Providing Semantic Interoperability of Clinical Information through an Ontology for the Electronic Patient Record

Newton MIYOSHI
Department of Physics and Mathematics
Faculty of Philosophy, Sciences and Letters of Ribeirão Preto
University of São Paulo
Ribeirão Preto, São Paulo, 14040-901, Brazil

and

Joaquim FELIPE
Department of Physics and Mathematics
Faculty of Philosophy, Sciences and Letters of Ribeirão Preto
University of São Paulo
Ribeirão Preto, São Paulo, 14040-901, Brazil

ABSTRACT

In recent years, there is an increasing demand for building interoperable information systems. Interoperability is crucial in the field of Health Care because the sharing of information may be essential to ensure a good treatment for the patient. The Clinical Evolution Record (CER) is an example of that, since it consists of a large document where we can find temporal information related to the whole history of patient's health conditions and medical procedures, exams, internments and treatments, among others, as part of the Electronic Patient Record (EPR). Aiming at providing solid and wide interoperability to CER information, in this article we propose an ontology based on a clinical data structure built in a previous work. To ensure the semantic interoperability we use the UMLS (Unified Medical Language System) Semantic Network as an upper-level ontology, so that the proposed ontology works as an extension of it.

Key-words: Ontology, Clinical Evolution, Interoperability, UMLS, Electronic Patient Record.

1. INTRODUCTION

One of the main topics of research in the field of Medical Informatics is the development of Electronic Patient Record (EPR) systems. The EPR aims to improve the organization and the quality of the information handled. The Clinical Evolution Record (CER) is a section of the Electronic Patient Record (EPR), where the whole history of patient's medical conditions and all procedures performed during the hospitalization are stored in a temporal way. The CER works as a communication channel among different health professionals. In most health institutions the CER is written by hand and stored in paper format. In a previous work, Felipe et al. [1] proposed a data structure for CER, through the identification of patterns of filling that allowed structured data entry instead of free entry. This identification resulted from a set of medical records analysis and interviews with physicians. However, the structured data entry itself isn't enough to ensure semantic interoperability, once different environments and contexts can demand different data formats.

An ontology can be seen as a mechanism to achieve the semantic interoperability. From the definition of concepts and the relationships in a domain of interest, the sharing of data between two different systems can occur. It is desirable and, in specific situations, even crucial to consider the reuse of existing ontologies during the creation process of a new one. When two ontologies are derived from the same upper-level ontology, it facilitates their integration, because they are related in a certain level of abstraction.

To allow the sharing of information stored in the CER, in this work we have developed an ontology based on the clinical data structure built in a previous work. This ontology covers the concepts of general clinical evolution as well as specific concepts of the following medical specialties: dermatology, oncopediatry and surgical gastroenterology. This ontology also extends the UMLS Semantic Network, which works as an upper-level ontology.

2. BASIC CONCEPTS

Ontology
The term ontology comes from the philosophical discipline that deals with the organization of reality, the study of the beings and their relationships [2]. In the field of Computer Science, ontology is defined by Gruber [3] as the "specification of a conceptualization". It can be seen as a mechanism to achieve semantic standardization and interoperability. Through an ontology it's possible to define, in a formal way, the concepts of a particular domain of interest, besides the relationships between these concepts.

By making the concepts and the relationships of a domain explicit, the information can then be considered machine-interpretable, enabling different information systems to communicate with each other. This is possible because it doesn't matter how each system represents the information, if both refer to the same domain through the same specification of concepts and relationships. With the development of an ontology it is also possible to reuse the knowledge of the domain of interest. In other words, ontologies can derive from each other, and then specialize in different areas and be more specific when necessary.

According to Bodenreider [4] we can categorize the ontologies based on the domain they represent and the level of detail they provide (Fig.1). Domain Ontologies are the ones which focus on specific areas of knowledge. General Ontologies represent an intermediary knowledge with a greater degree of independency between the specific areas. Upper-level ontologies map high-level abstract concepts that are common to all domain ontologies, allowing them to communicate with each other in a certain level of abstraction.

