



# INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

## DOMAIN ONTOLOGY BASED RECOMMENDATION USING MULTI-LABEL TEXT CATEGORIZATION APPROACH

<sup>1</sup>Barath Callistus.P,<sup>2</sup>Ramesh.P,<sup>3</sup>Dr.T.Nalini.

<sup>1</sup>Student,<sup>2</sup>Student,<sup>3</sup>Professor.

<sup>1</sup>Department Of Information Technology.

Dr. M.G.R Educational and Research Institute, Chennai, Tamilnadu, India.

**Abstract** - As deep web grows at a very fast pace, there has been increased interest in techniques that help efficiently locate deep-web interfaces. An exploratory search may be driven by a user's curiosity or desire for specific information. When users investigate unfamiliar fields, they may want to learn more about a particular subject area to increase their knowledge rather than solve a specific problem. A matching query style has significant limitations. Search results are satisfactory only when users give the right search words. To achieve more accurate results for an exploratory search, Smart Crawler ranks websites to prioritize highly relevant ones for a given topic. We evaluate the effectiveness of the proposed technique Ontology Recommendation Framework, and report the findings (e.g., the importance of context and content factors) in web revisitation Ontology Recommendation Framework delivers the best re-finding quality in finding rate. In proposed system the multi-key word search concept will be used, the system will be giving all the possible relevant links. This will be achieved in two ways the proposed framework by considering numerous ontologies from different domains and more user requirements while focusing on the other multi-label text categorization approaches. The query which is submitted to the application will be preprocessed, after pre-processing only root words will be taken and it will find Synonym, Hypernym and Hyponym and it will listed to the user so this is the reason that all possible links can be found related to search. Our dynamic management of context and content memories including decay and reinforcement strategy can mimic users' retrieval and recall mechanism If any words in that displayed list are selected then all the website links, images and news feeds will be given as final output to the user. Then the book mark concept is included that is the book marked link will be added to the application directly not to the browser so the bookmarked content will visible globally.

**Index Terms** - Synonym , Hypernym , Hyponym , Ontology, Pre-Processing.

## I. INTRODUCTION

To cluster the text documents over the web page based on the user typed key term. To enhance deep web search (ontology) and overcome grouping of unrelated documents into the same cluster. Aims to help Web users locate the best search tools for their search needs, resulting in faster and more accurate search results. We present work assumes that all user local instance repositories have content-based descriptors referring to the subjects, however, a large volume of documents existing on the web may not have such content-based descriptors. For this problem we strategies like ontology mapping and text classification/clustering were suggested. These strategies will be investigated in future work to solve this problem. The investigation will extend the applicability of the ontology model to the majority of the existing web documents and increase the contribution and significance of the present work.

