Bayes Nearest Neighbor Search for Top N-Query Retrieval in Web Document Mining

[I]Dr R.S. Vetrivel, [II]C.Senthuradevi, [III]P Jeyanthi Rani
[I]Professor, [II]Research Scholar, [III]Assistant Professor
[I,II,III]Computer Science, Subramanya College of Arts & Science, Palani

Abstract

Web mining is developed to automatically find out and extract information from web documents by using data mining methods. Mining process estimates accuracy in information of query retrieval. In ontology, the types of resources and semantic relationships among resources are identified through the semantic web search. Ontological information enhances the accuracy of information retrieval. Existing method was developed to achieve effective ranking and web search while considering semantic relationships. Retrieval of top-k query resource is done by effective ranking and search technique which are relevant to query keywords with semantic relationships. This process is used to shorten the unnecessary search space using length and weight thresholds of the semantic relationship path as well as enhance the searching efficiency of the document. However, semantic relationship extraction and retrieval of top-k query results need to be improved.

To overcome these problems, we propose a technique for top-N query retrieval in web document mining to provide relevant information for user request. In this method, the accurate search results can be produced by web document mining which enhances the speed of search. This process estimating relationship between words for providing similar contexts on web search by extort the conceptual content of the text. The results can be discovered through finding the most suitable query to related documents on web. This method uses the content based classification to rank the document on the basis of query request from the user.

Key Terms

Web document search, Keywords, Top N-Query Retrieval, Bayes nearest Neighbor Search

I. Introduction

A. Web document mining

Web mining is a process of extracting accurate information by web search using content mining, structure mining, and usage mining methods. Web content mining is called as web document mining. Web document mining is implemented to removing uncertainty through web search which helps to find out patterns from World Wide Web.

Traditional search engines are not appropriate to find items in web, due to incredible extension in quantity of data and with terrific growth of number of web pages. These search engines are not easy for people to maintain enormous data manually and it is more time consuming to extract the information.

Web document mining is the process of discovering unknown information from the web effectively with less time consumption. Web document mining consists of text, images, audio, video, such as lists and tables to determine the relevance of the content to the search query. Based on highest relevance to the keywords in the query, results provides by content mining. This type of data mining is used to classify web pages and to generate the relationship between web sites.

To find the most relevant information, user requesting information on a particular topic is searched among the thousands of results. Based on user request, web document mining performs to collect, categorize, manage and offers best probable information in the web. This process eliminates the noise and to improves the accuracy in information retrieval on the Web.

B. Semantic search in web mining

Semantic is the process of communicating with sufficient meaning to provide results. Sequence of symbols can be used to communicate. But it can easily affect the meaning and behavior of communication. Due to such difficulties in traditional web search, the semantic web search is developed in mining.

Semantic web provides best web service which coordinates and arranges all the data over web in a disciplined manner. Semantic search is established to provide sufficient meaning to search result. Semantic approaches are automatically utilizes the web resources. Semantic web is used to extract data from web and to perform data reuse process everywhere.

In semantic web search, sequences of symbols convey meaning. Semantic web is introduced in web search to overcome following main limitations

- According to representation of information, proper structure can be need for web content in search.
- In web search, due to less interconnection of information, data information will be uncertain.
- Poor in automatic transfer of information.
- Due to lack of a universal format, machines are not easy to recognize the given information.
- Difficult to deal with huge number of users and provides hope at all levels.

Semantic web mainly developed to perform the web content understandable not only by humans, but also machine understandable. During whole life cycle of information retrieval, semantics are necessary to enhance for improving performance level.

III. Bayes Nearest Neighbor Search

On uncertain web management, identification of nearest neighbor is a significant function. Authenticated Multi-step NN (AMNN) and Superseding Nearest Neighbor (SNN) search methods are able to perform only in single server. These searches do not perform from the distributed servers. Communication overhead also increased and not developed in to high dimensional data structure.

BNN (Bayes Nearest Neighbor) search is implemented for Nearest Neighbor (NN) search efficiently and to carry out similarity

search in web. BNN search is developed to retrieve the distance information from distributed servers is main advantage of this search which reduces the communication overhead. BNN search method is applicable for high dimensional data structure. The reliability of returning query result is high while searching is performed by BNN.

Bayes Nearest Neighbor Search is developed to extract the probable information from the web through searching which is related to the given request of user. Search system consists of three main subsystems such as document representation, user's requirements and method used to match user requirements to the document database.

From the figure 1.1 We analysis the Bayes Nearest Neighbor Search. When the user gives request to search queries, contents are transformed into document representation from the Document block which is related to user request queries. By using Bayes nearest neighbor search, these queries can be effectively retrieved to the user. .
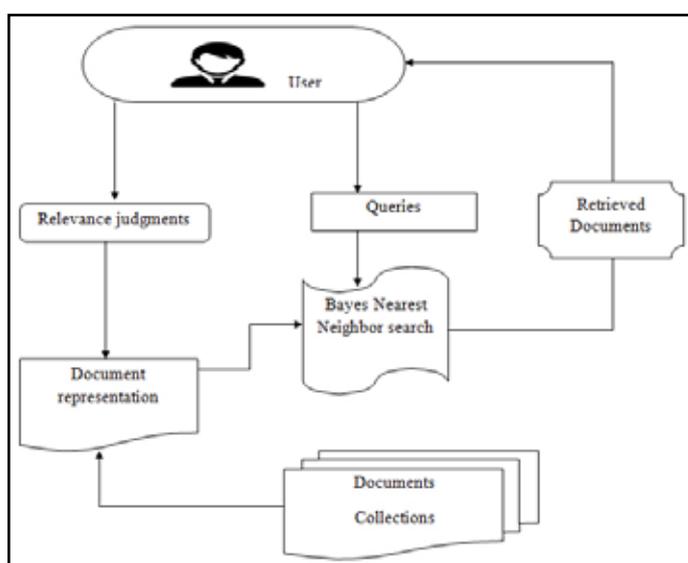

