Building and Reusing Medical Ontology for Tropical Diseases Management

Haikal Satria, Radha S. Priya, Lukman H. Ismail, Eko Supriyanto

Abstract—Currently, the use of ontologies in medicine is mainly focused on the representation and re-organization of medical terminologies. Moreover, there are only a small number of ontologies available in the domain of tropical diseases. Therefore, it is important to build and reuse medical ontology for tropical diseases management, which behaves as an information support system to represent the information related to tropical diseases in a uniform and organized manner. The building process of ontology in tropical diseases domain is fully described in this paper and has been implemented online using Protégé 4.1 and OWL DL syntaxes. In order to provide easy access to user, ontology browser was used as visualization tool. Functional testing of ontology has been done with involving medical data collected from faculty of tropical medicine, Mahidol University. Test result shows that the developed ontology is able to provide structured information about tropical diseases that are very useful for healthcare professionals. Since it is an online system, it can be accessed by the user anytime, anywhere and offer support for a better medical decision and indirectly helping in the management of tropical diseases.

Keywords—Tropical diseases, medical ontology, ontology browser, structured information, OWL language

I. INTRODUCTION

CURRENTLY, health care sector is experiencing a major crisis with information overload [1]. For instance, there are plenty of information available regarding the medical field in various sources such as journals, books, newspaper articles and internet, which is unorganized and unmanageable. Therefore, it is difficult to determine a reliable and trusted source of information, as there are new discoveries in the medical field, often. Clearly, there is a demand for re-use and sharing of patient data, their transmission and the need of semantic-based criteria for purposive statistical aggregation (i.e. different aggregation criteria for different purposes) to satisfy the new desiderate of sharing and reuse of knowledge across heterogeneous platforms. Therefore, ontologies are suitable to support a more effective data and knowledge sharing in the medical field [15]. Moreover, the increasing number of diseases and as the signs and symptoms are varied, it is a challenge for medical practitioners to clearly remember and recall all of the information related to the disease. This can lead to misdiagnosis of the disease and puts the patient’s life at stake. Clearly, an ontology-based approach is required to represent entities, ideas and events, along with their properties and relations, as a form of knowledge representation about the medical world [2].

Ontology is defined as a formal explicit representation of conceptualization of a domain that provides a platform for the sharing and reuse of knowledge across heterogeneous platforms. It contains semantic descriptions using concepts and relationship abstractions in a way that is readable by both human and machine [17].

Tropical diseases are infectious diseases that are prevalent in the tropical and subtropical regions (which is rare) or, more difficult to prevent or control [7]. Tropical diseases are the disease of the poor, and investments in control and research to develop more effective intervention tools and strategies have been minimal. Regardless, for some, effective intervention methods have been developed, and successful control has been achieved. Environmental conditions that facilitate the transmission include climatic, social and economic factors. Besides that, the lack of resources prevents affected populations from obtaining effective prevention and adequate care [8].

According to the Malaysian Society of Parasitology and Tropical Medicine (MSPTM), the warm temperatures and high humidity of South East Asia region are favorable condition to many medically important insects and arthropods to reproduce themselves and maintain their colony. So do the bacteria, parasites and viruses. They are able to survive with this sort of environment and can cause various diseases to human being, which is commonly known as tropical diseases [9].

In the current situation, although there are plenty of ontologies available related to the medical field, it is focused on the representation and reorganization of medical terminologies [6]. In addition, there is only a small number of ontologies available focusing on the tropical diseases domain. Therefore, it is important to formulate a Medical Ontology for Tropical Disease Management System to represent the information related to tropical diseases in a uniform, organized manner that can be accessed by the user anytime and anywhere.

