COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING
e-ISSN: 2582-7464

Computational Intelligence and Machine Learning
Vol-2, Issue-1, April 2021

A Framework for Ontology Based Semantic Search System in Ayurvedic Medicine

M.Gayathri[1], Dr.R.Jagadeesk Kannan[2]*

[1] Department of Computer Science and Engineering, SCSVMV Deemed University, Kanchipuram, India.
[2] School of Computer Science and Engineering, VIT University, Chennai, India.
*Corresponding Author Email: [2] mgayathri@kanchiuniv.ac.in

Abstract

India is known for its traditional medicinal system such as Ayurveda, Yoga, Unani, Siddha and Homeopathy. Ayurveda plays a significant role in curing the diseases without any side effects. Medicinal plants or herbs are considered as a major resource in meeting the need of people health care. Information about this medicinal knowledge must be preserved and digitized. There have been a massive number of publications and large number of articles on ayurvedic research in the form of unstructured textual data. Text mining approach is used to provide the solution to handle such voluminous of unstructured data. With the exponential growth of text based data, navigating the relevant information needed is the challenging task. Semantic understanding of document content forms the vital requirement for ensuring the quality of content retrieval. However, the current approaches are finding variation in textual classification in bringing the classification accuracy which may fail to understand the data during classification. Hence, an efficient model is required to search, classify and retrieve the most relevant data. The main objective of this research is to develop an effective and efficient framework and algorithm to search and retrieve the most relevant facts by including the application of ontology-based text mining approach. The current status of research is analyzed and reviewed in the area of semantic web retrieval, ontology-based approaches and various classification technique for building the framework. Text mining with the special emphasis on understanding the semantic meaning of content is achieved by using domain ontology called medicinal plant ontology construction. The challenges in finding the semantically related content for the given query are achieved through semantic web and ontology which enriched the data on web for structured representation thereby providing the strong semantics in knowledge representation. The methodology of information extraction is implemented by using medicinal plant ontology with semantic knowledge representation, an algorithm called OCEC (Ontology based Concept Extraction and Classification) was developed where each term is described semantically by mapping the terms and its related terms in the medicinal plant ontology. The web language called Web Ontology Language (OWL) is used for knowledge representation and is considered as richer semantic description language for describing unstructured and semi-structured content on the web thereby extracting the exact and relevant data and to offer a strong semantic search. To evaluate the performance of the proposed method, less relevant and most relevant documents were collected from online sources and digital libraries. Comparative study has been performed with various classification techniques. The experimental results show that the proposed method out performed. To further prove the efficiency of the model, experiments were conducted by giving different queries and the results are compared with other existing methods. The results show that the content retrieved by the proposed model improves precision and recall results.

Keywords

Ayurveda, Ontology, Semantic Web, Traditional Medicine, Web Ontology Language.

INTRODUCTION

In this digital world, the quantity of information produced everywhere leads to a data-intensive world. People in the world are connected by an internet which enables sharing of data with wide variety of interaction and offers global communication. Data is generated by every people and in almost all businesses and processes. This textual data is ubiquitous and growing rapidly. The major problem is that, data is available as unstructured messy data. So, making sense out of raw data and utilizing it for various applications is a challenging task of text mining process. Due to this, there was a need to design and develop methods and techniques which are used effectively to process a range of text based applications. In near past, the problem of text mining has increased much attention to process the huge amount of textual data and lot of research study was done to discover the useful knowledge from this voluminous content [1]. Ayurveda has gained much of significance in the last two decades. Many people are practicing this medicine in many parts of the world [2]. Ayurveda medicine can be followed and adapted in different countries with suitable modifications by considering its tradition, culture and constitutional profile of their population. Medicinal plants or herbs are considered as a major resource in meeting the need of people health care. A recent study pointed out that there are more than 13,000 plants has been examined. There are various medicinal plants possess high medical values [3]. The following table 1 shows some of the medicinal plants along with their usage.

