Applied Medical Informatics
Vol. 38, No. 2/2016, pp: 66-72

An Architectural Approach for Building Medical Ontologies

Mircea Adrian FOCŞA

Medical Informatics Department, Victor Babeş University of Medicine and Pharmacy Timisoara, 2nd E. Murgu Sqr., 300041 Timişoara, Romania
E-mails: mfocsa@umft.ro

Received: June 27, 2016 / Accepted: July 7, 2016 / Published online: July 29, 2016

Abstract
Starting from a use case scenario related to the Electronic Medical Record (EMR) information retrieval, this paper proposes several design principles for an ontology-based EMR system, which will allow the physician to perform a semantic search about a specific body structure or zone/space (affected, explored or treated). The research starts from analyzing the information requirements of the family doctors related to the medical history of their patients. By interviewing a group of family doctors, a set of common questions have been identified and modeled as one generic question. The key terms from this generic question then became the building blocks of an ontology, which was further built incrementally and constantly adjusted depending on the query response (design pattern by querying). Elements from other terminologies/ontologies have been reused thus facilitating interoperability with similar constructions. As a result of this research, we tested and proposed for adoption several basic modeling principles for medical ontologies, among them, the particularities of part-whole relations for body regions/spaces and anatomical components, the requirements for propagation of the medical activities from one class to an upper or subsequent class. These modeling principles have a good degree of generality and could be taken into consideration for building any medical ontology.

Keywords: Medical ontology; Semantic search; Electronic Medical Record (EMR); Architecture

Introduction

Healthcare systems today are confronted with many challenges, among them the management of an increased amount of data which need to be acquired, stored, processed and presented at the right time and place in an accessible form. Retrieving the information needed for medical decision is part of those challenges, the right clinical decision involving complete and accurate information for a specific clinical context [1]. Searching through a large amount of data remains a difficult task, even when working with advanced information systems [2].

In modern EMR systems the records are well organized so the doctor can easily find a specific event (visit, diagnostic decision, elaboration/modification of the treatment plan), granted that they know the time of occurrence or/and the exact name of it. In the case of patients with multiple clinical conditions and chronic diseases, for some simple questions like ”have you ever had an ophthalmology consult” or “do you have any kidney related problems ”, the answer could be difficult to obtain, both during the anamnesis and by searching the patient’s records. The Ontology-based information retrieval proved to be a solution, allowing the user to navigate through terms and relations or to elaborate general or nonspecific questions, taking advantage of the semantics behind the concepts [3].

This paper analyses a use case scenario and proposes the design principles for an ontology-based EMR system, which will allow the physician to extract information about a specific body structure or zone/space (affected, explored or treated) through a semantic search.

The Protégé framework was used for building the ontology and for checking its satisfiability, entailment and consistency by running the Fact ++ reasoner. This process was based on recent developments of World Wide Web Consortium (W3C) specifications, Web Ontology Language (OWL), OBO Foundry principles [4,5] and the Keet/Artale part-whole ontology [6]. OWL is the latest standard in ontology languages developed by the W3C, based on Resource Description Framework (RDF) and DAML+OIL language, providing a set of constructs which are in the same time human and machine readable and understandable. It has formal mathematical foundations in Description Logics (DL), which allows us to use a reasoner for the purpose of checking the ontology as we build it [7].

Material and Method

This research starts by analyzing the information requirements of the family doctors related to the medical history of their patients (what happened and when). Two groups of family doctors, 10 from Belgium (Walloon Region) and 10 from Romania (Timis County) were interviewed during the year 2014. Based on their most frequent answers, we produced a set of common questions that a search engine should be able to respond in a meaningful way. We can summarize them in the following syntax:

"Did the patient have a record of any

clinical condition/exploration/treatment

related to a [particular]

body region/system/structure (organ) or involving a [particular] *method/device/clinical specialty*?"

Searching by the [particular] word(s) should return the clinical condition or treatment, even though (quite frequently) the returned concept(s) do not contain the searched word(s), but are semantically related.

For example, if the physician wants to know if the patient has a health problem or an intervention (activity) related to a specific specialty – **Gastroenterology**.

