DS-Ontology: A Disease-Symptom Ontology for General Diagnosis Enhancement

Linda Mhadhbi
University of Tunis, Tunisia
linda.mhedhbi@gmail.com

Jalel Akaichi
King Khalid University, Saudi Arabia
J.akaichi@gmail.com

ABSTRACT

The diagnosis process is often challenging, it involves the correlation of various pieces of information followed by several possible conclusions and iterations of diseases that may overload physicians when facing urgent cases that may lead to bad consequences threatening people's lives. The physician is asked to search for all symptoms related to a specific disease. To make this kind of search possible, there is a strong need for an effective way to store and retrieve medical knowledge from various datasets in order to find links between human disease and symptoms. For this purpose, we propose in this work a new Disease-Symptom Ontology (DS-Ontology). Utilizing existing biomedical ontologies, we integrate all available disease-symptom relationships to create a DS-Ontology that will be used latter in an ontology-based Clinical Decision Support System to determine a highly effective medical diagnosis.

CCS Concepts

•Information systems → Web Ontology Language (OWL); Ontologies; •Computing methodologies → Knowledge representation and reasoning; •Applied computing → Health care information systems;

Keywords

Disease, Symptom, Ontology, diagnosis, Protègè.

1. INTRODUCTION

Firstly, inadequate response and bad decision made by mobile physicians may lead to bad consequences threatening people's lives. Moreover, there is growing information that overload physicians when facing urgent cases. This is why diagnostic errors are affecting a large number of patients and they are the most common cause of litigation against clinicians. The physician is asked to search for all symptoms related to a specific disease. To make this kind of search possible, there is a strong need for an effective way to store and retrieve medical knowledge from various datasets in order to find links between human disease and symptoms (Figure 1).

Moreover, the recent development of Semantic Web technology shows ontology as a promising tool in knowledge representation and reasoning. Ontology is "a model of a particular field of knowledge, concepts and their attributes, as well as the relationships between these latter" [1]. It is increasingly being used to bring order to data and to promote knowledge extraction and reasoning. In the last decades, the use of ontologies is enabling biomedical researchers to solve challenges and achieve their goals [2]. Therefore, we rely on ontology for its ability to provide a structured description and relationship between data, and to offer high expressivity and reasoning capability to establish the relationship between disease and symptoms.

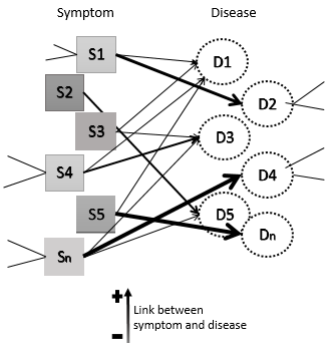

Figure 1: From symptom to disease: A simplified model

Indeed, biomedical ontologies are becoming increasingly important in order to provide effective biomedical knowledge representation to support clinicians and to give solutions for the diagnosis process. Examples include principally the Human Disease Ontology (DO) and Symptom Ontology (SYMP). Each of these ontologies has a particular focus, either it provides information on disease or symptom of a certain kind. These latter can be used in general diagnosis if the relationship between their concept is established. They are published in BioPortal, one of the outstanding references, the world's largest ontology on-line repository hosting over 450 biomedical ontologies and more than 6 million classes that define terms [3].

In this work, based on existing ontologies, principally on DO and SYMP, we propose a new disease-symptom ontol-

ACM 978-1-4503-4833-1/17/04 ...$15.00.
http://dx.doi.org/10.1145/3077584.3077586

ogy (DS-Ontology) that reveals the links between disease and symptoms. One could argue that these concepts are defined by other ontologies or models, why another ontology? Actually, there have been different attempts to incorporate main vocabularies for biomedical domain and to define semantic types and relationships between disease and symptoms whether by using ontologies matching algorithms, ontologies mapping, ontologies alignment, disease-symptom models (see related works in section 2). Still, the problem of disease-symptoms relations is not solved yet. Such a challenge requires further efforts and work.

In fact, this proposed ontology will be used in future works to create an ontology-based Clinical Decision Support System that assists the mobile physician in the diagnosis process and furthermore to find the appropriate health care institution for the patient state. If used, this ontology aims to improve, in an efficient way, the tasks performed by the mobile physician in emergency cases. The remainder of this paper is organized as follows; In the next section, we discuss the historical overview and practical application of ontologies and ontology-based models in general diagnosis. In section 3, we introduce the knowledge representation and then we provide a full understanding of the proposed solution. At last, the conclusion and the following research that will be made, are presented.