Fig. 1.1: Overview of Bayes Nearest Neighbor Search

II. Literature Survey

1. Techniques for Efficient Keyword Search in Cloud Computing

In this paper [1], the author plan to resolve the problem of providing exact keyword match through searching by fuzzy keyword .The author develops a method with two techniques to construct efficient fuzzy keyword sets through the cloud computing.

Gram based technique is used to enhance the performance through fast searching by considering substring from the given string which results as reduces the time. In Symbol-based tree traverse search scheme, where a multi way tree structure is built by symbols, which worked for more than one keywords entered by user. This proposed method is designed to provides security in search and privacy preserving

2. PIDGIN: Ontology Alignment using Web Text as Interlingua

This paper [2] is developed to solve the ontology alignment problem by utilizing very large natural language text corpus as an inter lingua which are related to different knowledge bases (KBs). The author can be introduced a PIDGIN for flexible, and scalable approach in this paper to automatic alignment of real-world KB

ontology's. PIDGIN analyzing both (1) shared relation instances across KBs, and (2) the verb phrases the text instantiations of relation instances, to align relations and categories across different KBs.

3. A Survey on Semantic Web Search Engine

In this paper [3] the author develops a technique to find best semantic search engines which utilizes different approaches in various ways to attain unique search for users. The author designs a most important part in this method, to analyze different between semantic search engines from the traditional searches. Search engine is the important tool to extract information in Web. This method is introduced to develop the advantages of important semantic web search engines.

4. Ontology Learning for the Semantic Web

For the use of complete and transportable machine understanding, ontology based semantic web search is developed in this paper [4], with the help of ontology's, both people and machines are communicated effectively through the web. The strong propagation of ontology is the only reason for the success of the Semantic Web search. The semantic web search requires quick and simple engineering of Ontology's. By using the proposed method, knowledge of acquisition bottleneck is avoided. Ontology learning technique is developed to carry out ontology import, extract, prune, refining which giving to the ontology engineer a organizing tools for ontology modeling.

Ontology engineer is helpful to Ontology Learning for extremely facilitates the construction of ontology's. The main objective of the ontology learning includes a number of complementary disciplines. And they provide different types of unstructured, semi-structured and fully structured data to support a semi-automatic, cooperative ontology engineering process.

5. Intelligent Semantic Web Search Engines: A Brief Survey

In this paper [5], the author developed to analysis the semantic search methodology, search engine generations and role of search engines in intelligent web. The WWW is helpful to permits the people to distribution of the information from the large database.

The quantity of data information grows billions of databases. While using of search engines in existing methods, retrieval of accurate information is not an easy task in web. The specializing tools are known by generically search engine. It is necessary for the searching accurate information on web. Semantic web technology is the most significant task for search engines to retrieve meaningful information efficiently.

6. Combining Semantic Search and Ontology Learning for Incremental Web Ontology Engineering

.This paper [6] develops a technique to perform semantic web search for ontology learning. Information retrieval (IR) has been discovered by some approaches as semantic web search by using ontology domain. The author was introduced an incremental approach for ontology learning based on ontological representation known as Meta ontology in previous methods.

The proposed method is based on two important schemes. The method is proposed in this paper, for developing a query transformation to granule ontology. The method is used to explain the process of semantic search for ontology learning from texts.

And describe how they make the process to be easy for web ontology engineering and semantic indexing from Web documents by using case based reasoning.

7. An Introduction to Latent Semantic Analysis

In this paper [7], the author proposes a Latent Semantic Analysis (LSA) technique by applying statistical computations for large quantity of text to extort and signify the contextual-usage meaning of words.

The main objective of this proposed work that combining all words contexts in that specified word does and does not appear offers a set of mutual constraints and It is helpful to verifying the similarity of meaning of words and sets of words to each other. The sufficient of LSA reflection of human being knowledge is estimated in different ways.

The LSA method is developed to establishing word and passage meaning representations. This method introduced to initiate the different of human cognitive phenomena, ranging from developmental acquisition of recognition vocabulary to word-categorization, sentence-word semantic priming, discourse comprehension, and judgments of essay quality.

8. Indexing by Latent Semantic Analysis

In this paper [8], the author proposes a method for automatic indexing and information retrieval. By using this proposed method, documents of higher-order structure (semantic structure) are easily determined. Based on the information found in queries, the determination of higher-order structure enhances the detection of relevant documents.

The author is develops a technique to perform singular-value decomposition. The decomposition of large term is done by document matrix is decomposition into set of ca. 100 orthogonal factors from the original matrix estimated by linear combination.

