Ontology Based Brucella Representation using Formal Concept Analysis

A. Velmurugan[a] and T. Ravi[b]

[a]Research Scholar, Department of Computer Science and Engineering, Sathyabama University, Chennai, India
[b]Professor, Department of Computer Science and Engineering, S.A. Engineering College, Chennai, India

Abstract: Brucellosis is a nationally noticeable disease. Many animal species are stricken, but internal pets are the most common source of infection to human. Brucellosis is still an uncontrolled severe health problem in many emerging countries including India. Brucellosis in India is very common, but yet often ignored disease. Distinct level of host Brucella interaction based on distinct host cell types and Brucella strains was defined. To systematically represent and analyze the host Brucella interaction, there is a need of ontology. The brucella ontology represents the species of brucella, brucella interaction with the host and a brucella pathogenesis mechanism. In this paper, formal concept analysis is used to extract Objects (G), Attributes (M) and Incidence relation (I) between G and M for data analysis and knowledge representation. By using these Objects, Attributes and Incidence relation, we propose a modeling brucella domain knowledge using Web Ontology Language (OWL), which will be encoded with Protégé to provide comprehensive information. The brucella ontology covers classes, properties, instances and a restriction on relationships and other description. Finally, HermiT reasoner used to check inconsistency and the SPARQL queries are used to attain required information.

Keywords: Ontology, brucella interaction, Lattice, formal concept analysis, OWL.

1. INTRODUCTION

Development of IDO set of ontologies is a community-wide effort concentrated on the development of a set of interoperable ontologies that together deliver coverage of the infectious disease domain. In recent years the benefits of ontology for the management, interaction, and processing of data and information have been realized, resulting in the surge of interest in the ontology within the biomedical research community and the NIH. IDOBRU ontology is in the domain of brucella and brucellosis. In the field of infectious disease, the study of interaction between pathogen and host is critically important. Brucella is a genus or gram negative, non-spore-forming, facultative, nonmotile, intracellular bacterium that causes chronic zoonotic brucellosis in human and variety of animal species. The different species of brucella are hereditarily much related, although each has slightly diverse host specificity.

In this session, some of the related concepts regarding Brucella interaction are defined. The comprehensive review of the omics [7] and Bioinformatics technologies explain about the Examination of brucella gene expression profiles inside host cells, Macrophage resistant responses to brucella infections founded on omics gene expression data examination, Brucella vaccine target forecast based on genome sequence analysis using vaxign reverse Vaccinology and Ontology based analysis of the manner of development of brucella disease, host immunity and vaccine targets. However, the translation of knowledge learned from the system and Bioinformatics studies into the generation of new vaccine and drugs infectious brucella infection, improvement is still challenging. The new global map of human brucellosis [2] depicts the incidence of human brucellosis worldwide with the details of incidence (annual cases per million of population). The following areas are currently listed as a high risk area: the Mediterranean Basin, Mexico, South and Central America, Eastern Europe, Asia, Africa, the Caribbean and the Middle East.

Literature based brucella gene annotation [3] would increase understanding of the complicated brucella pathogenesis mechanism. Brucella Bioinformatics Portal (BBP) is an entry for brucella researchers and it is developed to integrate existing brucella genome data and examination tools with literature mining and curation. Brucellosis Ontology [4] is an extensible ontology of the central core infectious disease ontology (IDO-core) and follows OBO foundry principles. IDOBRU is a first reported bacterial infectious malady ontology developed to represent a different disease aspect in a formal logical format. It serves as a brucellosis knowledge base and support brucellosis data integration and automated reasoning.

However, IDOBRU is still at the primary stage of its development as a knowledge base, the information about brucellosis in IDOBRU is still limited. Realistic approach [5] captures negative findings in an electronic health record system based on the BFO and RT. The classified phenomena described in terms of the various top-level classes and relations defined in BFO, taking into account the role of denial in the consistent descriptions. This model and intentional operators require second order treatment whose conversation falls beyond the scope of this paper. For the corresponding negative expressions do not mirror anything a reality, and to assume that they do is a confusion of the way language works.

