RESEARCH ARTICLE OPEN ACCESS

Building A Database Driven Reverse Medical Dictionary

Pragya Tripathi
Guide: Manjusha Deshmukh
Department of Computer Science Engineering
Pillai Institution of Information Technology, New Panvel
University of Mumbai, India

ABSTRACT

The project aimed to build a fully functional system called as reverse medical dictionary in order to achieve the efficiency in fast health treatment consultation system. Reverse medical dictionary allows users to get instant guidance on their health issues through an intelligent health care system. Simply we say that user can search their illness according to their symptoms at any point of time and get instant diagnosis.

Keywords :— NLP, WordNet, WSD (word sense disambiguation), Semantic Similarity.

I. INTRODUCTION

Information plays very vital and important role in this modern civilization to step forward in every sphere from earth to ocean, earth to planet, earth to sky. People are discovering amazing inventions at a rapid pace based on information. Therefore, it is important that the information we are using should be accurate.

The proposed system is fed with various symptoms and the disease/illness associated with those symptoms. The system allows user to share their symptoms and issues. It then processes user's symptoms to check for various illnesses that could be associated with it . Here we use some techniques to guess the most accurate illness that could be associated with patient's symptoms. Admin can add new disease details by specifying the type and symptoms of the disease into the database. Admin can view various disease and symptoms stored in database. This system will provide proper guidance when the user specifies the symptoms of his illness.

BASIC AND BACKGROUN KNOWLEDGE

Here we are going to explain the basic definition and strategies' for building reverse medical dictionary.

Natural Language Processing

Natural language can be any language which human can understand, like English, Marathi, Punjabi, Tamil, Hindi, etc but computer only understand machine language, So if we want computer to understand human language, we have to convert natural language into machine language. Natural Language Processing (NLP) will help us to interact between human and computer [7]. NLP process information contained in natural language text and make computers learns our language rather than we learn theirs. NLP is very important topic in todays world of internet.

WordNet

WordNet is a lexical database of English and is the product of a research project at Princeton University. It was designed to establish the connections between Parts of Speech (POS) which includes noun, verb, adjective, and adverb. The smallest unit in a WordNet is synset. Synset represents a specific meaning of a word. It includes the word, its explanation, and its synonyms. In this paper, we are only concerned about the similarity measure based on nouns, verbs and synonym relation of WordNet [1].

Semantic similarity

Semantic similarity which is sometimes called as topological similarity. Semantic similarity is calculated at many levels like document level, term level and sentence level. Here we are finding semantic similarity between two sentences, one sentence is input sentence which is symptoms which are going to match with other sentence, means the sentences which are in the databases.

II. PROPOSED APPROACH

In the proposed system we are finding various diseases according to user input which is number of symptoms. Here we use some techniques to guess the most accurate illness that could be associated with patient's symptoms.

System Architecture

Architecture of reverse medical dictionary is described in figure 1 .Where we describe our implementation architecture, with particular attention to design for scalability.

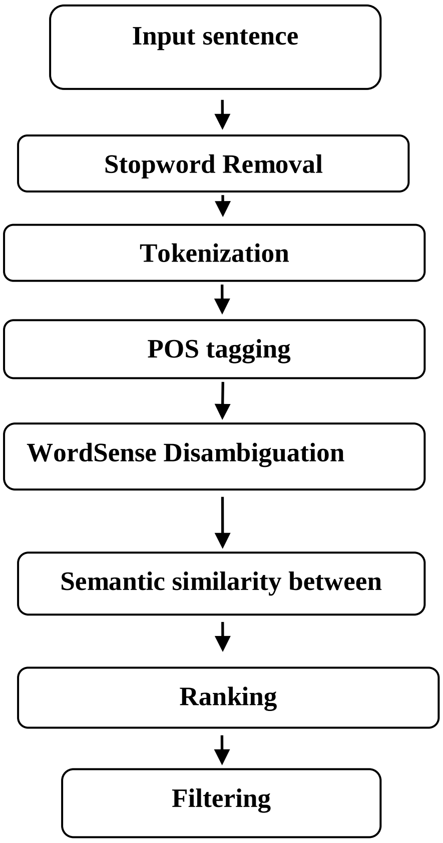

FIGURE 1: ARCHITECTURE OF REVERSE MEDICAL DICTIONARY

Modules For Building Reverse Medical Dictionary

We divide our reverse medical dictionary into number of modules.

Pre-processing

a) Stop word removal

Stop words are words which are filtered out after or before processing NLP text. We can say stop words are most common words in a language, although there is no single universal list of stop words used by all natural language processing tools. Here in project we are using a list of stop words.

b) Tokenization

Process of breaking stream of text into words, phrases and symbols or other meaningful elements called tokens.

These tokens become input for further processing.

