Hierarchical Conceptual Schema for Dengue Hemorrhagic Fever Ontology

Anisa Herdiani[1], Lily Fitria[2], Herika Hayurani[3], Wahyu C. Wibowo[4] and Saleha Sungkar[5]

**[1,2,3] Faculty of Information Technology, YARSI University
Jakarta, Indonesia**

**[4] Faculty of Computer Science, University of Indonesia
Depok, Indonesia**

**[5] Faculty of Medicine, University of Indonesia
Jakarta, Indonesia**

Abstract

Dengue is one of the most common infectious diseases and an enormous public health problem in Indonesia. In this paper, we discuss the development of hierarchical conceptual schema for Dengue Hemorrhagic Fever Ontology (DHFO) which contains general information on DHF and epidemiological information that can help in the formulation of effective DHF control policies in Indonesia. The DHFO is aimed at providing interoperability support for the knowledge management of DHF control initiatives, and serve as an open semantic web infrastructure for DHF research and treatment..

Keywords: *Dengue, Hierarchical Conceptual Schema, Knowledge Management, Ontology.*

1. Introduction

Ontology is an explicit specification of a conceptualization (Noy, 2001) that provides a platform for the sharing and reuse of knowledge across heterogeneous platforms. Ontology contains a coherent and interoperable suite of controlled structured representations of semantic descriptions of the domain's features using concepts and relationship abstractions so that it's 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 (Daramola, 2009).

Dengue is the most rapidly spreading mosquito-borne viral disease in the world. In the last 50 years, incidence has increased 30-fold with increasing geographic expansion to new countries and, in the present decade, from urban to rural settings (Fig. 1). An estimated 50 million dengue infections occur annually and approximately 2.5 billion people live in dengue endemic countries (WHO, 2009).

DHF pose a critical challenge due to a number of reasons: 1) the population awareness regarding to environmental cleanliness; 2) the complexity of dengue virus; 3) the complicated epidemiology through the vector; (4) lack of health education.

The problems prompted the need to complement existing biomedical approaches in an effort to control dengue fever by building ontology that provides knowledge management support for control of dengue. The goal of developing DHF ontology is (1) Provide an interoperable platform for accessing information on the epidemiology of DHF on the website (internet), (2) Provide information support for DHF control research and formulation of DHF control policy initiatives, (3) Provides interoperable platform for the sharing and reuse of knowledge related to dengue.

The outline of the rest of paper is given as follow. In the section 2 an overview of related research on medical ontologies is presented. Section 3 contains the methods used in developing the ontology. The fourth section contains a brief description of hierarchical conceptual schema for DHF ontology. The fifth section contains conclusions and future research plans.