The main objective of this research is to develop a medical ontology in the domain of tropical diseases, which functions as an information support system that contains the entire database of information related to tropical diseases with the highest occurrences in the South East Asia region in a uniform, organized manner. Besides that, it is to be able to retrieve the ontology and answer queries of users in the domain of tropical diseases within one interface.
This study is focused on the tropical diseases that have the highest occurrences in the South East Asia region, which includes Dengue, Malaria, Chikungunya, Melioidosis and Leptospirosis. The source of information is mainly covered form latest data and research of tropical medicine field from Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand. In this system, the complex relationships that exist between the different medical concepts in the tropical diseases context such as causes, signs and symptoms, types of transmission, diagnostics, prevention and control, management and death causes is covered by adopting an ontology-based approach. The users of this system include healthcare professionals such as medical physicians and nurses, medical surveillances, medical students, travelers and general patients.

II. LITERATURE REVIEW

A. Clinical Ontologies

Current developments in information technology and communication offer new opportunities in the implementation of high quality healthcare systems. These technologies ensure the support for better medical data processing, diagnoses that are more accurate and easier access to medical services [5]. Paper-based terminology systems cannot satisfy anymore the new desiderate of healthcare information system, as there is demand for re-use and sharing of patient data, their transmission and the need of semantic-based criteria for purposive statistical aggregation. An important feature of medical information system is the unambiguous communication of a complex and detailed medical concept. Therefore, ontologies are suitable to support a more effective data and knowledge sharing in medicine [15].

However, it is not an easy task to be achieved and a deep analysis of the structure and the concepts of medical terminologies are required. Such analyses can be carried out by adopting an ontological approach for representing medical terminology systems and for integrating them in a medical ontology [16].

Currently, the use of ontology have gained increasing relevance in the biomedical domain as 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 [17].

Medical ontologies have played important roles in facilitating the re-use, dissemination and sharing of patient information across disparate platforms. Besides that, they have also been used in semantic-based analysis of medical data. Examples of medical ontologies include GALEN, UMLS, MeSH, ON9, Tambis, The Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT), GALEN, MENELAS, ONIONS library or the highly complex UMLS. These ontologies were used to allow a standard, accurate exchange of data content between different systems and providers [5].

III. MATERIAL AND METHODS

First of all, the ontology is developed using Protégé 4.1 Beta software. Fig. 1 illustrates the step by step for ontology development.

![Workflow of Tropical Diseases Ontology Development](image)

**Fig. 1 Workflow of Tropical Diseases Ontology Development**

It is important to determine the scope of the ontology such as the domain that the ontology will cover, the usage of the ontology, the types of queries that the ontology should be able to answer and the users of the ontology. In the case of tropical disease ontology it was clear that the need on foundation inclusively for tropical disease to unify the information. This unification could maximize the availability of knowledge to construct decision support tools and provides common terminology for tropical disease. Therefore, we decided to develop global tropical disease ontology with prerequisites:

1) The ontology should be able to provides interoperability among current terminologies
2) The ontology should be an initial version to provide further continuous development.

The proposed tropical ontology not only covered medical process and background aspects, but also medical practice and management of disease to provide common clinical practice guidelines on controlling tropical disease.

A. Method

Protégé as is an open access and free ontology editor application. It was developed by Stanford Medical Informatics
group with the funding from various US government agencies in the past 15 years, which is now a core technology of the National Center for Biomedical Ontology [11]. Protégé have a community of about 50,000 users which includes research and industrial projects in more than 100 countries. It was initially developed to represent frame-based ontologies, in accordance with the Open Knowledge Base Connectivity (OKBC) protocol. Protégé has evolved, in collaboration with the University of Manchester, to represent ontologies in the OWL based on description logics.

Protégé allows the support of OWL and support for exporting Protégé ontologies into a variety of other formats such as RDF/S, OWL and XML Schema. Besides that, it has 69 plug-ins, which includes ontology visualization (OntoViz), ontology alignment (PROMPT) and interfaces with rule engines and formalisms such as SWRL (Semantic Web Rule Language).

Web based visualization and navigation of tropical disease ontology was done by using ontology browser library [30]. Ontology Browser behaves as the web interface, which allows the user to browse OWL Ontology and RDF (Linked Open Data) files. Ontology Browser functions in an Apache Tomcat 5.5 Server environment. Therefore, Apache Tomcat 5.5 server was installed.

