Relevant Information Retrival Using Keyword Query

A.V.Akhare[1], Dr. M. A. Pund[2]
[1, 2] Dept of CSE
[1, 2] PRMIT & R Badnera , Amravati, India

Abstract- *Information Retrieval is concerned with indexing and retrieving documents including information relevant to a user's information need. Relevance Feedback (RF) is a class of effective algorithms for improving Information Retrieval (IR) and it consists of gathering further data representing the user's information need and automatically creating a new query. In this paper, we propose a class of RF algorithms inspired by quantum detection to re-weight the query terms and to re-rank the document retrieved by an IR system. These algorithms project the query vector on a subspace spanned by the eigenvector which maximizes the distance between the distribution of quantum probability of relevance and the distribution of quantum probability of non-relevance. The experiments showed that the RF algorithms inspired by quantum detection can outperform the state-of-the-art algorithms.*

Keywords- Information retrieval, quantum mechanics, relevance feedback, quantum detection

I. INTRODUCTION

Finding relevant document is one of the hardest tasks. Information retrieval (IR) is the activity of obtaining information resources relevant to an information need from a collection of information resources. Searches can be based on full-text or other content-based indexing. Automated information retrieval systems are used to reduce what has been called "information overload" Many universities and public libraries use IR systems to provide access to books, journals and other documents. Web search engines are the most visible IR applications. An information retrieval process begins when a user enters a query into the system. Queries are formal statements of information needs, for example search strings in web search engines. In information retrieval a query does not uniquely identify a single object in the collection. Instead, several objects may match the query, perhaps with different degrees of relevancy. An object is an entity that is represented by information in a content collection or database. User queries are matched against the database information. However, as opposed to classical SQL queries of a database, in information retrieval the results returned may or may not match the query, so results are typically ranked. This ranking of results is a key difference of information retrieval searching compared to database searching. Depending on the application the data objects may be, for example, text documents, images, audio, mind maps or videos. Often the documents themselves are not kept or stored directly in the IR system, but are instead represented in the system by document surrogates or metadata. Most IR systems compute a numeric score on how well each object in the database matches the query, and rank the objects according to this value. The top ranking objects are then shown to the user. The process may then be iterated if the user wishes to refine the query. We propose to investigate the problem of keyword query routing for keyword search over a large number of structured and Linked Data sources. Routing keywords only to relevant sources can reduce the high cost of searching for structured results that span multiple sources.

II. RELATED WORK

Author : Ingo Frommholz Paper title: Supporting Poly representation in a Quantum-inspired Geometrical RetrievalFramework Proposed Methodology: Geometrical retrieval framework inspired by quantum mechanics can be extended to support poly representation . This system was unable to show that the well-motivated algorithms perform significantly better than the simple algorithms. Luis M. de Campos Implementing Relevance Feedback in the Bayesian Network Retrieval Mode Publication: JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 54(4):302–313,2003. present an approach for relevance feedback in the Bayesian Network Retrieval (BNR) model. It is very difficult to compare feedback methods when the retrieval engines are different General purpose propagation algorithms can´t be applied due to efficiency considerations. Automatically learning the relationships among terms could imply that some relationships are not strong enough.Retrieval effectiveness could be damaged. If the number of terms is very high, thelearning stage could be time consuming.M.Shanmugham BEARINGSTIMULATED ALGORITHMS INSPIRED BY QUANTUM DETECTION Publication: International Journal of Current Trends in Engineering & Research (IJCTER) e-ISSN 2455–1392 Volume 2 Issue 7, July 2016 pp. 249 – 255 Proposed Methodology: Present a class of RF algorithms inspired by the quantum detection has been proposed to re-weight query terms by projecting the query vector on the subspace represented by the eigenvector . Explicit RF ,

Pseudo RF and Implicit RF is based on observations that are proxies of relevance. The main problem with proxies is that they are not necessarily reliable indicators of relevance and thus should be consider noisy. These systems do no rely on only non-retrieval technology. CLAUDIO Automatic Query Expansion in Information Retrieval. ACM Computing Surveys, Vol. 44, No. 1, Article 1, Publication date: January 2012.

