Conceptual Graphs based information retrieval in HealthAgents

Madalina Croitoru[1], Bo Hu[1], Srinandan Dasmahapatra [1], Paul Lewis[1], David Dupplaw[1], Alex Gibb[2], Margarida Julia-Sape[3], Javier Vicente[4], Carlos Saez[4], Juan Miguel Garcia-Gomez[4], Roman Roset[5], Francesc Estanyol[5], Xavier Rafael[5], Mariola Mier.[5]

[1]IAM, University of Southampton, UK;
[2]University of Birghminham, UK;
[3] Universitat Autonoma de Barcelona, Spain;
[4]Instituto de Aplicaciones de las TIC Avanzadas, Spain;
[5] MICROART, Spain.

Abstract

This paper focuses on the problem of representing, in a meaningful way, the knowledge involved in the HealthAgents project. Our work is motivated by the complexity of representing Electronic Healthcare Records in a consistent manner. We present HADOM (HealthAgents Domain Ontology) which conceptualises the required HealthAgents information and propose describing the sources knowledge by the means of Conceptual Graphs (CGs). This allows to build upon the existing ontology permitting for modularity and flexibility. The novelty of our approach lies in the ease with which CGs can be placed above other formalisms and their potential for optimised querying and retrieval.

1 Introduction

HealthAgents [2] is an agent-based, distributed decision support system (DSS) that employs clinical information, Magnetic Resonance Imaging (MRI) data, Magnetic Resonance Spectroscopy (MRS) data and genomic DNA profile information. The aim of this project is to help improve brain tumour classification by providing alternative, non invasive techniques. A predecessor project, Interpret [20], has shown that single voxel MRS data can aid in improving brain tumour classification. HealthAgents builds on top of these results and further employs multi voxel MRS data, as well as genomic DNA micro-array information for better classification results. Moreover, HealthAgents is decentralizing the Interpret DSS by building a distributed decision support system (d-DSS). This way, the number of cases to be studied is increased, improving classifier accuracy.

In this paper we focus on the problem of representing, in a meaningful way, the knowledge involved in the HealthAgents project. We regard knowledge representation in the spirit of [9] where a knowledge representation is described in terms of the five roles that it should play. More precisely, knowledge representation is a (1) surrogate, (2) a set of ontological commitments, (3) a fragmentary theory of intelligent reasoning, (4) a medium for efficient computation and (5) a medium of human expression. We will explain why the choice of Conceptual Graphs [19, 18] fulfills these requirements and its relevance in the context of medical knowledge for HealthAgents.

The problem of representing healthcare information (Electronic Healthcare Record) about an individual has been a key research field in medical informatics for many years. This information [13] (which can include tests, observations, imaging information, diagnostics, patient identification, legal permissions) has either been stored in a structured document based format (e.g. relational databases

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE

etc.) or unstructured document based format (e.g. photocopied hard copies). Electronic Healthcare Records (EHR) are difficult to represent, in a consistent manner, due to their content complexity. However, information [1] interoperability [5] will benefit patient care as it will allow for exchange of data between multiple sites. This is important in the context of this project where we expect hospitals from different parts of the world to join the HealthAgents network.

To address the interoperability shortcoming a number of standards have been proposed in the literature. A few examples that attempt to represent EHR include [11], [10], [12] and [6]. The aim is to structure the knowledge (using markup techniques) so that the clinical content is precisely identified. The ability to uniquely refer to a piece of information is denoted, in the context of these standards, as "semantics" since it allows the identification of the meaning of the knowledge. In this paper, however, we claim that this representation expressiveness is not sufficient for information retrieval. In the spirit of [15] we define semantics as the capability of inferring (reasoning) implicit knowledge from the knowledge base (based on explicit knowledge and given rules). This is important for HealthAgents (brain tumor information could influence the patient diagnosis).