9. Probabilistic Latent Semantic Indexing

The author presents a technique as Probabilistic Latent Semantic Indexing approach in this [9]. Based on statistical latent class model, automated document indexing is done by this proposed method. The word perplexity is reduced by estimation of generative data model, based on the likelihood principle which aids to statistical standard methods for model fitting, over fitting control, and model combination. Probabilistic Latent Semantic Indexing is developed to achieve significant gains in precision among standard term matching and LSI. Training quantity of text documents are fixed by performing the Expectation Maximization algorithm which causes model has been contract with domain specific synonymy and polysemous word.

Performance of the proposed method is increased over direct term matching methods and LSI by retrieval experiments of test collections. In Probabilistic Latent Semantic Indexing, probabilistic variant has solid statistical foundation with proper generative data model.

10. Detecting Similar Areas of Knowledge Using Semantic and Data Mining Technologies

The author proposes a technique in this paper [10], to link multiple bibliographic sources. This work is a developed to identify common research areas among Ecuadorian authors and potential collaboration networks. The goal of this work achieved by combination of ontology's, vocabularies, and Linked Data

technologies with developing base data model. This method includes a process to extract, enrich, and represent bibliographical resources to discover patterns using data mining algorithms. The author has been employing a special prototype to offer centralized repository with bibliographic sources. Based on data mining techniques in the domain of Ecuadorian researcher's community, similar knowledge areas are discovered in the web.

11. Searches with Meanings: An Overview of Semantic Search Systems

In this paper [11], the author presents a technique to provide complete survey to gain an overall view of the current research status. Based on the most distinctive features the author classifies our considered systems into several categories. The categorization by no means prevents a system from classified into other categories.

To web and Intranet searching and browsing systems, the scope of the survey can be confined by the author. In this paper, a technique used to expand the scope and enhance the retrieval quality of conventional Information Retrieval techniques. The future research is concentrated on the trust and quality of the knowledge

12. Semantic Search meets the Web

In this work [12] the author presents a technique as semantic search approach which covers the whole information retrieval process. The author develops a method to deal with problem of novel semantic search system. a) through the ontology-based Question and Answering, provides capability to query Semantic Web information using natural language and b) according to QA process with a ranked list of documents in web, the specific answers can be retrieved.. Ontology-based semantic search provides results used to improve the keyword search technologies.

13. Web Graph and Page Rank algorithm

In directed graph [13], the web pages and hyperlinks of the WWW can be viewed as nodes and arcs. This graph contains a billion nodes today, some billion links, and emerges to grow with time which reviewed the known facts of macroscopic structure, diameter and in-degree and out-degree distributions. Web page may be of characterized by the Page Rank.

14. Topic-Sensitive Page Rank: A Context-Sensitive Ranking Algorithm for Web Search

In this paper [14], the author develops a technique as Topic-Sensitive Page Rank for web search. To capture more accurate notion with respect to particular topic, the author propose computing a set of Page Rank vectors, biased using a set of representative topics.

The author computing topic-sensitive Page Rank scores which satisfying the query using the topic of the query keywords for ordinary keyword search queries. In this paper we compute the topic-sensitive Page Rank scores using the topic of the context, while the search query is performed by highlighting words in a Web page. According to the topic-sensitive page rank scheme, we developing techniques to efficient implement of a large scale search system

15. Ranking Systems: The Page Rank Axioms

The author presents a paper [15] to deals with Page Rank, the most famous page ranking algorithm. In this paper, the author describes a set of simple (graph-theoretic, ordinal) axioms that are

fulfilled by Page Rank and coincide with Page Rank. The method developed to compute bridging the gap between page ranking algorithms and the mathematical theory of social choice.

16. An architectural design for effective information retrieval in semantic web

In this paper [16] the author propose a technique for a semantic web search. He present a method of Semantic Based Information Retrieval System to enhance the information retrieval in a more efficient way. This system architecture consider as input whisch list of plain keywords provided by the user.Then, the query is converted into semantic query. While using of the domain concepts of the pre existing ontology, this conversion is carried out and semantic relationship between them in runtime can bbe discovered. with the help of an improved algorithm, the relevant information is retrieved and ranked according to the relevancy.

17. Range query estimation with data skewness for top-k retrieval

In this paper [17],the author presents a technique as cost based query mapping method for top-k retrieval through incorporating data skewness in multi-dimensional histograms. The author develops a proposed method to provides robust performance across different types of data sets, query sets, distance functions, and histograms.

18. An effective query recommendation approach using semantic strategies For intelligent information retrieval

In this paper [18], the author proposes a technique for intelligent information retrieval. The query concepts can be extracted to find queries which are closer to each other. The bipartite graph is developed to identify the related queries, and the query similarity. For some typical topic-based pages, find out tokens appeared in URLs which are very meaningful, with help of analyzing the URLs clicked by users. These tokens provide a brief description from the subject of the URL.

19. Survey of Semantic Search Research

In this paper [19], the author deals with the research field of semantic search. The semantic techniques or search of formally annotated semantic contents are used in this paper for searching the document. The survey identifies and discusses various prevalent research directions in semantic search, as well as extracts common methodology used in them.

20. Extrapolation and minimization procedures for the Page Rank vector

In this paper [20] the author deal with an important problem in Web search is to determine the importance of each page. In this paper consists in computing, the left principal eigenvector (the Page Rank vector) of c which has to be chosen close to 1.

