

A Standard Methodology for the Construction of Symptoms Ontology for Diabetes Diagnosis

A.Jaya

Assistant Professor (SG)

Department of CA

BSA University, Chennai

ABSTRACT

Diabetes is a disorder that affects the way that the body uses food for energy. It is a common hormonal problem that if untreated can lead to diabetes complications such as diabetic neuropathy, kidney problems, heart problems, retinopathy and other disorders. Earlier diagnosis for diabetes can prevent the serious cases. This research work aims to diagnose or predict the diabetes in the earlier stage by using reasoner. The reasoner utilizes symptoms ontology for diagnosing the same. The symptoms ontology is a domain specific which helps to provide the possible meaningful factors which leads to diabetes. This paper aims to discuss the way of construction of Symptoms ontology for diabetes and utilization of same for Diagnosing diabetes.

General Terms

Methodology for ontology construction

Keywords

Ontology, diabetes, symptoms, diabetes diagnosis, Reasoner.

1. INTRODUCTION

Nowadays prevention of diabetes is major problem because of life style, food habits, work environment, tension which are major causes of the diabetes at early stage. Pre-diabetes is the state that occurs when a person's blood glucose levels are higher than normal but not high enough for a diagnosis of diabetes. This stage may lead to diabetes stage if patient has not followed the proper diet and exercise. In the symptoms ontology for diabetes consists related terms for various types like type1, type 2, prediabetes, gestational diabetes etc. It is a complex disorder that has many possible indicators. These include unexplained weight loss, excessive hunger (polyphagia), thirst (polydipsia), frequent urination (polyuria), dehydration, leg pain when walking (claudication), fatigue, dizziness and itching (pruritus). Awareness of these signs and symptoms can keep people alert to potential onset of diabetes. The above information can be processed intelligently using ontology. The symptoms ontology has the explicit specification of concepts of symptoms and their attributes as causes, which helps to provide the entire scenario of the diabetes diagnosis. Mainly, three basic elements in ontology are concepts, relations, and axioms to formalize the definitions. Concepts represent the underlying domain knowledge and the relation represents the relationship between the concepts and relations should be concentrated while constructing it. The reasons for construction of ontology are [5]:

- to share common understanding of the structure of information among people or software agents
- to enable reuse of domain knowledge

- to make domain assumptions explicit
- to separate domain knowledge from the operational knowledge
- to analyze domain knowledge

Ontology construction is an art which requires lot of creativity. The components, objects and attributes should be properly placed for effective retrieval. Every designer constructs the ontology in a different manner. There are no standard steps to be followed while constructing the same. Standard way of construction is necessary because;

- Information retrieval to be accurate – which helps for the accurate diagnosis
- Formal and Conceptualization - symptoms for diabetes is properly conceptualized
- Can use standard notations to represent the ideas – helps for diagnosing the same
- Developers can utilize the ontologies efficiently – ontology can be extended various types of diseases

This paper brings out the standard methodology for constructing the ontology in an efficient way. Section 2 discusses related works for the construction of ontology and section 3 depicts the idea of standard methodology for construction of ontology. Sections 4 and 5 shows the way of implementing the standard model for constructing the ontology and results obtained respectively. Section 6 derives the conclusion and future enhancements.

2. RELATED WORK

Blaz Fortuna et al [1] proposed the new technique for constructing the topic ontology for large collection of documents. Topic ontology is a set of topics connected with different types of relations. Each topic includes a set of related documents. Construction of such ontology from a given corpus can be a very time consuming task for the user. Text mining techniques are adapted to assign the topics for each document in construction of the semi automatic ontology which helps to reduce the time consumption.

Lee Gillam et al [2] discuss the method for extracting conceptual hierarchies from arbitrary domain- specific collections of text. These hierarchies can form a basis for a concept-oriented terminology collection, and hence may be used as the basis for developing knowledge-based systems via ontology editors .