A method [6] to provide definition patterns in relation among classes using OWL and describe a novel implementation of the RO based on this method. It provides the means to slowly move the representation of biomedical ontologies into formal knowledge representation language that incorporates an explicit semantics. This system implements a generic parser for the OBO flat file format. It represents the OBO ontologies as a list of ontological classes and a list of related types. To convert this representation to OWL, they use the Manchester OWL API. Cell type ontology is used for the evaluation. However, the performance of extracting relational patterns from OWL using automate designing. To make our approach is widely applicable within the semantic community. More sophisticated algorithms must be investing in to extract relational pattern assertion between classes.

OGG: a biological ontology [1] of genes and genomes to incorporate with other ontologies, OGG is aligned with the basic formal ontology (BFO). Each gene in OGG has over 10 explanation items, includes gene-associated gene ontology (GO) and PubMed article information. A URI dereferencing is defined as an act of recovering a representation of a resource blentified by a uniform resource identifier (URI). Since, different OGG OWL files exist and all OGG subsets use the same OGG namespace, for a given OGG team RI, Ontobee was not being able to identify which OGG OWL file to use for the URI dereferencing[14].

2. METHODOLOGY

The brucellosis required information is obtained from many sources on the web. Different level of host brucella interaction based on diverse host cell types and brucella strains was first defined ontologically. In this paper,

brucella ontology represents the species of brucellosis, signs and symptoms for brucellosis, transmission medium, spreading medium, diseases from brucellosis to human, brucellosis testing, brucellosis treatment and etc.

Traditionally, the discussion of words to concepts has been performed using a lexicon and for computational technique the tendency is to do the same. The lexicons used are either specially created for the task or a pre-existing language prototype, usually associated to Princeton's word net. The mapping of words to concepts is often ambiguous. Typically each word in a given language will relate to several concepts. Human uses the context to disambiguate the several meanings of a given piece of text where obtainable machine translation system cannot easily infer context. For the purpose of context mining, however, this ambiguity tends to less important than they are with machine translation for in large document the ambiguities tend to even, much as is the case with text mining.

Formal Concept Analysis (FCA)

Formal concept analysis is a way of deriving a concept hierarchy or formal ontology from a group of object and their properties. Each concept in the hierarchy denotes the set of object sharing the same values for a certain set of properties; and each sub concept of the objects in the concept above it. Formal concept analysis is a mathematical theory of data examination using formal contexts and concept lattices to knowledge processing. In this paper formal concept analysis is used to obtain conceptual knowledge discovery.

A formal concept analysis [13] is defined as below

Definition 1: A formal context is a tripartite (G, M, I) where G is a set of elements called object, M is a set of elements called attributes and I is a binary relation between G and M (i.e., $I \subseteq G \times M$). $(g, m) \in I$: to say that object g has attribute M. In this paper, Objects are classes, Attributes are Properties and Binary relations are individuals.

Definition 2: Formal concept of context object, attributes and binary relation (G, M, I) is a pair (A, B) with $A \subseteq G$, $B \subseteq M$, $A' = B$, $B' = A$ where if $A \subseteq G$ we define $A' = \{m \subseteq M \mid \forall g \in A: (g, m) \in I\}$ and if $B \subseteq G$ we define $B' = \{g \in G \mid \forall m \in M: (g, m) \in I\}$.

Definition 3: Sub-super concept, relation: $(A1, B1) \leq (A2, B2)$ $A1 \Leftrightarrow A2$ ($B1 \subseteq B2$ respectively) is defined as (A2, B2) is a super concept of (A1, B1).

For ontology, Classes are considered as Objects (G), Properties are considered as Attributes (M) and Individuals are considered as Incidence relation. In this brucella ontology, Objects are {Brucella species, Common reservoirs, Human disease, Signs and symptoms, Transmission, Treatment and etc.}, Attributes are {also_known_as, has, includes, contain, by, term_definition, entering_medium and etc}, Incidence relations are {anorexia, B_abortus, B_ovis, cattle, bone_marrow, depression, chronic_fatigue, rubber_gloves, computerized_ tomography_scan, bacteria_entering_the_body_through_membranes, malta_fever, pain_in_muscle}.

Calculating Formal Concepts

The formal concepts are calculated by using the derivation operator; we derive the formal concepts with the following procedure:

    (i)   Pick an object set (A)

    (ii)   Derive the attributes (A′)

    (iii) Derive (A′)′

    (iv) (A″, A′) is a formal concept

In this paper, Brucella ontology contains many objects. Each object includes their set of individuals as Object set. Features of objects are considered as attributes. Each object has attributes which may be a specific or common attributes.