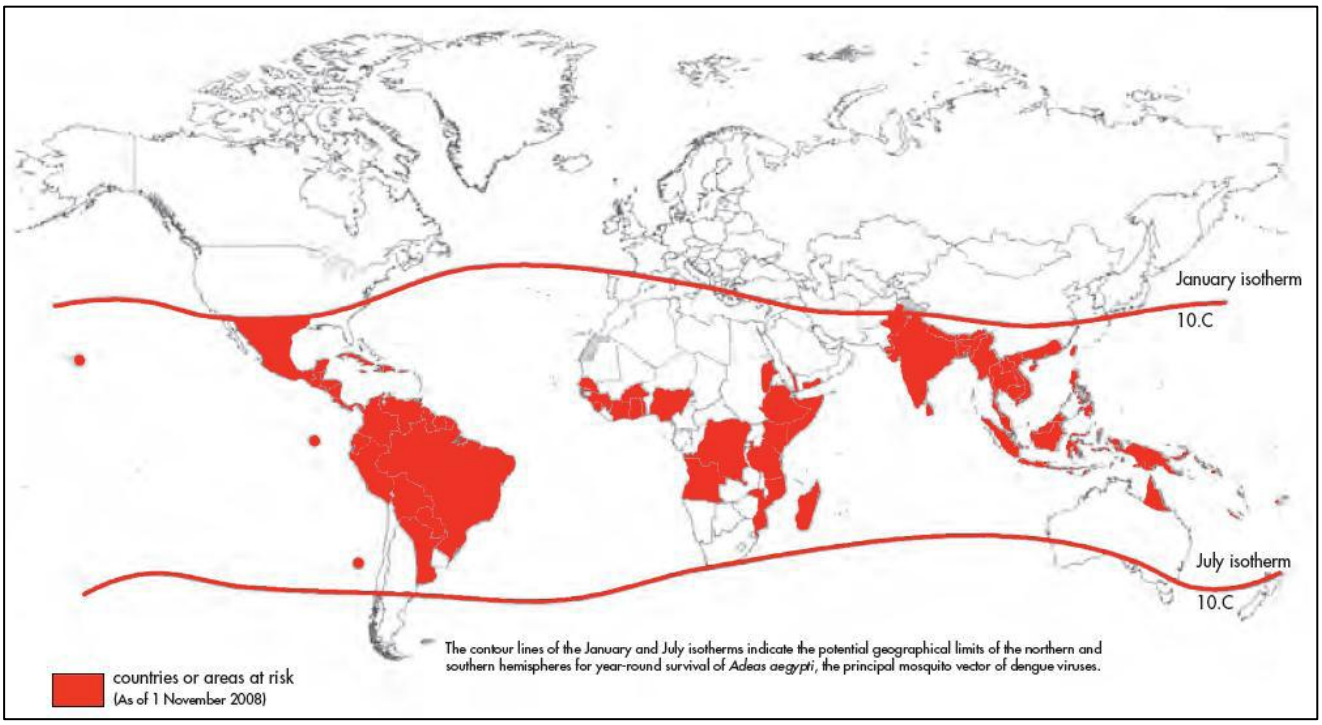

Fig. 1 Countries/areas at risk of dengue transmission (WHO, 2009)

2. Related Research

2.1 Vocabulary Resources

Medical vocabulary resources 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 (Daramola, 2009). Examples of medical vocabulary resources include (Cowell, 2010):

a. UMLS (Unified Medical Language System) Metathesaurus and Semantic Network (http://www.nlm.nih.gov/research/umls/) integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information systems and services, including electronic health records.
b. MeSH (Medical Subject Headings) vocabulary (http://www.nlm.nih.gov/mesh/meshhome.html), first published in 1954, is used to support literature indexing and document retrieval for the MEDLINE database of biomedical literature;
c. SNOMED (Systemized Nomenclature of Medicine), first released in 1965, was initially developed to support documentation of pathology data and is projected to become a worldwide reference vocabulary for structured clinical documentation;
d. International Classification of Diseases (ICD) (http://www.who.int/classifications/icd/en/), first published as the International List of Causes of Death in 1893, is the international standard for coding diagnostic information for health and vital records and is also commonly used for hospital billing purposes;
e. Gene Ontology (GO) (http://www.geneontology.org/), created in 1998, is a vocabulary resource for the annotation of gene and geneproduct data facilitating interoperability between a large number of diverse databases, especially in the domain of model organism research.

Vocabulary resources of this sort are standardly represented as graph-theoretical structures built up out of terms as the nodes of the graph and relations as edges. While there are a variety of other meanings associated with the term ‘ontology’, the usage here is consistent with that of large fluential ontology developer and user groups, including the Gene Ontology Consortium (http://www.geneontology.org/), the W3C community (http://www.w3.org/), and the OWL Web Ontology Language community (http://www.w3.org/2004/OWL) (Cowell, 2010).

2.2 National Center of Biomedical Ontology

The NBCO (National Center of Biomedical Ontology)"s Bio- portal (http://www.bioontology.org/ and http://www.bioontology.org/wiki/index.php/Main_Page) 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 DHF control.

