IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.6, Issue.10, Oct. 2018 E-ISSN: 2347-2693

Mapping Of Pulmonary Disease Ontology Terms Using Graph Stream

P. Radhika[1*], P. Suresh Verma[2], N. Lakshmi Kalyani[3], P. Rama Krishna[4]

[1]Dept. of Computer Science, Rayalaseema University, Kurnool, India
[2]Dept. of Computer Science, Adikavi Nannaya University, Rajahmundry, India
[3,4]Dept. of Computer Sciences and Engineering, VNR VJIET, Hyderabad, India

*Corresponding Author: radhika.pathi123@gmail.com

Available online at: www.ijcseonline.org

Accepted: 11/Oct/2018, Published: 31/Oct/2018

Abstract— Since long, organizations have been looking for information sources that can store data which can provide structured organization and focuses on meaning access of data. Sematic web helps a lot in this context to retrieve meaningful data. This paper emphasizes on mapping developed ontology terms and their retrieval with the help of Graph Stream. We developed an ontology for Pulmonary diseases which consists of classes , objects, relations, and their properties and division of diseases is given as sub classes and Levenshtein's Edit system algorithm has been used for similarity calculations. The generated ontology has been sent for preprocessing and is fed to Graph Stream for graph generation. The results produced are comparable to the results of human annotators.

Keywords— Ontology mapping, pulmonary diseases, similarity calculations, graph generation, human evaluation.

I. INTRODUCTION

Since the dawn of the mankind, knowledge has been the most vital resource and knowledge encoding has been the most important activity. With this knowledge, passed on by their ancestors, younger generations could evolve and develop new processes and techniques. As the time passed by and written languages were developed, this encoding technique transformed from symbols to language text. With the advent of computers, knowledge encoding has been transformed into the digital form. Mere encoding using computers was not a very satisfying, so researchers developed systems which could encode and reason with this knowledge. This gave way to a new research domain popularly known as Knowledge Representation[3] (KR). There have been numerous techniques to represent knowledge. From Minsky's frames[4] to predicate logic to concept hierarchies[5][6], various techniques have been developed to represent knowledge and thus develop knowledge based systems. These systems have been applied to solve various artificial intelligence (AI) problems.
Ontology Engineering is relatively a new approach in Knowledge Engineering (KE) framework. It has gained relative importance in the past fifteen years, as varied AI applications have started using ontologies to preserve their knowledge. A distinct advantage of ontologies[1][2] over other KR techniques is that it can systematically organize semantic knowledge. This facilitates fast and easy interpretation and exchange of desired information. As today's systems can use Web as a knowledge resource. Managing such large reservoir is a very daunting task. As web has transformed into semantic web, it is easier for ontologies to represent their meta-data. It provides semantic search that can be more conceptual and relevant then simple lexical search schemes employed by many retrieval systems and search engines.

A formal and widely accepted definition of ontology is given by Smith and Welty (2001) i.e. "An Ontology is generally regarded as a designed artifact consisting of a specific shared vocabulary used to describe entities in some domain of interest, as well as a set of assumptions about the intended meaning of the terms in the vocabulary." Ontologies manage information through creation of concepts, arranged in a hierarchy (often referred as concept hierarchies), where structured information is stored in different categories. Through this approach search, reuse and understanding of concepts/knowledge becomes easier and faster. Moreover, these concept hierarchies can be used to define relationships among data in an abstract and easy way. They are the very basis of any ontology as they specify axioms, rules and implications between facts in a more precise manner.

In the present scenario, medical information systems[7][8] lack proper organization of data to communicate information about medical activities for a patient information is collected as patient discharge summary or by using common terminologies. There has been variation across the terminologies and a proper organization of data is required. However there is an argument that automation of medical coding whose meaning should be written inside the structure itself within an ontology. The word ontology introduced in philosophy has been reused in the field of computer science from the beginning of 1990' s especially in artificial intelligence, knowledge representation ,information engineering[9][10], and today it has become a common item in information system modeling. The main aspect of ontology is to identify and classify a given set of domain items. For our mapping retrieval we developed a specific ontology for pulmonary diseases [11][12]where textual reports and internet resources have been used as main source of information. We sincerely believe that building a well formed ontology[13] require assigning clear meaning to concepts.

II. OBJECTIVES

Research objectives:
Here we developed a precise methodological process for knowledge engineering to build medical ontology based on semantic relationship recognition from the scrutiny of corpus sequences. Classes are defined and the relationships among the classes and sub classes have been defined to build the ontology. The contextual associations are defined in terms of properties which are nothing to the links to the corpus. The synonymy and hyperonymy links between the candidate terms are identified. The ontology is developed and if we supply ontologies in different formats to a matcher[14], for example one ontology in RDF format, while another in XML format, then they are unable to match them. Moreover, in one instance, if a matcher is required to match one type of ontologies (for example OWL ontologies) and in another instance, is required to match another type of ontologies (for example RDF or XML ontologies) then they might not be able to produce any results. Our matching system addresses both these issues, thus solves the problem of heterogeneity. The research objectives are broadly categorized into following steps.
1. Creation of pulmonary diseases ontology.
2. Mapping of ontology terms.
3. Graph based evaluation.