Table 1 : Some of the Medicinal Plants and its Use

Botanical Name	Common Name	Parts Used	Medicinal Uses
Swertia chirayita	Chirata	Whole plant	Skin disease, burning sensation, fever.
Ocimum sanclum	Tulsi	Leaves, root	Cough, cold, bronchitis
Grit kumari	Aloe vera	Leaves	Laxative, wound healing, skin burns and care, ulcer.
Azadirachta indica	Neem	Rhizome	Sedative, analgesic, epilepsy, hypertensive.
Saraca assoca	Ashok	Bark, flower	Menstrual pain, uterine, disorder, diabetes.
Commiphora wightii	Guggul	Gum rasine	Arthritis, paralysis, laxative.
Piper longum	long pepper	Fruit, root	Appetizer, enlarged spleen, bronchitis, cold, antidote.
Asparagus racemosus	Shatavari	Tuber, root	Enhance lactation, general weakness, fatigue, cough.
Terminalia chebula	Haridra	Seed	Wound, ulcers, leprosy, inflammation, cough.

PROBLEM DEFINITION

As pointed earlier, Ayurveda plays a major role in human life and it use herbs for treatment of diseases without causing any side effects. Information about this medicinal knowledge must be preserved because it exists from Vedic period. Number of publications and articles in ayurveda are published in the form of documents [4]. Hence, an efficient model is required to search and know the most relevant facts from these documents. In existing search engine, the searching is based on the keyword, it produces the result which may or may not be relevant to the user. For example, if the user gives the query as “Apple”. The results seem to be generic. It shows the contents about the fruit apple and also the apple product like I-phone, I-Pod, etc. The user intention may not be understood semantically. The main limitations of the web search today are,

- The searching results have high recall, low precision.
- Results are sensitive towards vocabulary.
- Results are single web pages.
- Contents on the web are not structured.

Hence, a technique is required to improve the efficiency of the model in order to search and retrieve the most relevant data. For understanding the information and to dig the useful and relevant ayurvedic content there should be a proper representation of knowledge. The main drawback of search engine today is that it does not understand the meaning of web content and it is not machined accessible. Although there are many tools available for retrieving text, the capabilities of such tool are still considered to be limited for retrieving the relevant information by understanding the content. It is quite difficult for machines to distinguish the meaning of “bank” when used in different context at different places. Recent days, Ontology is offering an explicit specification of a shared knowledge and conceptualization. Now, it becomes a modeling technique to represent information. Ontology typically consists of concepts, properties, instances and axioms. Different languages such as RDFS and OWL were utilized to represent ontology in a machine-readable form. Ontology ensures strong semantics in knowledge representation for many applications. Ontology and semantic web plays a vital role in deep understanding of information on the web. Ontology forms the backbone of the semantic web. It is considered to be important on the semantic clarity of concepts and entities which greatly improves and enhance the information analysis, sharing, retrieval and reuse [5] Therefore semantic web is used as an alternative approach for representing the web content which is easily machine accessible by using intelligent techniques.

EXISTING SYSTEM

[6] presented an application called semantic search. This was developed to overcome the drawbacks of existing search and designed to improve the traditional search. They discussed different supporting technologies to search for the data which is semantically related. [7] presented a new semantic search engine framework. It provides effective search by solving the issues on querying the semantic web. [8] gave an overview about semantic search engine. They developed semantic search engine based on ontology. Domain ontology for tourism was built with tourism information which makes use of synonym dataset by using WordNet to understand the meaning of the input queries in the searchable interface. [9] proposed an efficient semantic search engine framework. They developed engine by using new semantic ranking algorithm which operates over a sorting RDF triples. The framework typically includes four phases such as crawling, indexing, ranking and retrieval. It also included with enhanced crawling algorithm which crawls the relevant content from the web with minimal overhead. [10] Proposed an approach to enhance the performance of the information retrieval process from the digital library by using ontology. The retrieval system is based on ontology which uses ontology based annotations to retrieve the relevant result and recommend related topics. They evaluated and reviewed the performance of the ontology based search by comparing with keyword search. [11] proposed ontology based semantic search engine. It includes two types of search such as keyword-based search and a semantics-based search. This search engine works with different technologies such as domain ontology and RDF for the representation of data. They compared their search engine