For example,

For A: Brucella_species is an object and it contain a set of elements of {B_melitensis, B_abortus, B_canis, B_ovis, B_neotomae, B_pinnipedialis, B_ceti, B_microti, and B_inopinata}.

For A': Brucella_species object containing the attributes of {spread, occur common source of infection, contain human disease, includes common reservoirs, happen transmission, etiologic agent, at risk, and has testing}. Each attribute has their set of individuals.

For (A')': {contain human disease, etiologic agent, includes common reservoir, occur common source of infection, has testing}' = {B_melitensis, B_abortus, B_suis, B_canis}.

For (A'', A'): ({B_melitensis, B_abortus, B_suis, B_canis} {contain human disease, etiologic agent, include common reservoir, occur common source of infection, has testing}.

Lattice

Lattice concept is a method of transforming each formal concept into structure. Lattice context is used for Partial Order Classification as well as for separation of subset and super set concepts, which is derived from formal context of formal concept analysis. In Lattice context, the meet and join operation satisfy the axioms defining a lattice.
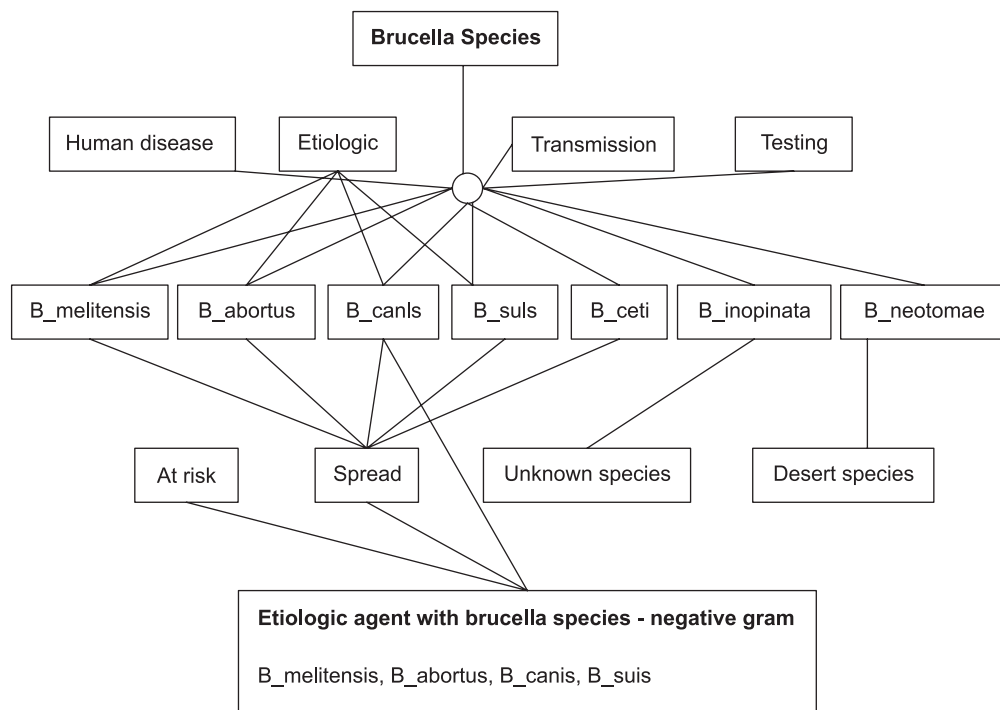

Figure 1: Lattice structure of brucella species. It produce output of etiologic output with brucella species – negative gram

The lattice structure of Brucella species classifies all the species and produces output of etiologic agent with brucella species – negative grams.

3. ONTOLOGY DEVELOPMENT

In this paper, formal concept analysis is used to extract knowledge in the form of Object, Attributes and Incidence relation by using context mining technique. After that, formal concepts are calculated by using some procedures. The procedures are described in the above section. By using the formal concepts, we propose a model for domain knowledge using Web Ontology Language (OWL), which will be encoded with protégé to provide comprehensive information with user friendly navigation. For brucella ontology development contain the following session:

    A.   Specification of requirement

    B.   Conceptualization of domain knowledge

    C.   Formalization

A. Specification of Requirement

The aim of this session is to obtain purpose of the ontology to the users. The main goal of the brucella ontology is to provide formalized knowledge for diagnosis process and to share a common understanding of structure of common understanding among people. Brucella ontology information is obtained from Formal concept analysis and web resources. Identification of important and salient terms is used to design ontology in an effective way for better understanding.