Motivation Information Retrieval is the process of obtaining relevant information from a collection of informational resources. It does not return information that is restricted to a single object collection but matches several objects which vary in the degree of relevancy to the query. So, we have to think about what concepts IR systems use to model this data so that they can return all the documents that are relevant to the query term and ranked based on certain importance measures. These concepts include dimensionality reduction, data modeling, ranking measures, clustering etc. These tools that IR systems provide would help you get your results faster. So, while computing the results and their relevance, programmers use these concepts to design their system, think of what data structures and procedures are to be used which would increase speed of the searches and better handling of data.

Aim To maintain the collection of documents according to different user search, To find query-document or document-document similarity. The reduction is not really substantial, To measure the performance relevance judgments more accurately and more quickly. Users can identify more relevant documents for each query, while at the same time make fewer mistakes, to Find the document according to content of the documents, to Implement the concept of relevant document suggestion.

Problem definition The information retrieval involved in so many sectors. Lot of location base service provider use this concept. The location based service providers such as google, yahoo, bing are used this concept. But when any user search for the specific document then the matching is perform on the basis of their text. So this is the problem, because if the name of that document is different and content is same then the result for that user will not get properly.

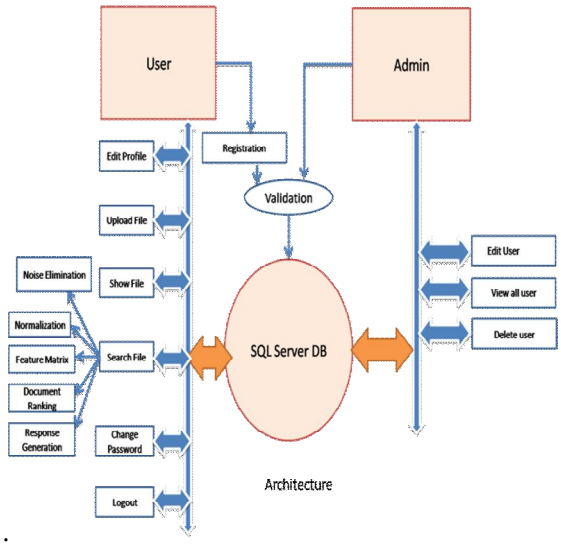
Proposed System We are going to propose a system using which the user can easily get the relevance document. When the user enter the query for search the document, then it directly compare within the data of the document file. So the relevant document will found by the system. We are also working to add feature, the system will recommend the keyword to the user for getting the best result or document. Objectives of this work is To maintain the collection of documents according to different user search. To find query-document or document-document similarity. The reduction is not really substantial, To measure the performance relevance judgments more accurately and more quickly, Users can identify more relevant documents for each query; while at the same time make fewer mistakes, also To Find the document according to content of the documents.

Algorithm Rocchio's Algorithm

The Rocchio algorithm is based on a method of relevance feedback found in information retrieval systems which stemmed from the SMART Information Retrieval System. Like many other retrieval systems, the Rocchio feedback approach was developed using the Vector Space Model. The algorithm is based on the assumption that most users have a general conception of which documents should be denoted as relevant or non-relevant. Therefore, the user's search query is revised to include an arbitrary percentage of relevant and non-relevant documents as a means of increasing the search engine's recall, and possibly the precision as well. Explicit Relevance Feedback It also called as Term relevance feedback. The system will suggest the term which types of term the user should add in search. Implicit Relevance Feedback It will find out the frequently search document easily.

