Building a Diseases Symptoms Ontology for Medical Diagnosis: An Integrative Approach

Osama Mohammed, Rachid Benlamri and Simon Fong*
Department of Software Engineering, Lakehead University, Ontario, Canada
*Department of Computer and Information Science, University of Macau, Macau, China

Abstract**—Medical ontologies are valuable and effective methods of representing medical knowledge. In this direction, they are much stronger than biomedical vocabularies. In the process of medical diagnosis, each disease has several symptoms associated with it. There are currently no ontologies that relate diseases and symptoms and only attempts at their infancy along with some simple proposed models. However, well established ontologies for diseases and for symptoms were already developed in isolation. In this article, we are proposing an alignment algorithm to align the diseases ontology (DOID) with the symptoms ontology (SYMP) creating a core diseases symptoms ontology (DSO) that can scale to any number of diseases and symptoms The core DSO links a few diseases to their symptoms.*** *(Abstract)*

Keywords**—Ontology Alignment, DOID, SYMP, DSO, Semantic Web**

I. ONTOLOGY AS A KNOWELDGE STRUCTURE FOR DISEASE DIAGNOSIS

Development of knowledge structures for healthcare is a research community-wide effort focused on the development of a set of interoperable knowledge modules that together provide solutions to many healthcare challenges of the 21st century. Such efforts encompass a wide range of activities, including ontology engineering, biomedical modeling, data mining, knowledge discovery tools and database development, simulation, and visualization. However, effective knowledge representation requires the use of standardized vocabularies to ensure both shared understanding between people and interoperability between information systems [1]. Internationally, there are countless existing biomedical vocabularies such as SNOMED-CT[1], LOINC[2] ICD-9 CM[3], MeSH[4] and UMLS[5]. Unfortunately, many of these existing biomedical vocabulary standards rest on incomplete, inconsistent, or confused accounts of basic terms pertaining to diseases, diagnoses, and clinical phenotypes [1]. There are indeed several attempts to harmonize such terminologies but such efforts are at their infant stage [2].

For this reason, there is a need for an effective way to store and retrieve knowledge related to human diseases. In this direction ontologies play a crucial role in defining standardized concepts. Beyond defining standards, medical ontologies are much more than biomedical vocabularies. They arrange concepts into ISA and sibling hierarchies, which effectively relate these concepts in a structural way that provides valuable inferences upon retrieval. Ontology hierarchies are valuable methods of knowledge representation. Figure 1 is an illustration of such hierarchies in disease ontology.

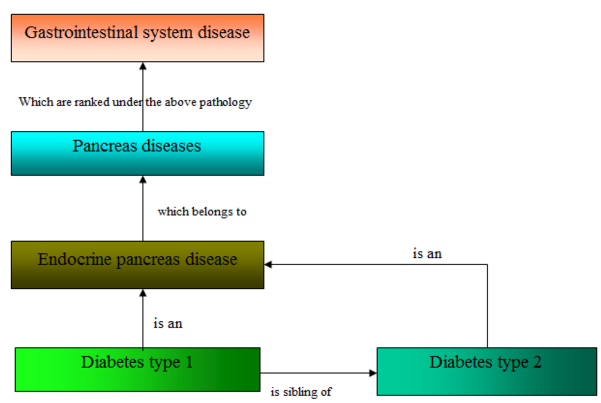

Figure 1. Disease ontology ISA and sibling hierarchies relate and classify diseases

[1] http://www.openclinical.org/medTermSnomedCT.html
[2] http://loinc.org/
[3] http://icd9cm.chrisendres.com/
[4] http://www.ncbi.nlm.nih.gov/mesh
[5] http://www.nlm.nih.gov/research/umls/

978-1-4673-5861-3/12/$31.00 ©2012 IEEE

A careful inspection of the attempts aimed at developing ontologies for disease diagnosis (e.g. TMO[6], Galen[7], HPO[8], IDO[9]) reveals that they are dedicated for specific purposes and cannot be used for disease diagnosis in general [1]. However, there are two OBO ontologies that can be used for general disease diagnosis if their classes and properties are aligned properly: DIOD[10] and the SYMP[11]. The following two sections briefly describe these ontologies.

