Available online at www.sciencedirect.com

ScienceDirect

Materials Today: Proceedings XX (2016) XXX–XXX

www.materialstoday.com/proceedings

PMME 2016

Ontology Based Indian Medical System

[1]M.Gayathri, [2]R.Jagadeesh Kannan

[1]*Assistant Professor,SCSVMV University,Kanchipuram-631501,India*
[2]*Professor,VIT University,Chennai,India*

Abstract

Abstract India is known for its traditional medicines like Ayurveda, siddha, unnani, homeopathy and yoga. Publishing of literatures on traditional medicines are expanding rapidly. Due to the exponential growth of articles and literatures existence, it is difficult for the user or the practitioners to find the relevant and useful information from this large amount of data. Text mining is applied on this data to find the useful facts. In this study, we concentrate on ayurvedhic medical system and we discussed on how to apply text mining techniques on ayurvedic literatures to find the most relevant ayurvedhic facts. Moreover we constructed ontology for the medicinal plants for better understanding of relations among such herbal plants with its usage in curing the diseases. Keywords: Traditional medicine, ontology, text mining

Selection and Peer-review under responsibility of International Conference on Processing of Materials, Minerals and Energy (July 29th – 30th) 2016, Ongole, Andhra Pradesh, India.

*K*eywords: Traditional medicine, ontology, text mining

Selection and Peer-review under responsibility of International Conference on Processing of Materials, Minerals and Energy (July 29th – 30th) 2016, Ongole, Andhra Pradesh, India.

1. Introduction

Traditional Medicines [4,3] derived from medicinal plants are used by about 60% of the world's population. India is the largest producer of medicinal herbs and is called as botanical garden of the world. In the last few years there has been an exponential growth in the field of herbal medicine and these drugs are gaining popularity both in developing and developed countries because of their natural origin and less side effects. These medicines are more popular in the world because it tries to cure the diseases without any side effects. India is known for its traditional medicines like Ayurveda, siddha, unnani, yoga and homeopathy. According to world Health Organization (WHO), approximately 80% of the people were following Traditional medicines for their curatives. Clinical Researches are finding the solution to use the traditional medicines effectively rather than using modern medicines. There are millions and millions of articles that has been published by them on traditional medicines. Ayurvedha focuses the use of plant based medicines. There are many number of medicinal plants which includes thulsi, ashwaganga, aloe vera, ginger, turmeric, garlic, cloves, cardamom etc which holds medicinal values in curing many diseases. Researchers and life scientist were publishing their findings in the form of literatures. We can find the large amount of literatures in knowledge bases. One such knowledge base is traditional knowledge digital library which is maintained by government of India. So finding the relevant information from the large amount of data is the biggest challenge. Data mining techniques are used to get the relevant facts. Before applying the techniques we need to preprocess the data because the data available in the literature is of unstructured (data may in pdf, word, notepad).so Text mining [7] become most important in the field of computer science because the data generated per day exceeds some tera bytes. It is applied on the unstructured data in order to yield useful information from this large amount of data facts. Here introduce the paper, and put a nomenclature if necessary, in a box with the same font size as the rest of the paper. The paragraphs continue from here and are only separated by headings, subheadings, images and formulae. The section headings are arranged by numbers, bold and 10 pt. Here follows further instructions for authors.

2. Objective

In this study, we discussed the tool which automatically identifies the information about an herbal plant and its salient features in curing diseases. The main motivation of our ayurvedic system is to reduce the time required for researchers to learn about an unfamiliar herb, or to find the herb related information from the traditional literature. The objective is to provide life scientist the important characteristics or functions of the herb and assist in understanding specific aspect of herb, by directing user to pertinent text passages.

3. Implementation

Implementation there are three main steps involved in order to carry out the mining task 1.Gather data 2. Process the data 3. Analyze the data.

3.1 Text Gathering

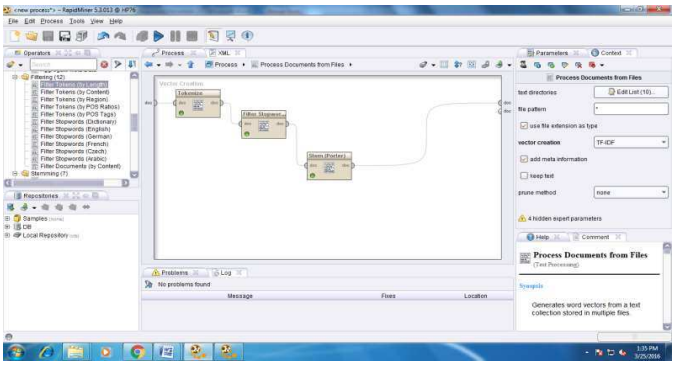
To mine the useful data, the first step is to gather the data as data which is scattered. In order to collect data we found internet, Traditional knowledge digital Library (TKDL) [8] and other literature sources. We collected around 500 ayurvedha based articles in various medical journals, NCBI (National Center for Bio Informatics) and other nuggets. Ayurvedha articles for our study includes for various medicinal plants like aloe vera, amla, ashwaganga, bitter melon, chirata, ginger, neem, tulsi, triphala and turmeric.