The brucella ontology[14] is developed for the purpose of knowledge sharing in the domain of brucella and brucellosis. This ontology includes the information on Brucella species, Human diseases, Signs and symptoms, a Common resource for infection, Transmission medium and testing[12] as well as treatment.

B. Conceptualization of domain knowledge

The main goal of this session is to detect, define and organize ontology concepts into hierarchies and attributes are used to refine the concept definition. During conceptualization acquired knowledge is structured and organized to obtain independent domain knowledge representation.

Class level Representation

In this brucella ontology, class level representation was constructed by using the top-down and bottom-up approach. We start to develop a class hierarchy from the class of Brucella and Brucellosis. Brucellosis contain the subclasses of Brucella_species, Areas, Common reservoirs, Common source of infection, Human disease, Incubation period, Prevention, Protection material, Signs and symptoms, Spread, Testing, Transmission, Treatment and Workers. Meta level attributes are applied to construct ontologies. Most of the classes have their own subclasses. For example, Human disease class includes the sub classes of clinical sign, Osteoarticular complication, Neurological symptoms, cardiovascular involvement, gastrointestinal signs, chronic symptoms, Mild flu like symptoms, genitourinary involvement. Likewise, each class has their specialized subclasses according to their needs.

Properties

Ontology properties are used to represent the relationship that exists in ontologies. The properties are commonly classified into two types. The two types of properties are data properties and object properties. The data properties are used to correlate data with their properties. The object properties are used to represent the relationship between one individual to another individual. In this brucella ontology, the data properties are entering medium, explanation, news and term definition.
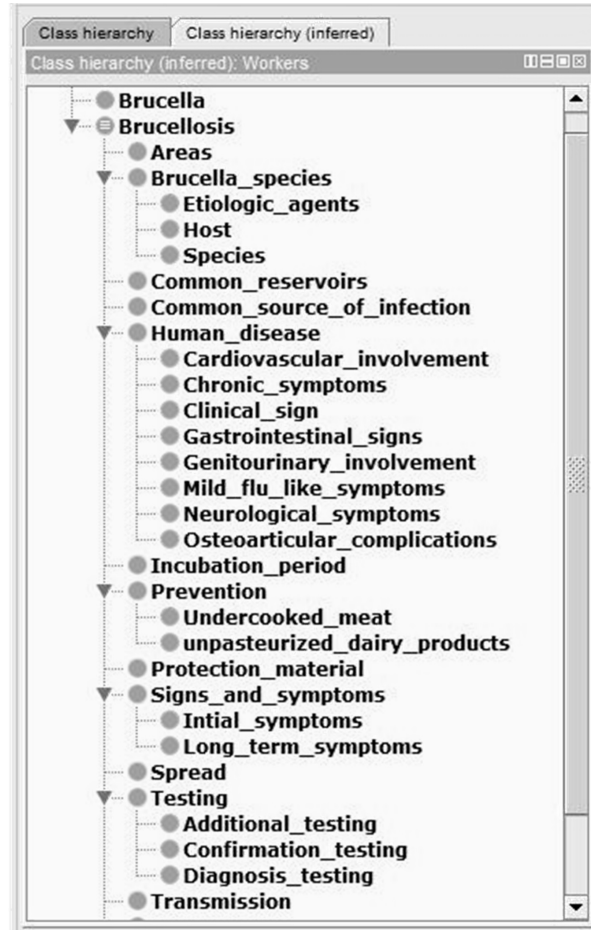
Figure 2: Class hierarchy representation of brucella ontology

The object properties of brucella ontology are also_known_as – property correlate the classes of brucella and brucellosis, are – property establish the connection between brucellosis and human disease symptoms, at_high_risk – property relate the classes of brucellosis and workers, at_risk – property describe the areas which is in risk about brucellosis, by – property correlates the classes of testing and human diseases, contain – establish the relationship between human disease and brucellosis, has – link the classes of etiologic agents with brucellosis, include – establish the relationship between common reservoirs and brucellosis, is – property is used to link incubation period with brucellosis, not_to_consumable_at – property establish the relationship between unpasteurized dairy products and prevention, occur – make a connection between common source of infection and brucellosis, suggest – property used to describe the possible treatments for brucellosis, through – used to make a relationship between brucellosis and transmission, and via – establish the relationship between brucellosis and the spreading medium. Instances can be created based on the properties and classes. Restrictions are defined based on the classes and their needs.