III. Methodology

The Bayes Nearest Neighbor Search technique is developed for Top N-Query Retrieval in Web Document Mining. This proposed work to improve efficiency of information retrieval which provides accurate search results.

The effective ranking and searching techniques for Web resources given in ontology method was developed for retrieving top-k resources of most related to query keywords through semantic

relationships. For semantic relationships, weighting measure allocates higher weight to significant relationships. They give higher level contribution to discriminate the results.

Ranking method is developed with weighting measure to indicate the number of meaningful semantic relationships between resource and keywords coverage. It also used for discriminating power of keywords. The author develops a method to shorten the unnecessary search space using length and weight thresholds of semantic relationship path. Threshold Algorithm based on extended inverted index is developed to answer top-k results.

The effective ranking and search technique have the following disadvantages.

- Existing method provided the low level of accuracy in the information retrieval.
- Does not consider the issues of semantic relationship.
- Fails to produce the page ranking performance in search quality of the document results.

To overcome these problems we propose a method in web document search to discover require information through keywords which are related to the results. Initially, the user entered the information which he need from the web. The data inserted into web with some keywords. The web document search generated the most related results to the user request.

Here, filtered data are produced which are relevant to results through the keywords. The clustering and classification techniques are used in next level of processing of search. The cluster based techniques which clustering the data items into groups with the help of keywords. Based on keyword from user, classification techniques are organizing the groups. Then, Bayes nearest search is developed in this proposed work. This BNN search helps to user to find accurate results on web which related to query request.

Advantages

- The proposed method is used to improve the accuracy of search results.
- The performance search speed also increased by this technique.
- The efficiency of query retrieval is successfully enhanced.

From fig 3.1 when the user inserted the text (information) with keyword, the web document search generates the most related documents to user request which is taken from the document database. Then queries with related keywords are combined to clustering and classification block. Clustering and classification carried out by clustering and classification techniques. Finally the BNN search is used to discover the best nearest neighbor information which results are improving the performance and top –N queries are retrieved to the user.
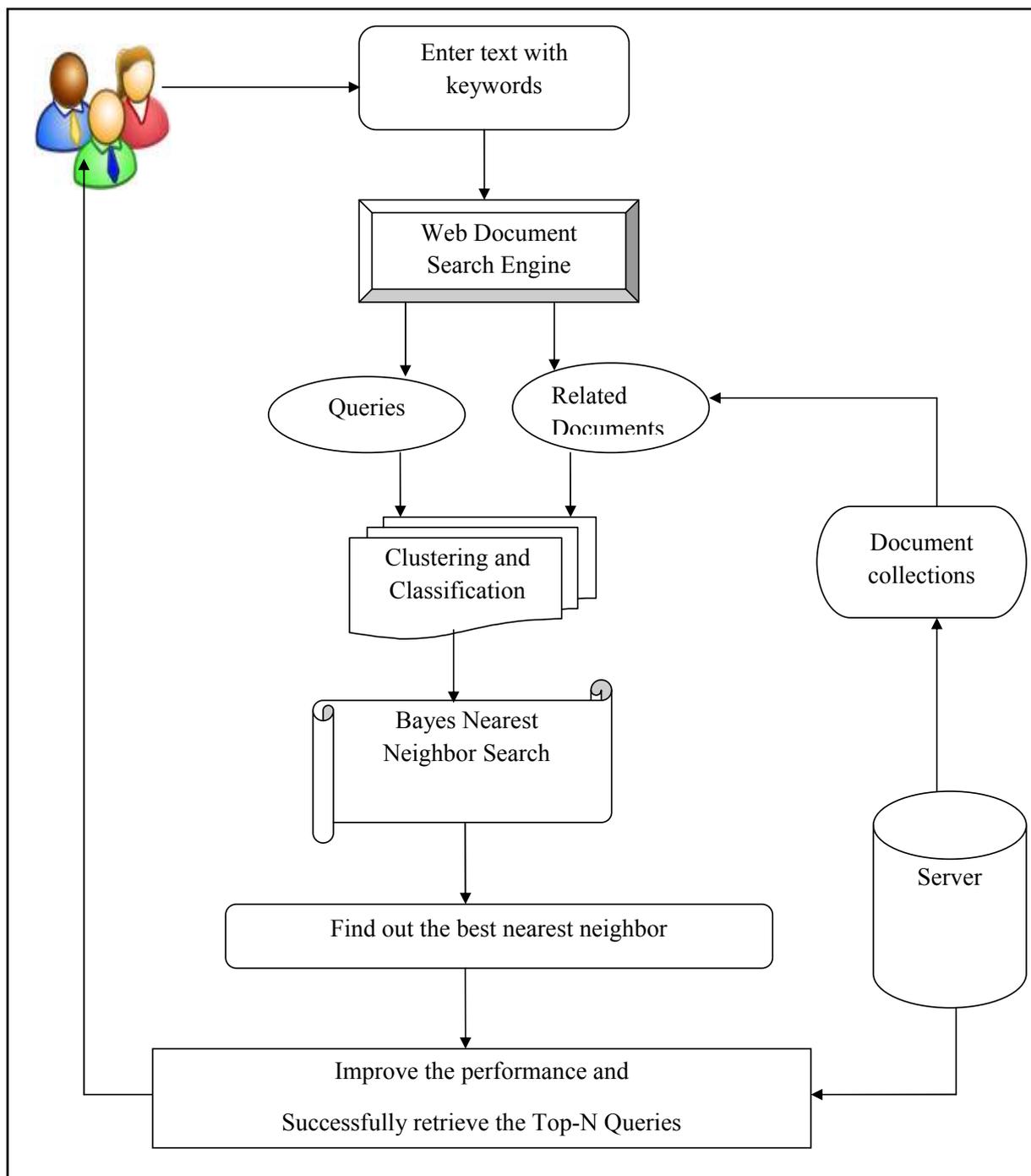
Fig. 3.1: Architecture of Bayes Nearest Neighbor Search method