III. METHODOLOGY

The architecture for the entire frame work for developing ontology and calculating similarity metrics and applying string matching algorithms and developing final mappings is shown in Figure1.

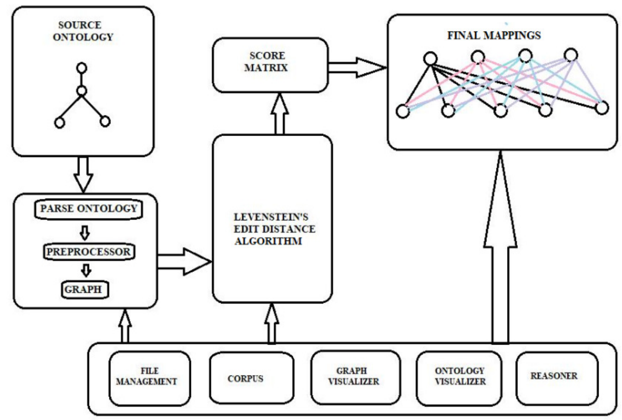
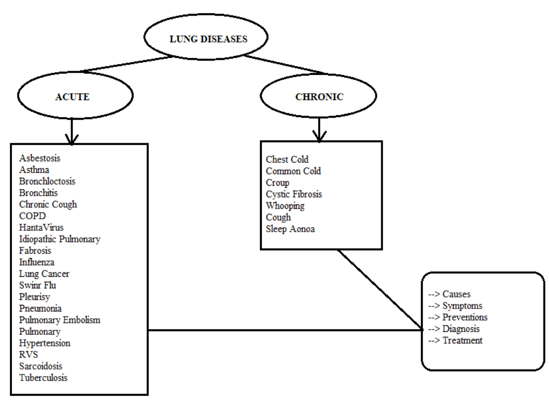


Figure 1: Architecture of pulmonary disease ontology

We first need to define the specification of the ontology. In order to do so, we need to define the purpose and the scope of the ontology and who are going to be the intended users of the ontology. Once these are addressed, we need to create classes/concepts, subclasses/sub-concepts, properties etc. for ontologies. Identifying classes for a particular ontology is a very tricky task. Thus, we need to devise a mechanism for this. For the Lung Diseases, first we need to identify the scope and purpose of ontology. We need to identify the intended users for the ontology. The intended users for the ontology can be:

1. Doctors
2. Victims of the diseases
3. Patients
4. Common People
5. Students
6. Hospital Staff and Management
7. Researchers working on research projects

A. *Identification of terms*
All the data required for the ontology is to be gathered and arranged such that it is easy for us to identify the classes and the sub classes. Here is the how we divide our data:



.

Figure 2: Division of data

B. Defining classes and a class hierarchy and building an OWL File

Here in this ontology we give all the data in the form of classes and sub classes. This hierarchy is developed in the tool called Protégé.

C. Conceptualization

While generating concepts/classes, we identified the main classes and their corresponding sub-classes. It was found that at times some of the classes had similar properties. For example, faculty and staff had most of the common properties. Thus, these are termed as equivalent classes. Figure 3.8 shows the structure of ontology.

Moreover, there were classes which did not have any similarity at all. These were termed as disjoint classes. This demarcation of equivalence and disjoints was done because while inferring about facts or while applying reasoning on the ontologies, we would be using this information.

IV. RESULTS AND DISCUSSION

The following are the snapshots of the development of OWL file:

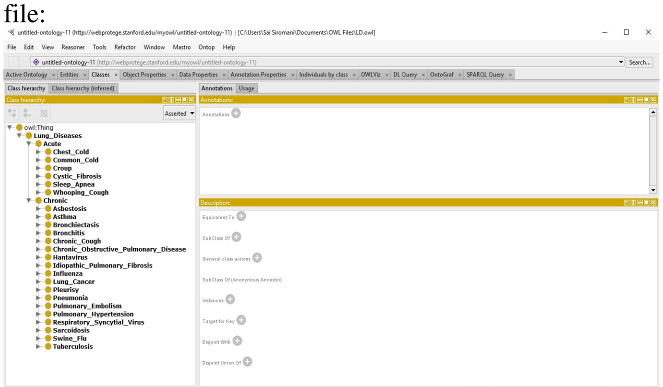


Figure 3: Division of the diseases and giving them as Sub Classes

After developing ontologies the problem of finding semantic mappings between given ontologies can be determined by finding correspondences between two concepts we have used string based method Levenshtein distance algorithm[15] for this purpose. Levenshtein distance is popularly used string metric that measures the differences between the two strings. The distance is calculated from a smallest number of insertions, deletions, substitutions required for changing one string to another. These operations are performed while changing a class name from first to second that are in same domain. The sub polynomial estimate calculation is utilized to enhance the execution of alter remove calculation. Just because of this, algorithm runs in nonlinear time. To enhance performance first step is to compress and then compute edit distance between the strings. A software, SEDIL is used for this. In string matching, two different input strings represent the same context are considered as equal. Mathematical representation

