that solves it calls for two domains of knowledge to be fully defined: the domain of medicine and the domain of computer science. It is usually rather difficult for an expert to describe both those two domains properly for the reason that he/she should have a wide knowledge of both sciences to achieve that.

To find and utilize a specific algorithm that will solve a particular problem, one must first know in detail what the problem is. In order to achieve a manageable and efficient search of appropriate algorithms, which will be able to solve medical informatics problems, the creation of a framework, where medical practitioners will describe the problems' structure and informatics practitioners will generate the algorithms that solve these problems, will be suggested. To provide a structure framework for modeling the problems and algorithms in medical informatics two domains of knowledge are required: the domain of medicine and the domain of computer science. Building ontologies, which will offer the possibility to describe medical informatics problems and the algorithms that solve these problems, would possibly provide with an answer to the need of combining, exchanging and reusing knowledge representation of both domains.

MedProAlg Ontology

Initially, we implemented a prototype ontology, which has been built in Protégé-2000 [10]. This ontology was named as MedProAlg Ontology (Medical Problem Algorithm Ontology) and the classes as well as the slots are illustrated in Figure 1. There are two important matters that should be noted before presenting MedProAlg ontology: the first one is the utilization of UMLS when categorizing problems (class UMLS_ROOT), and the second one is the fact that the class pragmatics, which describes certain elements of problems and algorithms, has been based on Dublin core elements metadata.

limited network of 134 semantic types, and the Specialist Lexicon which corresponds to lexical resources.[8,9].

A short description of MedProAlg is given below:

The primary classes of MedProAlg are ProblemDescription and AlgorithmDescription. These are the classes in which expert users can describe the medical informatics problem and the algorithm that solves the problem respectively.

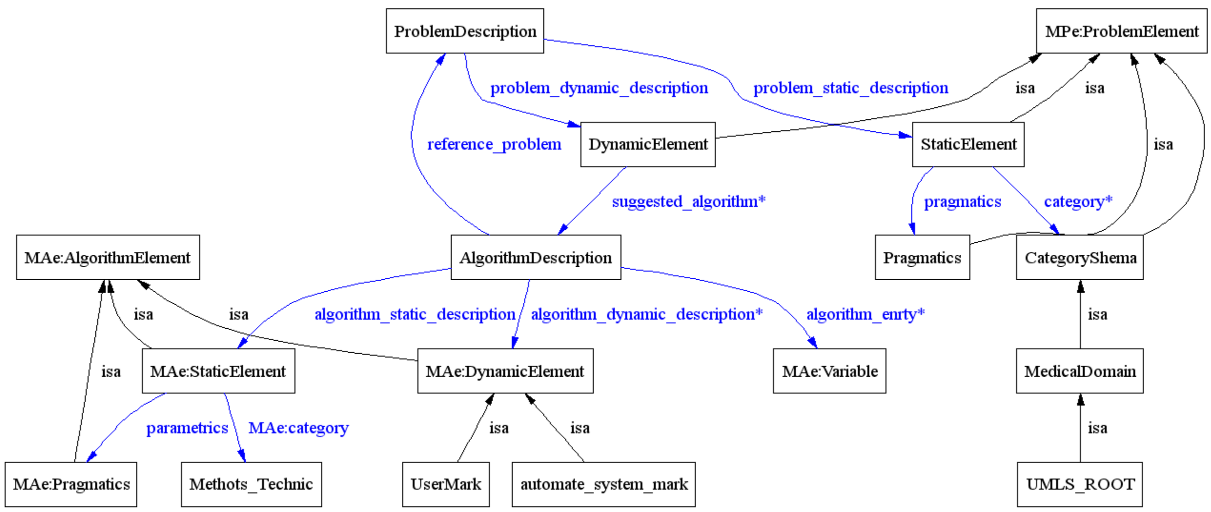
Every problem contains its static and its dynamic elements, which are classes named as StaticElement and DynamicElement correspondingly. The static elements of a problem are the ones that do not alter as time progresses, whereas the dynamic elements are changing in course of time.

The problem description (ProblemDescription class) consists of two slots, namely the problem_static_description, and the problem_dynamic_description.

The static description (problem_static_description slot) of each problem is formulated when expert users enter the problem for the first time. The value type of the problem_static_description slot is instance and allows the StaticElement class, which in turn contains the following subclasses: CategorySchema class and Pragmatics class.

The category of each problem (CategorySchema class) derives from the description of the medical

Figure 1:MedProAlg Ontology Schema



The Dublin Core (DC), conceived in 1995 in Dublin, Ohio, is a 15-element metadata set similar to a library catalogue card and intended to aid discovery of electronic resources [7]. The 15 elements (covering resource content, intellectual property and instantiation) are: title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage and rights.

The Unified Medical Language System (UMLS), which has been developed by the US National Library of Medicine since 1986 [8]. It is intended to help health professionals and researchers use biomedical information from different sources and is made by mapping many existing terminologies within a unifying framework. It comprises three knowledge bases: the Metathesaurus, a large repository of concepts, the Semantic Network, a

section (e.g. electrocardiography, anesthesiology etc.) or subsection (e.g. estimation of heart rate) based upon UMLS. Thus, the CategorySchema class includes the UMLS_ROOT subclass.