Table 1. The set of DL queries running for searching a specialty

DL Query	Result's Heading
”Activity and (isExaminating some (Structure and (isPart some (System and hasSpecialty some **Gastroenterology**))))” ”Activity and (isExaminating some (Space and (LocationFor some (Structure and (isPart some (System and hasSpecialty some **Gastroenterology**))))))”	Clinical Examinations
”Activity and (isExploring some (Space and (LocationFor some (Structure and (isPart some (System and hasSpecialty some **Gastroenterology**))))))”	Paraclinical Examinations
”Activity and (isActingOn some (Structure and (isPart some (System and (hasSpecialty some **Gastroenterology**)))))” ”Activity and (isActingOn some (Space and (LocationFor some (Structure and (isPart some (System and (hasSpecialty some **Gastroenterology**)))))))”	Therapeutic Procedures
”Clinical_Condition and (isAffecting some (Structure and (isPart some (System and (hasSpecialty some **Gastroenterology**)))))” ”Clinical_Condition and (isAffecting some (Space and (LocationFor some (Structure and (isPart some (System and (hasSpecialty some **Gastroenterology**)))))))”	Clinical conditions
”Finding and hasProjection some (Space and (LocationFor some (Structure and (isPart some (System and (hasSpecialty some **Gastroenterology**))))))”	Findings

As a result of a semantic search by "Gastroenterology" (Table 1), a headed list of Examinations, Therapies and Findings should be displayed on EMR user interface (Fig. 1).

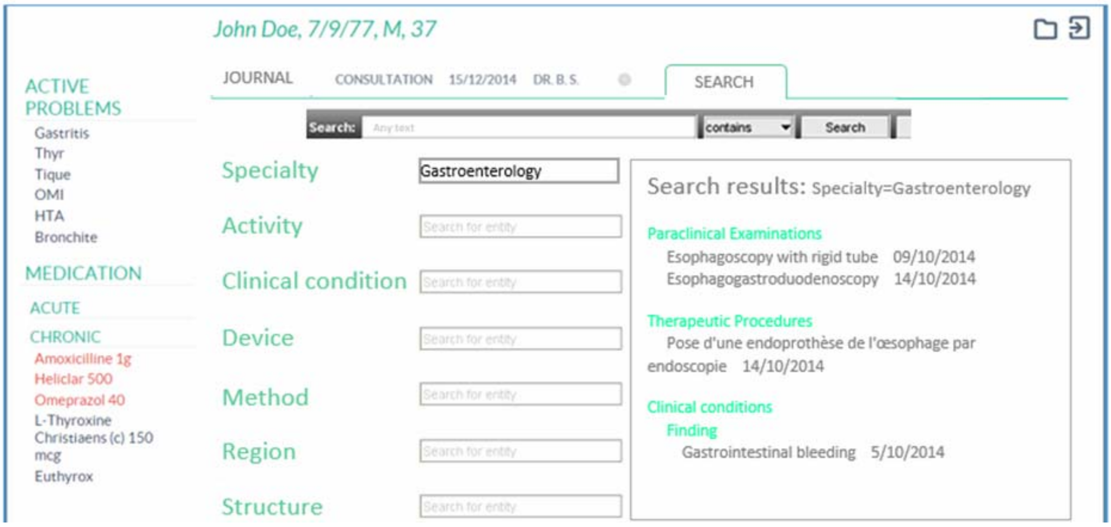

Figure 1. Example of a semantic search

The construction of the ontology started from a number of building blocks (marked in italics on the generic question described above: clinical specialty, body region, system, structure, clinical condition, exploration, treatment, method, device), by declaring them as primitive concepts, and further defining all the clinical significant subsequent concepts, the relations between them and their attributes [8], [9]. The appropriate proprieties (transitivity, reflexivity, symmetry and opposite), domain, range and disjunction were also defined for most of the relations.

The DL queries' syntax was used to refine:

- atomic concepts
- basic categories and the position of the atomic concepts within the hierarchy
- relations for linking the basic categories of the atomic concepts

During the incremental development of the ontology, by repeatedly running the DL queries we were able to check if the current design results in the expected answers.