$$d_{(a,b)}\ (i,j) = \begin{cases} max(i,j) & \\ \begin{cases} d_{(a,b)}\ (i-1,j)+1 \\ d_{(a,b)}\ (i,j-1)+1 & if\ min(i,j)=0 \\ d_{(a,b)}\ (i-1,j-1)+1\ (a_i \neq b_j) \end{cases} \end{cases}$$

Where

$1_{(a_i \neq b_j)}$ - indicator function equal to 0 if ai = bj, otherwise 1

Comparative mapping result using Levenshtein distance algorithm is shown in Figure4

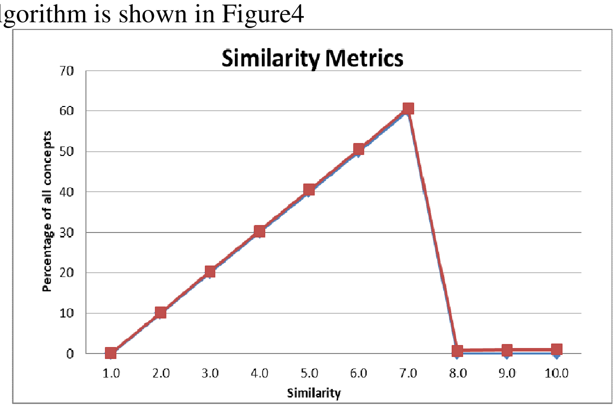


Figure 4 Similarity Metrics with Levenshtein distance showed an average score of 0.7

With Levenshtein Edit Distance algorithm we could show an average similarity of 0.7 on the whole. The Graph can be generated using Graph Stream and the required data along with the related information is given in the form of a graph.

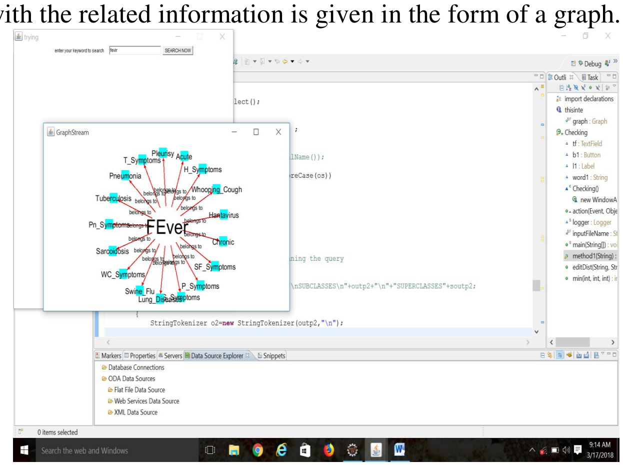


Figure 5 Symptoms and their matching terminologies

V. CONCLUSION

In this paper, we have shown Semantic Relationships between disease names and symptoms with the incorporation of Semantic Web. We supplied ontologies in different formats to a matcher, one ontology in RDF format, while another in XML format, then also we could retrieve semantic relationship thus solving the problem of heterogeneity. We can navigate from symptoms to diseases associated and vice versa making it simpler for use as well as detailed understanding. In the future we could make use of advanced Machine learning algorithms in place of string matching algorithm to display the results dynamically by aligning the clusters' space. We could make further improvements to the search engine being used. Semantics matching could me made even more accurate with incorporation of efficient and effective algorithms consuming lesser time.

REFERENCES

[1] A Kivela, E Hyvonen ."Ontological theories for the Semantic Web", Helsinki: HIIT Publications, 2002, pp.111- 136

[2] T Gruber, "Towards principles for the design of ontologies used for knowledge sharing", International Journal of Human-Computer Studies .Vol. 43(5/6), pp.907-928. 1995

[3] Harmelen, Frank & Lifschitz, Vladimir & Porter, Bruce. (2007)." The Handbook of Knowledge Representation". Elsevier Science San Diego, USA. 1034.

[4] Marvin Minsky. 1974. A Framework for Representing Knowledge. Technical Report. Massachusetts Institute of Technology, Cambridge, MA, USA.

[5] J. Han and Y. Fu. "Dynamic generation and renement of concept hierarchies for knowledge discovery in databases". In Proc. AAAI'94 Workshop on Knowledge Discovery in Databases(KDD'94), Seattle, WA, 157-168, 1994.