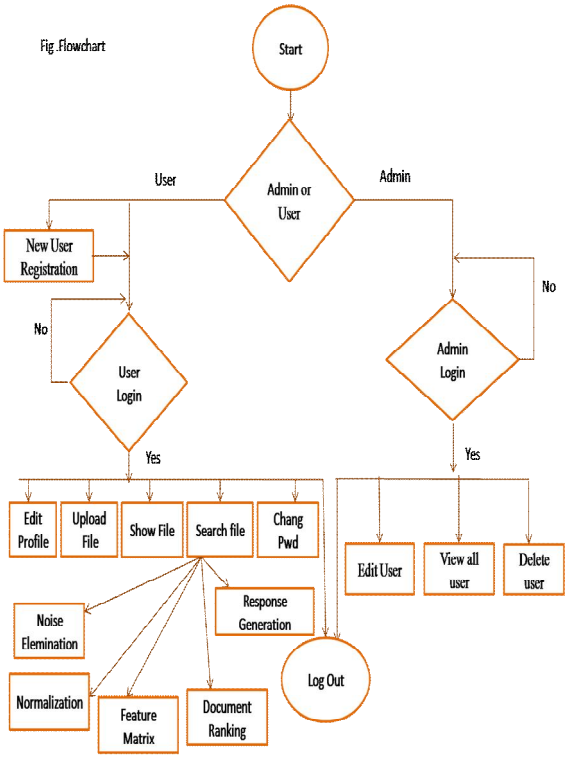
Proposed System Architecture

We are going to propose a system using which the user can easily get the relevance document. When the user enter the query for search the document, then it directly compare within the data of the document file. So the relevant document will found by the system. We are also working to add feature, the system will recommend the keyword to the user for getting the best document.



Architecture

Fig .Flowchart



III. CONCLUSION

Main objective of relevant information retrieval using keyword query is to To measure the performance relevance judgements more accurately and more quickly also To maintain the collection of documents according to different categories and according to different user search. The result given by system is according to the priority or the sequence of the more no of word occurred in each document , the document having more no of word entered during keyword search is placed at the first place of search result along with the time required for each document

REFERENCES

[1] V. Hristidis, L. Gravano, and Y. Papakonstantinou, "EfficientIR-Style Keyword Search over Relational Databases," Proc. 29th Int'l Conf. Very Large Data Bases (VLDB), pp. 850-861, 2003.

[2] F. Liu, C.T. Yu, W. Meng, and A. Chowdhury, "Effective KeywordSearch in Relational Databases," Proc. ACM SIGMOD Conf.,pp. 563-574, 2006.Y. Luo, X. Lin, W. Wang, and X. Zhou, "Spark: Top-K KeywordQuery in Relational Databases," Proc. ACM SIGMOD Conf.,pp. 115-126, 2007.

[3] M. Sayyadian, H. LeKhac, A. Doan, and L. Gravano, "EfficientKeyword Search Across Heterogeneous Relational Databases,"Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 346-355, 2007.

[4] B. Ding, J.X. Yu, S. Wang, L. Qin, X. Zhang, and X. Lin, "FindingTop-K Min-Cost Connected Trees in Databases," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 836-845, 2007.

[5] B. Yu, G. Li, K.R. Sollins, and A.K.H. Tung, "Effective Keyword-Based Selection of Relational Databases," Proc. ACM SIGMODConf., pp. 139-150, 2007.

[6] Q.H. Vu, B.C. Ooi, D. Papadias, and A.K.H. Tung, "A GraphMethod for Keyword-Based Selection of the Top-K Databases,"Proc. ACM SIGMOD Conf., pp. 915-926, 2008.

[7] V. Hristidis and Y. Papakonstantinou, "Discover: Keyword Searchin Relational Databases," Proc. 28th Int'l Conf. Very Large Data Bases(VLDB), pp. 670-681, 2002.

[8] L. Qin, J.X. Yu, and L. Chang, "Keyword Search in Databases: ThePower of RDBMS," Proc. ACM SIGMOD Conf., pp. 681-694, 2009.

[9] G. Li, S. Ji, C. Li, and J. Feng, "Efficient Type-Ahead Search on Relational Data: A Tastier Approach," Proc. ACM SIGMOD Conf.,pp. 695-706, 2009.