B. Chandrasekaran et al [3] discuss the importance of ontology. Ontological analysis clarifies the structure of knowledge. Given a domain, its ontology forms the heart of

any system of knowledge representation for that domain. Thus, the first step in devising an effective knowledge representation system, and vocabulary, is to perform an effective ontological analysis of the field, or domain. Weak analyses lead to incoherent knowledge bases. Since Ontologies enable knowledge sharing, they require more analysis.

Nicola Guarino et al [9] discuss the qualities of formal ontology and its properties. The methodology for ontology design is depends on formal ontology properties which are to be built based on core set Meta properties. These meta-properties are formalizations of the basic notions of identity, rigidity, and dependence. They have considerable value in understanding ontology, as they form a subset of all the properties in the ontology and carry relevant structural information. The backbone carves the domain into useful segments through the categories, identifies every kind of entity in the domain through the types, and contains the most useful groupings of entities through the quasi-types.

Noy et al [5] proposed the standard way to create the ontology with all its formal properties. Since ontology design is a creative process, no two ontologies developed by different people will be the same. But the user can capture this procedure as a guideline for constructing the new ontology.

Jones et al [6] provides the survey of various methodologies that specifically address the issue of the development and maintenance of ontologies. The user can adapt the methods for the construction of ontology as per the needs and desires.

Olavo Mendes et al [7] discuss the ontology development process that involves many activities can present a high level of complexity, depending on the intended scope, size and level of detail of the ontology under construction. The steps for development of software are incorporated in software engineering process.

3. ARCHITECTURE FOR CONSTRUCTION OF ONTOLOGY

Designing ontology is a creative process; the designer may view the ontology in different structures, concepts and relations. Ontology development process involves various phases and procedures. There is no standard approach for constructing the same. The new approach is to discuss about the construction of domain ontology using standard software development model in an efficient way. Our approach to construct the ontology adapts the different phases of waterfall model [8] for the first version of ontology. Since ontology can be reusable, rapid prototyping model is utilized for the subsequent versions for the refinement. The Figure –1 depicts the ontology development life cycle for the construction of symptoms ontology.

3.1 Phases involved in the Ontology construction

Construction of ontology adapts the water fall model [8] which requires following the set of phase described below. Each phase which has its own purpose and it is discussed below:

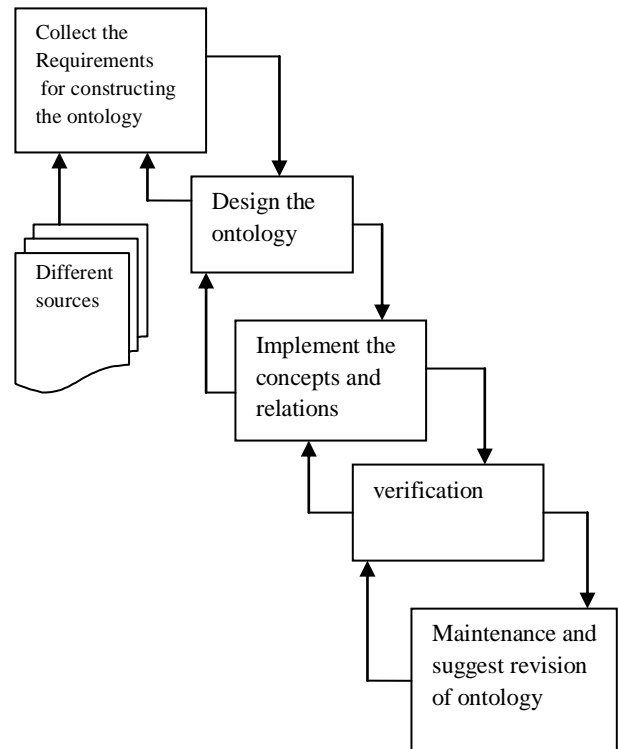


Fig –1 Waterfall model for ontology construction

3.1.1 Phase 1 – Requirements analysis

Acquire the knowledge about the diabetes domain in order to analyze the necessary requirements and to identify scope of the domain for the construction of ontology. During this phase, the purpose of the symptoms ontology is clearly defined. A well-characterized requirements specification is significant to the design, evaluation and re-use of ontology. Symptoms and Medical Knowledge for ontology can be acquired from, standard textbooks; research papers, doctors, patients and other encyclopedias.