## II. LITERATURE REVIEW

SEMANTICS-BASED AUTOMATED SERVICE DISCOVERY [1] Aabhas V. Paliwal, Basit Shafiq, Jaideep Vaidya, Hui Xiong and Nabil Adam A large number of web services structure a service oriented architecture and facilitate the creation of distributed applications over the web. These web services offer various functionalities in the areas of communications, data enhancement e-commerce, marketing, utilities among others. Some of the web services are published and invoked in-house by various organizations. These web services may be used for business applications, or in government and military. However, this requires careful selection and composition of appropriate web services. Hongyan Liu, Jun He, Dan Zhu, Charles X. Ling, and Xiaoyong Du [2] MEASURING similarity between objects is a fundamental task in domains including data mining, information retrieval, and so on. Many applications require operations such as “find similar objects.” We are expected to recommend items that the user may prefer. Then, if we know the similarity between users, we can recommend items that similar users may also prefer. The application of similarity measures is not limited to this task. Changjun Jiang, Haichun Sun, Zhijun Ding, Pengwei Wang and MengChu Zhou [3] Internet data are heterogeneous, redundant, disordered, and exponentially growing. Finding the right information from them becomes an ever-challenging issue. an indexing network model is proposed that organizes information in webpages at three levels: words, webpages, and categories, thereby leading to a semantic association graph. Words are used as the description of webpages and categories. Webpage classification is used to gather similar web pages together. HaiChun Sun, ChangJun Jiang, ZhiJun Ding, PengWei Wang and MengChu Zhou [4] An exploratory search may be driven by a user’s curiosity or desire for specific information. When users investigate unfamiliar fields, they may want to learn more about a particular subject area to increase their knowledge rather than solve a specific problem. However, a matching query style has significant limitations. Search results are satisfactory only when users give the right search words. Jian Wu, Liang Chen, Yipeng Feng, Zibin Zheng, Meng Chu Zhou, and Zhaohui Wu [5] Quality-of-service-based (QoS) service selection is an important issue of service-oriented computing. A common premise of previous research is that the QoS values of services to target users are supposed to be all known. However, many of QoS values are unknown in reality. SOC attracts industry’s attention and is applied in many domains. With a growing number of alternative Web services that provide the same functionality but differ in quality properties, the problem of selecting the best performing candidate service is becoming more and more important. Xiaocao Hu, Zhiyong Feng, Shizhan Chen, Keman Huang\*, Jianqiang Li and Mengchu Zhou [6] ONTOLOGY has been proven effective for providing the formal, explicit, and shared conceptualization of common knowledge of a particular domain, and has been widely used in many disciplines, such as biomedical engineering, knowledge engineering, Web services and Restful APIs .Maria-Luiza Antonie, Osmar R. Zai’ane [7] Automatic text categorization has always been an important application and research topic since the inception of digital documents. The text classification task can be defined as assigning category labels to new documents based on the knowledge gained in a classification system at the training stage. In the training phase we are given a set of documents with class labels attached. and a classification system is built using a learning method. Aabhas V. Paliwal, Nabil R. Adam, Hui Xiong and Christof Bornhövd [8] A large number of web services are being developed as an emerging standard to construct distributed applications in the web. Service requesters have access to a choice of descriptions to various services that provide similar service functionality. Most current approaches for web service discovery cater to semantic web services, i.e., web services that have associated semantic descriptions. Pierluigi Plebani and Barbara Pernici [9] A SERVICE-ORIENTED Computing aims to provide a set of methods and tools to support the design and the execution of applications based on Web services. According to this paradigm, at design time, developers identify which activity is to be performed and then try to find and select the Web services closest to such requirements. PengCheng Xiong, YuShun Fan and MengChu Zhou [10] SERVICE-ORIENTED architecture has evolved to become a promising technology for the integration of disparate software components using Internet protocols. Web Services Description Language (WSDL), Universal Description, Discovery, and Integration (UDDI), and Simple Object Access Protocol are three core standards for Web service description, registration, and binding, respectively. Most of the time, multiple Web services need to invoke each other dynamically to accomplish service requestor’s manifold requirements.

### III. PROPOSED MODEL

This paper proposes a practically efficient and flexible searchable scheme which supports both multi-keyword ranked search and synonym based search. To address multi-keyword search and result ranking, Vector Space Model (VSM) is used to build document index, that is to say, each document is expressed as a vector where each dimension value is the Term Frequency (TF) weight of its corresponding keyword. A new vector is also generated in the query phase. The vector has the same dimension with document index and its each dimension value is the Inverse Document Frequency (IDF) weight. Then cosine measure can be used to compute similarity of one document to the search query. To improve search efficiency, a tree-based index structure which is a balance binary tree is used. The searchable index tree is constructed with the document index vectors. So the related documents can be found by traversing the tree.

Framework employs text categorization and unsupervised learning techniques. The benefits of the proposed framework are twofold: 1) ontology organization according to the opinion of domain experts and 2) ontology recommendation with respect to user requirement. Moreover, an evaluation model is also proposed to assess the effectiveness of the proposed framework in terms of ontologies organization and recommendation. The main consequences of the proposed framework are 1) ontologies of a corpus can be organized effectively, 2) no effort and time are required to select an appropriate ontology, 3) computational complexity is only limited to the use of unsupervised learning techniques, and 4) due to no requirement of context awareness, the proposed framework can be effective for any corpus or online libraries of ontologies.

### IV. SYSTEM ARCHITECTURE

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods

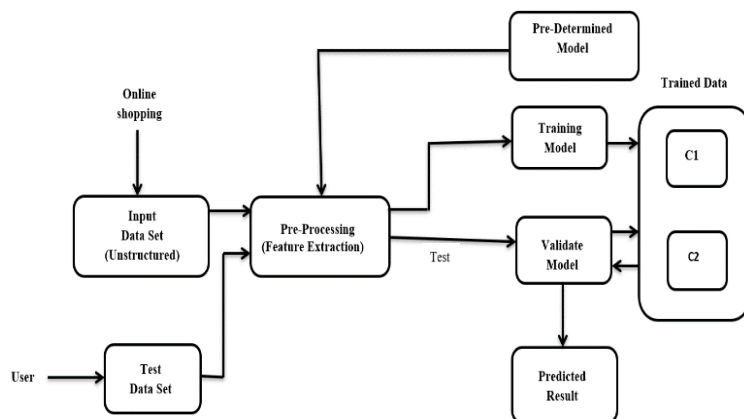


Figure 1: Architecture Diagram

### V. MODULE DESCRIPTION

- Blind User Interface
- Search space
- Input from User
- Data Preprocessing
- Stop word Removal
- Ontology Clustering
- Multi-term Search
- Cluster the Most Relevant Content

## USER INTERFACE

### SEARCH SPACE

- After user login process, web user can enter the search space page.
- This is the environment for user to search the content from the web server.
- This Search Space is the interface for user and web servers.