3.2 Text Processing

In order to process the collected data NLP [2] Techniques are involved because of the text in unstructured format. Various NLP tasks forms the processing. It includes Tokenizing, POS tagging, stop words and stemming. Tokenization is the process of breaking a stream of text up into words, phrases, symbols, or other

meaningful elements called tokens. part-of-speech tagging (POS tagging or POST), also called grammatical tagging or word-category disambiguation, is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition and its context—i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. Stop words are which does not add meaning to the sentence such as a, the, in of, which etc.Stemming is the process for reducing inflected (or sometimes derived) words to their word stem, base or root form for example consider the word use the stem words includes uses, used, user, used by etc.In order to carry out this task we used Rapid miner, the tool for text processing.

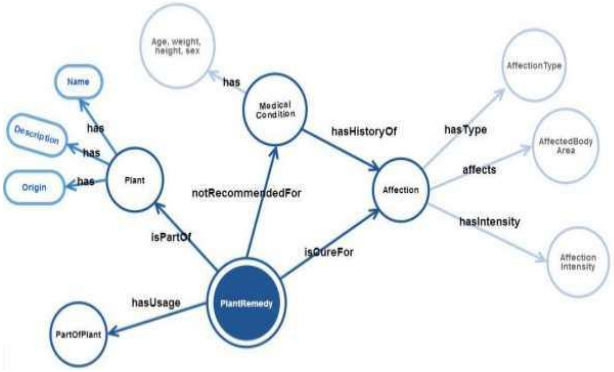
Figure 1: Text preprocessing



3.3 Text analyzing

Once the data is processed, It is ready to analysis those data, analyzing the data involves applying data mining techniques on those processed data. In our method, before analyzing the data we constructed ontology for the processed data. Ontology [5] is the explicit specification of knowledge that specifies the relation among data. To find the meaningful and semantic integration of data Ontology is constructed. Ontology plays an important role in bringing the relevant hits and it can be used to offer electronic inference. This is the relevant perception for management of knowledge. We used protégé tool to design our medicinal plant ontology in spite we used existing plant and disease ontology in our study for effective construction of knowledge base. We used plant classification as the base for constructing the ontology. We considered which part of the plant holds the medicinal value for treating which type of diseases. For example considering the herbal plant thulsi; part of the plant used is leaf, and it is used to cure common cold and cough. For analyzing we applied novel data mining [1] technique and the documents are ranked using the tf * idf ranking algorithm.

Figure 2: Schematic view of plant ontology



Our system includes four major parts which are explained below.

i. Retrieving documents for an herb.
ii. Identifying important terms and its relevant.
iii. Categorization of those terms.
iv. Visualize the result.

i. Retrieving document for an herb

The following steps are used

a. Herb identifier or official name forms an input to the system.
b. Synonyms of the herb are gathered (the herb can be called with multiple names).
c. All names, synonyms and variations of these are used to retrieve the TKDL abstracts.

We aim at retrieving the abstracts because they are readily available and also possess the condensed description of the whole study. This makes our system falls in efficient retrieval by saving time.

Disambiguation of herb names

While the use of synonyms increases the recall, it can lead to drop in precision if we did not eliminate irrelevant retrievals. For an example amla is the plant which can be called with multiple senses of words. The basic step to determine the abstracts for a given herb are.