results by submitting queries to the system and measured their performance and efficiency by submitting the same set of queries to other existing search engine like Kngine, Wolfram Alpha and Google. The limitations of the existing work includes the following 1.There was not much work focused in the domain - Ayurveda for retrieving the relevant information related to this area of medicine. 2. Construction of medicinal plant ontology plays a significant role in understanding the concepts and its related concepts in that particular domain. Most of the researchers used the existing ontology which may fail to understand their point of focus and indent of knowledge discovery. 3. Text classification is considered as one of the main tasks included in text mining process for information retrieval. It forms the supervised machine learning process for document classification. Semantic understanding of document content forms the vital requirement for ensuring the quality of content retrieval. However, the current approaches are finding variation in textual classification in bringing the classification accuracy which may fail to understand the data during classification.

PROPOSED METHOD

In order to define an effective model to retrieve the most relevant content from ayurvedic document and to provide semantic description to the document by mapping the concept from medicinal plant ontology and to improve the performance of the classification task, ontology based semantic retrieval system is proposed and viewed as layer. The architecture of proposed framework is shown in “Fig. 1”. It consists of five phases,

1. Data collection
2. Pre-processing and classification
3. Ontology development and reasoning
4. Knowledge representation
5. User search query

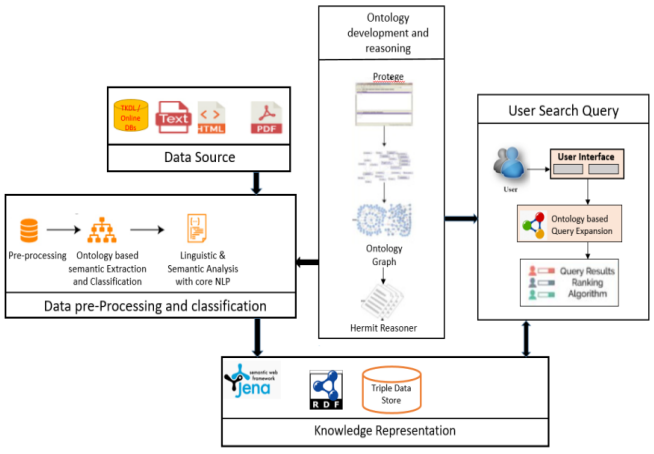

Figure 1. Architecture of Proposed System

The Indian herb based documents were collected from the TKDL, PubMed, NCBI and other online digital repositories. Pre-processing is done to transform the text into something an algorithm can digest is a complicated process. It forms the important task before classification, as this process is used to remove unwanted information from the document and to reduce the document size and appropriate content will be considered for the processing [12]. The pre-processing steps such as Tokenization, POS tagging, stop word removal, Stemming are done on the collected documents. Domain ontology plays a vital role in integrating the semantic features with the information retrieval. This kind of integration is needed to extract synonyms, polysemy words that are the term relevant to the important terms. Domain Ontology called medicinal plant ontology was constructed [13]. Based on this, the ontology based classification algorithm is proposed which forms the basis for semantic data classification. We used algorithm called “Ontology based Concept Extraction and Classification (OCEC)” for learning the semantic aspect of the classification using medicinal plant ontology and for classification of every new document [14]. To represent the document and for extracting the features TF-IDF weighting was used. Before identifying the weighting, it includes the semantic description of documents. This helps in identifying the terms as concepts in the document. Thus the documents are represented as vector of concepts. Each term is described semantically by mapping the terms and its related terms in the medicinal plant ontology. Thus it involves the process of checking whether the concept is present in medicinal plant ontology. Now the concept is extracted and concept weighting is calculated for finding how important and relevant the document is. Classification process is carried out by using kNN classifier to assign documents to predefined classes. Hence the ontology based extraction and classification system able to understand the data during retrieval process, thereby enhancing the quality and consistency of the overall classification process.