II. DISEASE ONTOLOGY (DOID)

Organising diseases in an ontology hierarchy is extremely useful as it forms a pathological classification of diseases for use in medical systems. Such an undertaking is massive given the number of known human diseases let alone this ontology has to be updated as time passes and more diseases develop and are discovered. The most prominent disease ontology developed to date is the Human Disease Ontology (DOID). Started in 2003 as part of the NUgene project at Northwestren University, it has been published in several versions over several years and contains to this date over 8600 known human diseases and 14,600 terms. DOID is currently a standard ontology adopted by the OBO Foundry[12] (See figure 2).

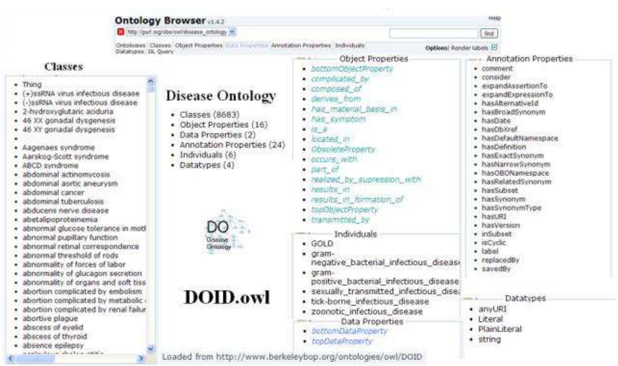

Figure 2. DOID ontology as visualized by the Manchester University Ontology Browser[13]

III. SYMPTOMS ONTOLOGY (SYMP)

Symptom Ontology (SYMP) was developed in 2005 by the Institute for Genome Sciences (IGS) at the University of Maryland. Today it contains more than 900 symptoms. SYMP's hierarchy categorizes symptoms under certain headings for example categorizing all types of pain (arm, leg, headache, back pain, chest pain, etc) under physical pain. Figure 3 illustrates such categorizing.

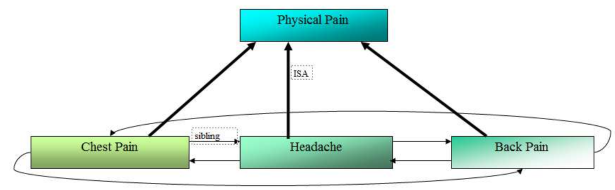

Figure 3. Symptoms ontology defines and relates symptoms

SYMP became a standard ontology and was adopted by the OBO Foundry during 2008.

IV. RELATING SYMPTOMS TO DISEASES IN AN ONTOLOGY

Currently, there is no ontology that defines disease class hierarchies, symptom class hierarchies, and establishes relations between disease and symptom classes. Such ontology would be very useful for diagnosis recommendation systems. There are in-progress attempts to modify the human disease ontology to include symptoms and relations between those symptoms to diseases.[10] Also, there were models proposed for such undertaking. GHDO [3] proposes an ontology model that relates diseases to symptoms (phenotypes) and to the other three elements that uniquely identify a disease: disease type, causes, and treatment. However, no GHDO ontology has been published from the proposed model. In the next section we are presenting our attempt to develop an ontology for disease diagnosis based on aligning the two standard ontologies (DOID and SYMP).

V. MEDICAL ONTOLOGY ALIGNMENT

Ontology alignment is the idea of combining two (or more) ontologies into one and defining relationships between the concepts of the ontologies forming a new ontology in the process. Alignment between ontologies is a critical challenge for semantic interoperability [4] as well as for producing hybrid ontologies [5]. As the medical domain is represented by multiple ontologies, there is a need for creating mappings among these ontologies elements in order to facilitate the integration of data and reasoning across these ontologies. There are two main approaches to alignment: *Ontology Matching* and *Ontology Linking*. Ontology matching techniques are

[6] http://code.google.com/p/translationalmedicineontology/
[7] http://www.co-ode.org/galen/
[8] http://www.human-phenotype-ontology.org/index.php/hpo_browse.html
[9] http://infectiousdiseaseontology.org/page/Main_Page
[10] http://do-wiki.nubic.northwestern.edu/index.php/Main_Page
[11] http://symptomontologywiki.igs.umaryland.edu/wiki/index.php
[12] http://www.obofoundry.org/
[13] http://owl.cs.manchester.ac.uk/browser

for relating ontologies on the same domain or on partially overlapping domains. For example, ontology mapping works if two disease ontologies are to be aligned. In such case, disease classes from both ontologies are matched. Special mapping constructs are used to indicate how elements from different ontologies are semantically related or equivalent. Ontology linking, in contrast, allows elements from distinct ontologies to be coupled with links [6]. A strict requirement is that the domains of the ontologies that are being combined are disjoint. This means that the classes/concepts of both ontologies must be separate for ontology linking to be applied. For example, ontology linking is appropriate for aligning disease and symptom ontologies as diseases and symptoms are separate concepts. In the next section we are developing a diseases symptoms ontology (DSO) using ontology alignment by linking two notable existing ontologies: DOID and SYMP.

VI. LINKING SYMP & DOID TO CREATE DSO

In this section, we propose an ontology linking algorithm for combining SYMP & DOID to create DSO ontology. Linking all the SYMP & DOID classes is a massive undertaking. However, if the linking process is simple & repetitive, then a core ontology can grow with time to connect all diseases classes to their symptoms classes. The following is our algorithm of this process:

```
Module linkAlignment (SYMP, DOID) {
    for all diseases D € DOID {
        copy D into DSO
        fetch symptoms of D from a health website or server, a database, etc
        for all symptoms names S of D {
            for symptoms S1 € SYMP {
                if(S == S1.name) {
                    1. copy S1 into revised DOID under a base symptoms class
                    2. define a new has_symptom object property for D where D
                       has_symptom S1
                }
            }
        }
    }
    output DSO
}
```

Although the above algorithm has been designed to be performed manually, it is a systematic one which can be automated for a large number of diseases. If done manually, a collaborative community-wide effort can result in the creation of a large DSO including a large number of diseases and symptoms.

VII. CONNECTING SYMPTOMS & DISEASES CLASSES

Establishing relations between symptoms classes and diseases classes is done by defining a *has_symptom* object property. The domain of the property is the set of diseases classes and the range is the set of symptoms classes. In the diagnosis process, each disease is known to have a number of symptoms. Correspondingly in DSO, each disease class needs to have a number of has_symptom properties where each property links the disease class to one of its symptoms classes. The following table illustrates this linking process for the hypertension disease class.

Table 1. Hypertension disease class (from DOID) linked to symptoms classes (from SYMP) representing the symptoms of hypertension

Disease Class Name	Disease Class DOID Ontology Code	Object Property	Symptom Class	Symptom Class SYMP Ontology Code
Hypertension	DOID_10763	has_symptom	blurred vision	SYMP_000012
Hypertension	DOID_10763	has_symptom	drowsiness	SYMP_000024
Hypertension	DOID_10763	has_symptom	tinnitus	SYMP_0000393
Hypertension	DOID_10763	has_symptom	nosebleed	SYMP_0000448
Hypertension	DOID_10763	has_symptom	headache	SYMP_0000504
Hypertension	DOID_10763	has_symptom	flushing	SYMP_0000511
Hypertension	DOID_10763	has_symptom	Nausea	SYMP_0000458
Hypertension	DOID_10763	has_symptom	palpitation	SYMP_0000530
Hypertension	DOID_10763	has_symptom	frequent urination	SYMP_0000563
Hypertension	DOID_10763	has_symptom	urgency of urination	SYMP_0000590
Hypertension	DOID_10763	has_symptom	nocturia	SYMP_0000564
Hypertension	DOID_10763	has_symptom	dizziness	SYMP_0000610
Hypertension	DOID_10763	has_symptom	breathing difficulty	SYMP_0019153
Hypertension	DOID_10763	has_symptom	fatigue	SYMP_0019177

Using the protégé editor[14], this linkage can be made simple. As illustrated in figure 4, the *has_symptom* properties can be added under the superclasses tab within the original DOID. For each symptom class related to the hypertension disease class, we need to create a new *has_symptom* object property.

[14] http://protege.stanford.edu/

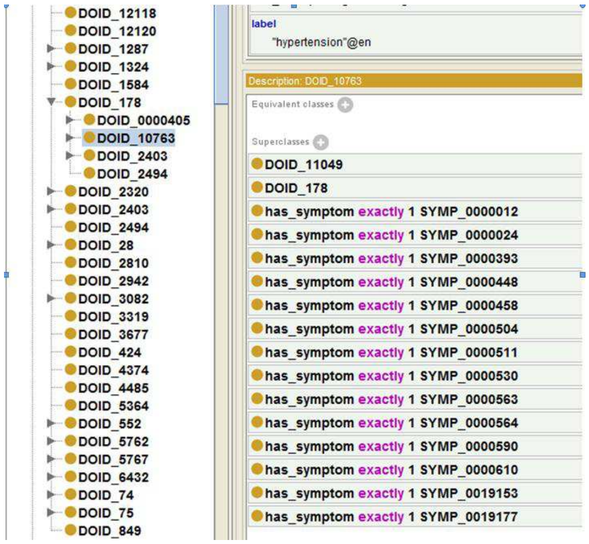

Figure 4. DSO has_symptom property relates a disease class to its symptom classes

The linking result is a rich DSO with disease hierarchies, symptoms hierarchies, and relations between disease and symptom classes. Figure 5 is an example of such hierarchies/relations for anemia disease class.

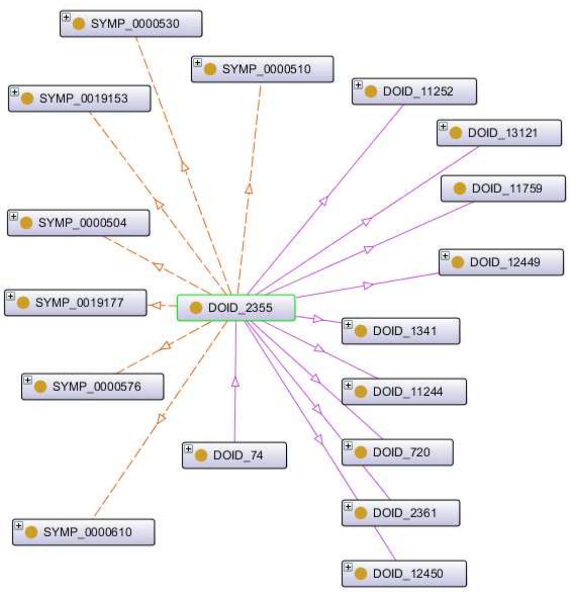

Figure 5. Anemia (DOID_2355) ISA hierarchy and has_symptom relations to its symptoms classes as shown by OntoGraph tab in Protégé

The next two figures are an illustration of how the has_symptom property, which was added to the DOID, connects diseases of DOID to symptoms of SYMP. Figure 6 shows connections between several DOID diseases and SYMP symptoms, while figure 7 shows the connections between the hypertension disease class to some of the corresponding symptoms classes in SYMP as an example of such connections.