2. STATE OF THE ART

Nowadays, information systems are omnipresent in the area of health care [4]. They stimulate exchanges between different medical communities, to improve practices in the health sector and to understand the mechanism and the interpretation of medical reasoning, abstraction and the development of knowledge. Quality of care and patient safety are major issues in the context of automation of health information systems [5]. The development of semantic web technologies reveals that ontology is a promising tool in knowledge representation and reasoning [6] [7]. Ontology is widely used as a model to represent the relations between concepts within a specific domain [8] in order to facilitate information integration, data search and exchange, and other critical knowledge-intensive tasks.

In addition, the emergence of ontologies in various fields has captured the attention of medical informatics researchers who seek formal knowledge representation and reasoning in the health care domain. Indeed, structuring knowledge related to emergency systems becomes a must in order to Progress. Therefore, the use of ontology insures a better knowledge representation and efficient semantic reasoning. With it we aim to provide common, unambiguous semantics and vocabularies for the proposed purpose. Applying such a domain in emergency management not only improves the efficiency and effectiveness of the decision-making process during emergency cases, but also facilitates the integration and the manipulation of various concepts. This contributes is to a better emergency response, management and provision of timely medical emergency care [9] [10].

Professionals and clinicians have observed the variations in patient‘s health and stored this information using different methodologies (databases of patient histories). The problem of standardizing medical languages and terminologies has been a major concern of medical research. Authors in [11] details why it is hard to develop such terminologies.

Over time, and with the evolution of science, there have been new technologies to store knowledge of different types of disease, symptoms, signs, etc. Scientists tried to organize these information in a useful manner in order to store and retrieve knowledge related to human disease. Effective knowledge representation necessitates the use of standardized vocabularies to guarantee both shared understanding between people and interoperability between information systems. Yet, there have been countless biomedical vocabularies integrated in the Unified Medical Language System (UMLS) [12] such as SNOMED CT [13], RxNorm, ICD, etc. However, many of these latter are ether incomplete, inconsistent or confused accounts of basic terms pertaining to diseases, diagnoses and clinical phenotypes [14].

Up to now, there is no biomedical vocabulary that enables us to query symptoms for disease. Thus, there is a strong need for establishing the relationship between these two concepts to overcome querying difficulties. For instance, one might be interested in finding all the symptoms that indicate a certain disease but it is difficult to uncover this kind of knowledge due to the lack of data categorization in spite of the multiple biomedical vocabularies on BioPortal. In fact, there have been considerable efforts to propose improvements in applying ontologies to incorporate main vocabularies for biomedical domain and define semantic types and relationships between them but there is still a lot of work to be done. Examples includes Human Disease Ontology (DO) [15], Symptom Ontology (SYMP) etc.

In [16], authors specified that there is no ontology that defines disease class hierarchies, symptom class hierarchies, and establishes relationship between disease and symptom class in order to provide a general diagnosis of patient. Therefore, they proposed an alignment between DO and SYMP. Accordingly, this alignment has been repeated to only 11 diseases. Still, the major drawback appeared when they were based on a health website, server or a database to create the missing link. Authors in [17] tried to find likely disease of a patient basing on annotation. They built an initial ontology containing lymphoma-related diseases and symptoms as a prototype implementation, but this work needs several necessary enhancement.

In [18], an initial Model for Clinical Information (MCI) based on the Ontology of General Medical Science (OGMS) and other OBO ontologies is proposed. MCI integrates structured data about diagnosis (ICD-10), procedures (OPS), demographics, administered drugs (ATC), laboratory values (LOINC) and health care providers. Yet, MCI is in a very initial stage and they plan to extend it in future works. Rather than concepts mappings, authors in [19] focused on the improvement of concept proprieties and the extraction of semantic relationship. They tried to arrange data around specific concepts (disease and symptom) from different ontologies using existing BioPortal mappings. This work is different from general ontology alignment, it tries to understand the semantics of data across ontologies in order to build a model that relate disease classes and symptom classes. However, there is no effective ontology created using this proposed model.

To summarize, existing ontologies that provide a well structured biomedical vocabulary do not have enough relationships between symptoms and disease. In fact, we canâĂŹt use existing ontologies to infer likely disease of a patient. Regarding these issues, we propose a disease-symptom ontology (DS-Ontology) that covers the process of general diagnosis.

Indeed, knowledge about disease-symptom relationship are found in common clinical knowledge resources and confirmed by our expert in medical domain. Such ontology would be the core of our Clinical Decision Support System so as to support mobile physicians in finding out patient's diagnosis quickly.