B. Reuse Consideration

Various medical vocabulary resources and specific ontologies have been developed in the last decade. These independent and different perspectives create difficulties to communicate among systems. Review on common medical vocabulary resources has been done in [16]. ICD-10 as de-jure vocabulary standard from WHO has been used widely in term public health services. Specific purpose on diseases ontologies is shown in Table 1. Some parts of projects provide specific association to the available tropical disease.

<table>
<thead>
<tr>
<th>Project</th>
<th>Purpose</th>
</tr>
</thead>
</table>
| Infectious Disease Ontology (IDO) [18] | • Dengue Decision Support System (DDSS).  
• Common physiological processes of infectious diseases, covers malaria and dengue. |
| Human Malaria Control Ontology (HMCO) [20] | • Interoperability support for the knowledge management of malaria control initiatives  
• Open semantic web infrastructure for malaria research and treatment. |
| OBO Foundry - DO [38] | • Open semantic web infrastructure for human diseases. |

The medical ontology by Hadzic and Chang (2005) was purely based on providing interoperability support for research and diagnosis of human disease [18]. Besides that, a prototype Generic Human Disease Ontology (GenDO) that provides common general information regarding human disease was created. This information was provided in four dimensions. However, the dimension of disease control was excluded [26].

Other than that, the Infectious Disease Ontology (IDO) was designed to make infectious diseases-relevant data derived from different sources, which are comparable and computable. In addition, it also provides coverage of entities that are common to many infectious diseases [25]. The Vector-borne Disease Ontology 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 disease. In addition, MalIDO (Malaria IDO) is a first step. It incorporates several information dimensions such as gene models for A.gambia, Anatomy of mosquito, insecticide-resistance and physiological processes of mosquito [20].

Several tropical disease domains are relevant to these various disease ontologies. Reusing terms and reasoning from this vocabulary and specific ontology for disease could reduce strenuous effort in the process of candidate terms and its relations.

Specific ontologies were developed by aid of locally database schemes and vocabulary resources. Such database schemes and vocabulary resources could give ambiguities and overlapping resources when it reuse. Mapping into standard and universal reference such as ICD was needed to provide formal terminology and interoperability.

Next step was formulation of candidate terms. It was separated in terms of noun, which is basis for class names and verbs, which is basis for property names. Based on these terms, the relevance terms were organized in taxonomic (subclass) hierarchy as shown in Fig. 2, which all nouns were defined as classes. OWL classes were known as sets that contain individuals.

C. Source of candidate terminology

The complex knowledge structure of tropical disease should include clinical manifestation, therapeutic approach and epidemiology. The main objectives of tropical disease ontology also covered on wide disease treatment, control and prevention to drive health policy strategy based on each particular disease. Scope and domain knowledge of information was based on provided data and discoveries from Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand.

Formalization of domain knowledge was harvested from indexed articles, classification naming and terminology from World Health Organization as well as public health domain textbooks.

It has been considered that determination candidate was continuous and repetitive process and not persistence through time. Parthood process between terms has followed occurrences categories where phases and temporal parts which correspond to intervals through which they produce [38]. Entities candidate of initial tropical disease considered objective aspects. These aspects were:

a. Tropical disease should covers highest occurrence diseases and represents a diagnosis

b. Tropical diseases pathology should accompany pathological process
c. Symptoms and sign of suffering individuals from each disease as well as post severity classification and pre/post stage
d. Investigation, risk factors, prevention, and treatment support for individuals suffers from tropical disease
e. Use of evidence-based medicine based on clinical Malaysia practice guidelines provided to manage and control tropical disease

D. Taxonomy Definition

The high level graphical presentation of tropical disease ontology is shown in figure 2. Analysis of evidence-based from main source and reuse of available disease terminology gives 10 selected term candidates. These terms, were identified and written down in an unstructured list and showed as summary representation.