Host Brucella Interaction

Brucella bacteria are able to attack and contaminate both professionals and non-professionals phagocytes. The interaction between the host cell and these hosts dictate the outcome as infection [8]. At least three types of host cells are recognized: macrophages, dendritic cell and epithelial cell. We have two types of brucella strains (Smooth brucella interaction, Rough brucella interaction).

Brucella interactions are:

(i)   Macrophage – Smooth brucella interaction.

(ii) Macrophage – Rough brucella interaction.

(iii) Dendritic cell – Smooth brucella interaction.

(iv) Dendritic cell – Rough brucella interaction.

(v)   Epithelial cell – Smooth brucella interaction

(vi) Epithelial cell – Rough brucella interaction

Pathogenesis of Brucellosis

The ability of brucella cause disease requires a few critical steps during infection. Brucella can invade epithelial cells of the host, allowing infection through mucosal surface: M cells in the intestine have been identified as a portal entry for brucella. Once brucella has invaded, usually through peptic or breathing tract, they are capable of living intracellularly inside phagocytic or non-phagocytic host cells. Brucella has the ability to interfere with intracellular trafficking, preventing fusion of brucella containing vacuole(BCV). The outcome infection is dependent on the species of Brucella and host. The brucella that infect livestock are host restricted. For example, Brucella melintensis, B. abortus, B. suis, B. ovis infect preferentially small ruminants such as, cattle and sheep. With the exception of B.ovis, these brulla have zoonotic potential, with B. melitensis being a most pathogenic for human.

Brucellosis bacterial virulence factors (exotoxins, cytolysins, a capsule, fi mbriae, fl agella, plasmids, lysogenic phages, endotoxic lipopolysaccharide (LPS) and activators of host cell apoptosis) are plays an important role. LPS plays important role brucella virulence because it prevents complement mediated bacteria assassination and delivers resistance contrary to antimicrobial peptides such as defensins and lactoferrin. Another significant virulence procedure of Brucella is the BvrR/BvrS two-component controlling system, which is essential for modulation of the host cell cytoskeleton upon Brucella attack, and for regulation of expression of outside membrane proteins, some of which are essential for full virulence.

Host Immune Response to Brucella Infection

Virulent Brucella is a silent bacterium that captures many host immune mechanisms to serve its own survival and repetition in inner host [9]. Virulent Brucella is able to replicate inside macrophages which are normally powerful instinctive immune cells. Brucella can live in replicative phagosomes inside macrophages where nutrients are hard to achieve. The Brucella-comprising phagosome does not fuse by bactericidal lysosomes. Virulent brucella prevents the programmed macrophage cell death. Rough attenuated brucella strains induce caspase-2-mediated pro-inflammatory cell death in macrophages through the release of cytochrome from mitochondria [10, 11].

4.   RESULT AND DISCUSSION

The brucella ontology was developed by protégé version 4.3. It covers an important class to define the concepts of brucella ontology and also it covers object properties and data properties to define the relationship. This brucella ontology describes the restrictions and value constraints to obtain an exact result. For this work we are getting resources from web/internet like Center for Disease Control and Prevention. This ontology explains the pathogenesis of brucellosis, brucella interaction with host, brucella species and related details are represented ontologically. The formal concept analysis is used to obtain and also represent relevant information. The main

advantages of formal concept analysis are understood quickly. In the existing brucella ontology inference rules are used for reasoning.