3. REALIZATION

In order to provide a full understanding of the whole process of DS-Ontology construction, we describe the different steps (Figure 2) as follows: First, in the data acquisition stage and to identify the ontology main concepts (i.e. classes and their subclasses), we have collected data from various existing ontologies describing a part of the large domain: Human Disease Ontology (DOID) and Symptom Ontology (SYMP).

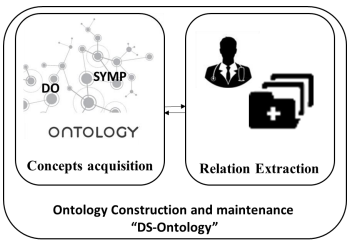

Figure 2: A simplified model of DS-Ontology construction

First, DOID is currently the most prominent disease ontology to be developed. It is an ongoing project that proposes to create a single structure for the classification of disease. It was created in 2003 as part of the NU gene project at Northwestern University, Center for Genetic Medicine and the University of Maryland School of Medicine, Institute for Genome Sciences. It has been published in several versions for many years. The content of DOID has had 192 revisions since 2012, including the addition of 760 terms [15]. Actually, it contains 11,402 classes and 36% of all terms now include definitions (Figure 3). DOID is currently a standard ontology adopted by the OBO Foundry [20].

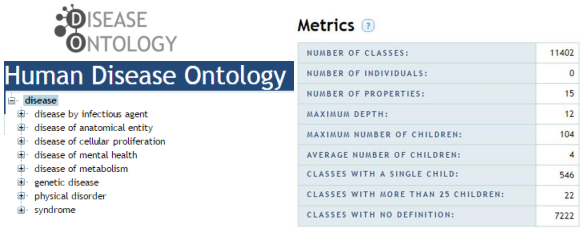

Figure 3: Human Disease Ontology (DOID) as represented in BioPortal

As regards Symptom Ontology (SYMP), it is an OBO Foundry ontology developed in 2005 by the institute for Genome Sciences at the University of Maryland. Actually, it contains 936 symptoms that are organized by body regions. We created the DS-Ontology using Protègè an open-source ontology editor and framework. It is a flexible and robust development environment. Using this, researchers and domain experts can easily build effective knowledge-based systems [21]. In the diagnosis process, each disease is known to have a number of symptoms. Establishing the relationship between disease and symptom become extremely necessary. Thus, starting from DOID ontology, we extended it to encompass not just disease but also the different related symptoms. To make this kind of link possible, we start by combining DOID and SYMP in one global ontology. To make this happen, we loaded both DOID and SYMP in Protege 5.0 and then we merge these two ontologies using the refactor parameters. The next step is to define an object property "has-symptom" to relate every disease to the appropriate symptom as shown in Figure 4.

Figure 4: DS-Ontology and the main object property "has-symptom"

The relation between disease and symptoms is concluded from diagnosis documents, books and in collaboration with medical experts in order to capture their knowledge. The domain of the property is the set of disease classes and the range is the set of symptom classes. Therefore, the different diseases of the existing ontology, DOID, are connected to the added symptoms to have as a result an ontology that covers links between disease and symptom. Figure 5 is an example of such a link of a heart attack as visualized by OntoGraph in Protègè.

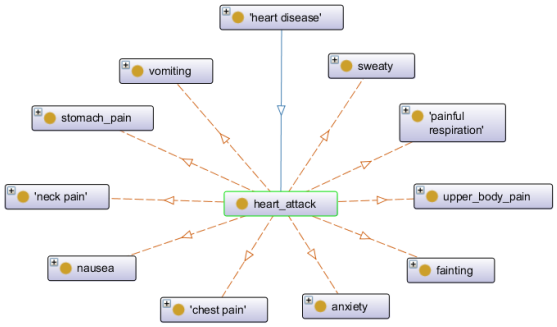

Figure 5: The symptoms of a heart attack in DS-Ontology as visualized by OntoGraph in Protègè

Compared to other ontologies, our DS-Ontology has more

advantages. It provides symptoms for 200 diseases and there is an ongoing work to complete the rest of them.

4. CONCLUSION

In this article, we propose a DS-Ontology (Disease-Symptom Ontology) that covers the links between disease and symptom. A careful survey of the current research on disease symptom ontologies reveals that there is no effective ontology that one can use in general diagnosis. As future work, we plan to optimize further the proposed DS-Ontology to be the core of a Clinical Decision Support System dedicated to mobile physician when facing critical and urgent situations. It will be used to extract the exact disease of a patient in case of emergency.

5. REFERENCES

[1] Kaushal Giri. Role of ontology in semantic web. *DESIDOC Journal of Library & Information Technology*, 31(2), 2011.