[6] J. Han. Mining knowledge at multiple concept levels. In Proc. 4th Int. Conf. on Information and Knowledge Management (CIKM'95), Baltimore, Maryland, 19-24, 1995.

[7] Bali, R.K., 2000. Towards a qualitative informed model for EPR implementation: Considering organizational culture. Proceeding of the IEEE International Conference on Information Technology Applications in Biomedicine (ITABITIS), Arlington, USA., pp: 353-358.

[8] Bose, R., 2003. Knowledge management-enabled health care management systems: Capabilities, infrastructure and decision-support. Exp. Syst. Appl., 24: 59-71

[9]A. Gómez-PérezOntology evaluationHandbook on Ontologies (first ed.), Springer, Berlin (2004)pp. 251–275

[10] Zweigenbaum, J. Bouaud, B. Bachimont, J. Charlet, B. Séroussi, J. - Boisvieux From text to knowledge: a unifying document-oriented view of analyzed medical language Methods Inf. Med., 37 (4–5) (1998), pp. 384-393

[11] Audrey Baneyx, Jean Charlet, Marie-Christine Jaulent, Building an ontology of pulmonary diseases with natural language processing tools using textual corpora, International Journal of Medical Informatics,Volume 76, Issues 2–3,2007.

[12]J. Simon, M. Dos Santos, J. Fielding, B. SmithFormal ontology for natural language processing and the integration of biomedical databases Int. J. Med. Inf., 75 (3–4) (2006), pp. 224-231

[13]J. Charlet, B. Bachimont, M.-C. JaulentBuilding medical ontologies by terminology extraction from texts: a proposalComput. Biol. Med. (2005)

[14] J. Euzenat and P. Shvaiko Ontology Matching, 2007, ISBN: 978-3 540-49611-3, Springer Berlin Heidelberg, New York

[15] Rishin, Haldar, Debajyoti Mukhopadhyay, Levenshtein Distance Technique in Dictionary Lookup Methods: An Improved Approach, arXiv preprint arXiv: 1101.1232, 2011.

[16] C. Dharma Devi, S. Thaddeus "Information Retrieval Mechanism for Dynamic Health Care" International Journal of Computer Sciences and Engineering Vol.-6, Issue-9, Sept. 2018

Authors Profile

P.Radhika pursued Master of Technology in Computer Science from JNTUH in year 2008. She is currently pursuing Ph.D. and c urrently working as Assistant Professor in Department of Computer Science and Engineering VNR VJIET since 2008. She is a member CSI and ISTE. He has published more than 5 research papers in reputed international journals including Thomson Reuters and it's also available online. Her main research work focuses on Ontology,Semantic Web, Service Orinted Architecture, E-learning She has 10 years of teaching experience and 6 years of Research Experience.

P.Suresh Varma presently working as a Professor of Computer Science and Engineeing & Dean, Faculty of Engineering and Technology, Adikavi Nannaya University. He has served as Principal, University Colleges from 4th Dec 2009 to 13th June 2012. Professor Varma has been engaged in teaching, research and research supervision at Adikavi Nannaya University, Rajahmundry since Oct 2008. He pursued his M.Tech in Computer Science and Technology from Andhra University in 1999 and secured Ph.D in Computer Science and Engineering from Acharya Nagarjuna University in 2008. He has supervised the research work of a number of Ph.D and M.Phil scholars and published and lectured extensively on Communication Networks, Data mining, Cloud Computing, Big Data and Image Processing. In 2010 Government of Andhra Pradesh honored with Best Teacher Award in the occasion of Teacher Day. He has published over 100 papers in journals and conferences. 9 Ph.Ds are awarded and 3 Ph.Ds are submitted under his guidance and over 100 M.Tech./MCA thesis are supervised.

N.Lakshmi Kalyani pursued her Bachelor of Technologies and Master of Technologies from Jewaharla Nehru Technological University of Hyderabad, Telangana, India. in year 2013. Currently working as assistant Professor in Department of Computer Science and Engineering, VNR VJIET, Hyderabad, Telangana, India Since 2015. She is a life member of ISTE & CSI since 2016. She has published more than 2 research papers in reputed international journals. Her main research work focuses on Internet of Things, Ontology, Virtual Reality. She has 3 years of teaching experience and 5 years of Research Experience.

P RAMAKRISHNA, Completed B.Tech & M.Tech in Computer Science and Engg from JNTUH, Hyd, having an experience of 13 years in Teaching & Research. Presently working as an Asst. Professor in VNRVJIET, Hyd. He is a life member of ISTE & CSI since 2016. He has published more than 2 research papers in reputed international journals. His Area of Interests- Ontology, Network Security, Data Mining, Cloud Computing.