Symptoms are classified Internal and external factors. Internal symptoms are blood sugar level, Polydipsia, Polyuria, Polyphagia, etc, and external symptoms are weight loss, Blurry vision, Irritability, Infections, Poor wound healing etc.

3.1.2 Phase 2 Ontology design

Identify the key concepts that exist in the diabetes domain, their properties and the relationships that hold between them; Structure the knowledge in the form of concepts which are arranged in an ordered format. For example, in the diabetes domain, blood level, food control; exercise for the body and etc are the explicit concepts. Each can be expanded with its own attributes.

Using Object oriented approach, the concepts are properly crafted the ontology in the form of attributes and its relation. . For designing ontology, the following sub steps have to be followed,

- Enumerate important terms related to ontology
- Define the classes and class hierarchy using identified terms

- Define the properties of classes and concepts – slots
- Define the Facets of each

Table –2 shows the comparative studies of ontology and Unified Modeling language (UML). Some of the terms are showing the similar concepts between them. The ontology for the domain can be referred as package in UML which consists of the classes describing about the domain. Similarly, the property of the ontology can be reflected as attributes, associations between the classes in UML. Retrieval of classes, attributes from ontology is discussed in the section 4.

Table –1 comparative studies of ontology and UML

DAML Concept	UML Concepts
Ontology	Package
Class	Class
Type	“Classifier” role
Hierarchy	Class generalization relations
Property	Aspects of attributes, Associations and classes
Subclass of	Specialization of classes
Restriction	Constraint association ends, including multiplicity and roles
Cardinality	Multiplicity
Y = minCardinality, Z = maxCardinality	Multiplicity range Y..Z

3.1.3 Phase 3 - Implementation:

There are many possible languages can be used, including general logic programming languages like Prolog for implementing the ontology. Identify the suitable language or tools for the ontology construction based on the application selected. It represents the entire well-known concepts in formal languages. For example, in the semantic web application, entire data can be represented in the Resource description Frame work (RDF) and is used for information retrieval application effectively. Here we have constructed the symptoms ontology in RDF Schema as a language layer, XML Schema for data typing, and RDF to assert data.

3.1.4 Phase 4 - verification:

In this verification phase, it is necessary to check whether the major concepts are incorporated in ontology with its proper attributes etc. If not, upgrade it consequent version ontology. Since, many people may have different symptoms for diabetes and it may vary from person to person also as discussed in the

section 1. In order to bring out the updated knowledge about the domain, ontology is needed to be revised to make it as an efficient one. Updating the symptoms in the ontology, will helps to retrieve the more exact terms for diagnosing the diabetes. Initially the most common symptoms for the diabetes are identified and constructed and later it is revised.

3.1.5 Phase 5 Ontology Maintenance:

Maintain the ontology for the diabetes diagnosis application and suggest the revision if required. It takes care of the existing ontology concepts and records the unavailability of the terms, keep tracks the retrieval time and can be improved in the next version of the ontology. Ontology can be maintained in two phases such as

- Initial construction phase – (it intends to construct new ontology to provide a clear view about the concepts)
- Operation phase – (upgrading ontology with new concepts, attributes, time for retrieval of terms, quality etc.)

Symptoms ontology is used for medial diagnosis in order to diagnose the diabetes. But in some situations, ontology is required to be changed as per the application used for end users. Ontology helps for reasoning the cases semantically. Since necessary versioning is required for perfect diagnosis, construction methodology integrates with rapid prototyping model for the enhancement of the ontology. Prototyping is especially good for designing good human-computer interfaces. Rapid prototyping has been used as a tool for iterative user requirements engineering and human-computer interface design. [9] In the prototyping model, the product can be revised as per the dynamic change of requirement in every iteration. Similarly, the ontology can also be revised the as per the requirements.