A. Modules

a) Web Document Search by keywords
b) Clustering
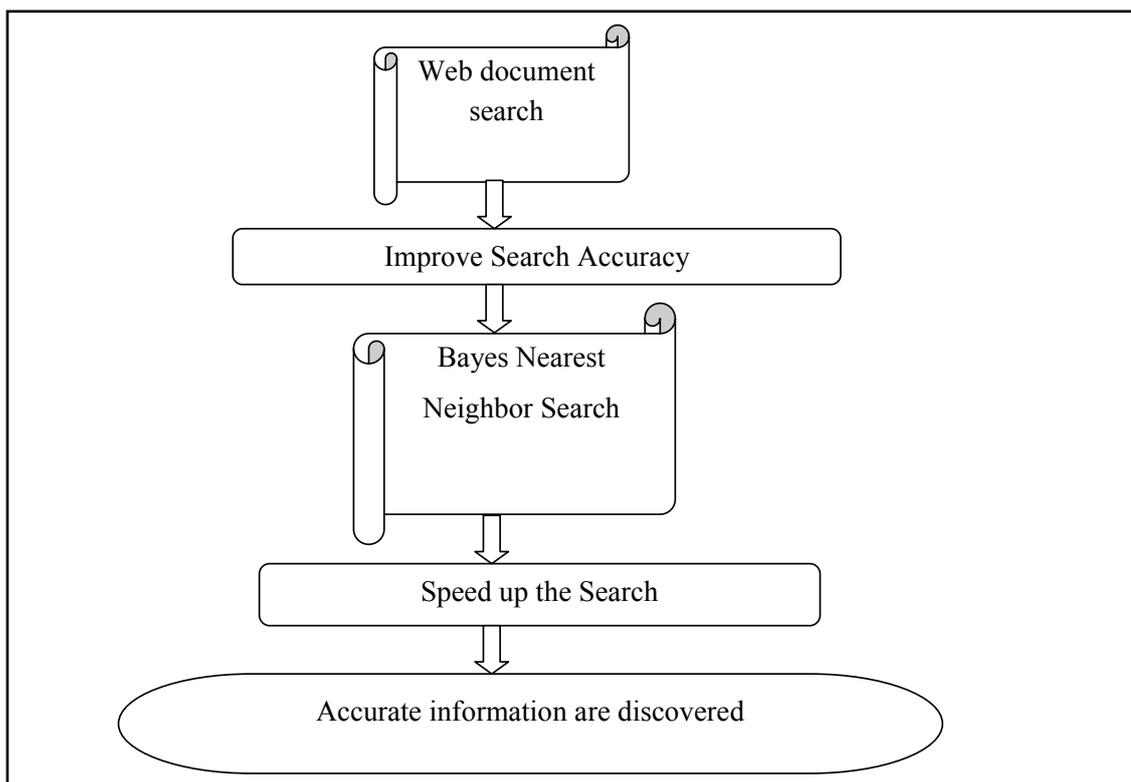c) Top N-Query Retrieval by using Bayes Nearest Neighbor Search

Fig. 3.2 : Principle of proposed method

B. Modules Description

1. Web Document Search

The user can be search the required information through the internet. Initially, the user is needed to send the request with some keyword for finding results. The web document search is used to generate the most relevant results for query request from the user. The web document search helps to enhance the search accuracy by understanding searcher intent and appropriate meaning of terms in the web document in the web. The suitable results are produces through the most relevant results, comparing with keywords.

The web document search reflects on framework of search, location, intent, and variety of words, synonyms, generalized and specialized queries. The relevant search results are provided by concept matching and natural language queries.

In web document search, searching of required information is an easy task with the help of keywords. The web document mining is discovering the exact information for query request which are relevant to results. The most related results generated automatically while consider web document search. The web document search is deal with semantic search in ontology.

It permits users to search for ontology's by specifying ontology title or the topic of interest. Users can specify an URL of a web document containing information dealing with the topic. Linguistically derived set of relevant keywords is extracted automatically and used in the search. Augment results of traditional keyword search with data retrieved from the semantic web.

2. Clustering and classification

Clustering and classification techniques are developed for organizing documents into groups which are related to the keywords through web. Clustering mechanism is introduced to create groups automatically based on calculation of the degree of association between items and groups. This clustering and classification mechanisms examines document collection as group which aids to analyzing other documents containing some of those same words.

Clustering is the process of document indexing process through keywords associated to a document. Cluster based analysis is the very suitable method for query retrieval. Cluster analysis has been used to assigns documents into groups for the information retrieval in web search.