2.3 Infectious Disease Ontology

The IDO ontologies are designed as a set of interoperable ontologies that will together provide coverage of the infectious disease domain (IDO, 2012). At the core of the set is a general Infectious Disease Ontology (IDO-Core) of entities relevant to both biomedical and clinical aspects of most infectious diseases. Sub-domain specific extensions of IDO-Core complete the set providing ontology coverage of entities relevant to specific pathogens or diseases. The sub-domain specific IDO extensions currently under development are:

a. IDO - Brucellosis
b. IDO - Dengue fever
c. IDO - infective endocarditis
d. IDO - influenza
e. IDO - malaria and other vector-borne diseases
f. IDO - Staphylococcus aureus
g. IDO - tuberculosis
h. IDO – Vaccines

The IDO ontologies are being developed in accordance with the principles of the Open Biomedical Ontologies (OBO) Foundry and with extensive use of its member ontologies. This approach ensures that IDO and its sub-domain-specific extensions have sufficient underlying formalism to support computational analyses and automated reasoning and that they are interoperable with other relevant biomedical and clinical ontologies, including those outside the domain of infectious diseases (IDO, 2012).

The DDSS project (http://www.rams-aid.org) is the most developed and best demonstrates the long-term potential of computing with ontologies. The goal of the DDSS is to guide the implementation of locally appropriate Dengue and Dengue vector control programs. The DDSS makes use of the Mosquito Insecticide Resistance Ontology (http://www.obofoundry.org/), the Vector Surveillance Ontology, the Vector Control Ontology, and the Dengue ontology (Cowell, 2010).

However, the work by RAMS-AID Research that develop ontology-based DDSS (Decision Dengue Support System) should be enhanced by the latest data that has been established by WHO in a publication titled "Dengue Guidelines for Diagnosis, Treatment, Prevention and Control" in (WHO, 2009). This publication defined Dengue – diagnosis, Dengue – therapy, Dengue - prevention and control, Endemic Diseases - prevention and control, Fluid therapy, Diagnosis, differential, Disease outbreaks - prevention and control, and Mosquito control.

3. Method

The methodology would include the ontology development life cycle that occurs during the development process, guidelines, principles that influence each stage of the life cycle. Development life cycles that are common for most ontologies: specification, knowledge acquisition, Implementation (includes Conceptualization-Integration-Encoding), and evaluation (see Fig. 2). Hierarchical conceptual schema resulted in the conceptualization phase, after the knowledge acquisition phase.

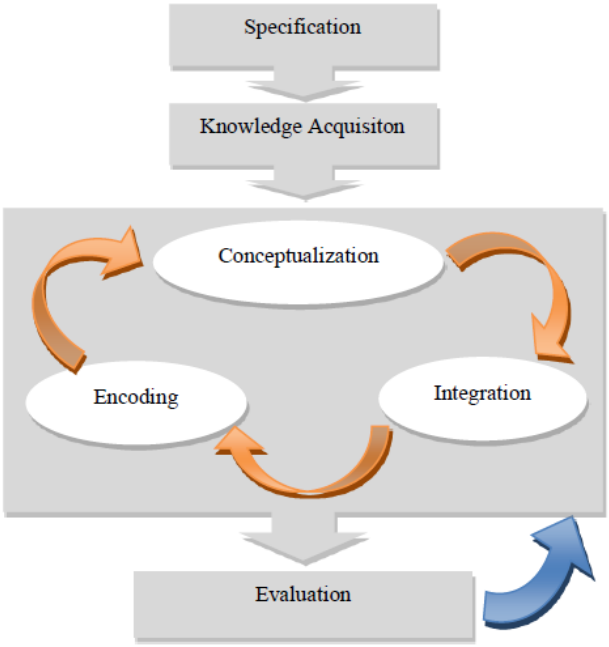

Fig. 2 Ontology Development Cycle

3.1 Specification

Specification is a phase where the purpose, scope and granularity of ontology are determined. This phase determines the type and coverage of data sources (databases, bibliographic information and reusable ontologies) needed to build ontology that supports a specific purpose, application or task.

The scope and granularity of DHF ontology described as follow:

1. The design of DHF ontology aligned to Infectious Disease Ontology (IDO) and based on the principles of Open Biomedical Ontologies (OBO) Foundry.
2. The Ontology implemented in OWL (Web Ontology Language)
3. The DHF ontology cover eight areas as follow:
   a. Dengue – diagnosis;
   b. Dengue – therapy;
   c. Dengue - prevention and control;
   d. Endemic Diseases - prevention and control;
   e. Fluid therapy;
   f. Diagnosis, differential;
   g. Disease outbreaks - prevention and control;
   h. Mosquito control.