4 ROLE OF ONTOLOGY IN DIAGNOSIS SYSTEM

Medical diagnosis system is to accept the symptoms of a patient as an input to the and These are related to patient’s body conditions, heredity status, working environment (having physical movement or not), food habits etc. The Figure –2 shows the architecture of medical diagnosis system using ontology. In the Analyzer module, every word has been tokenized and the required key terms are retrieved from the text input and that will be parsed to the Reasoner module. Reasoner evaluates the questionnaire by applying weights for each symptom. Separate numeric values are assigned for ‘yes’ or ‘no’ type questions.

For example, in the Questionnaire,

1. Does your parents has diabetes (yes/ no)
2. How many times you have rice during a day (none/ 1 / 2 / 3/above)
3. Intake of water in liters (1/2/3)
4.

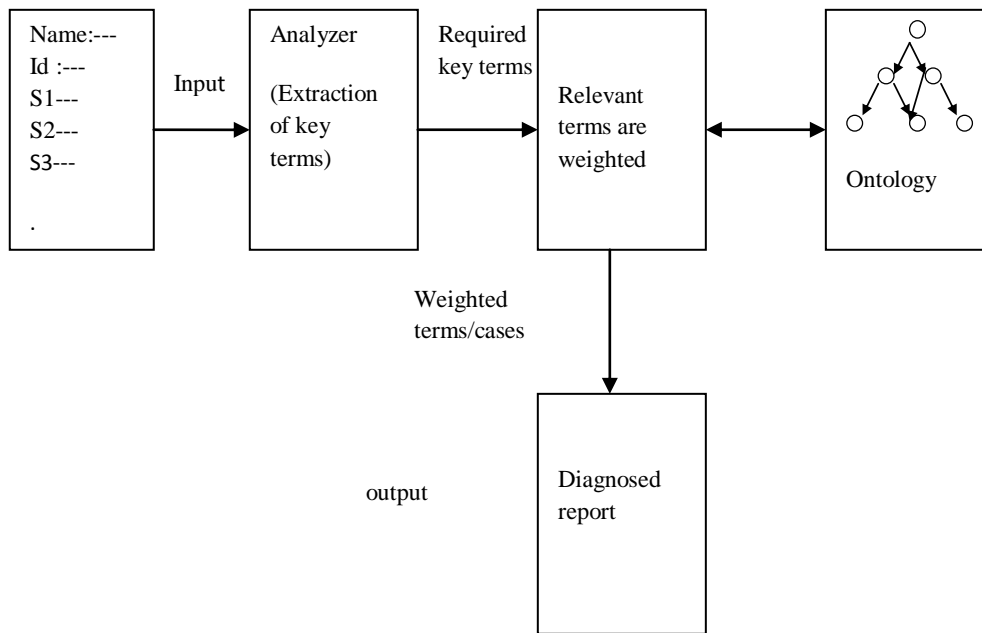


Fig – 2 Architecture of medical diagnosis system

For the ‘yes’ answer questions the weight age is given as one and for ‘no’ answered questions, the weight age is given as zero. Similarly in multi option questions, based on the patients’ answers selection of the answers, the appropriate weight is awarded like in the question 2, if the person takes rice for 1 time it is 0, 0.25, and 0.5,75,1 respectively. Rule based reasoning is applied for evaluating quantified questionnaire. Likewise, each symptom related questionnaire are answered by the patients.

Based on the weightage of the result, their relevant type of diabetes is identified. Apart from the reasoning on weightage, each symptom is checked with consistency of the ontology concepts for symptoms. Since it provides semantics of the concept, which will be useful for reasoning the given inputs. On the relevant count of the symptoms, ontology is analyzed with inputs; the relevant diagnosis report is retrieved from the ontology. Figure 3 depicts portion of symptom ontology.

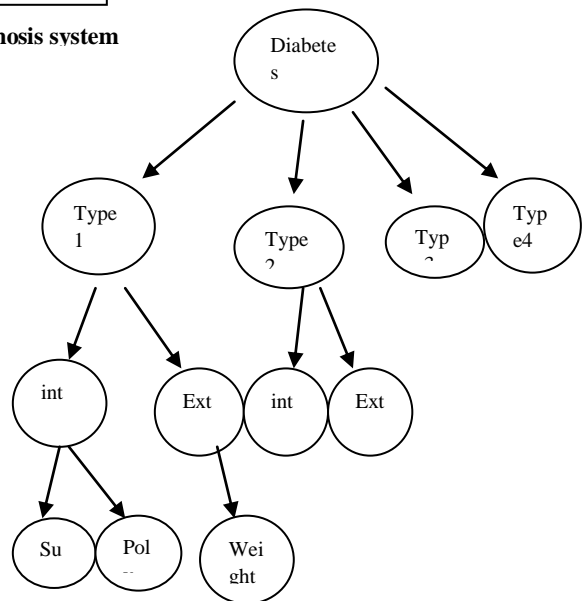


Fig 3 portion of symptom ontology

5. IMPLEMENTATION

The primary use of the symptoms ontology is to reason the patient’s symptoms with ontology semantic concepts in order to diagnose the diabetes. Figure 4 shows the protégé tool environment for owl files construction for classes, their objects and attributes. Figure 5 describes the portion of ontology. The pellet reasoner methods help to check the consistency, classification, realization and concept Satisfiability of the ontology which are helpful for evaluation.

➤ Consistency checking

Ontology should not contain any contradictory facts to perform the consistency. The OWL direct semantics provide the formal definition of ontology consistency used by Pellet [4].

➤ Concept Satisfiability

It determines whether a class has any instances. If a class is unsatisfiable, then defining an instance of that class will cause the whole ontology to be inconsistent.

➤ Classification

It computes the subclass relations between every named class to create the complete class hierarchy. The class hierarchy is used to answer queries such as symptom (type 1).

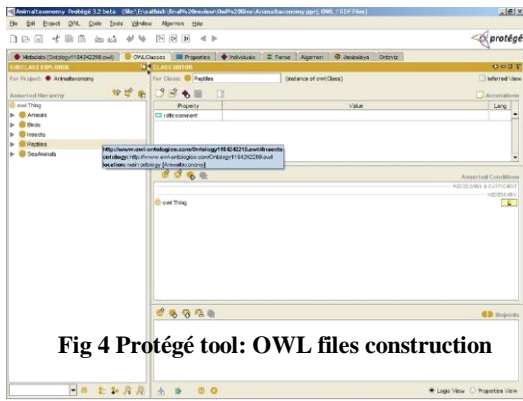


Fig 4 Protégé tool: OWL files construction

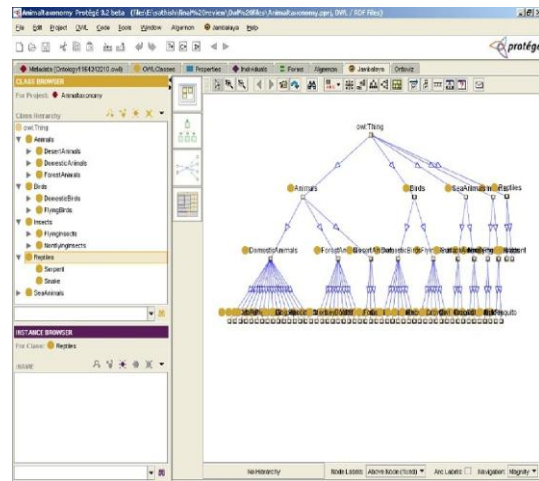


Fig 5 Protégé tool: Ontology visualization

Table 2 - results of the terms retrieved from ontology

S.No	Nature	No of terms attempts	No of retrieved terms	% of retrieval
1.	Appropriateness	120	54	45
2.	Relevant	120	42	35
3.	Unrelated	120	24	20

➤ **Realization**

It is used to find the most specific classes that an individual belongs to; i.e., realization computes the direct types for each of the individuals. Realization can only be performed after classification since direct types are defined with respect to a class hierarchy. Using the classification hierarchy, a list of all types for each individual can be retrieved from the ontology.

6. RESULTS

The retrieval of related terms from the ontology is the primary task for the diabetes diagnosis application. Since the symptoms ontology has many concepts, terms and relations, are needed to fetch the appropriate terms in existing ontology. Checking for the unavailability of terms or symptoms from the base is mandatory. The constructed symptoms ontology was tested for the retrieval of terms with accuracy. Out of 120 number of testing, the details about the retrieval of terms are listed below: 54 attributes are retrieved appropriately (exact match), 42 are approximately retrieved and 24 are retrieved

not relevantly or saying that the search is unsuccessful. Table 2 shows the results obtained from the constructed ontology. The Appropriate retrieval can be improved by concentrating the retrieval of terms with their synonyms and their semantic. The semantic retrieval of terms from the ontology helps for improving the accuracy of diabetes diagnosis. The ontology can be enhanced further by considering the terms and their semantics efficiently.

7 CONCLUSIONS AND FUTURE WORK

The ontology was constructed based on the described methodology and in each phase some of the raised issues have been solved and utilized successfully for the medical diagnosis. Since we have not shown much emphasis on language, our approach lacks in efficient retrieval. Symptoms can be enhanced in nature; the ontology keeps going for refinement for the next level. In order to make efficient in refinement, ontology merging and alignment can also be concentrated in the ontology life cycle development. The semantic retrieval should be forced much in order to bring out the efficient retrieval. By implementing these ideas, the percentage for appropriate retrieval can be improved and we can adapt for the standard method to construct the ontology.

8 REFERENCES

[1] Blaz Fortuna , Marko Grobelnik , Dunja Mladenič , “ System for semi- automatic ontology construction “ European semantic web conference ESWC 2006 .
 [2] Lee Gillam, Mariam Tariq and Khurshid Ahmad, “Terminology and the construction of ontology “, Ingenta Connect, volume 11., pp 55 – 81 , November 1, 2005
 [3] B. Chandrasekaran , John R. Josephson, V. Richard Benjamins, “What Are Ontologies, and Why Do We Need Them?“ IEEE intelligent system , , pg 20 – 25, Feb 1999 .
 [4] Evren Sirin , Bijan Parsia , Bernardo Cuenca Grau, Aditya Kalyanpur, Yarden Katz , “Pellet: A Practical OWL-DL Reasoner
 [5] N. Noy, D.L. McGuinness, Ontology Development 101: A guide to Creating your First Ontology, Stanford Medical Informatics Technical Report No. SMI-2001-0880, 2000
 [6] D. Jones, T. Bench-Capon and P. Visser, *Methodologies for Ontology Development*, Proceedings of the IJCAI-99 workshop on Ontologies and Problem -Solving Methods, 1999
 [7] Olavo Mendes , Alain Abran , “ Software Engineering Ontology: A Development Methodology “ TechRepublic , White papers nov 2004.
 [8] Roger S Pressman, R. S. Pressman & Associates, “Software Engineering: A Practitioner's Approach, 7/e 2010
 [9] S. P. Overmyer: Revolutionary vs. Evolutionary Rapid Prototyping: Balancing Software Productivity and HCI Design Concerns. Center of Excellence in Command, Control, Communications and Intelligence (C3I), George Mason University, 4400 University Drive, Fairfax, Virginia 1991.
 [10] Nicola Guarino and Christopher Welty , LADSEB-CNR , “A Formal Ontology of Properties “ , *Proceedings of 12th Int. Conf. on Knowledge Engineering and Knowledge Management* , Lecture Notes on Computer Science, © Springer Verlag 2000