In HealthAgents we developed HADOM (HealthAgents Domain Ontology) which conceptualises the parameters of the employed techniques (MRI, MRS, DNA microarrays etc.), the clinical information (age, sex, tumor location etc.) and the known brain tumor classes compliant to WHO (World Health Organisation) [2]. For example the structure "medical control" contains the information related to different MRI, MRS etc tests underwent by the patient. The HADOM ontology provides the basic terminology for the HealthAgents database schema and allows for interoperability at the terminological level. This is illustrated in Figure 1. For information retrieval and query capabilities we propose a Conceptual Graph based description of the data sources. In this paper we give the motivation for this knowledge representation formalism, informally describe Conceptual Graphs and show why they are suitable in the context of HealthAgents.

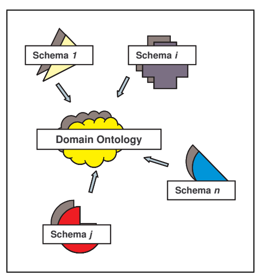


Figure 1. Ontological interoperability of HealthAgents database schema

2 Motivation

In HealthAgents we need to integrate medical knowledge from different sites and retrieve it in an intelligent manner. We thus need a flexible mechanism for data representation and querying.

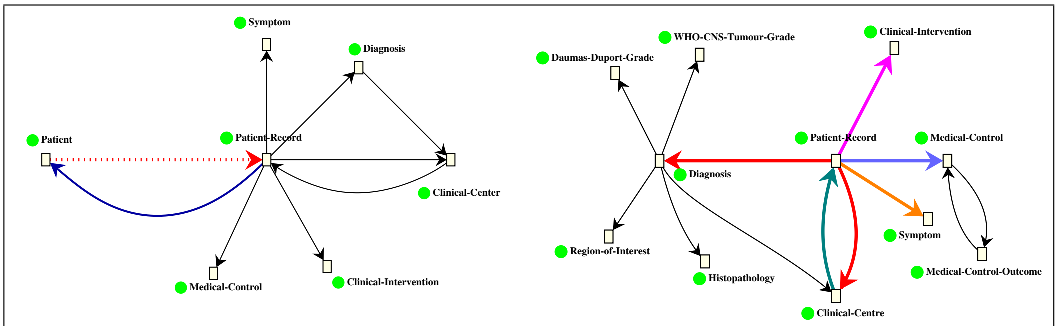
At the moment the data in the HealthAgents system is stored in relational databases at the various participating European clinical centers. A uniform vocabulary needed for interoperability reasons is provided by means of HADOM. The patient concept is at the centre of HADOM (see Figure 2(a)). Each visit of a patient is given a unique ID to be differentiated from other EHR regarding the same person. A particular patient instance, therefore, has several associated patient records. Tissue focus defines instances of the concerned areas under two sub groups, namely Primary_Focus and Secondary_Focus. A particular focus is related to one visit of a patient via Patient_Record (see Figure 2(b)). Many medical instruments and methods have been developed to diagnose brain

[1]In this paper we follow the work of [1] to distinguish between data, information and knowledge.

[2]Available from Harvard Medical School at: http://neurosurgery.mgh.harvard.edu/newwhobt.htm

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE

tumour. In HADOM, we enumerate the following approaches and define them as sub-concepts of Medical_Control: Biopsy, HRMAS, Magnetic_Resonance and Microarray.



(a) Patient and Patient-Record (b) One visit of a patient

Figure 2. Conceptual view of HealthAgents HADOM

The problem with representing the EHR in this format is that certain rules that can help retrieve implicit knowledge are hard to represent. Indeed, mutual understanding among software agents is partially rooted in a commonly agreed vocabulary/terminology in the brain tumour domain when such agents need to communicate with each other to express things like "retrieve cases of all patients under age 5" and "fetch a case of glioma from Hospital A" where underlined words are concepts from HADOM. That is to say, the domain ontology captures only the static model rather than the inference procedures. We would like to be able to express statements like "due to the fact that ... the tumour is malignant" or "all peak areas with ... characters suggest ...". Such separation (static model rather than inference procedures) is based on both theoretical and practical considerations. On the one hand, such inferences are built using rules, machine learning techniques, etc. which, currently, are not ready to be combined with major knowledge representation and reasoning formalisms, e.g. Description Logic, Frames, Entity-Relationship Model, etc. On the other hand, a medical diagnosis is normally a complicated process with ambiguity and uncertainty which cannot be entirely and precisely formalised in an inference model based on taxonomic knowledge. This, however, does not deny the merit of building a reasoning system on top of HADOM to provide moderate suggestions and warnings to clinicians. Such reasoning capability would be more appropriate to perform simple and specific tasks. This sort of extra reasoning power will also allow one to check for consistency within the HealthAgents ontology.