### INPUT FROM USER

- Get the input text from the user for the search process.

### DATA PREPROCESSING

#### STOP WORD REMOVAL:

Stop words are words which are filtered out prior to, or after, processing of natural language data (text). It is controlled by human input and not automated. These are some of the most common, short function words, such as the, is, at, which and on.

### ONTOLOGY CLUSTERING

#### Nym's Group:

Words ending in **nym's** are often used to describe different classes of words, and the relationships between words.

- **Hypernym:** A word that has a more general meaning than another.
- **Hyponym:** A word that has a more specific meaning than another.
- **Synonym :** One of two (or more) words that have the same (or very similar)

#### Text Analysis:

The Artificial-Intelligence literature contains many definitions of ontology (Word net).

- It includes machine-interpretable definitions of basic concepts in the domain and relations among them.
- The featured results produced by the sentence-based, document-based, corpus-based, and the combined approach concept analysis have higher quality than those produced by a single-term analysis similarity.

### MULTI-TERM SEARCH

Get the multi-term input from the user and it will search the keyword one by one and get the relevant content from the web servers. Our system get the search result deeply from the search engines and its search the terms randomly till last key term in that multi-term list.

#### 1. CLUSTER THE MOST RELEVANT CONTENT

From the multi-term search result we cluster the more relevant content based on the relationship user input term. And we classify the cluster and give the final output like most relevant content comes first and out comes next output screen.

## VI. CONCLUSION

This paper has presented an OTMM for grouping of research proposals. Research ontology is constructed to categorize the concept terms in different discipline areas and to form relationships among them. It facilitates text-mining and optimization techniques to cluster research proposals based on their similarities and then to balance them according to the applicants' characteristics. The experimental results at the NSFC showed that the proposed method improved the similarity in proposal groups, as well as took into consideration the applicants' characteristics (e.g., distributing proposals equally according to the applicants' affiliations). Also, the proposed method promotes the efficiency in the proposal grouping process.

## REFERENCES

- [1] A. V. Paliwal, B. Shafiq, J. Vaidya, H. Xiong, and N. Adam, "Semantics based automated service discovery," IEEE Trans. Services Comput., vol. 5, no. 2, pp. 260–275, Apr./Jun. 2012.
- [2] H. Liu, J. He, D. Zhu, C. X. Ling, and X. Du, "Measuring similarity based on link information: A comparative study," IEEE Trans. Knowl. Data Eng., vol. 25, no. 12, pp. 2823–2840, Dec. 2013.
- [3] C. Jiang, H. Sun, Z. Ding, P. Wang, and M. Zhou, "An indexing network: Model and applications," IEEE Trans. Syst., Man, Cybern., Syst., vol. 44, no. 12, pp. 1633–1648, Dec. 2014.
- [4] H. Sun, C. Jiang, Z. Ding, P. Wang, and M. Zhou, "Topic-oriented exploratory search based on an indexing network," IEEE Trans. Syst., Man, Cybern., Syst., vol. 46, no. 2, pp. 234–247, Feb. 2016.

- [5] J.Wu et al., “Predicting quality of service for selection by neighborhoodbased collaborative filtering,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 43, no. 2, pp. 428–439, Mar. 2013.
- [6] X. Hu et al., “Accurate identification of ontology alignments at different granularity levels,” IEEE ACCESS, 2016, doi: 10.1109/ACCESS.2016.2614759.
- [7] M.-L. Antonie and O.R. Zaane, “Text Document Categorization by Term Association,” Proc. IEEE Int’l Conf. Data Mining (ICDM ’02), 2002.
- [8] A.V. Paliwal, N. Adam, H. Xiong, and C. Bornhoevd, “Web Service Discovery via Semantic Association Ranking and Hyperclique Pattern Discovery,” Proc. IEEE/WIC/ACM Int’l Conf. Web Intelligence, 2006.
- [9] P. Plebani and B. Pernici, “Urbe:Web service retrieval based on similarity evaluation,” IEEE Trans. Knowl. Data Eng., vol. 21, no. 11, pp. 1629– 1642, Nov. 2009.
- [10] P. Xiong, Y. Fan, and M. Zhou, “A petri net approach to analysis and composition of web services,” IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 40, no. 2, pp. 376–387, Mar. 2010.

