



Semantic Retrieval Using Fuzzy C Means Clustering for Geography Ontology

¹Chetna Devkar , ²Ms. Preeti Jain, (Sr. Assistant Professor)

^{1,2}Department of Computer Science & Engineering
Acropolis Institute of Technology & Research
Indore, India

Abstract: A research for semantic information retrieval has been accessible. The Semantic Web is a novel technique for systematize information and it correspond to a huge interest area for the worldwide research society, except it is at rest far as of a significant accomplishment. In this research work we have anticipated a novel algorithm for information retrieval base on ontologies, lively semantic network with lexical chains important a technique for score and ranking consequences by means of a narrative metric to determine semantic relatedness among words. Our technique has a number of newness, in exacting concerning of a common knowledge base from which we take out specific domain ontologies. Furthermore, the consequences of our experiment are capable and support novel efforts in this direction. In exacting we are permit for the prospect of instigate several form of normalization for the semantic module with deference to the length of lexical chains and we are improving the investigation accuracy of our system evaluate with other technique. At nearby, other associated topic could be investigate:

- Essential a prescribed linguistic replica to signify our ontologies.
- With standard language to illustrate ontologies and modified their reusability and distribution.
- Infer applicable documents and associated terms to have particular ontologies.

Keywords: MCL, RDF, FCM, HTML

I. INTRODUCTION

Huge quantity of data are habitually composed and store in the collection of lots of organization. But data in dissimilar organizations are multifaceted in nature and frequently inadequately prepared and reproduce and be present in dissimilar formats. Successfully develop these huge volumes of data is attractive a main confront for this category of industry. They exhibit the require of warehousing and data mining technologies in the geography domain[1]. Ontology is a knowledge base which describes conception, attributes and associations in a precise domain with open stipulation that feature interoperability among human and machine able to resolve the problems of ambiguity in information sharing and reprocess. Appropriate to its strength in attractive knowledge illustration sharing and reusing [2], ontology has establish extensively applications in region like knowledge management [3], database design [4], in sequence retrieval and in sequence integration [5][6]. Ontology is an necessary measurement of several applications and the expansion and relevance of ontology technology open a novel way near knowledge sharing and reuse. Sustain by ontology together the user and the system can converse with each other with a frequent considerate knowledge of a domain. For the significance of ontology, ontology assembly has been regard as a important concern. Ontology creation is a extensive, expensive and contentious work. Consequently, numerous revise for semi-automatic or automatic ontology creation technique have emerged. For instance anticipated an technique to semi-automatically create domain ontology base on indian word divider and data mining. The proposed method is prove to be effectual in construct domain ontology, and as well certain the quality of the ontology at a positive level based technique for automatic building of domain ontology from domain quantity. In this technique, every document in the compilation is representing by a graph. Following the production of document graphs 'n' the problem. For a number of application this might be significance the cost, i.e. invest a group of time and attempt in construct a ideal ontology, except in a lot of cases this is not possible. The benefits of have a knowledge based or ontology based request have to be reasonable next to the attempt of constructing and preserve the ontology in

difficulty. Though ontologies arbitrary march term weighting is working to estimation the significance of the information of a expression to the quantity from both local and inclusive perception. Subsequently, the Markov Clustering (MCL) algorithm is use to disambiguate stipulations with different meanings and collection similar terms to generate concepts. After that, an enhanced Fuzzy C Means (FCM) algorithm controlled by both vertices and in formativeness is oppressed to discover chance latent relatives amongst these concept. In conclusion, the domain ontology is productivity in the Resource Description Framework (RDF) format.

Recommend a thought relation searching technique that combine the individuality of middle-out and top-down technique in a procedure that remind you of snowflakes crystallization stand on the crystallizing thought examination technique, this study implements an ontology building method that can repeatedly mine domain perception out of domain document,

decide relationships among thought, and create the domain ontology. Therefore, though, despite the theories and technologies organism used, automatic ontology building have forever concerned three most important building processes, perception extraction, and concept relations exploration and document preprocessing. Document preprocessing refers to filter out noises in documents to keep significant terms, concept extraction refers to mine domain thought out of vocabulary and perception relations examination refers to mining relations among concepts and organize them to conclude the ontology building process. In the processing of ontology creation, relations among concept and the behavior concept are organized by their relations pressure the ontology structure, which in turn influence the accuracy of domain knowledge. Therefore, perception relations investigation is the almost all significant process of ontology conception. The available perception relation searching process mostly follows the three advances: top-down, bottom-up and middle-out. Every of these technique has its possess strengths and weaknesses. A bottom-up technique identify first the nearly all outstanding perception and generalizes them into added abstract concepts, though a bottom-up technique discover it hard to spot unity among correlated concepts. A top-down technique starting at the top can consequence in choose and magnificent arbitrary high level group. A middle-out technique identifies the core of essential terms, and then specifies and generalize them. The technique by difference strikes a balance in conditions of the level aspect and requires less re-work, which as well leads to less in general effort.