3.2 Knowledge Acquisition

In the knowledge acquisition phase, domain knowledge were acquired from domain experts, database metadata, other ontologies and other re-usable information such text book information and research papers.

WHO has organized most of DHF knowledge in the Dengue Guidelines For Diagnosis, Treatment, Prevention And Control (New Edition, 2009). The guideline has covers eight areas mentioned in the specification phase.

DHF Ontology is in the domain of disease ontology that using medical vocabularies by Bioinformatics Core Facility in collaboration with NuGene Project at the Center for Genetic Medicine.

The vocabularies designed to facilitated the process of disease mapping and condition associated with particulars medical codes such as ICD9CM, SNOMED, and so on. In the development of DHF Ontology we consider to reuse existing ontology such as disease ontology (DO), Dublin Core (DC), Environment Ontology(EnvO), Foundational Model of Anatomy (FMA), Gazetteer (GAZ), Infectious Disease Ontology (IDO), Ontology for Biomedical Investigations (OBI), Ontology for Clinical Investigations (OCI), and Pathogen Transmission (TRANS) (see Table 1).

Table 1 : Reusable Ontology

Ontology	***Description***
Diseases Ontology (DO)	Human disease, DHF Ontology is a subset of this ontology
Dublin Core (DC)	Interoperable online standard metadata
Environment Ontology (EnvO)	Habitat and environment of an organism
Foundational Model of Anatomy (FMA)	Structure of mammals and part of human body.
Gazetteer (GAZ)	Geographic location, place, and name of place, and relationship among them.
Infectious Disease Ontology (IDO)	Biomedical and clinical aspect of infectious disease.
Ontology for Biomedical Investigations (OBI)	Design, protocol, instrumentation, and analysis that implemented in biomedical investigation.
Ontology for Clinical Investigations (OCI)	Clinical testing and related clinical study.
Pathogen Transmission (TRANS)	How pathogen transmitted from a host, reservoir or other sources, to another host.

4. Hierarchical Conceptual Schema

DHF Hierarchical Conceptual Schema arranged based on knowledge acquisition associated to DHF. The schema consists of 328 class abstraction that covers eight dimensions as follow: (1) Dengue – diagnosis, (2) Dengue – therapy, (3) Dengue - prevention and control, (4) Endemic Diseases - prevention and control, (5) Fluid therapy, (6) Diagnosis, differential, (7) Disease outbreaks - prevention and control, (8) Mosquito control.

Sixteen disjoint subclasses comprising `epidemiology_info`, `type`, `symptom`, `virus_serotype`, `vector`, `host`, `phase`, `treatment`, `control`, `advocacy`, `diagnostic_method`, `surveillance`, `planning_and_response`, `programme_assessment`, `vaccines`, and `year_data` were modelled as constituents of the superclass **`dengue`** using "`belongTo`" object property.

Concepts relationships among classes (concepts) in the DHFO class hierarchy were represented using object property abstractions that define the nature of association between the classes. These include associations between `virus_serotype` and `vector` ("`hasVector`"), `type` and `symptom` ("`hasSymptom`"), `type` and `treatment` ("`hasTreatment`"), `type` and `control` ("`hasControl`"), `type` and `virus` ("`isCausedby`"), `virus` and `treatment` ("`isCuredby`"), `virus` and `control` ("`isPreventedby`"), `vector` and

`continent` (“`isFrom`”), `epidemiology_info` and `year_data` (“`hasEpidemyData`”) etc.

Hierarchical conceptual schema can changes or evolves based on knowledge that has been acquired or by collaborate with other ontologies. Graph representation of DHF hierarchical conceptual schema can be seen at fig. 3.