The inference rule contains the following disadvantages, (i) the alteration of knowledge base can be complicated, (ii) when introducing new knowledge to solve some specific problem, we might introduce contradictions with the previous rules, (iii) inference rules require pattern matching and computational cost of rule based system can be very high. To avoid these types of problem, in this paper, we used formal concept analysis method. Formal concept analysis is a useful technique to identify group of software entities and easy to understand. After completing the development of the ontology HermiT reasoner was used to check inconsistencies and unsatisfiability. If there is any error occurred and it is often corrected until the reasoner does not provide any discrepancies. The outcome of the reasoner did not provide any exception or error for brucella ontology. Now, the brucella ontology is ready to use. The DL and SPARQL queries are used to obtain required information. When we execute SPARQL queries, we need to define prefixes.

5. CONCLUSION

In this paper, we onto logically show several brucella host interaction, pathogenesis of brucellosis. A formal definition of host brucella interaction was represented in OWL format. A formal concept analysis used for deriving concept hierarchy from object and properties. This paper projected ontology gradual development of framework to help people to meet knowledge based and intelligent information (i.e. conceptual knowledge discovery). The brucella ontology covers objects (G), attributes (M) and incidence relations (I) to represent domain knowledge in Web Ontology Language.

REFERENCES

[1]   Yongqun He, Yue Liu, Bin Zhao, "OGG: a biological ontology for representing genes and genomes in specific organisms", ICBO Proceedings, 2014.

[2]   Georgios Pappas, Photini Papadimitriou, Nikolaos Akritidis, Leonidas Christous and Epameinondas Tsianos, "The new global map of human brucellosis", LANCET infectious diseases, DOI: 10.1016/S1473-3099(06)70382-6, 2006.

[3]   Zuoshuang Xiang, Wenjie Zheng and Yongqun He, "BBP: Brucella genome annotation with literature mining and curation", BMC Bioinformatics, Doi: 10.1186/1471-2105-7 47, 2006.

[4]   Yu Lin, Zuoshuang Xiang and Yongqun He, "Brucellosis Ontology (IDOBRU) as an extension of the Infectious Disease Ontology", Journal of Biomedical Semantics, 2011.

[5]   Werner Ceustersa, Peter Elkinb, and Barry Smith, "Negative Findings in Electronic Health Records and Biomedical Ontologies: A Realist Approach", International journal of Medical informatis, 2007.

[6]   Robert Hoehndorf, Anika Oellrich, Michel Dumontier, Janet Kelso, Dietrich Rebholz-Schuhmann and Heinrich Herre, "Relations as patterns: bridging the gap between OBO and OWL", BMC Bioinformatics, 2010.

[7]   YongqunHe, "Analyses of Brucell a pathogenesis, host immunity, and vaccine targets using systems biology and bioinformatics", Doi: 10.3389/fcimb.2012.00002, 2012.

[8]   R. Martin Roop, Jennifer M. Gaines, Eric S. Anderson, Clayton C. Caswell, and Daniel W.

[9]   Martin, "Survival of the fittest: how Brucella strains adapt to their intracellular niche in the host", Doi: 10.1007/s00430-009-0123-8, 2009.

[10]   Analysis of E-commerce challenges in INDIA by using Weka Tool, Published in IJCTA. [Scopus Indexed].

[11]   Perspectives on Educational Data Mining – A Study Published in Man in India. [Scopus Indexed]

[12]   Bronner D, O'Riordan M and He Y, "Caspase-2 mediates a Brucella abortus RB51-induced hybrid cell death having features of apoptosis and pyroptosis. Front Cell Infect Microbiol". Doi:10.3389/fcimb.2013.00083, 2013.

[13] Albert Mayan J and Ravi .T, "Structural Software Testing: Hybrid Algorithm for Optimal Test Sequence Selection during Regression Testing", International Journal of Engineering and Technology (IJET), Vol. 7, Issue. 1, 2015.

[14] Aloui Amira and Grissa Touzi Amel, "An Extension of Protégé for an Automatic Fuzzy-Ontology Building Using Clustering and FCA", International Journal of Computer Science & Information Technology, Vol. 7, No. 2, 2015.

[15] Mary S. Prince, Dr. Baburaj. E , "Performance enhancement in session identification", IEEE Conference, ICCICCT-2014, pp. 837-840.

[16] Velmurugan A, Ravi T, "Alleviate the Parental Stress in Neonatal Intensive Care Unit using Ontology", Indian Journal of Science and Technology, Vol. 9, No. 28, July 2016, pp. 1-6.