Therefore, we propose describing the knowledge contained in the sources by the means of Conceptual Graphs. This allows us to build upon the existing ontology while not overcomplicating the ontology with rules to describe data extraction techniques that can employ different parameters which greatly influence the outcome data. An immediate advantage of our Conceptual Graphs choice is their graph based reasoning mechanisms which allow versatile querying algorithms [7]. In the next section we informally introduce Conceptual Graphs and explain our choice of knowledge representation formalism in the context of the five roles of [9] enumerated in Section 1.

3 Conceptual Graphs based Representation

During the past 30 years, a wide variety of knowledge representation schemes have been developed, each of which have their own benefits and drawbacks. Expressiveness and efficiency are the key factors that greatly affect the competence of a representational scheme. In general, the term semantic network encompasses an entire family of graph-based visual representations. Initially, they have been introduced [17] for processing the semantics of natural languages. The system KL-ONE [4] and its descendants are the main representative descendants of this kind of semantic networks. The lack of

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE

a clear formal semantics of the first members of KL-ONE family has been successfully repaired by the most prominent KR languages, Description Logics (DLs) [3]. John Sowa developed Conceptual Graphs (CGs) on the basis of semantic networks and Peirce's Existential Graphs [16]. These graphs can be viewed as a diagrammatic system of logic, with the purpose "to express meaning in a form that is logically precise, humanly readable, and computationally tractable" [19].

Conceptual Graphs represent background knowledge, i.e. basic ontological knowledge, in a structure called support, which is implicitly used in the representation of factual knowledge as labelled graphs. A support consists of a concept type hierarchy, a relation type hierarchy, a set of individual markers that refer to specific concepts and a generic marker, denoted by *, which refers to an unspecified concept. The support defines the main concepts and relations that exist in the world we are trying to describe. These concepts and relations are going to be linked together by the means of a ordered bipartite graph that will describe the facts we are interested in. The ordered bipartite graph is going to represent the "stencil" which is going to be "filled in" with the concepts / relations taken from the support. A CG can be viewed as a bipartite graph that provides a semantic set of pointers to two ontologies. This means that we can reuse sources' ontologies, database schemas etc. for the purpose of describing those sources by the means of a CG. Moreover, the attached semantics of Conceptual Graphs make them a powerful reasoning knowledge representation and reasoning formalism. CG reasoning mechanisms can be viewed as a powerful tool for the querying process.

Due to the nature of the framework, the CG description can be placed on top of other knowledge sources or integration systems. This is called a Knowledge Oriented Specification of a source (KOS). Once the conceptual graph describing a source is built and integrated with other sources' CGs, the system is able to retrieve the answers to user's queries in a fully automated way. The conceptual graphs that integrates information from several KOSs and directs user querying is called a CG Mixer. This architecture is represented in Figure 3.