The main intension of the Clustering in the information retrieval field which enhance the efficiency and effectiveness of the results. The structure of the literature of a field is determined by the clustering. For achieving recovered user interference the clustering techniques are used in web search. After achieving clustering, the classification also performed to provide query based on keywords. There are two main types of categories involved in cluster classification analysis.

i. Non hierarchical type

The Non hierarchical type of clustering which divides the documents such as N items into M clusters. It requires less computation for clustering. However, hierarchical type of clustering is preferred for query retrieval.

ii. Hierarchical type

The Hierarchical type of clustering provides the nested data items effectively which are sequentially linked. This type of clustering is perfect for cluster-based document retrieval in the web. The mostly used hierarchical methods such as single link, complete link, group average link, and Ward's method.

The hierarchical clustering is successfully performed in the field of large data sets. The Hierarchical based clustering is particularly useful for high dimensional data in the web search. These hierarchical methods are requiring more time and space for clustering the items.

To resolve these problems we proposes a techniques such as SLINK and minimal spanning tree algorithms for single link method, Voorhees algorithm for group average link, and reciprocal nearest neighbor algorithm for Ward's method.

3. Top N-Query Retrieval by using Bayes Nearest Neighbor Search

BNN search technique is used to choose the best NN object in the web. BNN search process is developed to discover the nearest neighbor for high dimensional data structure. It automatically reduces the communication overhead. Initially the queries are requested by the clients to server. The each query is examined and the output to the clients is produced by the BNN search.

Existing searches are not sufficient for multiple users have to process in real time at the same time. To avoid scalability problems we propose a Bayes Nearest Neighbor Search for query retrieval. BNN search is implemented for NN search and similarity search in web.

BNNS performs NN search efficiently and retrieve the distance information not only from single server but also from distributed servers. The query result returned by BNN will be a reliable one.

BNN search is presented to perform the NN search efficiently for both high dimensional data and retrieve the information from the distributed servers. Elimination of the server which is not involved in the result producing contribution is done by false hit (FH) reduction algorithm. The BNN search will produce the reliable query and also it consumes less time to perform the NN search across distributed server.

Input: Database 'D', Number of queries 'Q'
Output: Retrieval of the top-N Queries
Step 1: **Begin** Step 2: Enter the required text with keywords Step 3: Web document search is generates the related documents to queries from server Step 4: Clustering the related queries into groups Step 5: Classify the queries based on relevant to keywords Step 6: Bayes Nearest Neighbor Search find out the best nearest neighbor query Step 7: the top –N Queries which related to keywords are retrieved to the user Step 8: **End**

Fig. 3.3 : Algorithm for Top N-Query Retrieval by using Bayes Nearest Neighbor Search

IV. Experimental Evaluation

In this paper worked on performance evaluation in terms of Query retrieval efficiency, Accuracy of search retrieval results, Work-document relation, Searching speed. The performance metric to evaluates and analyze the value in java environment simulations. This technique improves the effectiveness of top –N query retrieval. The performance measures of the proposed work are analyzed with following metrics:

• Query retrieval efficiency
• Accuracy of search retrieval results
• Word-document relation
• Searching speed

1. Query retrieval efficiency

Query Retrieval Efficiency () of the BNNS technique is defined as the rate at which the required information is retrieved based on the users requested queries. The QRE is measured in terms of percentage (%).

Query Retrieving Efficiency (QRE) is measured based on the number of user queries used in the experimental evaluation.

Table 4.1: Tabulation for Query retrieval efficiency

No of queries (Q)	Query retrieval efficiency (%)	
	ERST (Existing)	BNNS (Proposed)
10	76	85
20	78	87
30	80	89
40	82	91
50	84	93

Table 4.1 illustrates the query retrieval efficiency is measured based on the number of queries used in the experimental evaluation. From the table value, the proposed BNNS technique is compared with the existing ERST technique. The results reported here it confirms that the BNNS technique provides the improved results than the other existing method.
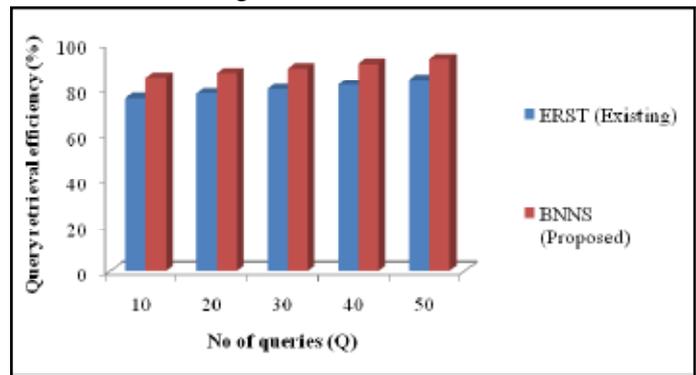

Fig. 4.1: Measure of Query retrieval efficiency

Figure 4.1 demonstrates the query retrieval efficiency .The queries taken for the experimental consideration is varied from 10 to 50. From the figure X axis represents the document size whereas Y axis denotes query retrieval efficiency using BNNS method. From the figure it is clearly evident that the proposed BNNS technique improves the query retrieval efficiency than the Existing ERST method. Hence, the retrieval rate is improved up to 11% by the proposed BNNS technique than the existing ERST method. When the number of queries increased, the query retrieval efficiency increases accordingly.