[2] Daniel L Rubin, Nigam H Shah, and Natalya F Noy. Biomedical ontologies: a functional perspective. *Briefings in bioinformatics*, 9(1):75–90, 2008.

[3] Christopher Ochs, Zhe He, Ling Zheng, James Geller, Yehoshua Perl, George Hripcsak, and Mark A Musen. Utilizing a structural meta-ontology for family-based quality assurance of the bioportal ontologies. *Journal of biomedical informatics*, 61:63–76, 2016.

[4] Jean Charlet, Gunnar Declerck, Ferdinand Dhombres, Pierre Gayet, Patrick Miroux, and Pierre-Yves Vandenbussche. Construire une ontologie médicale pour la recherche d'information: problématiques terminologiques et de modélisation. In *23es journées francophones d'Ingénierie des connaissances*, pages 33–48, 2012.

[5] René Amalberti, Yves Auroy, Don Berwick, and Paul Barach. Five system barriers to achieving ultrasafe health care. *Annals of internal medicine*, 142(9):756–764, 2005.

[6] Fouad Zablith, Grigoris Antoniou, Mathieu d'Aquin, Giorgos Flouris, Haridimos Kondylakis, Enrico Motta, Dimitris Plexousakis, and Marta Sabou. Ontology evolution: a process-centric survey. *The Knowledge Engineering Review*, 30(01):45–75, 2015.

[7] Leo Obrst, Michael Gruninger, Ken Baclawski, Mike Bennett, Dan Brickley, Gary Berg-Cross, Pascal Hitzler, Krzysztof Janowicz, Christine Kapp, Oliver Kutz, et al. Semantic web and big data meets applied ontology. *Applied Ontology*, 9(2):155–170, 2014.

[8] Harry Chen, Tim Finin, and Anupam Joshi. Using owl in a pervasive computing broker. Technical report, 2005.

[9] Mijal Mistry, Dipti Shah, Pranav Pathak, and Abu Sarwar Zamani. Ontologies: Need, usage and attainment of health care system. In *Intelligent Systems and Signal Processing (ISSP), 2013 International Conference on*, pages 381–386, 2013.

[10] Werner Ceusters, Barry Smith, and Jim Flanagan. Ontology and medical terminology: Why description logics are not enough. *Towards an Electronic Patient Record (TEPR 2003), Boston, MA*, 2003.

[11] Alan L Rector et al. Clinical terminology: why is it so hard? *Methods of information in medicine*, 38(4/5):239–252, 1999.

[12] Olivier Bodenreider. The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*, 32(suppl 1):D267–D270, 2004.

[13] Dennis Lee, Nicolette de Keizer, Francis Lau, and Ronald Cornet. Literature review of snomed ct use. *Journal of the American Medical Informatics Association*, 21(e1):e11–e19, 2014.

[14] Richard H Scheuermann, Werner Ceusters, and Barry Smith. Toward an ontological treatment of disease and diagnosis. *Proceedings of the 2009 AMIA summit on translational bioinformatics*, 2009:116–120, 2009.

[15] Warren A Kibbe, Cesar Arze, Victor Felix, Elvira Mitraka, Evan Bolton, Gang Fu, Christopher J Mungall, Janos X Binder, James Malone, Drashtti Vasant, et al. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic acids research*, page gku1011, 2014.

[16] Osama Mohammed, Rachid Benlamri, and Simon Fong. Building a diseases symptoms ontology for medical diagnosis: an integrative approach. In *Future Generation Communication Technology (FGCT), 2012 International Conference on*, pages 104–108. IEEE, 2012.

[17] Heiner Oberkampf, Sonja Zillner, Bernhard Bauer, and Matthias Hammon. Interpreting patient data using medical background knowledge. *ICBO*, 897:3, 2012.

[18] Heiner Oberkampf, Sonja Zillner, Bernhard Bauer, and Matthias Hammon. An ogms-based model for clinical information (mci). In *ICBO*, pages 97–100, 2013.

[19] Heiner Oberkampf, Turan Gojayev, Sonja Zillner, Dietlind Zühlke, Sören Auer, and Matthias Hammon. From symptoms to diseases–creating the missing link. In *European Semantic Web Conference*, pages 652–667. Springer, 2015.

[20] Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J Mungall, et al. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11):1251–1255, 2007.

[21] John H Gennari, Mark A Musen, Ray W Fergerson, William E Grosso, Monica Crubézy, Henrik Eriksson, Natalya F Noy, and Samson W Tu. The evolution of protégé: an environment for knowledge-based systems development. *International Journal of Human-computer studies*, 58(1):89–123, 2003.