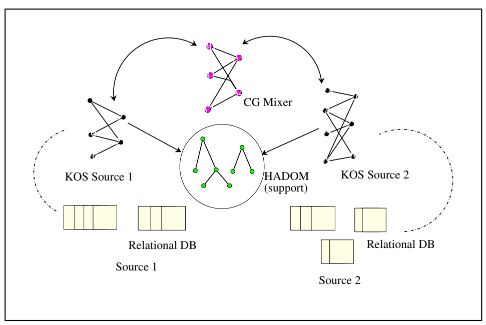
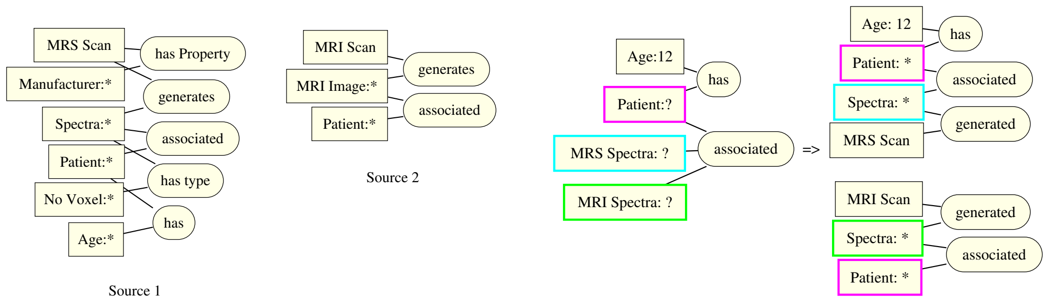


Figure 3. Conceptual Graph Description of Knowledge

More precisely let us look at two examples depicted in Figure 4 (a) and (b). In Figure 4 (a) we present a simplified example of two KOSs for MRS and MRI data sources. Once the sources are described with Conceptual Graphs they can be integrated in a CG Mixer. In this "global view" of the system the domain expert specifies exactly what queries can be posed in terms of this integrated schema. Once the query is posed, the relations from the CG Mixer are rewritten to direct the query to the appropriate data sources. Querying a CG Mixer is intuitively depicted in Figure 4 (b). A CG Mixer has the ability to focus on certain aspects of the data (considered at that time prevalent) but also is of dynamic nature (the changes will only affect the graph representation and not the data/wrappers).

To conclude, the five roles of R. Davis, H. Shrobe and P. Szolovits ([9]) – to view a knowledge representation (KR) in terms of their five important and distinctly different roles that it plays – helps highlighting the advantages of using Conceptual Graphs in HealthAgents. *A KR is a fragmentary theory of intelligent reasoning.* The projection checking algorithms (which correspond to logical deduction) can be optimized for practical applicability [7]. A KR is *a medium for pragmatically efficient computation.* A Conceptual Graph based description of data sources allows for representa-

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE



(a) Knowledge Oriented Specification (b) Query rewriting

Figure 4. Conceptual Graph based knowledge description and querying

tion of HealthAgents knowledge in a clear, organised manner. The CG query algorithms adapt the basic projection operation, offering a new pragmatic view to the knowledge integration problem. *A KR is a medium for human expression* and *a surrogate, substitute for the thing itself.* The visual capabilities of Conceptual Graphs means that the domain expert and the user are able to clearly see the knowledge they are working with. *A KR is set of ontological commitments.* Conceptual Graphs depict knowledge based on a "support" which encode the ontological knowledge (via the node labels). This is essential in a domain as complex as the medical domain where is crucial to be able to reuse existing information.

4 Conclusions

CGs are an intuitive, visual way of creating a semantically sound representation of knowledge; this makes conceptual graphs particulary suitable for knowledge description / querying in an interoperability scenario [14]. Since reasoning is essential for querying another one of the most important feature of CGs we plan to exploit is their reasoning capabilities. Mechanisms for reasoning can be computationally improved for data retrieval [7]. This visual, reasoning aspect, clearly differentiates our approach from existing work. Moreover, CGs allow reuse (by means of wrappers) of existing ontological knowledge expressed in different languages. This is very important as it allows us to reuse existing standards and / or available medical ontologies.

At the moment the CG describing a source is manually built by the domain expert. However, this step can be automated both from an information extraction and from a graph combination view point. In any case this issue is out of the scope of this paper since several wrappers [21] and / or suitable conceptual graph extensions [8] have been already proposed. The novelty of our approach lies in the ease with which CGs can be placed above other formalisms and in their potential for optimised querying and retrieval.

References

[1] A. Aamodt. Knowledge-intensive case-based reasoning in creek. In *Advances in case-based reasoning, 7th European Conference, ECCBR 2004, Proceedings*, volume 3155 of *Lecture Notes in Artificial Intelligence*, pages 1–15. Springer, 2004.