2. Accuracy of search retrieval results

Accuracy of search retrieval results is measured as the ratio of the number of the queries retrieved to user which is related to keywords. Web document search process investigates to provides better search accuracy by accepting searcher decided and the dependant meaning of conditions as they occur in the searchable data space, whether on the Web or inside a blocked system, to make additional relevant results.

The accuracy of search retrieval result is measured in terms of percentage (%).

Table 4.2: Tabulation for accuracy of search retrieval results

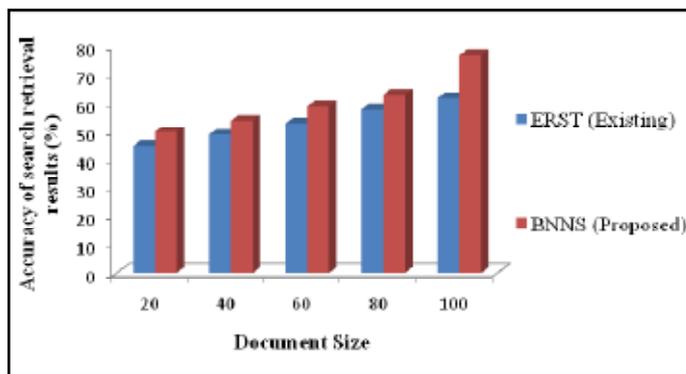| Document Size	Accuracy of search retrieval results (%)	
	ERST (Existing)	BNNS (Proposed)
20	45	50
40	49	54
60	53	59
80	58	63
100	62	77

Fig. 4.2: Measure of accuracy of search retrieval results

Figure 4.2 demonstrates the accuracy of search retrieval results. From the figure X axis represents the document size whereas Y axis denotes query retrieval efficiency using BNNS method .The documents taken for the experimental consideration is varied from 20 to 100. From the figure it is clearly evident that the proposed BNNS technique improves the accuracy of search retrieval results than the Existing ERST method. Hence, the accuracy of search retrieval result is improved up to 13% by the proposed BNNS technique than the existing ERST method. When the document size is increased, the accuracy of search retrieval results are increases accordingly.

3. Word document Relation
Most of the retrieval methods are based on simple word matching strategies to describe the rank of relevance of a document with respect to a query.
The word document relation is measured in terms of percentage (%).

Table 4.3: Tabulation for Word document Relation

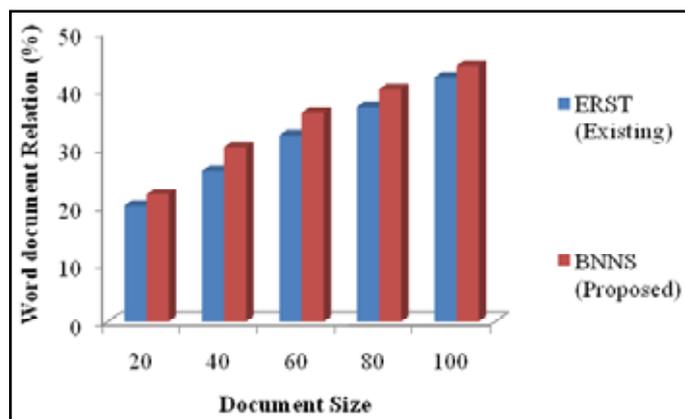| Document Size	Word document Relation (%)	
	ERST (Existing)	BNNS (Proposed)
20	20	22
40	26	30
60	32	36
80	37	40
100	42	44

Fig. 4.3: Measure of Word document Relation

Figure 4.3 demonstrates the word document relation results. From the figure X axis represents the document size whereas Y axis denotes word document relation result using BNNS method .The documents taken for the experimental consideration is varied from 20 to 100. From the figure it is clearly evident that the proposed BNNS technique improves the word document relation results than the Existing ERST method. Hence, the word document relation result is improved up to 10% by using the proposed BNNS technique than the existing ERST method. When the document size is increased, the word document relation results are increases accordingly.

4. Searching speed
Searching speed is defined as the rate of searching the required document to user request in the web which is related to the keywords. Searching Speed is measured in terms of milliseconds (ms).

Table 4.4: Tabulation for Searching speed

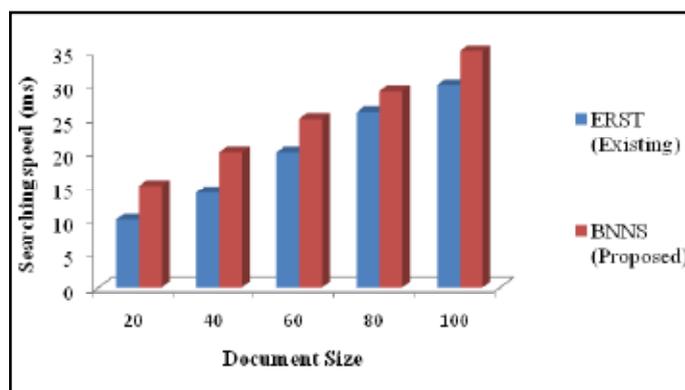| Document Size	Searching speed (ms)	
	ERST (Existing)	BNNS (Proposed)
20	10	15
40	14	20
60	20	25
80	26	29
100	30	35

Fig. 4.4: Measure of searching speed

Figure 4.3 demonstrates the searching speed. From the figure

X axis represents the document size whereas Y axis denotes searching speed using BNNS method .The documents taken for the experimental consideration is varied from 20 to 100. From the figure it is clearly evident that the proposed BNNS technique improves the word document relation results than the Existing ERST method. Hence, the searching speed is improved to 29% by using proposed BNNS technique than the existing ERST method. When the document size is increased, the searching speed is increased accordingly.