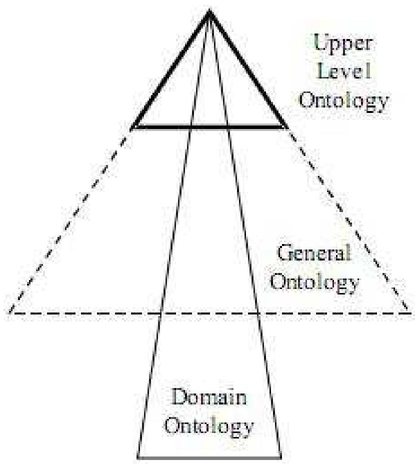

Figure 1: Types of ontologies [5]

Unified Medical Language System
The Unified Medical Language System (UMLS) is a project of the National Library of Medicine (NLM) which aims to support the development of information systems that act as they understand the language of biomedicine and health [5]. The UMLS consists basically of three knowledge sources: the Metathesaurus, the Semantic Network and the Specialist Lexicon.

The Metathesaurus is a huge vocabulary database where different biomedical concepts and relationships from many vocabulary sources (thesauri, classifications, code sets, etc) are stored, in a unique format. It is organized by meaning and each concept in the Metathesaurus is attached to a semantic type in the Semantic Network providing a consistent categorization for all the concepts.

The UMLS Semantic Network can be seen as a conceptual framework because it categorizes all concepts and relationships in the UMLS. It contains 135 semantic types and 54 semantic relationships which, together, form the complete structure of the Semantic Network. Each semantic type is represented by a node and the relationship is a link between two nodes. The "is-a" relationship form the hierarchical structure of the Semantic Network (Fig.2).

By categorizing the biomedical concepts and relationships, the UMLS Semantic Network can be considered an upper-level ontology for the biomedical domain [6]. Although the Semantic Network is already an abstraction of the biomedical concepts, sometimes an even smaller group of semantic types can be desirable. In this case we can use the Semantic Groups, which are 15 groups formed by different semantic types from the Semantic Network. These groups were made to reduce the semantic complexity of the Network, allowing an easier categorization of the concepts [7].

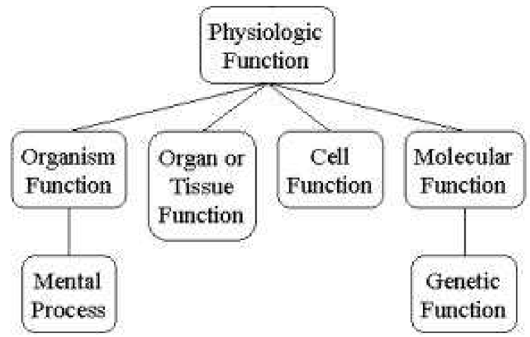

Figure 2: Part of the UMLS Semantic Network corresponding to the semantic type "Physiologic Function"

The Specialist Lexicon is a lexicon of general English words and also contains a large number of biomedical terms. It provides lexical information necessary for the Natural Language Processing (NLP) Specialist Systems.

3. STRUCTURE OF THE CLINICAL EVOLUTION RECORD

The School Hospital of the Faculty of Medicine of University of São Paulo at Ribeirão Preto, Brazil (SH-FMRP) is a large teaching health institution which performs more than 2500 appointments per day, generating a huge amount of information. Part of this information is stored in paper format and it includes the Clinical Evolution Record (CER).

At SH-FMRP the record of clinical evolution is done in a non-structured way and stored in Portuguese by means of two paper documents: the Clinical Evolution File (CEF) and the Clinical Observation File (COF). The COF is filled out the first time the patient is attended. In the CEF, information such as anamnesis, drugs used by the patient, laboratory and test results, physical exam, among others, is stored. The simple digitalization of the information inside the CEF does not make it computationally interpretable and it is difficult to use this information to support decision making. To solve this problem, Felipe et al. [1] proposed a methodology to structure the information related to the CER. It's important to notice that each medical specialty has its own particularities, so the resulting structure from this process is closely linked with each specialty.