RESULTS AND DISCUSSIONS

Different search techniques like keyword-based search and ontology-based Search were compared with the proposed method. Keyword-based search algorithm works based on syntactic search that contains deficiencies. However, the algorithm does not support for semantic information retrieval effectively. It concentrates on enriching the information retrieval semantic annotation but the input herb query are not explicitly available in the text. Therefore the method are unable to display the results. It also lacks in deep retrieval inference in library to extract the content. For example, searching for herb “Triphala” and “Terminalia bellirica” often gives different information although both queries are semantically related. Even though, the method tries to focuses on semantics model and semantic search, it does not address knowledge representation system. Ontology-based search algorithm by introducing the method which works with the integration of semantic search and ontology based reasoning. Where, semantic web search are widely used in digital libraries for information retrieval systems. This method work to expresses that how information retrieval system can be improved. Here, the method displays the result

in ontology class representation and as well as keyword-based information retrieval systems. It is used for prioritizing disease and symptoms complexes for an herb interest and the algorithm relies on prior phenotypic information, and able to find herb similarity based on disease symptoms. The method only retrieved the information which is linked with the ontology instance. It is unable to recognize synonyms of keywords. However, it does not consider the herb knowledge representation for mapping the relationship between herbs and disease. Proposed method finalizes the semantic evaluation process through utilization of system relevance to the users by mapping with medicinal plant ontology. In detail, each document is displayed by two similarity values. The proposed method applies the ranking of retrieved content based on distance of the document and the query where the lowest distance indicates that retrieved content has most relevant result. The ranking function states that optimum retrieved result is given if the retrieved herb information resources are prioritized according to their probability that resource is relevant to the user query. The results proved that, the proposed system bring most promising and prominent output compared with other different kinds of search as shown in table 2. From "Fig 2", it indicates that, number of retrieved contents by the proposed method was more relevant and the number of relevant contents retrieved for the given queries was high when compared with other existing techniques.

Table 2 : Comparison of Recall and Precision Value on Various Search Techniques

S.No	Queries	keyword-based search		ontology-based search		proposed system	
		Recall	***Precision***	***Recall***	***Precision***	***Recall***	***Precision***
1	Indian gooseberry	0.92	0.82	0.82	0.79	1	0.85
2	Arjuna	0.45	0.55	0.9	0.42	0.97	0.75
3	winter cherry	0.29	0.29	0.7	0.62	0.95	0.87
4	Fenugreek	0.87	0.42	0.94	0.44	1	0.75
5	Lemon Grass	0.94	0.1	1	0.65	1	0.8
6	Neem	0.66	0.29	0.83	0.66	0.99	0.79
7	Tulsi	0.89	0.38	0.78	0.73	0.92	0.88
8	Caraway	0.59	0.26	0.79	0.68	0.83	0.79
9	Clove	0.95	0.78	0.95	0.47	1	0.65
10	Ginger	0.4	0.1	0.89	0.67	0.9	0.97
11	Garlic	0.9	0.23	0.93	0.95	0.75	0.99
12	Lemon Balm	0.5	0.45	0.89	0.76	0.95	0.59
13	Mints	0.19	0.13	0.89	0.64	0.79	0.47
14	Cinnamon	0.85	0.24	0.5	0.63	0.95	0.73
15	Rosemary	0.49	0.29	0.9	0.82	0.94	0.83
16	Ashwagandha	0.78	0.14	0.92	0.56	1	0.83
17	Fennel	0.85	0.69	0.86	0.7	0.95	0.85
18	turmeric	0.95	0.5	0.83	0.93	1	1
19	Triphala	0.83	0.57	0.83	0.82	0.95	0.93
20	Asafoetida	0.93	0.5	0.95	0.6	0.79	1
Average Recall & Precision		**0.711**	**0.386**	**0.855**	**0.677**	**0.931**	**0.816**