POS tagging

The task of finding correct part of speech (POS like noun,verb, pronoun, adverb ...) of each word in the sentence is known as tagging. The algorithm takes a sentence as input and a specified tag set (a finite list of POS tags). The output is a single best POS tag for each word.

Word Sense Disambiguation (WSD)

WSD means the task of finding meaning of an ambiguous word in the given context or the task that identify appropriate meaning(sense) to a given word in a text.

For example Bank can have two senses:

1. Edge of a river 2. Financial institution that accepts money

Semantic similarity between two sentences

After doing WSD our last step is finding semantic similarity which is sometimes called as topological similarity. Here we are finding semantic similarity between two sentence, one sentence is input sentence which is symptoms which are going to match with other sentence, means the sentences which are in the databases.

Let us consider the given two sentences as an input to this process; first the words of two sentences are compared. If the two words of the sentences are matched, its similarity score is calculated which are based on syntactic level. If the words of the two sentences are not matched, then synsets of the word is extracted from sentnce1and compared with the other word of the sentence2. If the words are matched at synset level then return the score as 1, otherwise return 0. Even the words are not matched, then consider the definition of the word sense of the sentences and compare the similarity score of the sentences which are totally based on semantics. This way we compute how two sentences are similar semantically.

Ranking

In this module we ranked the output according to the score which we have from previous step.

Filtering

This is the last step in which we take a threshold value (let say 0.9) and do the filtering according to this value. We eliminate those output which are below the threshold value and take only those which are above it.

III. EXPERIMENT RESULTS

We use databases consisting of 1000 number of diseases.

We ran our algorithm on this database. The various similarity score is computed and compared. As pointed out in the introduction, feature based measure give best semantic similarity score between pair of sentences.

Table 1. Sample pair of sentences

Inputs	Obtained Output	Test Status
pain in stomach	abdominal rigidity, diarrhea, ulcer, q fever.	Identified
redness in eye	Conjunctivitis, corneal abrasion, foreign object in the eye, fungal eye infections.	Identified
bleeding gums	Leukemias, painful gums, talon rodenticide poisoning.	Identified
aching, tiredness, fever	Chicken pox	Identified
pain in legs	lymphatic filariasis	Identified
fever,vomiting, headache	strep throat	Identified

IV. RESULT ANALYSIS

To test how our proposed system performs, Actual result analysis of proposed system in terms of precision, recall.

Precision is nothing but how many of the returned results are correct, that means the ratio of correct positive observations.
Precision = TP / (TP+FP)

Recall is nothing but how many of the positives the system will return, that means the ratio of correctly predicted positive events. As recall increases precision decreases and vice-versa.
Recall = TP / (TP+FN)

Accuracy is perhaps the most intuitive performance measure. It is simply the ratio of correctly predicted observations.

Accuracy = (TP+TN) / (TP+FP+FN+TN)

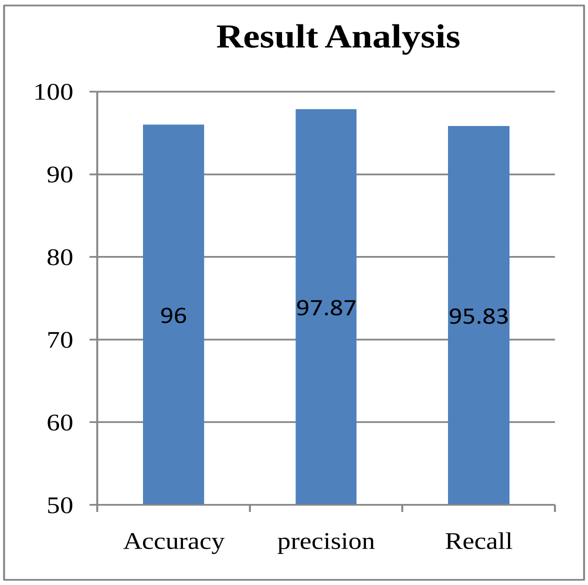

Figure 2 Result analysis of Reverse Medical Dictionary

There are systems available which can perform same work as reverse medical dictionary. By doing rigorous testing with those systems and with proposed system, comparison is shown below in figure 3. Here we are comparing our system by isabel symptom checker and mayo clinic.