[2] C. Arús, B. Celda, S. Dasmahapatra, D. Dupplaw, H. González-Vélez, S. van Huffel, P. Lewis, M. Lluch i Ariet, M. Mier, A. Peet, and M. Robles. On the design of a web-based decision support system for brain tumour diagnosis using distributed agents. In *WI-IATW'06: 2006 IEEE/WIC/ACM Int Conf on Web Intelligence & Intelligent Agent Technology*, pages 208–211, Hong Kong, December 2006. IEEE.

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE

[3] F. Baader et al., editors. *The Description Logic Handbook*. Cambridge Univ. Press, 2003.

[4] R. J. Brachman and J. G. Schmolze. An Overview of the KL-ONE Representation System. *Cognitive Science*, 9(2):171–216, 1985.

[5] N. Brown and M. Reynolds. Strategy for production and maintenance of standards for interoperability within and between service departments and other healthcare domains. Technical report, CEN/TC 251 Health Informatics, Brussels, Belgium, 2000.

[6] D. Clunie. *DICOM Structured Reporting*. PixelMed, 2000.

[7] M. Croitoru and E. Compatangelo. Conceptual graph projection: a tree decomposition-based approach. In P. Doherty, Mylopuolos, and C. Welty, editors, *Proc. of the 10th Int'l Conf. on the Principles of Knowledge Representation and Reasoning (KR'2006)*, pages 271–276. AAAI, 2006.

[8] M. Croitoru, E. Compatangelo, and C. Mellish. Hierarchical knowledge integration using layered conceptual graphs. In *Proc. of the 13th Int'l Conf. on Conceptual Structures (ICCS'2005)*, number 3596 in Lect. Notes in Artif. Intell., pages 267–280. Springer, 2005.

[9] R. Davis, H. Shrobe, and P. Szolovits. What is a Knowledge Representation? *AI Magazine*, 14(1):17–33, 1993.

[10] E. C. for Standardization. Health informatics - electronic health record communication: Reference model, 2004.

[11] http://www.hl7.org. Health level 7.

[12] http://www.openehr.org. Openehr.

[13] I. Iakovidis. Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare records in europe. *International Journal of Medical Informatics*, 52(128):105–117, 1998.

[14] G. W. Mineau. A First Step toward the Knowledge Web : Interoperability Issues among Conceptual Graph Based Software Agents, Part I. In *Lect. Notes in Comp. Sci.*, volume 2393, pages 250–260. Springer, 2002.

[15] M.-L. Mugnier. Knowledge Representation and Reasonings Based on Graph Homomorphism. In *Proc. of the 8th Int'l Conf. on Conceptual Structures (ICCS'2000)*, pages 172–192, 2000.

[16] C. S. Peirce. Manuscript 514, 1909. Available at http://www.jfsowa.com/peirce/ms514.htm.

[17] M. Quillian. Semantic memory. In M. Minsky, editor, *Semantic Information Processing*, pages 227–270. MIT Press, 1968.

[18] J. Sowa. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co., 2000.

[19] J. F. Sowa. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, 1984.

[20] A. R. Tate, J. Underwood, D. M. Acosta, M. . Julia-Sape, C. Majos, A. Moreno-Torres, F. A. Howe, M. van der Graaf, M. M. Lefournier, F. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. L. Bosson, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. Wilkins, A. C. Bell, C. Remy, A. Heerschap, D. Watson, J. R. Griffiths, and C. Arus. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. *NMR Biomed*, 19:411–434, 2006.

[21] M. Willems. A Conceptual Semantics Ontology for Conceptual Graphs. In *Proc. of the 1st Int'l Conf. on Conceptual Structures (ICCS'93)*, pages 312–327, 1993.

Acknowledgements

This work is supported under the OpenKnowledge and HealthAgents STREP projects funded by EU Framework 6 under Grant numbers IST-FP6-027253 and IST-FP6-027213.

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE