

Developing Ontology Support for Human Malaria Control Initiatives

Olawande Daramola, Segun Fatumo
Department of Computer and Information Sciences,
Covenant University, Ota, Nigeria.

Abstract

Malaria is one of the most common infectious diseases and an enormous public health problem in Sub-Sahara Africa, Asia and parts of America. In this paper, we discuss the development of the Human Malaria Control Ontology (HMCO) which contains general information on Malaria and epidemiological information that can help in the formulation of effective malaria control policies. The HMCO is aimed at providing interoperability support for the knowledge management of malaria control initiatives, and serve as an open semantic web infrastructure for malaria research and treatment.

Introduction

Ontology is a formal explicit representation of the conceptualization of a domain that provides a platform for the sharing and reuse of knowledge across heterogeneous platforms. An ontology contains semantic descriptions of the features of a domain using concepts and relationship abstractions in a way that is readable by both man and machine. In recent times, the use of ontology have gained increasing relevance in the biomedical domain in that it enables researchers to stay abreast of current biomedical knowledge and promotes the understanding of such information. They also facilitate the sharing and reuse of biomedical knowledge across heterogeneous platforms for the delivery of medical services and implementation of health-related policies ¹.

Malaria is one of most worrisome vector-borne diseases that affect humans. It is caused by the parasite *Plasmodium falciparum*. Malaria is endemic in the tropical regions and sub-tropical regions of the world which are mainly in the South East Asia, Middle East Asia, Central and South America, and Sub-Sahara Africa. Millions of

malaria cases are reported each year, killing over one million per year in Sub-Sahara Africa². Generally, the control of a vector-borne disease such as malaria pose a critical challenge due to a number of reasons: 1) the complex nature of its transmission which involves three entities, which are the host (human), vector (female anopheles mosquito) and pathogens (*Plasmodium specie*); 2) the complicated epidemiology through the vector; and 3) social problems (poverty), geography, and resistance of pathogens to insecticides. All of these challenges compel the need to complement existing biomedical approaches of tackling the spread of vector-borne diseases with readily accessible, interoperable, and semantically-rich knowledge management support. This challenge motivated our pursuit of developing an ontology-based support for human malaria control. The Human Malaria Control Ontology (HMCO) contains information on human malaria that can be leveraged for the formulation of human malaria control initiatives in Sub-Sahara Africa. In terms of benefits the HCMO is expected to: 1) provide an interoperable platform for accessing malaria epidemiology information over the web; 2) provide information support for malaria control research and formulation of malaria control policy initiatives; and 3) Create an interoperable platform for the sharing and reuse of knowledge on malaria.

The outline of the rest of this paper is given as follows. In Section 2 an overview of related work on medical ontologies is presented. Section 3 is a short description of the design and implementation of the HMCO. Section 4 is a discussion of the possible application of the HMCO, while the paper is concluded in Section 5 with an outlook of future work.

Related Work

Medical Ontologies have played useful roles in facilitating the re-use, dissemination and sharing of patient information across disparate platforms. Also, they have been used in semantic-based statistical analysis of medical data. Examples of medical ontologies include GALEN³, UMLS⁴, MeSH⁵, ON9⁶, Tambis⁷, The Systematized Nomenclature of Medicine^{8,9}, Foundational Model of Anatomy⁹, MENELAS ontology⁹, Gene Ontology¹⁰ and LinKBase¹¹. The NBCO's Bio-portal¹² consist of more than 50 bio-ontologies that span several aspects of biomedicine including diseases, biological processes, plant, human, bio-medical resources etc. However, none of the ontologies in the bio-portal is specifically dedicated to malaria control. The work by Hadzic and Chang¹ was based on providing interoperability support for research in, and diagnosis of human disease using ontology-based approach. A prototype Generic Human Disease Ontology (GenDO) that contains common general information regarding human diseases was created which captured the information in four dimensions. However the dimension of diseases control was not included.

The Infectious Diseases Ontology (IDO)¹³ is designed to make infectious diseases-relevant data derived from different sources comparable and computable. It also provides coverage of entities that are common to many infectious diseases. The Vector-borne disease ontology¹⁴ is an ongoing project that is designed to provide an integrated interoperable platform for the sharing and reuse of knowledge about a group of vector-borne diseases in which the MalIDO (Malaria IDO) is a first step. The MalIDO incorporates several information dimensions such as gene models for A.gambie, Anatomy of mosquito, insecticide-resistance, and physiological processes of mosquito. As a contribution our work in the HMCO specifically focuses on the creation of an interoperable platform that gives access to epidemiological information on malaria in Sub-Sahara Africa that can be used for the formulation of malaria control policies.

Description of the HMCO

The HMCO captures information on human malaria in 7 dimensions. These are: (1) Malaria vectors (2) Malaria types, (3) Malaria

parasites, (4) Malaria Symptoms, (5) Malaria treatment (prevention, therapy), (6) Epidemiology data on malaria, and (7) Malaria Control. The design of HMCO was based on the Open Biomedical Ontologies (OBO) foundry principles¹⁵, while OBO foundry naming conventions were also adopted significantly in naming its concepts. It was implemented as an OWL ontology using the Protégé 3.4 Ontology tool. The conceptual taxonomy of the HMCO consists of 97 class abstractions that cover the seven dimensions of our interest. Nine disjoint subclasses comprising *vector*, *treatment*, *continent*, *type*, *parasite*, *epidemiology-info*, *symptom*, *year_data*, and *control* were modelled as constituents of the superclass *human_malaria* using 'belongTo' object property. The subclasses for the three classes: *symptom*, *treatment*, and *continent* were modelled as OWL value partitions (viz. each of the classes was represented as comprising disjointed subclasses that cover all known instances of each class). Concepts relationships among classes (concepts) in the HMCO class hierarchy were represented using object property abstractions that define the nature of association between the classes. These include associations between *parasite* and *vector* ('hasVector'), *type* and *symptom* ('hasSymptom'), *type* and *malaria_therapy* ('hasTherapyDrug'), *type* and *malaria_prevention* ('hasPreventionDrug'), *type* and *parasite* ('isCausedby'), *parasite* and *malaria_therapy* ('isCuredby'), *parasite* and *malaria_prevention* ('isPreventedby'), *vector* and *continent* ('isFrom'), *epidemiology_info* and *year_data* ('hasEpidemydata') etc. The class hierarchy and description of entities in the HMCO is shown in Table 1 (appendix). Also, specific object properties and datatype properties in the HMCO have appropriate cardinality restrictions imposed on them in order to effectively capture the semantics of relationships among the classes in the HMCO. Figure 1(appendix) is a view of the class hierarchy of the HMCO.

Application of the HMCO

Currently, the HMCO knowledgebase is being populated with available data (at present we have epidemiological data on malaria from the year 1999 to 2003 for 16 countries in

Sub-Saharan Africa), while the data gathering process is still ongoing. At the completion of the first version of the HMCO, it is expected to serve as a web-based repository for accessing epidemiology information on malaria in Sub-Saharan Africa. As a first step to attaining this, a prototype semantic web application has been built that can be used to query the HMCO knowledgebase.

The Java programming language implementation technology was engaged in building the prototype semantic web application using the NetBeans Java IDE. The Web GUI that facilitates client interaction with the HMCO knowledgebase was implemented using Macro Media Flash and Dream Weaver web design tools, and Java Server Pages (JSP). The business logic for querying the HMCO was implemented as an Enterprise Java Beans (EJB) component that is invoked from a Java Servlet class running on Sun Application Web Server 9.0. The EJB makes use of Protégé ontology Java APIs to access the HMCO knowledgebase for information retrieval. The Pellet 1.5 Descriptive Logics (DL) reasoner¹⁶ to facilitate semantic web reasoning (entailment, subsumption, and ABox reasoning) on the classes and individuals in the HMCO.

Conclusion

In this paper, a description of the Human Malaria Control Ontology (HMCO) has been presented. The HMCO offers as its contribution, an interoperable platform for accessing epidemiological information on malaria as viable knowledge management infrastructure for malaria control policy formulation and research in Sub-Saharan Africa. Though still an ongoing work, a preliminary test of the usability of the HMCO has been undertaken with promising results. Subsequently, the HMCO will incorporate other dimensions of information on malaria by importing relevant ontologies like the IDO and the Vector-borne diseases ontology (particularly the MalIDO) and will be presented for submission to the bio-ontology portal for open access and evaluation.

Acknowledgements

This research was undertaken under the Covenant University Postgraduate Research Award.

References

1. Hadzic, M. and Chang, E. Medical ontologies to support human disease research and control International Journal of Web and Grid Services. 2005; 1:139 – 150.
2. Snow R.W., Guerra C.A., Noor, A.M., Myint, H.Y., Hay, S.I. (2005). "The global distribution of clinical episodes of Plasmodium falciparum malaria". Nature 434 (7030): 214–7.
3. Rector, A., Solomon, W., Nowlan, W. and Rush, T. A Terminology Server for Medical Language and Medical Information Systems. Methods of Information in Medicine. 1995; 34:147-157.
4. Pisanelli, D., Gangemi, A. and Steve, G.. An Ontological Analysis of the UMLS Methathesaurus. Proceedings of AMIA Conference, 1998.
5. Stuart, N., Aronson, A., Doszkocs, T., Wilbur, J. Bodenreider, O., Chang, F., Mork, J., and McCray, A. Automated Assignment of Medical Subject Headings. Poster presentation at: AMIA Annual Symp., Washington, DC, 1999.
6. Pisanelli, D., Gangemi, A., and Steve, G., WWW-available Conceptual Integration of Medical Terminologies: the ONIONS Experience. Proceedings of AMIA Conference, 1997.
7. Stuckenschmidt, H., van Harmelen, F., Serafini, L., Bouquet, P., Giunchiglia, F. Using C-OWL for the Alignment and Merging of Medical Ontologies, proceedings of the First International Workshop on Formal Biomedical Knowledge Representation (Co-located with KR 2004), Whistler, Canada 2004.
8. SNOMED International, a division of the College of American Pathologists (CAP) (<http://www.snomed.org>).
9. Bodenreider, O, Burgun, A. "BIOMEDICAL ONTOLOGIES", Medical informatics: Advances in knowledge management and data mining in biomedicine. Springer-Verlag. 2005; 211- 236.

10. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genet. 2000; 25: 25- 29
11. LinkBase:
<http://www.landcglobal.com/pages/linkbase.php>
12. National Center for Biomedical Ontology Bio- Portal, <http://www.bioontology.org> (accessed 04/03/2009)
13. Infectious Diseases Ontology, <http://www.infectiousdiseaseontology.org> (accessed: 03/03/09).
14. Vector- Borne Disease ontology, http://www.infectiousdiseaseontology.org/IDO_Extensions.html (accessed 04/03/2009)
15. OBO Foundry Principles <http://www.obofoundry.org/crit.shtml> (accessed 04/03/2009)
16. Pellet DIG Reasoner, (available at: <http://pellet.owldl.com/>)

- human_malaria
 - control
 - drainage
 - mosquito_net
 - insecticide
 - epidemilogy_info
 - Drc_Data; Kenya_Data; Angola_Data; Nigeria_Data; Mozambique_Data; Namibia_Data; Ghana_Data; Sudan_Data; CAR_Data; Mali_Data; Uganda_Data; Malawi_Data; Tanzania_Data; Zambia_Data
 - Parasite
 - P. falciparum
 - P. ovale
 - P. malariae
 - P. vivax
 - Symptom
 - rigor
 - vomiting
 - convulsion
 - haemoglobinuria
 - anemia
 - fever
 - arthralgia
 - shivering
 - abnormal_posturing

- retina_damage
 - treatment
 - malaria_prevention
 - hydroxyxhloriquine;
 - chloroquine; proguanil;
 - doxycycline; mefloquine;
 - atavoquone- proguanil
 - malaria_therapy
 - artesunate- sulfadoxine- pyrimethamine;
 - artesunate- amodiaquine;
 - artesunate- mefloquine;
 - artemether- lumefantrine;
 - quinine; cotrifazid;
 - sulfadoxine- pyrimethamine; primaquine
 - type
 - severe_malaria
 - chronic_malaria
 - vector
 - A.gambiae; A.freeboni, A.culicifacios; A. fluviatilis; A. minimus; A. phillipinesis; A. stephensi; A. leucosphyrushave
 - Continent
 - Africa; Antarctica; Asia; Australia; Europe; North_America; South_America
 - year_data
 - _1990 ... _2008

Figure 1. A view of class hierarchy of the HMCO

Class name	Description	Subclasses
------------	-------------	------------

1	human_malaria	Main ontology superclass	control, epidemiology_info, parasite, symptom, treatment, type, vector, continent, year_data
2	control	Defines the concepts under malaria control	severe_malaria, chronic_malaria
3	epidemiology_info	Captures information on epidemiology data of 16 countries in Sub-Saharan Africa	Drc_Data, Kenya_Data, Angola_Data, Nigeria_Data, Mozambique_Data, Namibia_Data, Ghana_Data, Sudan_Data, Central African Republic (CAR)_Data, Mali_Data, Uganda_Data, Malawi_Data, Tanzania_Data, Zambia_Data
4	parasite	Defines the different types of human malaria pathogens	P. falciparum, P. ovale, P. malariae, P. vivax
5	symptom	Defines the different types of malaria symptoms as a set of disjointed classes.	rigor, vomiting, convulsion, haemoglobinuria, Anemia, Fever, arthralgia, shivering, abnormal_posturing (children), retina_damage (children)
6	treatment	Defines the types drugs for prevention and treatment of human malaria	malaria_prevention, malaria_therapy
7	type	Defines the types of human malaria	severe_malaria, chronic_malaria
8	vector	Defines the different types of human malaria vectors	A.gambiae, A.freeboni, A.culicifaciens, A. fluviatilis, A. minimus, A. philippinensis, A. stephensi, A. leucosphyrus
9	continent	Defines continental regions to which specific malaria vectors belong. Instances of this class the maps to a property of the vector class	Africa, Antarctica, Asia, Australia, Europe, North_America, South_America
10	year_data	Defines information on malaria endemics on a yearly basis for 19 years.	_1990 ... _2008

Table 1. An Overview of Classes in the HMCO