This study recommends a context-based ontology construction technique for extract geography examination domain ontology from unstructured indian text documents. The proposed technique consist of the steps of domain documents preprocessing

- Conception clustering based on the fuzzy c-means.
- Environment extraction.
- Domain ontology creation.

According to the context-based ontology structure technique, the planned technique of ontology construction method in this study that can repeatedly mine domain thought out of domain document establish relationships among concept and build the domain ontology. Therefore, thus reducing cost and load that would be incur in a guide building process.

Ontology is a understandable explanation which is used to construction domain knowledge and set up unified terminologies and the relationships among the terms associated to applicable domain. With this category of unified terminologies and associations can appreciate data sharing and reusing in dissimilar systems and the realistic and valuable storage of fuzzy ontology can make sure ontology allocation.

On the conditions of knowledge intend at problems such as complex in order sources, scattered storage, unsaved fuzzy acquaintance and so on which live in the field of knowledge supervision at present, this editorial accept the fuzzy ontology technique to store knowledge. Specified the compensation of relational database in data association and supervision, in this work, a storage technique using fuzzy c-means clustering ontology base on relational database is planned. Primary, we shall for a moment bring in illustration technique of fuzzy data types by extend the RDF data type and after that explain the retrieve methods of ontology and the storage technique of structure and illustration of fuzzy ontology in the relational database.

II. MOTIVATION

When ontologies are create the similar knowledge achievement bottleneck transpire as it has in every one knowledge gaining efforts since the materialization of knowledge based system and ontologies. It is source and time intense to build ontologies and together the ontology engineers and domain specialist concerned require being extremely skilled and if possible having knowledge concerning every others view have been construct since the early on days of artificial intelligence, such ontologies were not general in typical software systems. At present additional and additional semantic application emerge and what in progress out as researchers construction 'geography' ontologies has currently develop into a group that is concerning to appreciate the vision of the Semantic Web. The pathway of a huge deal computer science investigate is to build improved and improved application that resolve novel problems or resolve problems enhanced. Now a days standard web users and software developers desire to construct ontologies at the same time the request are added complex than presently a few years ago. This is particularly true in the Semantic Web context anywhere the entire web is the request area of ontology-based systems and concern such as scalability and efficiency situate higher requirements as well on the ontologies. The amount of the ontologies to be build can as well be a problem. In answer to this disseminated development methodologies, modularization and decoupling of mechanism as well as technique for support the users in the nearly all recurring and time-consuming tasks of ontology engineering have emerged, still additional vital is the topic of ontology preservation and development. In a greater part the world around the function i.e. the source for the circumstance of the ontology is not constant it is continually varying and developing. To advantage from an ontology-based request the ontology have to change at the similar rate as the environment. If the alter rate is comparatively deliberate then manual evolution and change tracking of the ontology capacity be a rational option other than in highly agile environment where alter happen often and are not forever easy to detect and track merely guide methods are almost certainly not sufficient.

III. PROPOSED METHODOLOGY

In our collection Fuzzy Clustering technique we assume a fuzzy membership function that will attend to the degree of uncertainty for everyone consequential alignments. At this time an alignment will have a relationship value for every of the classes and all the relationship values of a meticulous association will be further up to one which designate the valid

continuation of the alignment. Figure 1 represent our replica of collection Fuzzy Clustering. In arrange to construct MFs repeatedly clustering was added capable than classification technique as there was no class label connected among the alignments. Also, MFs necessitate fuzzy or extend beyond class limitations which are unavailable in generic clustering methods. We establish Fuzzy C Means (FCM) clustering give overlap class precincts where a data occurrence can fit in to added than one clusters. We useful FCM on the connection calculate values provide by the alignment tools and get the membership values for every of the alignments. We as well get the center values of individuals clusters. Figure 1 illustrate how the membership functions are creature generate from 'n' alignment tools. If there are three cluster, we can simply differentiate the cluster with highest center importance as the cluster of nearly all eminence alignments. In figure 1 Membership functions (p) produce from every of the tools. For a extracting cluster we can merge the 'n' MFs create from 'n' tools by be appropriate the period Type 2 Fuzzy Logic theory. In our technique for exacting cluster we create, everywhere the membership degree stage of a single understanding was the period of its practical maximum and bare minimum membership values. We will observe illustration of building the case study segment the present the difficulty is how we can conclude the most excellent alignments from the residential in Fig 1. An alignment have longer distance represent that the association tools did not concur in the lead its might, so the alignment that has longer interval is additional uncertain than the association with shorter interval. This instinct