Fig. 3 DHFO Hierarchical Conceptual Schema

5. Conclusion

This paper describes the development of hierarchical conceptual schema for dengue hemorrhagic fever ontology (DHFO). DHF Hierarchical Conceptual Schema arranged based on knowledge acquisition associated to DHF.

The goal of developing DHFO is (1) Provide an interoperable platform for accessing information on the epidemiology of DHF on the website (internet), (2) Provide information support for DHF control research and formulation of DHF control policy initiatives, (3) Provides interoperable platform for the sharing and reuse of knowledge related to dengue.

Furthermore, DHFO will integrate other DHF dimensions by importing relevant ontology such as IDO and will be submitted to the bio-ontology portal that can be accessed and evaluated by anyone.

Acknowledgments

This study is one of research roadmap of the Faculty of Information Technology YARSI University, and funded by the DIPA Directorate General of Higher Education Ministry of National Education through Grant named "Hibah Pekerti".

References

[1] Cowell, L. G, and Smith, B," Infectious Disease Ontology", in Vitali Sintchenko, Infectious Disease Informatics, New York: Springer, 2010, 373-395.
[2] Daramola, Olawande, F. Segun. 2009. Developing Ontology Support for Human Malaria Control Initiatives. Department of Computer and Information Sciences, Covenant University, Ota, Nigeria.
[3] Dengue Guidelines for Diagnosis, Treatment, Prevention and Control. New Edition. 2009. World Health Organization.
[4] Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genet. 2000; 25: 25-29.
[5] Infectious Disease Ontology, (1 June 2011), (http://www.infectiousdiseaseontology.org).
[6] National Center for Biomedical Ontology Bio-Portal, http://www.bioontology.org (diakses September 2011)
[7] OBO Foundry Principles, (1 June 2011), (http://www.obofoundry.org/crit.shtml).
[8] Sang, Low Hong. 2007. Knowledge Representation and Ontologies for Lipid and Lipidomics. Thesis Report. Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore.
[9] Vector-Borne Disease Ontology, (1 June 2011), (http://www.infectiousdiseaseontology.org/IDO_extensions.html)
[10] Gruber, T. R. 1993. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199-220, Stanford.
[11] Gruber, T. R. 1993. Toward Principles for the Design of Ontologies. International Workshop on Formal Ontology, March, Padova, Italy.
[12] Noy, NF., McGuinness, DL. 2001. Ontology Development 101: A Guide to Creating Your First Ontology. Knowledge Systems, AI Laboratory, Stanford University (KSL-01-05).

Anisa Herdiani holds BEng from School of Electrical Engineering and Informatics, Bandung Institute of Technology in 2006. She holds MEng from the same Institute in 2009. She is currently an academic, research staff, and Head of Data and Information Management Research Group at Faculty of Information Technology YARSI University. Her research interests are learning technology, consumer health informatics, and knowledge based systems. She is also a member of YARSI E-Health Research Center (YEHRC); has won some research grants; published a number of national and international papers in proceeding and journal.

Lily Fitria holds BSc and MSc from Faculty of Computer Science University of Indonesia in 2007. She is currently an academic and research staff of Faculty of Information Technology – Universitas YARSI. She is also a member of YARSI E-Health Research Center (YEHRC). Her research interests are information retrieval, and consumer health informatics.

Herika Hayurani holds BSc and MSc from Faculty of Computer Science University of Indonesia in 2008. She is currently an academic and research staff of Faculty of Information Technology – Universitas YARSI. She is also a member of YARSI E-Health Research Center (YEHRC). Her research interests are information retrieval, and e-Health.

Wahyu C. Wibowo graduated from of ITB (Bandung, Indonesia), Indiana University (Bloomington, IN, USA), and RMIT University (Melbourne, Australia). He is currently a lecturer at Computer Science Faculty, University of Indonesia. Prevously, he was working for University of Indonesia as the head of Information System Development Division in the Directorate of Information System Services and Development (PPSI). His research interests are knowledge representation and processing, knowledge extraction, and spatial databases

Saleha Sungkar is a Professor in Faculty if Medicine University of Indonesia. She graduated from University of Indonesia, Faculty of Medicine. Her research interests are parasitology and virus.