By reusing parts of other open access OWL ontologies [10], we performed the import of elements from:

- Foundational Model of Anatomy (FMA) [11] - Anatomy taxonomy, for structures
- Classification Commune des Actes Médicaux (CECAM) [12] for Exploratory and Therapeutic interventions
- Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) core [13] for Clinical conditions (Findings - groped by symptoms and signs, Diseases and Syndromes)
- Unified Medical Language System (UMLS) [14] – for harmonizing between different concept representations

Following the organization of concepts in CCAM, the evaluation, treatment interventions and clinical conditions were grouped on systems – organs – (method), defining for each grouping category the equivalent class (necessary and sufficient condition) – Figure 2. For the very granular concepts (leafs) only the necessary conditions were defined, except the case when they referenced two or more zones/structures and the distinctions between them was considered important for the DL query result.

Each class was annotated with the English and French name and, where possible, with the synonyms, the preferred term, the abbreviation or some comments/descriptions. For interoperability reasons, the UMLS_CUI and SNOMED_CID were added for most of the concepts. For the most common clinical conditions, online references from the Merck manual – professional version [15] and/or Medscape [16] were added as attributes.

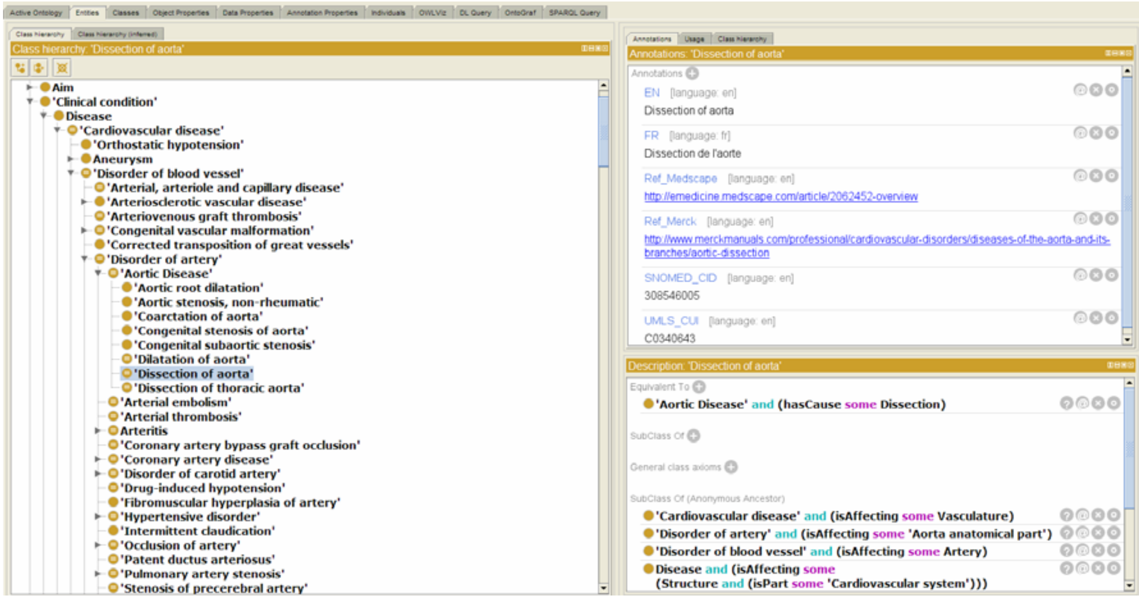

Figure 2. The hierarchy of main classes, the definition and attributes of the concepts

Results and Discussion

Despite the focus on the very basic set of concepts needed by the physician to express most frequent diagnostic and therapeutic procedures and clinical conditions, the resulted ontology accommodates almost 20000 classes, 30000 logical axioms (3500 equivalent class axioms), and 80 object properties.

By analyzing the DL query results, we constantly adapted the semantic model, concluding in some important modeling decisions, described in the following sections.

1. Modeling the anatomical parts

In respect to the **taxonomy of part-whole relations**, the distinction should be made on different types of mereological parthood relations and meronymy (cognitive and linguistic perspectives) [17]. The main semantic difference is related to transitivity of the part-whole relation. While in most cases the part-whole relation is transitive, in case of collections/sets and tree-like structures it should be declared intransitive. The following examples illustrate both cases:

A) **Transitive part-whole properties**

An anatomical structure (ex.: an organ) is compound from several components (genetic or arbitrary defined) of two types:

- Constitutional parts (structural/histologic) – linked to the upper class by isS-Part/hasS-Part property (S = Structural) (1)
- Spatial (morphologic) parts = spatial subdivisions of that anatomical structures – linked to the upper class by isA-Part/hasA-Part property. (A = Anatomical) (2)

Stomach hasA-Part only (Body_of_stomach or Cardia_of_stomach or Cavity_of_stomach or Stomach or Fundus_of_stomach or Pyloric_antrum or Pyloric_canal or Pylorus)
(1)

Stomach hasS-Part only (Mucosa_of_stomach or Muscle_layer_of_stomach or Serosa_of_stomach or Submucosa_of_stomach) (2)

All the organs belonging to a system should be grouped under a generic class (ex. Digestive Organ). Each organ will be linked with its system through isP-Part (is participating) property (3).