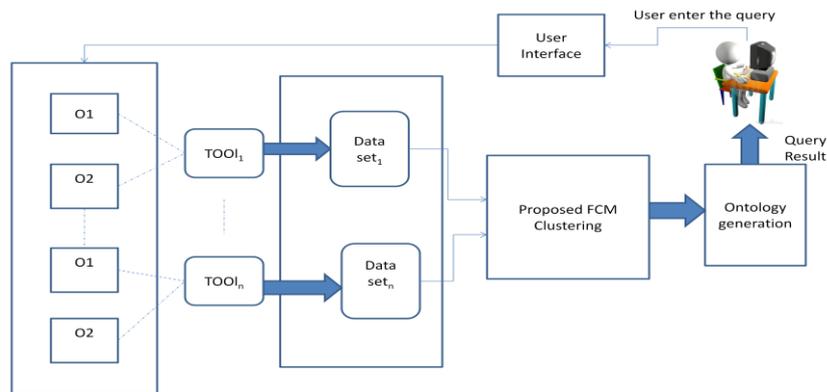


Figure:1 Our Proposed System Architecture.

make membership functions by design by apply FCM on the consequences of the alignment tools. Here n is the numeral of position tools, p is the quantity of clusters or the relationship functions. O1 and O2 are two ontologies organism associated gives us a technique of evaluate the alignments base on their reservations. Yet again, the alignment that has superior interval denote is stronger as it indicate that the majority of the alignment ontology recommended this single elevated correspondence determine. So this alignment is nearly every one probable to be a sturdy one. At this top we desirable to expand an equation that stability the two perception of indecision (interval width) and strength (interval mean). We begin probable as a superiority metric of an configuration. The equation 1 is a biased harmonic denote method that give additional weight on the gap mean than the interval (width). The motive of with weighted harmonic mean in its place of generic single is that instinctively it is extra important to have superior interval mean quite than have lower indecision. If an configuration show higher interval signify it is additional probable to be in the strong cluster.

$$P_i = (1 + \beta^2) \frac{\text{mean}_i * \text{interval}_i}{\text{mean}_i + \beta^2 \text{interval}_i} \quad \text{eq.(1)}$$

Wherever Pi is the possible of the arrangement with index i, β is a steady, Intervali is the gap of the ith alignment, Meani is the reckoning mean of Intervali. The possible value reflects both the indecision and the strength of an association. Consequently, the best alignment has the maximum mean and slightest interval(width).In our apparition we can have web search improvement with a hybrid technique that take into description together syntactic and semantic in sequence in a system that has as a prospect of knowledge ontology. We recommend with a query structure created by a catalog of terms to retrieve subject keywords and area of interest domain keyword to enhanced correspond to the dissimilar workings of the information retrieval process (customer interests, objects to retrieve).For example, if a abuser requirements to obtain information regarding the geography we have as subject matter keywords:=weather, and domain keyword:=drought. This scheme can repossess pages that are attractive from the user's perception, exclusive of consent to for ones connected to the drought weather which pertains to the geography domain.

In our scheme the prospect of information is WordNet, a common knowledge base prepared from a linguistic position of analysis. A concise explanation of this knowledge source is specified in the subsequent subdivision. Still if WordNet have numerous deficit in a few intangible domains it is one of the nearly everyone used linguistic possessions in the study area. The primary objective of our occupation is to intend a system competent of retrieving and ranking consequences, attractive into description the semantics of the pages. This system must be proficient to achieve the subsequent tasks, fetching of searching web documents include the keywords particular in the query. This undertaking can be proficient using conventional search engines. Preprocessing eliminate from web documents every those essentials that do not indicate supportive information (HTML tags, scripts, applets, etc.). Mining an study of the documents satisfied from a semantic point of analysis, transfer a score with esteem to the query. Reporting and habitual the