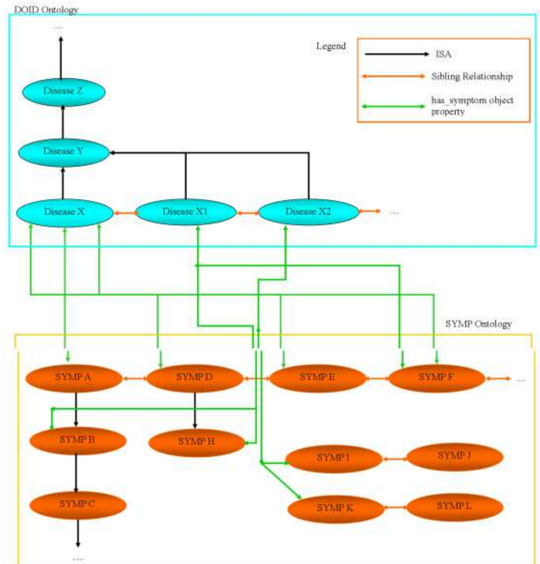

Figure 6. DSO connects SYMP & DOID terms via the has_symptom Property

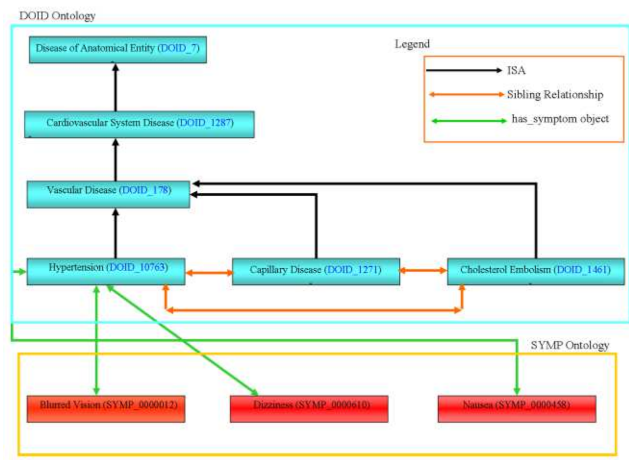

Figure 7. Hypertension has_symptom Connections to some of its symptoms in SYMP

VII. CONCLUSIONS

In this article we introduced our method to engineer ontology for disease diagnosis. Careful survey to the current research on disease diagnosis ontologies reveals no effective ontology available for general disease diagnosis. However, there are two OBO standard ontologies that can be used for general disease diagnosis: DIOD and SYMP. We proposed an alignment algorithm to align these two ontologies. The proposed alignment method has been repeated for 11 inter-related diseases to produce a core ontology for disease diagnosis-- *Diseases Symptoms Ontology (DSO).* Indeed our method and process can be repeated for any number of diseases to create a larger version of the DSO. Our core DSO ontology has been published on our university Flash server (http://flash.lakeheadu.ca/~omohamme/DSO.owl) and can be used for analysis by other semantic web applications.

ACKNOLEDGMENT

The first author would like to thank Dr. Sabah Mohammed and Dr. Jinan Fiaidhi of the Computer Science Department, Lakehead University for their valuable comments that significantly improve the first draft of this article.

REFERENCES

[1] R.H. Scheuermann, W. Ceusters, and B. Smith, "Toward an Ontological Treatment of Disease and Diagnosis," Proceedings of the 2009 AMIA Summit on Translational Bioinformatics, pp. 116-120, San Francisco, CA, 2009, http://ontology.buffalo.edu/medo/Disease_and_Diagnosis.pdf

[2] R. Hamm, S.E. Knoop, P. Schwarz, A.D. Block, and W.L. Davis, "Harmonizing clinical terminologies: driving interoperability in healthcare," Loinic Stud Health Technol Inform Report, vol. 129, pp. 660-663, 2007, http://loinc.org/articles/Hamm2007a

[3] M. Hadzic, E. Chang, "Ontology-based Multi-agent Systems Support Human Disease Study and Control," Proceeding of the conference on Self-Organization and Autonomic Informatics (I), In Frontiers in Artificial Intelligence and Applications, Edited by H. Czap, R. Unland, C. Branki, H. Tianfield, IOS Press, vol. 135, pp. 129-131, Glasgow, UK, 2005

[4] T.C. Hughes, B.C Ashpole, "The Semantics of Ontology Alignment," Approved for public release, August, 2004, http://www.atl.external.lmco.com/projects/ontology/papers/SOA.pdf

[5] B. Zhang, "Alignment of multiple ontologies of anatomy: deriving indirect mappings from direct mappings to a reference," Proceedings for AMIA Annual Symposium, pp. 864-868, 2005, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560629/

[6] M. Homola, L. Serafini, "Towards Formal Comparison of Ontology Linking, Mapping and Importing," Proceeding of 23rd Int. Workshop on Description Logics (DL2010), CEUR-WS 573, Waterloo, Canada, 2010, www.cs.uwaterloo.ca/conferences/dl2010/papers/paper_26.pdf