The Pragmatics class is based on Dublin Core Metadata Elements (for example the problem or algorithm title, the date of the problem's or algorithm's insertion, who enters the problem or algorithm and where from, the problem's or algorithm's description, the problem's or algorithm's subject and keywords, the date of the last modification, etc.). Whenever the Pragmatics class is a subclass of MPe:ProblemElement refers to problems, and whenever is a subclass of MAe:AlgorithmElement refers to algorithms.

ISOFormat is a class containing: ISO31166, ISO 639-2, ISO8601 subclasses. The instances of these subclasses include the codes for the representation of names of countries, names of languages, and dates and times encoded with the W3C encoding rules, respectively.

The dynamic description (problem_dynamic_description slot) of each problem refers to the suggested algorithms (suggested_algorithm) and to various statistics (statistic_element). The value type of the problem_dynamic_description slot is instance and allows the DynamicElement class, which in turn contains two slots, namely the suggested_algorithm and the statistic_element.

The value type of the suggested_algorithm slo is instance and allows the AlgorithmDescription class, whereas the value type of the statistic_element slot is string.

The algorithm description (AlgorithmDescription class) also includes dynamic and static elements (algorithm_static_description slot, algorithm_dynamic_description slot), as well as the entry variables of the algorithm (algorithm_entry slot) and the specific problem that solves (reference_problem slot).

The value type of the algorithm_static_description slot is instance and allows the MAe:StaticElement class. MAe:StaticElement class in turn includes two slots: MAe:category and pragmatics. The value types of MAe:category slot and the pragmatics slot are instances that allow the MAe: CategorySchema and the Pragmatics classes respectively.

The MAe:CategorySchema class contains the Methods_Technic subclass.

The value type of the algorithm_dynamic_description slot is instance and allows the MAe:DynamicElement class. MAe:DynamicElement class contains three subclasses, namely the UserMark, the When_Where_Be_Used and the Automate_System_Mark.

The value type of the algorithm_entry slot is instance and allows the MAe:Variable class. The MAe:Variable class includes the following slots: description, measurement, name, variable_type, variable_name. The value type of description, measurement and name slots is string, whereas the value type of the variable_type and the variable_name slots is instance and allow the BasicDataType and PrimitiveDataType classes respectively.

The value type of the reference_problem slot is instance and allows the ProblemDescription class.

Test Case – Example

The MIP in test case is: how we can calculate, from the ECG strip, since it is printed at a constant speed, the heart rate (figure 2). The algorithm that solves it is: For regular rhythm the heart rate can be calculated by:

Dividing the number of 0.4 second time lines (1 mm) in a P-P or R-R interval into 1500.

The problem and the algorithm that solves it used in this test case are rather simple so that the utilization of the MedProAlg ontology can be comprehended as simply as possible.

The steps followed are:

Firstly, the Medical Domain Expert describes the MIP's static characteristics. More specifically:

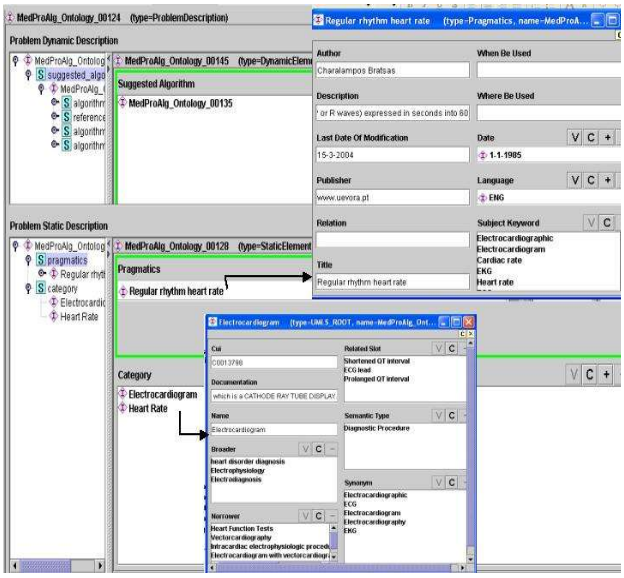
The MIP's Pragmatics are:

- ❖ Author: Charalampos Bratsas.
- ❖ Problem description: The heart rate can be calculated from the ECG strip since it is printed at a constant speed, for regular rhythm the heart rate can be calculated by: dividing the number of 0.4 second time lines (1 mm) in a P-P or R-R interval into 1500.
- ❖ Publisher: http://www.uevora.pt.
- ❖ Last date of modification: 15-3-2004. Date of the problem's input into the system: 1-3-2004. Language: English (ENG).
- ❖ Title: Calculate heart rate from ECG Keywords: ECG, EKG, Electrocardiographic, Electrocardiogram, Cardiac rate, Heart rate.