document significant to the query. Currently we illustrate an instance to bring in our framework and its associations through the proposed planning possessions of the system interface. The topic keywords is use in the fetching stride where a quantity of pages are fetched from conventional search engines Google and then preprocessed by the section. The proposed system is based on numerous services. In this circumstance every software module performs events illustrate in the previous to semantic meaning of the web documents. The search engine wrapper obtain the query and get used to it to the exact syntax of the search engines with the query adapter element thus generate the query string for the single selected search engines. In categorize to accomplish a high level precision the search engine wrapper present the modified query to the search engines by resources of the search engine submitter in arrange to get the web links page. Subsequent to this phase, the parser examine this page in arrange to retrieve the links that are restricted in it. The web fetcher retrieves the pages connected to the relations and stores them in the web repository. The pages are recovered by the web catcher even as the repository planner inserts them in the web repository. The formation of a web page frequently has a arrangement page that is collected of simulation. At present we understand that these objects do not give constructive in sequence to our system. The web fetcher repossess as defaulting the primary two level in the site structure and stores them with the similar hierarchy, we contain use to crawl the web. It ask for the preliminary address to establish the crawling with the depth for crawling. This is the beginning position for crawling. When known with URL of a few web pages and one as profundity, it crawls that page next to with the pages that are related from this kernel point.

IV. RESULT ANALYSIS

To generate utilize of knowledge available in ontology, we necessitate access the ontology. Parser to entrance RDF files is not accessible in C but is accessible in C#. Our application is written in C# and ASP.net. So alternative obtainable with us were to write our possess parser in C# or to create the ontology information accessible program. We count frequency-of-occurrences of every web document that is downloaded during crawling and stored this in sequence in database. through the assist of this information and knowledge from ontology, we produced document vectors .Every these document vectors are next stored in database in sequence as regards every one content and their source documents is stored in database. We use MSSQLserver as the database.



Figure:2 Ontology search system

OntologySearch	Recommendation%	New Approach(ms)	Old Approach(ms)
my%20test	3	7	10
test	7	45	240
web%20dunia	1	42	45

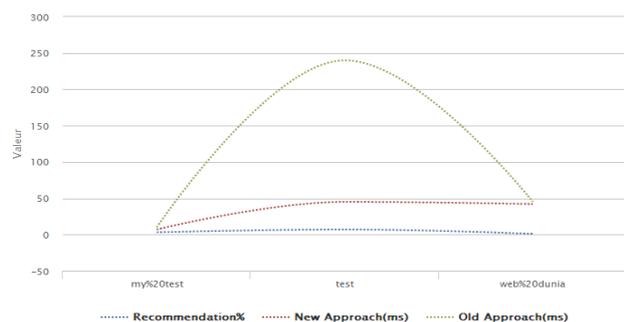


Figure: 3 Ontology graph showing timing between old & new approach.

Ontology Search	Recommendation%	New Approach(ms)	Old Approach(ms)
my%20test	3	7	10
test	7	45	240
web%20dunia	1	42	45

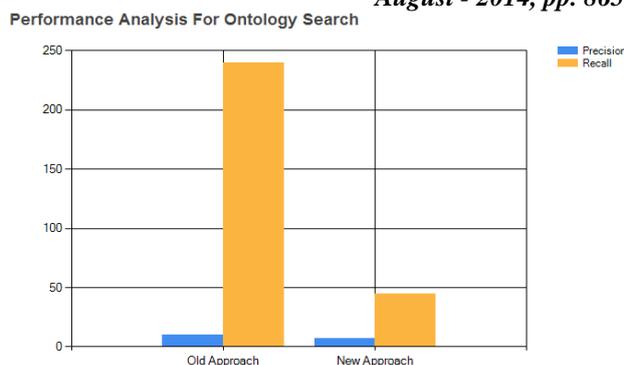


Figure: 4 Performance analysis for ontology search

Recommendation Search	Recommendation%	New Approach(ms)	Old Approach(ms)
dainik bhaskar	6	23	130
freejob alert	1	39	210
study forum	1	43	230
technology	1	43	230
test2	3	7	50

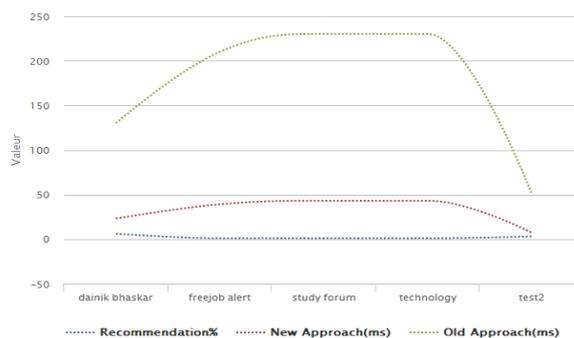


Figure: 5 Comparative study showing timing between old and new approach in semantic web search.