1. Identify the non-ambiguous abstracts.
2. Identify all possible synonyms for the given herb
3. Create language model for the set of abstracts identified in 1 and 2
4. Choose the language model that best represents the abstract.

ii. Identifying important terms ant it's relevant

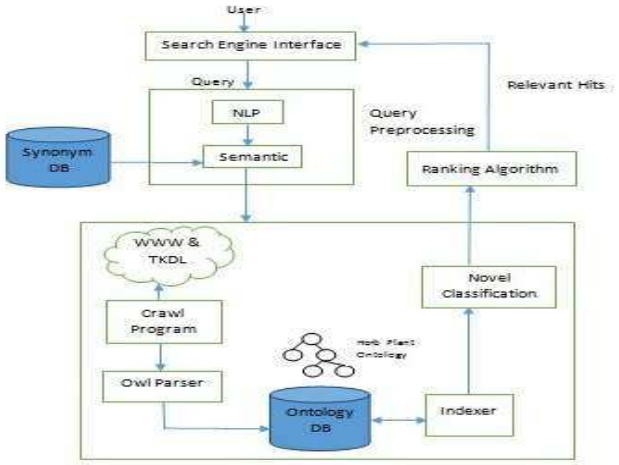
Our system assign scores to unigrams (single word),bigrams(two-word) as well as set of biomedical terms that we extracted from different knowledge bases including TKDL,NCBI ,plant ontology, disease ontology. Important terms are identified by calculating the frequency of occurrences of words. These words are considered to be the keywords .Each and every keywords, its conceptual word their properties and its relation are identified .This is done with the automatic construction of the plant ontology designed.

iii. Categorize the important terms

People have different interest. Researchers might be interested in looking for learning about diseases in which the herb plays an important role. Conversely life scientists might be interested in functional terms processes and its technique. We divide our system into primary and secondary categories.

Primary category is similar to search engine where the user of the system submit the query and get the response. In the secondary category we listed all the herbs, official or scientific name, family name, parts used in curing diseases (its usage) and how to prepare it.To do so, we used novel classification data mining algorithm for effective classification and mining of relevant terms [6].

Figure 3: Structure of the ayurvedhic system



iv. Visualize the result

Now our Database contains the precompiled keywords and their categorization. If the researcher wishes to learn about the herb they can start by specifying the herbal plant names, aliases, scientific name or synonyms as an input the system. The resultant have the most relevant pages are ranked and retrieved for the user.

4. Results and discussion

Traditional medicines have its own value even in the digital era. We presented the ayurvedic system which is most particular in retrieving the ayurvedhic based literatures. Our system is built based on the domain ontology called herbal plant ontology. Therefore we hope that our system tries to search for the documents semantically, retrieves the most relevant pages and the total time is also limited. It can be compared with other popular search engines like yahoo, Google etc, but all tries to retrieve the irrelevant hits and the time which is taken to retrieve those document is also high. This can be measured in terms of precision and recall.

Table 1. Comparative study with other system

System	Recall	Precision	f-measure
Other system without semantic integration	78	32	72.3
Ontology based semantically integrated system	89	23	89.9

5. Conclusion

This system is built to retrieve the ayurvedhic literatures. It include various component which is not found in many of the system which includes the module which gathers abstracts that mention the given herb name or synonyms and automatically filters out abstracts that are irrelevant to the herb, mining of multi-word terms, including methods to avoid redundancy and problems with low frequency. In future we can extend the use of this system for other Indian traditional medicines.

References

[1]. Lihua Jiang, Hong-bin Zhang, Xiaorong Yang, Nengfu Xie, Research on Semantic Text Mining Based on Domain Ontology, Springer, 2013.

[2]. Hsin-Chang Yang ; Chung-Hong Lee, Mining open source text documents for intelligence gathering, Vol. 2, 2012 , Page(s): 969 – 973.

[3]. Shih-Wei Chen, Yu-Ting Tseng, Tsai-Ya Lai, The design of an ontology-based service-oriented architecture framework for traditional chinese medicine healthcare, 2012 , Page(s): 353 – 356.

[4]. Pooja Goyal, Tanu Midha, RP Sharma, DS Martolia, Amarnath Gupta, Utilization of Indian Systems of Medicine and Homeopathy(ISM&H): a cross-sectional study among school students of kanpur city, Indian Journal of Pre, Vol. 42 No.3, 2011.

[5]. XONTO: An Ontology-based System for Semantic Information Extraction from PDF Documents.

[6]. Catalina O Tudor, Carl J Schmidt, K Vijaya-Shanker. eGIFT: Mining Gene Information from the Literature, BMC Bioinformatics 2010, 11:418.pdf

[7]. Irena Spasic, Sophia Ananiadou, John McNaught and Anand Kumar, Text mining and ontologies in biomedicine: Making sense of raw text, Henry Stewart publications, 1467-5463. Briefings in bioinformatics. Vol 6. No 3. 239–251. September 2005.

[8]. M Hirwade, Protecting Traditional Knowledge Digitally: A Case Study of TKDL, Article from eprints.rclis.org/14020/1/TKDL_paper.pdf - 2010.

[9]. Web Resource: http://en.wikipedia.org/wiki/AYUSH