Recognized candidates for this high stage are needed to be normalized and grounded to the canonical names before instantiation of their individual terms. This process was done by overlook synonym terms then changes the label name with the simplest name.

All selected candidates then reorganized to describes their structures based on the high level terminology. As in figure 3, this structure creates a taxonomical hierarchy, organized by super-sub concept relationship. Individuals were ground term instance from concept object. As shown in figure 4, individual was an object perceived from human semantic. Inheritance relationship process for each individual was determined by property similarity based on main source and reuse consideration. Relationship constrains between distinct individual-level has been considered in the next stage.

The computable terminology subset in this tropical disease ontology constitutes a decidable fragment of first-order logic, which is supported by reasoner, such as HermiT. This reasoned are able to test whether the ontology contains contra dictionary assertion, a sub of another concept or whether the concept is satisfiable. This test makes it possible for a reasoner to compute the inferred ontology concept hierarchy. Besides that, it also does consistency checking. For instance, based on the description (conditions) of a concept, the reasoner can check whether or not it is possible for the concept to have any individuals.

The reasoned used in our method is Hermit 1.3.3. Once the inferred hierarchy has been computed, an inferred hierarchy window will pop open on top the existing asserted hierarchy window. If there are inconsistencies present in the inferred hierarchy, its icon will be highlighted in red [31].

As mentioned above, choosing the right candidates for tropical disease term concept or individual should consider all objective aspects. Within this initial knowledge acquisition, the resulting of baseline tropical disease ontology consists of 235 concepts/individuals with a deep up to four levels. These entities have sufficient enough to follow all objective aspects.

E. Tropical Disease Relation

The construction of tropical disease ontology should obey formal ontological principles to prevent interpretation bias on each individual terminology. In this stage, relationship constrains between each individual terminology were developed. Constrains relationship in tropical disease is based on Mereological relations and are applied between terminologies. This constrain relationship is also inherited relationship, which means relationship are also defined in their individuals and concepts. This relationship type is sufficient for the purpose of indexing and annotating initial tropical disease ontology, especially for clinical coding purposes. Focus of individual-relations of this initial phase is on Named_Tropical_Disease concepts to others listed concepts.

The relationship properties as shown in table 2 represent the relationships property between two individuals. Constrains of this property is a value constrain, restricted into their concepts
descriptions or concepts name then added their declarative hypernym/hyponym (is-a) and meronym/holonym (has-of).

Table 2: Summary of tropical disease property

<table>
<thead>
<tr>
<th>Properties</th>
<th>Inverse</th>
<th>Individual Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasCause</td>
<td>isCauseOf</td>
<td>22</td>
</tr>
<tr>
<td>hasControlStrategyOf</td>
<td>isControlStrategyOf</td>
<td>3</td>
</tr>
<tr>
<td>hasDeathCauseOf</td>
<td>isDeathCauseOf</td>
<td>4</td>
</tr>
<tr>
<td>hasGradingSeverity</td>
<td>isGradingSeverityOf</td>
<td>5</td>
</tr>
<tr>
<td>hasLaboratoryFindingOf</td>
<td>isLaboratoryFindingOf</td>
<td>12</td>
</tr>
<tr>
<td>hasManagement</td>
<td>isManagementOf</td>
<td>11</td>
</tr>
<tr>
<td>hasModeOfTransmission</td>
<td>isModeOfTransmissionOf</td>
<td>16</td>
</tr>
<tr>
<td>hasPositiveResultOf</td>
<td>withPositiveResultIndicates</td>
<td>2</td>
</tr>
<tr>
<td>hasRiskFactor</td>
<td>isRiskFactorOf</td>
<td>7</td>
</tr>
<tr>
<td>hasSignAndSymptomOf</td>
<td>isSignAndSymptomOf</td>
<td>62</td>
</tr>
<tr>
<td>hasStage</td>
<td>isStageOf</td>
<td>1</td>
</tr>
<tr>
<td>hasTransmissionTypeOf</td>
<td>isTransmissionTypeOf</td>
<td>4</td>
</tr>
<tr>
<td>hasType</td>
<td>isTypeOf</td>
<td>6</td>
</tr>
<tr>
<td>hasVector</td>
<td>isVectorOf</td>
<td>4</td>
</tr>
<tr>
<td>isToDiagnose</td>
<td>isDiagnosedBy</td>
<td>1</td>
</tr>
</tbody>
</table>