Recommendation Search	Recommendation%	New Approach(ms)	Old Approach(ms)
dainik bhaskar	6	23	130
freejob alert	1	39	210
study forum	1	43	230
technology	1	43	230
test2	3	7	50

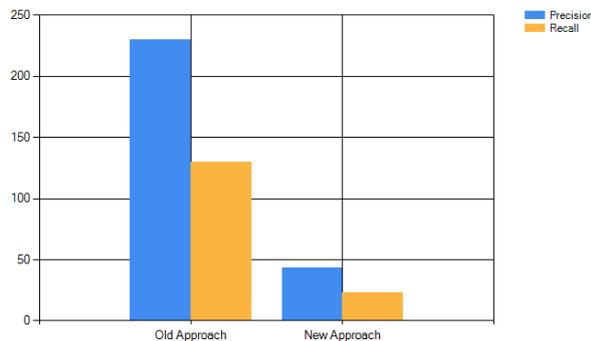


Figure:6 Performance analysis for semantic web search.

The create ontology was contrast with ontology construct by expert according to the indicators of Precision & Recall, to assess the accuracy rate of ontology produced by anticipated context-based ontology structure. Precision refers to the percentage of accurate concepts extracted, where Recall refers to percentage of expert-defined thought extracted. Precision and Recall can be review into a different metric recognized. This conduct test means to examine the dissimilarity of precision and recall among the proposed geography examination geography domain ontology building technique and domain experts.

V. CONCLUSION

This reading proposes a geography examination domain ontology building replica that features and addition technique. The technology like context mechanism in this work a storage technique using fuzzy c-means clustering ontology base on relational database is planned. Primary, we shall for a moment bring in illustration technique of fuzzy data types by extend the RDF data type and after that explain the retrieve methods of ontology and the storage technique of structure and illustration of fuzzy ontology in the relational database. The superiority of ontology resultant from proposed ontology building technique might not be as superior as what can be constructing by specialist still it can serve as an primitive ontology that help experts in collect and organize thought connected to domain knowledge and the relations between these concepts. The proposed ontology building method can be used to go faster the process of construct domain ontology and diminish the cost for simply manual construction of domain ontology though, for several special cases such as a domain with fast changing terms and perception or with complex semantics it is extremely complex to construct suitable domain ontology.

REFERENCES

- [1] Elena Demidova, Xuan Zhou, Wolfgang Nejdl, “Efficient Query Construction for Large Scale Data”. SIGIR’13, July 28–August 1, 2013.
- [2] YakupYildirim, Adnan Yazici, Senior Member, IEEE, and TurgayYilmaz Student Member, IEEE, “Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model”. IEEE Transactions on Knowledge and Data Engineering, VOL. 25, NO. 1, January 2013.
- [3] Lixin Han, Guihai Chen, “A fuzzy Clustering Method of Construction of Ontology-Based User Profiles”. L. Han, G. Chen / Advances in Engineering Software 40 535–540, 2009.
- [4] Chunchen Liu, Jianqiang Li, “Semantic-based Composite Document Ranking”. IEEE Sixth International Conference on Semantic Computing, 2012.
- [5] Adriana Maria Figueiredo, Julio C. dos Reis, Marcos A. Rodrigues, “Improving Access to Software Architecture Knowledge An Ontology-based Search Approach”. International Journal Multimedia and Image Processing (IJMIP), Volume 2, Issues 1/2, March/June 2012.
- [6] Jun Zhai, Jianfeng Li and Yan Lin, “Semantic Retrieval Based on SPARQL and Fuzzy Ontology for Electronic Commerce”. Journal of Computers, VOL. 6, NO. 10, October 2011.
- [7] Rim Faiz, Boutheina Smine’9, Jean-Pierre Descles, “Relevant Learning Objects Extraction Based on Semantic Annotation of Documents”. WIMS’12, June 13–15, 2012.
- [8] Deaton, C., Shepard, B., Klein, C., Mayans, C., Summers, B., Brusseau,A., and Witbrock, M.: “The Comprehensive Terrorism Knowledge base in Cyc”. In Proceedings of the 2005 International Conference onIntelligence Analysis, 2005.
- [8] Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R.S., Peng, Y., Reddivari,P., Doshi, V. and Sachs, J.: “Swoogle: “A Semantic Web Search and Metadata Engine”. In Proc. 13th ACM Conf. on Information and Knowledge Management, Nov. pp. 652--659, 2004.
- [9] Gregory, M., McGrath, L., Bell, E., O’Hara, K., and Domico, K.:, “Domain Independent Knowledge Base Population From Structured and Unstructured Data Sources”. In Proc. of the 24th International Florida.