Stomach isP-Part some 'Digestive system' (3)

Anatomical spaces are partitioned with isR-Part (regional part) (4). hasR-Part/isR-Part could be applied only on spaces with the same dimensions (bi or three-dimensional space). The two types of spaces are linked by hasBoundary property. In this relation, the range and domain differ in dimensionality by one degree. The relation between the 3D region and the organ (or organ parts) is based on isLocated/LocationFor property (also transitive) (5).

Gastrointestinale_space is R-Part some 'Abdominal region' (4)

Gastrointestinale_space LocationFor some Stomach (5)

All the four partonomic relations, subtypes of isPart/hasPart are transitive:

- isP-Part (participating /functional criteria)
- isA-Part (anatomical/morphological criteria)
- isS-Part (structural/histological criteria)
- isR-Part (spatial delimitation)

B) **Non transitive part-whole properties**

An injury of a finger is not an injury of all the fingers which would be obtained if a transitive part-whole relation will be used between finger and the set of fingers (6).

Finger isMemberOf some Set_of_Fingers, isMemberOf
(6)

A branch of a vessel or nerve is not a part of the vessel/nerve which would be obtained by a transitive part-whole relation between a branch and the vessel/nerve it originates from (7).

Left_gastric_vein isBranch some Esophageal_vein isBranch (7)

isMemberOf and isBranch are intransitive types of isPart relation and will be defined accordingly as distinct relations, and not as subclass of isPart relation.

2. Modeling the exploration activities in relation with structures or spaces

A) **The exploration of a structure** (structural or regional organ part) should involve the exploration of the parent structure (organ to which it belongs) [18]. Ex. gastric mucosa exploration involves the stomach exploration. For this inference to work (to be found by the Reasoner), the following conditions must be met:
- chaining isExploring and isPart (8) - SuperProperty of (Chain):

isExploring **o** isPart **SubPropertyOf** isExploring
(8)

- applying the closer axiom for parts of an organ (under the hasPart relation)
- declaring isPart relations for all the involved parts, as we could not reason over the inverse relationship hasPart (Reasoner limitation)

The *SuperProperty of (Chain)* considerably slow down the reasoner, on big ontologies resulting in the freezing the application. That is why we found to be a better solution modeling the DL query instead of using the chaining mechanism. The following DL query (9) produce the same result as chaining:

Activity and isExploring some (Structure and isPart some Stomach) (9)

B) **The exploration of a body region** implicitly explores all the organs contained in that region.

Similar to the previous approach (having in this case isLocated/LocationFor instead of isPart/hasPart), we can use:

- the relations concatenation (10)- SuperProperty of (Chain)

isExploring **o** isLocated **SubPropertyOf** isExploring (10)

- an adapted DL query syntax (11). In this case, for each region we should declare all the organs which are located there and for each organ, the region where it is located.

Activity and isExploring some (Structure and isLocated some 'Abdominal region') (11)

Conclusions

We demonstrated one of the basic principles of ontology building, which is that any ontology should serve to a specific purpose. The study focused on a very basic need of the user – the information retrieval. During incremental building of the ontology, running DL queries and observing the expected and the actual result, some basic modeling principles had been identified: the two types of part-whole relations, differing by the transitivity property, and the requirements for propagation of activities (from a subclass to its parent class – in case of structure's exploration - or from the parent class to subclasses in case of region's exploration).

These modeling principles have a good degree of generality and could be taken into consideration for building any medical ontology.