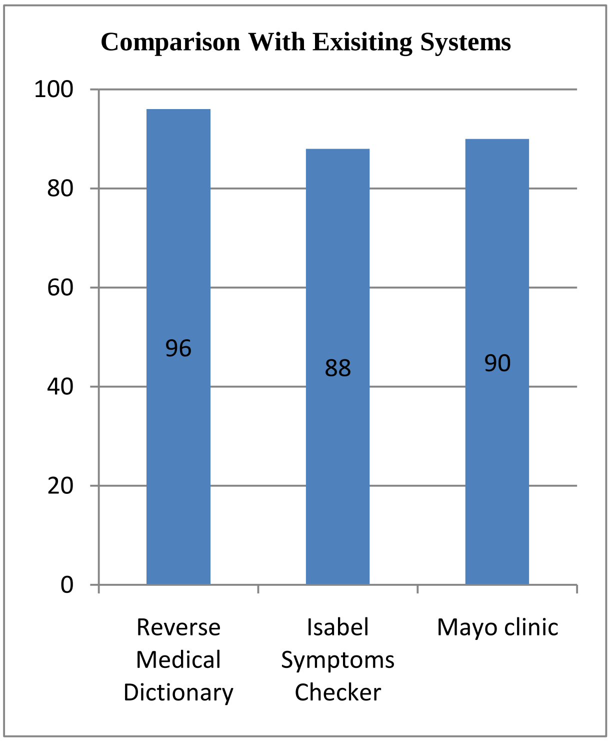

Figure 3 Comparison of accuracy of proposed system with existing systems

V. APPLICATIONS

The reverse medical dictionary allow user to share their symptoms and issues ,then the system processes user's symptoms to check for various illnesses that could be associated with it. So this system concern with many applications in medical field. Few of them are listed below:

1. Reverse medical dictionary allows users to get instant guidance on their health issues through an intelligent health care system. Simply we say that user can search their illness according to their symptoms at any point of time and get instant diagnosis.

2. Reverse medical dictionary can be used by all patients or their family members who need help in emergence.

3. Reverse medical dictionary puts the world's medical knowledge at your fingertips and enables you to make sense of your symptoms. It will change the way you speak to your doctor forever.

4. Reverse medical dictionary is the system that is relied on by doctors and nurses to help with diagnosis and is acknowledged as the clear leader in its field.

5. Medical dictionary include smart medical service and operation management system which are built for the patients living in urban and rural, have the uniform technical standards and wide range of applications, the aim is to achieve application of remote medical treatment and basic health care network.

VI. CONCLUSIONS

As we know that Reverse medical dictionary has vital role in medical field, the medical dictionary is a system which supports end user and on the other hand it is a consultation project in which users can get immediate guidance on their health issues through an intelligent health care system. Here the system contains a database that database contains diseases name along with their various symptoms. It also has an option for users or we can say patient to sharing their symptoms and issues.

The system processes those symptoms to check for various diseases that can be associated with it. If user's symptoms do not exactly match any disease in the database, then it is shows the diseases user could probably have based on his/her symptoms. We propose a set of methods for building reverse medical dictionary. Here the main concept is finding the similarity measure between the two sentences.

Although medical dictionary is available online using data mining technique, but we proposed reverse medical dictionary by using NLP concepts which is very different from others medical dictionary. It is a new concept and we describe a set of experiments that show the quality of our results, as well as the runtime performance under load.

VII. FUTURE SCOPE

We can increase our efforts to develop machine learning methods, to exploit information intelligently and extract the best knowledge and we can also extend the WSD algorithm with supervised learning with such methods as the Naive Bayesian Classifier model which will improve our reverse medical dictionary.

ACKNOWLEDGMENT

I am thankful to Prof. Manjusha Deshmukh for giving valuable inputs required for the topic, and her invlaluable support.I am thankful to Prof. Satishkumar Varma for providing me the idea of this topic, and for his invaluable support.I would like to thank M.E. Coordinator Prof. Sharvari S. Govilkar and all staff members for their direct and indirect support.I would like to thank the H.O.D. of computer engineering department Dr.Madhumita Chatterjee for her invaluable support.I would like to thank Principal Dr. R. I. K. Moorthy for their invaluable support.

REFERENCES

[1] Ryan Shaw, Anindya Datta, Debra Vander, and Kaushik Dutta, Member, IEEE ," Building a Scalable Database-Driven Reverse Dictionary" IEEE tranction , Vol. 25, no 3, March 2013.

[2] E. Gabrilovich and S. Markovitch, "Wikipedia-Based Semantic Interpretation for Natural Language Processing", Journal of Artificial Intelligence Research, vol. 34, no. 1, pp. 443-498, 2009.

[3] N.Segata and E.Blanzieri," Fast Local Support Vector Machines for Large Datasets", Proc. Int'l Conference,

Machine Learning and Data Mining in Pattern Recognition, July 2009.

[4] J.P Sutton "Smart medical systems" Nat. Space Biomed. Research. Inst., Houston, TX, USA, 06 January 2003.

[5] Sujatha R, "A Survey of Health Care Prediction Using Data Mining" International Journal of Innovative Research in Science, Engineering and Technology Vol. 5, Issue 8, August 2016

[6] Sona Baby, Ariya T.K "A survey paper of data mining in medical diagnosis" International journal of research in engineering and technology , 2014

[7] Thanh Ngoc Dao, Troy Simpson "Measuring Similarity between sentences", 2005

[8] Thabet Slimani, "Description And Evaluation Of Semantic Similarity Measures Approaches", 2013.