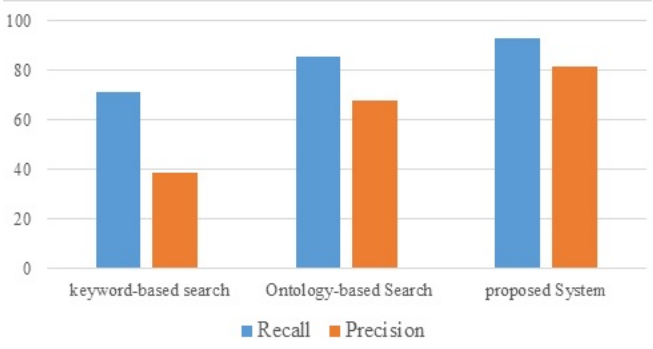

Figure 2. Comparative Analysis of Different Search Techniques with Proposed Method

CONCLUSION

An effective and efficient framework is proposed to search and retrieve the most relevant facts by including the application of ontology based text mining approach. Medicinal plant ontology is constructed which greatly helps in understanding the knowledge in that domain. This knowledge is stored in the OWL which greatly results in strong semantic representation of that knowledge. New method of classification is followed to automatically identify the herb related information from the ayurvedic biomedical literature. The documents are semantically described which enhance the classification technique thereby producing the

most appropriate results. To provide human like understanding of system, deep learning algorithm may be applied to further improve the performance and accuracy. Many open source latest tools like orange, KNIME can also be used to apply different algorithm with different perspective. A maximum number of text mining measures can be implemented and graphical visualization can be enhanced in a better way. Current technology such as cloud and mobile computing are used increasingly for a wide variety of applications. These technologies bring a new set of challenges like security and privacy. Text mining techniques can be extended to profit from the positives and avoid or workaround the drawback of these systems.

REFERENCES

[1] Feldman, Ronen and Sanger, James. The Text Mining Handbook: Advanced Approaches, 2006.
[2] Kurup, P. N. V. 'Ayurveda - A potential Global Medical system', In Mishra LC editor, Scientific Basis for Ayurvedic Therapies, New York, CRC Press, pp 1-15, 2004.
[3] Dahanukar, S. A., Kulkarni R.A., and Rege N. N. (2000) 'Pharmacology of Medicinal Plants and Natural Products, Indian Journal of Pharmacology, Vol 32, No.4, S81–S118.
[4] Chaussabel D. 'Biomedical Literature Mining', American Journal of Pharmacogenomics, Vol 4, No.6, PP 383-393, 2004.
[5] Grimm S., Abecker A., Volker J and Studer R. Ontology and Semantic Web, 2011.
[6] Guha, Ramanathan, Rob McCool, and Eric Miller. 'Semantic Search', In Proceedings of the 12th International Conference on World Wide Web, ACM, pp 700-709, 2003.
[7] Fatima, Arooj, Cristina Luca, and George Wilson. 'New Framework for Semantic Search Engine', UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, IEEE, pp 446-451, 2014.
[8] Laddha, Shilpa S., and Pradip M. Jawandhiya. 'Semantic Tourism Information Retrieval Interface', In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, pp 694-697, 2017
[9] El-gayar, M. M., N. Mekky, and A. Atwan. 'Efficient Proposed Framework for Semantic Search Engine Using New Semantic Ranking Algorithm', International Journal of Advanced Computer Science and Applications, Vol 6, No.8, 2015.
[10] AlAgha, Iyad M, Mohammed GH Al-Masri. 'An Ontology Based Approach to Enhance Information Retrieval from Al-Shamelah Digital Library', IUG Journal of Natural Studies, Vol 24, No.1, 2016.
[11] Sayed, Awny, and Amal Al Muqrishi. 'IBRI-CASONTO: Ontology-based Semantic Search Engine', Egyptian Informatics Journal, Vol 18, No.3, pp 181-192, 2017
[12] Mohbey, K. K., and Tiwari, S. 'Pre-processing and Morphological Analysis in Text Mining', International Journal of Electronics Communication and Computer Engineering, ISSN, 2011.
[13] Gayathri, M., and R. Jagadeesh Kannan, 'Construction of Domain Ontology for Traditional Ayurvedic Medicine', In Intelligent Computing in Engineering, pp. 417-426. Springer, Singapore, 2020.
[14] Gayathri, M. and Kannan, R.J., 'Ontology based Concept Extraction and Classification of Ayurvedic Documents', Procedia Computer Science, 172, pp.511-516, 2020.