DL language syntax is used to describe classification and scoring process for inter-individual level relationship of tropical disease ontology. Table 3 shows the restriction and Boolean concept syntaxes between description logic and properties of its Manchester OWL. There are only two types of restrictions used in this research project, some and only. Keyword “some” or existential restrictions describes classes of individuals that participate in at least one relationship along a specified property to individuals that are members of a specified class. Only or universal restriction is used to describe classes of individuals that for a given property only have relationships along this property to individuals that are members of a specified class.

Figure 5 shows the snapshot of inter individual level relations in Chikungunya individual of tropical disease ontology.

Table 3 : Description logic syntax and its axiom

<table>
<thead>
<tr>
<th>Description Logic</th>
<th>Manchester Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restrictions</td>
<td></td>
</tr>
<tr>
<td>some</td>
<td>∃</td>
</tr>
<tr>
<td>only</td>
<td>∀</td>
</tr>
<tr>
<td>value</td>
<td>∃</td>
</tr>
<tr>
<td>min</td>
<td>≥</td>
</tr>
<tr>
<td>exactly</td>
<td>=</td>
</tr>
<tr>
<td>max</td>
<td>≤</td>
</tr>
<tr>
<td>Boolean Concept Constructors</td>
<td>⊓</td>
</tr>
<tr>
<td>and</td>
<td></td>
</tr>
</tbody>
</table>

Last stage on building tropical disease was adding annotation. Annotation was added to provide information (metadata – data about data) to classes, individuals and object/datatype properties. Tropical disease ontology provides for individual annotation both from World Health Organization terminology and Infectious Disease Ontology descriptions.

Overall process stage of formal ontology has been done to provide tropical disease ontology. This initial tropical disease ontology provides 32 relationship properties with 54 individual annotations. Figure 6 shows overall structures of tropical disease using OWLViz tool in Protégé 4.1.

F. Ontology Visualization

Based on our method, visualization is done by using Ontology Browser. Ontology Browser required the ontology based on OWL file to be based on a domain address. This needs an upload process the OWL file to a web-hosting provider such as Limeweb as it is important for the file to be
publicly accessible for the server to find and browse them. Fig. 7 illustrates the file manager in the web hosting provider, Limewebs, which contains the TropicalDisease1.owl file.

Fig. 7 File manager in the web hosting provider, Limewebs

Next, the Ontology Browser war file is downloaded and renamed to “browser.war”. This folder is copied into the CATALINE_HOME/webapps/folder. After that, Apache Tomcat 5.5 Server service is started and the Ontology Browser can be accessed off the root of the tomcat server (http://127.0.0.1:8080/browser/). Fig. 8 illustrates the interface of the Ontology Browser.

Fig. 8 The interface of the Ontology Browser

III. RESULT ANALYSIS

Testing was carried on the functions in the system, and the databases created. The testing was conducted to ensure the functionality of the system such as the ontology retrieval, the expandable class hierarchies, interrelationships, and the ability of the system to find entities and answer queries of the user by using auto complete and dropdown.

In the Ontology Browser, it can be seen that there are eight tabs across the top of the screen that are ontologies, classes, object properties, annotation properties, individuals, data types and DL query, available and can be browsed except for one tab, which is faded out that is data properties since it has not been defined in the ontology development process as shown in Fig. 8.
Fig. 9 shows that the tree can be expanded and collapsed as required by using the + sign on the left of the entities. When a particular entity is selected, a new page will automatically appear to show the current selection. The current selection can be identified easily as it is highlighted in yellow. In the right hand side, the super classes show the interrelationship that have been described or defined in which the font in turquoise represents the object properties and purple represents the restrictions. The Disjoints title represents an individual that cannot be an instance of more than one of these classes. The usage title represents the usage of the entity throughout the ontology.