While there was a lot to learn from this, there are things yet to be clarified, like the limitations of the isA relation in a medical taxonomy, the device/method distinction (as many methods includes on their name the device), or the particularities of Topography (a kind of part-whole relation). The Medication, Clinical Evaluation, Laboratory tests and Clinical processes should also be included in order to obtain a complex and comprehensive ontology serving as a basis for Semantic integration of EMR, Clinical workflow Management and Clinical Decision Support.

List of abbreviations

CECAM – Classification Commune des Actes Médicaux
DAML – DARPA (Defense Advanced Research Projects Agency) Agent Markup Language
DL – Description Logics
EMR – Electronic Medical Record
FMA – Foundational Model of Anatomy
OIL – Ontology Interchange Language
OWL – Web Ontology Language
RDF – Resource Description Framework
SNOMED_CID – Concept Identifier of the SNOMED-CT concept
SNOMED-CT – Systematized Nomenclature of Medicine – Clinical Terms
UMLS – Unified Medical Language System
UMLS_CUI – Concept Unique Identifier of the UMLS
W3C – World Wide Web Consortium

Acknowledgements

The author is indebted to his colleagues from the Belgian Society for General Practitioners/Family Physicians, especially to Dr. Silviu Braga, for open minded attitude,

cooperation and valuable support.

References

1. Edinger T, Cohen AM, Bedrick S, Ambert K, Hersh W. Barriers to Retrieving Patient Information from Electronic Health Record Data: Failure Analysis from the TREC Medical Records Track. AMIA Annu Symp Proc. 2012:180-188.
2. Focsa M. Knowledge-based EHR Systems. Proceedings of the 31st Romanian National Conference on Medical Informatics "Solution-based Medical Informatics". 2010:64-68
3. Schulz S, Hanser S, Hahn U, Rogers J. The Semantics of Procedures and Diseases in SNOMED CT. Methods Inf Med. 2006; 45(4):354-8.
4. OBO Foundry [Internet]. [cited 2016 June 27] Available from: http://www.obofoundry.org/principles
5. Stevens R, Nogradi A. OBOPedia: An Encyclopaedia of Biology Using OBO Ontologies. Semantic Web Applications and Tools for Life Sciences (SWAT4LS), Cambridge, 2015.
6. Keet C.M, Artale A. Representing and reasoning over a taxonomy of part–whole relations. Applied Ontology 2008;3.1-2:91-110.
7. Obitko M. Introduction to Ontologies and Semantic Web [Internet]. [cited 2016 June 27] Available from: http://www.obitko.com/tutorials/ontologies-semantic-web
8. Quinn MM. Mannion J. Improving Patient Safety Using Interactive, Evidence-Based Decision Support Tools. Jt Comm J Qual Patient Saf. 2005; 31(12):678-83.
9. Blobel B. Architectural approach to eHealth for enabling paradigm changes in health. Methods Inf Med. 2010;49(2):123-34.
10. Ceusters W, Smith B, De Moor G. Ontology-Based Integration of Medical Coding Systems and Electronic Patient Records; IFOMIS Reports [Internet]. 2004 [cited 2016 June 27] Available from: http://ontology.buffalo.edu/medo/CodingAndEHCR.pdf
11. Foundational Model of Anatomy [Internet]. [cited 2016 June 27] Available from: http://si.washington.edu/projects/fma
12. CCAM online [Internet]. [cited 2016 June 27] Available from: http://www.ameli.fr/accueil-de-la-ccam
13. IHTSDO: SNOMED CT [Internet]. [cited 2016 June 27] Available from: http://www.ihtsdo.org/snomed-ct
14. Unified Medical Language System (UMLS) [Internet]. [cited 2016 June 27] Available from: https://www.nlm.nih.gov/research/umls
15. Merck manual – professional version [Internet]. [cited 2016 June 27] Available from: http://www.merckmanuals.com/professional
16. Medscape [Internet]. [cited 2016 June 27] Available from: http://emedicine.medscape.com
17. Seyed P, Rector A, Sattler U, Parsia B, Stevens R. Representation of Part-Whole Relationships in SNOMED CT. Cornet R, Stevens R, editors. Proceedings of the 3rd International Conference on Biomedical Ontology. In CEUR Workshop Proceedings - KR-MED: ICBO'12. Vol. 897. CEUR-WS.org. 2012.
18. Vigo M, Jay C, Stevens R. Constructing conceptual knowledge artefacts: Activity patterns in the ontology authoring process. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems 2015:3385-94.