Chapter 1.13
Ontologies in the Health Field

Michel Simonet
Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France

Radja Messai
Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France

Gayo Diallo
Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France

Ana Simonet
Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France

ABSTRACT

Health data and knowledge had been structured through medical classifications and taxonomies long before ontologies had acquired their pivot status of the Semantic Web. Although there is no consensus on a common definition of an ontology, it is necessary to understand their main features to be able to use them in a pertinent and efficient manner for data mining purposes. This chapter introduces the basic notions about ontologies, presents a survey of their use in medicine and explores some related issues: knowledge bases, terminology, and information retrieval. It also addresses the issues of ontology design, ontology representation, and the possible interaction between data mining and ontologies.

INTRODUCTION

Ontologies have become a privileged and almost unavoidable means to represent and exploit knowledge and data. This is true in many domains, and particularly in the health field. Health data and knowledge had been structured through medical classifications and taxonomies long before ontologies had acquired their pivot status of the semantic web. In the health field, there are still more than one hundred classifications (e.g., ICD10, MeSH, SNOMED), which makes it very difficult to exploit data coded according to one or the other, or several of these classifications. The UMLS (Unified Medical Language System) initiative tries to provide a unified access to these classifications, in the absence of an ontology of the whole medical domain - still to come.

In order to apprehend the interest of ontologies in data mining, especially in the health domain, it is necessary to have a clear view of what an ontology is. Unfortunately, there is no consensus within the scientific community on a common definition of an ontology, which is somewhat paradoxical, as one of the characteristics of an ontology is to represent a consensus of a community on a given domain. However, one does not need to enter the specialists' debate on ontologies to understand their main characteristics and therefore be able to use them in a pertinent and efficient manner for data mining purposes.

On a first level, one can think of an ontology as a means to name and structure the content of a domain. Among the numerous definitions that have been given, there is some kind of agreement that an ontology represents the concepts of a domain, the relationships between these concepts (IS-A and other relationships), the vocabulary used to designate them, and their definition (informal and/or formal). The IS-A relationship plays a central role, as it provides the (tree-like) skeleton of an ontology. This structure need not be a tree, as a concept may specialize several upper concepts, contrary to a taxonomy. Compared with a thesaurus, an ontology is freed from a particular language: an ontology deals with concepts, independently from the (natural) language that is used to designate them, while a thesaurus deals with terms that are expressed in a particular language. Moreover, a thesaurus does not enable the creation of new relationships between terms, whereas ontologies do.

There is no strict boundary between taxonomies, thesauri and ontologies, and a taxonomy may be considered as a particular case of an ontology. In practice, most ontologies rely on a taxonomic skeleton which is enriched with ontology-specific features. One can also notice that the conceptual schema of a database, expressed in object form, is close to an ontology (a micro-ontology) of the application domain of the database. Therefore, most people dealing with health data have been dealing with ontologies, either explicitly or implicitly – most often implicitly. However, making explicit the notion of ontology has made it possible to formalize and unite various formalisms and practices. The current ontology standard in the web universe, namely OWL[1], might not be the final standard for ontologies, but it has initiated a movement towards the need for an agreement for such a standard.

Ontologies have their roots in Aristotle's categories, and particularly in Porphyry's tree-like representation (3rd century), which laid the foundations for modern ontologies. This tree-like structure is still present in ontologies and in most knowledge representation systems through the IS-A relationship. The attributes in object or frame-based systems and the roles in Description Logics provide the other relationships of a possibly corresponding ontology. However, the introduction of ontologies in the field of Computer Science by Gruber in the 90's was not motivated by philosophical considerations but by the need of a representation in first-order logic of knowledge-based systems in order to facilitate their interoperability (Gruber, 1991). Today's ontologies are still strongly linked to first-order logic, either through Description Logics, which constitute the main stream in the ontology domain, or through conceptual graphs, which also have a strong logic background. Ontologies have also become an unavoidable support to knowledge and data integration.

In order to provide a level of understanding of ontologies that enables the reader to comprehend their interest in data mining, we first present two examples to introduce the basic notions related to ontologies. We then make a short historical presentation of the notion of ontology, with its philosophical background and its introduction in the computer field in the 90's for knowledge sharing purposes. In the following section, we make a survey of classifications and ontologies in the medical and biological fields. We then address some issues related to ontologies: knowledge

Related Content

The Social Role of Technology in Healthcare Quality Improvement
Anastasia K. Kadda (2012). *Quality Assurance in Healthcare Service Delivery, Nursing and Personalized Medicine: Technologies and Processes (pp. 178-186).*
www.irma-international.org/chapter/social-role-technology-healthcare-quality/58732

Designing Medical Research Web Sites
Jonathan P. Grady, Michael B. Springand Armando J. Rotondi (2009). *Handbook of Research on Information Technology Management and Clinical Data Administration in Healthcare (pp. 852-869).*
www.irma-international.org/chapter/designing-medical-research-web-sites/35818

Early Intervention and Health Promotion in the Management of Chronic Liver Disease
Shaila Hussain (2018). *International Journal of User-Driven Healthcare (pp. 23-39).*
www.irma-international.org/article/early-intervention-and-health-promotion-in-the-management-of-chronic-liver-disease/214995

Patient-Doctor Interconnectivity: Improving Health Care Management and Patient Compliance with Web Technology
Jerry B. Weinberg, Steven P. Klein, Robert Klepper, Bernard Waxman, Xudong Yuand Daniel K. Anderson (2000). *Managing Healthcare Information Systems with Web-Enabled Technologies (pp. 186-204).*
www.irma-international.org/chapter/patient-doctor-interconnectivity/25831

A Comparative Study of FFT, DCT, and DWT for Efficient Arrhytmia Classification in RP-RF Framework
Tea Marasoviand Vladan Papi (2018). *International Journal of E-Health and Medical Communications (pp. 35-49).*
www.irma-international.org/article/a-comparative-study-of-fft-dct-and-dwt-for-efficient-arrhytmia-classification-in-rp-rf-framework/191122