The methodology consisted of the following stages: 1) to raise the characteristics of the CEF from each specialty, 2) identify and propose a unified structure, 3) validate the unified structure with medical specialists. In stage 1, 50 medical records were analyzed from three medical specialties: dermatology, surgical gastroenterology and oncopediatry. Physicians from each medical specialty were interviewed to identify patterns and to clarify the question raised during the analysis of the medical records. Also in stage 1, a pre-structure from CER for each specialty was defined.

In stage 2 it was possible to identify parts of structures that were commonly used in the three specialties and then to determine a unified structure called the common-core structure (Fig3).

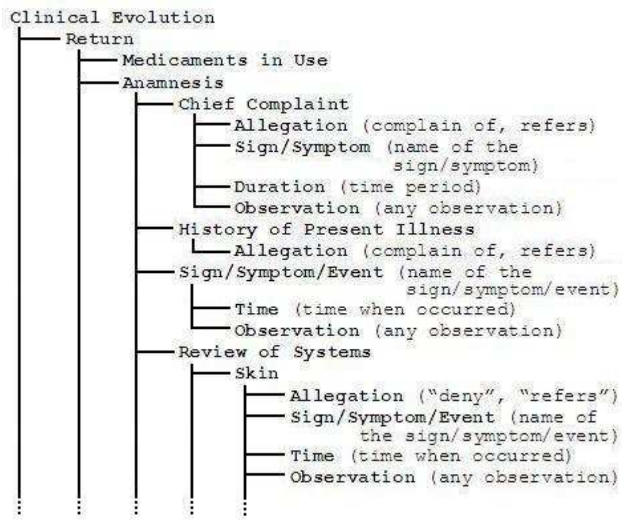

Figure 3: Part of the common-core structure

In stage 3 the validation of the common-core structure was made with the medical specialists together and the specific structure from each medical specialty was studied. This whole vision allowed to identify possible mistakes on the structure and served as a final validation of the common-core structure and the three specific structures.

4. METHODOLOGY

It is important to emphasize that there is no single correct way for building ontologies. Existing methodologies result from the large experience of their authors in the construction of ontologies and should be viewed as guidelines. There are several proposed methodologies for building ontologies, such as the Uschold methodology [8], Methontology [9] and the guide 101 [10]. We have chosen to rely on the guide 101 by Noy and McGuiness. The reason for this choice is that the guide 101 is a method rather didactic, practical and objective for the construction of ontologies. For the construction of our ontology we used the UMLS Semantic Network as an upper-level ontology. To achieve this we followed these steps:

Step 1 – Defining the Scope and Domain
In this step we define the scope and domain of our ontology. This is done by answering a series of basic questions: "What domain will the ontology cover?", "What are we going to use the ontology for?", "What types of question should the ontology provide answers for?", "Who will maintain and use the ontology?"

Step 2 – Enumerating Important Terms for the Ontology
It is useful to obtain a list of terms that may become the concepts of our ontology. These terms are, in most cases, used to describe the domain of interest. We reused the clinical data structure described in Section 3 for the CER, and from the structures we listed the key terms to form our ontology. We identified the important terms for each medical specialty as well the important terms that are common to medical areas.

Step 3 – Identifying the Concepts
With the listing of the important terms we carry out the identification of concepts that comprise our domain of interest and compound our ontology. The concepts are usually related to objects (physical or logical). They can be nouns or verbs in the sentences that describe the domain.

Step 4 – Extending the UMLS Semantic Network
To make easier the mapping of the concepts identified in step 3 in the UMLS Semantic Network we used the semantic groups. They reduce the complexity of the Semantic Network. We can subdivide this step as follows:

Phase 1: Mapping the concepts to one of the 15 semantic groups;

Phase 2: Mapping each concept of our ontology to one or more semantic types among the semantic types that belongs to the same semantic group that the mapped concept belongs to.

Step 5 – Validating the Ontology
To verify the mapping, we performed the search of the concepts in the Metathesaurus. For each concept found in the Metathesaurus, we identified its semantic group and the semantic type and thus we could identify possible errors in mapping. Finally we talked with medical specialists to review and clarify some doubts.

Ontology Editor
For the construction of our ontology we used the Protégé Ontology Editor because it is a free software, open source, used by several projects to build ontologies in biomedical field and also has a lot of plug-ins developed by various groups, and, in addition, it is well accepted in the knowledge engineering community [11]. As ontology specification language we use the Web Ontology Language (OWL) because it is endorsed and maintained by the W3C consortium and largely used by the scientific community [12].

5. RESULTS

In step 1 the domain of our ontology was identified as the record of the patient's clinical evolution stored in the clinical evolution file, and this record has its own peculiarities for each medical specialty, but it is possible to identify a common structure to all of them. We decided specify the ontology to cover 3 medical specialties: oncopediatry, surgical gastroenterology and dermatology. The ontology can be used to build applications that can acquire and store the information that compound the CER. The ontology will also serve to ensure the semantic interoperability of the systems that uses that information. It is possible to discover what the result was of patient history, what tests were made in the patient, the diagnosis and other information acquired during clinical evolution.

During the execution of step 2, we divided the terms into four groups: common core, which are the terms that compound the common-core structure, the specific terms of dermatology, oncopediatry and surgical gastroenterology. We identified 90 terms in the common-core, 18 in oncopediatry, 20 in surgical gastroenterology and 88 in dermatology (Table 1).

Table 1: Examples of terms per analyzed area

Area	Term
Common-Core	Diagnosis
Common-Core	Motive
Oncopediatry	Painless to palpation
Dermatology	Infiltration
Surgical Gastroenterology	Purulent

In Step 3, keeping the previous division into groups, we identified a total of 66 concepts among the 90 terms of the common core, 17 in surgical gastroenterology, 15 in the oncopediatry and 79 in dermatology.

We then mapped the identified concepts into the Semantic Groups of the UMLS Semantic Network (Fig.4), during the execution of step 4.

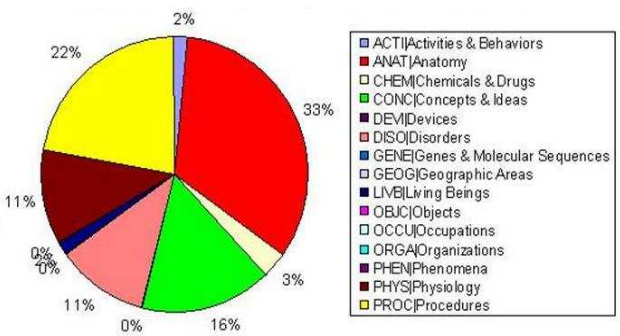

Figure 4: Distribution of common-core concepts into the semantic groups of the UMLS Semantic Network

For each mapped concept we assigned a semantic type of the Semantic Network using the Protégé tool (Table 2). We used the implementation of the Semantic Network provided by the TKGB group (Temporal Knowledge Bases Group) [13].

Table 2: Examples of mapped concepts and their respective medical specialty

Concept	Area	Semantic Group	Semantic Type
Active Ingredient	Common-core	Chemicals & Drugs	Pharmacologic Substance
Height	Common-core	Physiology	Organism Attribute
Nuchal Rigidity	Oncopediatry	Disorder	Sign or Symptom
Lividity	Dermatology	Disorder	Sign or Symptom
Blood Pressure	Surgical Gastroenterology	Physiologic Function	Organism Function

In step 5 we searched for the existence of mapped concepts in the Metathesaurus to check the mapping. Most of the identified concepts were found in the Metathesaurus, both for the common core and to the medical specialties. In Fig.5 we show the distribution of the concepts found in the Metathesaurus into the semantic groups.

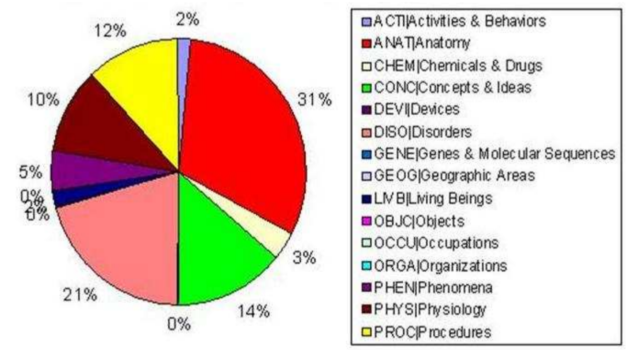

Figure 5: Distribution of the concepts found in the Metathesaurus into the corresponding semantic groups

The search of the concepts in the Matathesaurus helped to identify mapping errors and the interviews with the medical specialists clarified the remaining doubts. A summary of the results is shown in Table 3. Part of the resultant ontology can be seen in Fig.6. We can see the modeled concepts identified by the prefix 'OR', such as 'OR_Height' and 'OR_Body_Height' that are important in the clinical evolution domain and was not covered in the UMLS Semantic Network.

Table 3: Quantity of terms, concepts and concepts found in the Metathesaurus per medical specialty

Area	No. Terms	No. Concepts	No. Found in Metathesaurus
Common-Core	90	66 (73,3%)	58 (87,9%)
Dermatology	88	79 (89,8%)	39 (49,4%)
Oncopediatry	18	15 (83,3%)	10 (66,6%)
Surgical Gastroenterology	20	17 (85%)	14 (82,4%)
Total	216	177 (81,9%)	121 (56%)

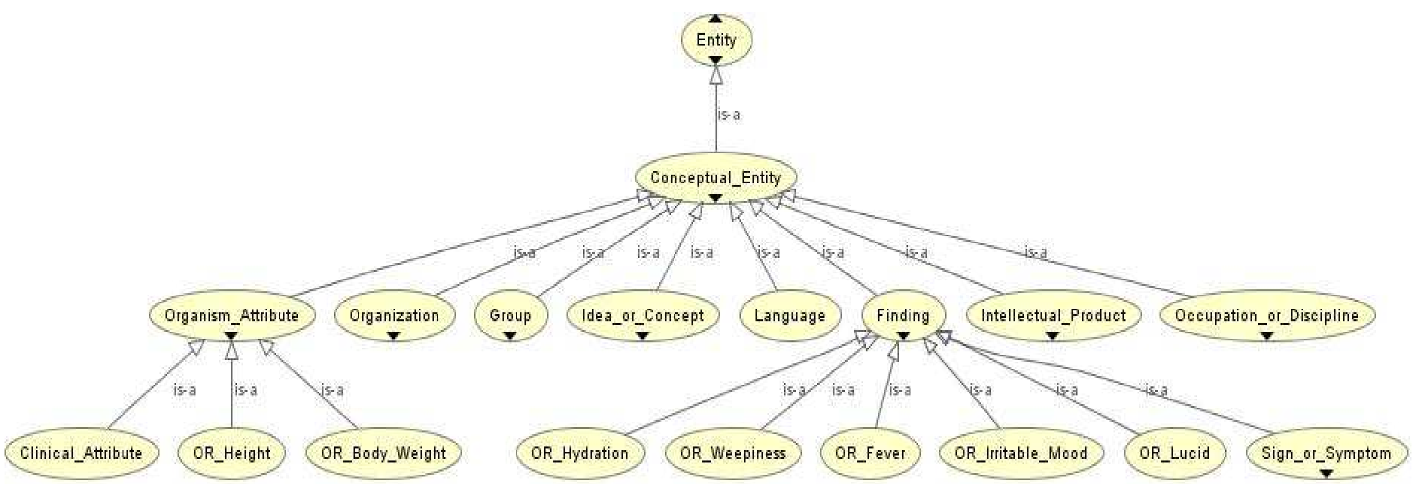

Figure 6: Part of the built ontology, the proposed concepts are identified by the prefix 'OR'

6. CONCLUSION

A major difficulty faced in the process of building an ontology is the complexity of the domain of interest. In the area of biomedicine this complexity is even greater due to the nature of the handled information. We found that the structure of information was a very important point for the development of the ontology since it reduces the difficulty inherent in the format of free text information. The use of semantic groups as the entry point to the extension of the UMLS Semantic Network was also a particularly important issue, because it facilitates the mapping of concepts. The use of the Metathesaurus as a way for verifying the mapping helped in the identification of several errors and finally the interview with professionals helped to clarify the remaining questions. Therefore we develop an ontology for the registration of clinical course of patients using the Semantic Network ontology as high level, thus providing an extension of it to the field of clinical evolution.

As a next step we will use two different systems that have distinct representation of the patient's clinical evolution data to corroborate the use of the ontology as an efficient way to share information between these systems.

7. ACKNOWLEDGMENT

We thank the Foundation for Research Support of São Paulo State (FAPESP) for supporting this work by grant number 08/53940-2.

8. REFERENCES

[1] J. C. Felipe, et al. **Estruturação da Informação para Registro Eletrônico da Evolução Clínica do Paciente**. In: Álvaro Rocha. (Org). Informática de Saúde – Boas Práticas e Novas Perspectivas. 1a ed. Porto – Portugal. Editora da Universidade de Fernando Pessoa. 2007.

[2] N. Guarino and P. Giaretta. **Ontologies and Knowledge Bases.** Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing, pp. 25-32. 1995.

[3] T. R. Gruber. **A translation approach to portable ontology specification.** In: Knowledge Acquisition, vol. 5, pp. 199-200. 1993.

[4] O. Bodenreider and A. Burgun. **Biomedical Ontologies**. In: Medical Informatics: Advances in Knowledge Management and Data Mining in Biomedicine. Springer-Verlag. 2005.

[5] NLM. **Unified Medical Language System**. <Available in http://www.nlm.nih.gov/pubs/factsheets/umls.html>. <Accessed in 18th February of 2009>.

[6] A. T. McCray. **An Upper-Level Ontology for the Biomedical Domain.** Comparative and Functional Genomics. pp. 80-84. 2003.

[7] A. T. McCray, A. Burgun and O. Bodenreider. **Aggregating UMLS Semantic Types for Reducing Conceptual Complexity.** IMIA. Medinfo. Amsterdam. 2001.

[8] M. Uschold and M. King. **Towards a Methodology for Building Ontologies.** Workshop on Basic Ontological Issues in Knowledge Sharing. 1995.

[9] A. Gómez-Peréz. **A Framework to Verify Knowledge Sharing Technology**. Expert Systems with Application**.** Vol. 11, No. 4. pp. 519-529. 1996.

[10] N. F. Noy and D. L. McGuiness. **Ontology Development 101: A Guide to Creating Your First Ontology.** Stanford Knowledge System Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Report SMI-2001-0880, March. 2001.

[11] SCBIR. **Protégé**. Stanford Center for Biomedical Informatics Research. <Available in http://www.protege.stanford.edu> <Accessed in 19th February of 2009>.

[12] W3C. **Web Ontology Language (OWL)**. World Wide Web Consortium. <Available in http://www.w3.org.br/2004/OWL/>. <Accessed in 19th February of 2009>.

[13] E. J. Ruiz. **The UMLS Semantic Network in OWL.** Temporal Knowledge Bases Group. Universitat Jaume I. <Available in http://http://krono.act.uji.es/people/Ernesto/UMLS_SN_OWL>. <Accessed in 19th February of 2009>