Secondly, the Medical Domain Expert can define the category of a problem with a word or words contained within the category. The system is linked to the Unified Medical Language System Knowledge Source Server (UMLS KSS) and searches for the word or words. The characteristics received are the following:

- ❖ Names of the categories: Heart Rate, Electrocardiogram.
- ❖ Concept Unique Identifier (CUI) respectively C0018810, C0013798.
- ❖ Definition of Heart Rate: The number of times the HEART VENTRICLES contract per unit of time, usually per minute. (MeSH) DISPLAY. Definition of Electrocardiogram as documentation: Measurement and interpretation of electrical manifestations of heart activity. (CRISP Thesaurus) Recording of the moment-to-moment electromotive forces of the HEART as projected onto various sites on the body's surface, delineated as a scalar function of time. The recording is monitored by a tracing on slow moving chart paper or by observing it on a cardioscope, which is a CATHODE RAY TUBE.
- ❖ Broader relationships of Heart Rate: Cardiac function, etc. Broader relationships of Electrocardiogram: heart disorder diagnosis, Electrophysiology, Electrodiagnosis, etc.
- ❖ Narrower relationships of Heart Rate: RHR heart reflex etc. Broader relationships of Electrocardiogram: Heart Function Tests, Intracardiac, procedure with ECG, etc.
- ❖ Semantic type of Heart Rate: Finding, Organism, etc. Semantic type of Electrocardiogram: Diagnostic Procedure, etc.
- ❖ Synonymous of Heart Rate: HR - Heart rate, Cardiac rate, etc and Synonymous of Electrocardiogram: EKG, ECG, etc.

Figure 2: The MIP test case description in protégé-II.



Thirdly, the domain expert of informatics connects the algorithm to the MIP and describes the algorithm's static characteristics as well as its input variables. More specifically:

- Reference Problem: Calculate heart rate from ECG.
- Input variables: dt. Name of variable:dt, Description: Length in seconds between 2 successive R-R or P-P Measurements: seconds, Variable type: numerical, Variable Value 0.4 (float).
- The algorithm's static characteristics are the pragmatics and the category. The algorithm's category is calculation, whereas the pragmatics are the following: Author: Friedman HH, Description: Heart rate = 1500/dt, Date: 16-3-2004, Last Date of Modification: 30-3-2004, Publisher McGraw-Hill Book Co, Language: English (ENG), Title: Estimation of Heart Rate, Keywords: heart rate calculated from the ECG algorithm for rythm of heart rate.

Fourthly, at the time when the MIP's algorithm enters the suggested algorithm is automatically completed into the dynamic description of the algorithm.

Finally, when Medical Researchers or Practitioners are using the algorithm they have the possibility to validate it by completing the dynamic description of the algorithm. In particular, the dynamic description consists of:

- The name of the algorithm's user: Paulo Quaresma. When from has the algorithm been used: 16-3-2004, and where: Evora. User status: Professor. The rating given to the algorithm by the user (Handily Mark: 10, Efficient Mark: 9).

Issues and conclusions

In this paper a framework able to describe problems and algorithms in medical informatics via ontologies was presented. This particular framework provides with the ability to combine, exchange and reuse knowledge representation of the medicine domain and of the computer science domain.

The MedProAlg Ontology along with a set of instances-descriptions, will constitute a medical informatics problems and algorithms knowledge base. A step further, the algorithms' description will provide the framework of the program that implements the algorithm. The informatics practitioners will then complete this framework, so that the algorithm will be ready to use by medical practitioners.

The finally aim, is to built an integrated intelligent system. The medical researchers will be able to search the appropriate algorithm, which will solve a particular MIP. This MIP or a similar one will be the Knowledge Base of MIPs and their respective algorithms. Medical researchers will also be able to download and validate the implemented algorithm.

References:

1. Svirbely J, Sriram MG. MEDAL, a compendium of medical algorithms for access over the Internet. Proc AMIA Symp. 1999;p1172.
2. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, and Richardson WS. Users' guides to the medical literature. XXII: Howto use articles about clinical decision rules. JAMA. 2000; 284: 79-84.
3. Gruarino N. Formal Ontology and Information Systems, in Guarino N(ed.). Formal Ontology and Information Systems. Amsterdam. IOS-Press.1998.
4. Davies J, Fensel D, Harmelen F, Towards the semantic web, Ontology-driver Knowledge Management , Wiley &Sons Ltd, 2003.
5. Magkanaraki, S. Alexaki, V. Christophides, and Dimitris Plexousakis. *Benchmarking RDF Schemas for the Semantic Web*. In Proceedings of the First International Semantic Web Conference (ISWC'02), Sardinia, Italy. June 9-12, 2002.
6. Merriam-Webster's Collegiate Dictionary, http://www.m-w.com/dictionary.htm.
7. Dublin Core Element Metadata, http:// www.dublincore.org
8. Humphreys BL, Lindberg DA, Schoolman H, Barnette G. The Unified Medical Language System: An Informatics Research Collaboration. J Am Med Inform Assoc. 1998; 5(1):1-11.
9. Achour Ls, Dojat M, Rieux C, Bierling P, and Lepage E. A UMLS-based Knowledge Acquisition Tool for Rule-based Clinical decision Support System Development, J Am Med Inform Assoc 2001; 68(1):73-85.
10. Protégé-2000 an ontology tool, http://protege.stanford.edu