As shown in Fig. 10, it is possible to click the classes that are the font in black under the Superclasses, Disjoints and Usage title as it is linked using inverse properties, as well.

Upon clicking the object properties tab, the object properties defined in the ontology appears as shown in Fig. 11. The object properties are in turquoise font. In the other hand, when the individuals tab is clicked, the 44 individuals that are defined in the ontology appear as shown in Fig. 12.

When the annotation properties tab is clicked, the 53 annotation properties that are defined in the ontology appears as shown in Fig. 13. The annotation properties are in green font.
When the classes that is the font in black is clicked, a new page prunes out to provide full information related to the class including the annotations, superclasses and usage as shown in Fig. 14.

In figure 15, it can be seen that there is an alternative to do a search that is by using the search box at the top right of the screen and start typing the name of the entity that the user is looking for. The search allows auto complete and dropdown is implemented to help the user to find the entity that they are interested in.

Based on the functional testing, it can be seen that the system shows a successful result. It can be seen that the whole ontology can be retrieved by using the seven tabs across the top. Besides that, the classes are expandable to show its class taxonomies (subclass) hierarchies. In addition, all the objects and properties in the ontology are interrelated appropriately and lead to the correct one when selected. Finally, besides searching by clicking and expanding the classes taxonomy, the alternative search box can be used to find entities and answer queries of the user using auto complete and dropdown in a fast, efficient manner. This enable the system to provide information related to tropical diseases in a uniform, organized manner. Therefore, the system fulfills the objective of the research project.

In addition, a pilot study was developed and questionnaires are distributed to three medical doctors, two nurses, five medical students and twenty Biomedical Engineering students to obtain response towards the system in
several scopes such as user friendly, efficiency and effectiveness of the system as an information support and future use of the system. Fig. 17 illustrates the results of the pilot study.

![Fig. 17 Results of the pilot study to obtain user feedback](image)

Based on the overall results, it can be seen that the system has obtained a positive user feedback as all the respondents have rated it as 4 and 5 that is agree and most agree only.

IV. CONCLUSION

A new medical ontology for tropical diseases management has been built with reuse of existing ontology tools and medical data. This work was done as an effort to help medical practitioner and society to manage the tropical diseases epidemiology within South East Asia region. New terminologies, taxonomical hierarchy, and medical data interrelationship have been formulated in this work and implemented in ontology server. A comprehensive tropical diseases data with 53 annotation properties have been implemented to provide useful structured information for user. Performance test result shows that the built ontology system is user friendly, enough information, and recommended to be used as reference for tropical diseases management.

ACKNOWLEDGMENT

The authors would like to express our appreciation to Faculty of Tropical Medicine, Mahidol University for contributing information related to tropical diseases existent and emerging in the South East Asia region.

Our appreciation also goes to the advanced diagnistics and progressive human care research group members for their ideas and comments on this paper.

REFERENCES


[34] Maja Hadzic, Elizabeth Chang, “Ontology-based Multi-agent Systems Support Human Disease Study and Control”


Haikal Satria is a computer engineer. He has been working for four years in the diagnostics research group, biotechnology research alliance, Universiti Teknologi Malaysia. He obtained his master in computer engineering from Duisburg-Essen University, Germany. His research interest is computer programming and telehealth. He has some international publications in the area of computer and telehealth.

Prof. Supriyanto is a biomedical engineer. He is head of diagnostics research group, Universiti Teknologi Malaysia. He obtained his PhD in medical electronics from University of Federal Armed Forces, Hamburg, Germany. His research interest is engineering application in medicine. He has more than 100 international publications in the area of medical electronics, medical image processing and medical computing. He has 18 national and international innovation awards and more than 10 patents of biomedical products.