B. Summary

This paper provides an elaborate and overview of some of the best semantic search engines that handles different methods in various ways to yield unique search experience for users. It is summarized that searching the internet today is a challenge and it is estimated that nearly half of the issue questions go unanswered. Web document search has the force to develop the conventional web search. The proposed technique is used for discovering exact information through web search which are most related to keywords for query request from user.

V. Conclusion & Future Scope

1. Conclusion

 In this paper, the author proposes a technique as Bayes Nearest Neighbor Search for Top N-Query Retrieval in Web Document Mining to provide accurate results. BNN search process is used for query to find out the nearest neighbor for high dimensional data structure with the minimal communication overhead. The goal of BNN search is to retrieve top-N results which are exactly relevant to query keywords. In web search, the performance of accuracy is increased by using Bayes Nearest Neighbor Search based top-N Query Retrieval.
In further extend the ranking performance, which is especially useful for high dimensional data more effectively. So improve the query performance over the previous approaches.

References

[1] P.Niranjan Reddy and Y.Swetha, "Techniques for Efficient Keyword Search in Cloud Computing", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (1), 2013, 66 – 68.

[2] Derry Wijaya, Partha Pratim Talukdar and Tom Mitchell, "PIDGIN: Ontology Alignment using Web Text as Interlingua", CIKM'13, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.

[3] G.Sudeepthi, G. Anuradha and Prof. M.Surendra Prasad Babu, "A Survey on Semantic Web Search Engine", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012.

[4] Alexander Maedche and Steffen Staab, "Ontology Learning for the Semantic Web". Karlsruhe, Germany.

[5] G.Madhu, Dr.A.Govardhan and Dr.T.V.Rajinikanth, "Intelligent Semantic Web Search Engines: A Brief Survey", International journal of Web & Semantic Technology (IJWST) Vol.2, No.1, January 2011.

[6] Nesrine Ben Mustapha, Hajer Baazaoui Zghal, Marie-Aude Aufaure, and Henda Ben Ghezala, "Combining Semantic Search and Ontology Learning for Incremental Web Ontology Engineering", CHATENAY-MALABRY Cedex, France.

[7] Thomas K Landauer, Peter W. Foltz and Darrell Laham, "An Introduction to Latent Semantic Analysis", Discourse Processes, 25, 259-284.

[8] Scott Deerwester; Susan T Dumais; George W Furnas; Thomas K Landauer, "Indexing by Latent Semantic Analysis", Journal of the American Society for Information Science (1986-1998); Sep 1990; 41, 6; ABI/INFORM Global pg. 391.

[9] Thomas Hofmann, "Probabilistic Latent Semantic Indexing", Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval.

[10] Xavier Sumba, Freddy Sumba, Andres Tello, Fernando Baculima, Mauricio Espinoza and Victor Saquicela "Detecting Similar Areas of Knowledge Using Semantic and Data Mining Technologies", Elsevier,Electronic Notes in Theoretical Computer Science,Volume 329,2016, Pages 149-167.

[11] Wang Wei, Payam M. Barnaghi and Andrzej Bargiela, "Search with Meanings: An Overview of Semantic Search Systems", Jalan Broga, 43500 Semenyih, Selangor, Malaysia.

[12] Miriam Fernandez, Vanessa Lopez, Marta Sabou, Victoria Uren, David Vallet, Enrico Motta and Pablo Castells, "Semantic Search meets the Web", Walton Hall, Milton Keynes, MK7 6AA, UK.

[13] Danil Nemirovsky, "Web Graph and Page Rank algorithm", Petersburg State University, Russia.

[14] Taher H. Haveliwala, "Topic-Sensitive Page Rank: A Context-Sensitive Ranking Algorithm for Web Search", Stanford University.

[15] Alon Altman and Moshe Tennenholtz, "Ranking Systems: The Page Rank Axioms", June 5–8, 2005, Vancouver, British Columbia, Canada.

[16] M. Thangaraj , G. Sujatha " An architectural design for effective information retrieval in semantic web", Elsevier,Expert Systems with Applications,Volume 41,July 2014,Pages 8225-8233.

[17] Anteneh Ayanso , Paulo B. Goes , Kumar Mehta "Range query estimation with data skewness for top-k retrieval", Elsevier, Decision Support Systems,Volume 57,Pages 258-273.

[18] Wei Song , Jiu Zhen Liang , Xiao Long Cao , Soon Cheol Par "An effective query recommendation approach using semantic strategiesfor intelligent information retrieval", Elsevier, Expert System with Applications, Volume 41,2014,Pages 366-372.

[19] Eetu Makela, "Survey of Semantic Search Research", Helsinki University of Technology, Media Technology, and University of Helsinki.

[20] Claude Brezinski and Michela Redivo-Zaglia,"Extrapolation and minimization procedures for the PageRank vector", Dagstuhl Seminar Proceedings 07071.