Volume No:4, Issue No:6 (November-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

Relevance Feedback Algorithm Inspired By Quantum Detection

Ch. Anil
Department of Computer Science & Systems Engineering,
College of Engineering (A), Andhra University, Visakhapatnam, Andhra Pradesh 530003, India.

Prof. P. Srinivasa Rao
Department of Computer Science & Systems Engineering,
College of Engineering (A), Andhra University, Visakhapatnam, Andhra Pradesh 530003, India.

ABSTRACT

Information Retrieval (IR) 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.

INTRODUCTION

Information Retrieval (IR) is concerned with indexing and retrieving documents including information relevant to a user's information need. Although the end user can express his information need using a variety of means, queries written in natural language are the most common means. However, a query can be very problematic because of the richness of natural language. Indeed, a query is usually ambiguous; a query may express two or more distinct information needs or one information need may be expressed by two or more distinct queries. Text Retrieval Conference (TREC) [1] test collection1 from which the query Mexico City has the worst air pollution in the world. Pertinent Documents would contain the specific steps Mexican authorities have taken to combat this deplorable situation. is submitted to an IR system based on the Vector Space Model (VSM). This system would return both relevant documents and irrelevant documents as shown in Although the number of relevant documents in the top ten document list is quite high, there are some irrelevant documents – for example, LA062790-0048 is irrelevant because it is about a very specific case of river pollution at the Mexican border – and the Mean Average Precision (MAP) is only 15.2%. An IR system addresses the problems caused by query ambiguity by gathering additional evidence that can be used to automatically modify the query. Usually a 1. http://trec.nist.gov/ query is expanded because the queries are short and it cannot exhaustively describe every aspect of the user's information need; however, some irrelevant documents may be retrieved or relevant documents may also be missed when a query is not short as shown in the previous example. The automatic procedure that modify the user's queries is known as Relevance Feedback (RF); some relevance assessments about the retrieved documents are collected and the query is expanded by the terms found in the relevant documents, reduced by the terms found in the irrelevant documents or reweighted using relevant or irrelevant documents. RF has a long history: It was proposed in the 1960s; it was implemented in the SMART system in the 1970s in the context of the VSM; it was investigated at the theoretical level; it eventually attracted interest from other researchers because of the consistent effectiveness improvements observed in many experiments [2-5].

Cite this article as: Ch. Anil & Prof. P. Srinivasa Rao, "Relevance Feedback Algorithm Inspired By Quantum Detection", International Journal of Research in Advanced Computer Science Engineering, Volume 4 Issue 6, 2018, Page 7-12.

Volume No:4, Issue No:6 (November-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

EXISTING SYSTEM

The automatic procedure that modify the user's queries is known as RF; some relevance assessments about the retrieved documents are collected and the query is expanded by the terms found in the relevant documents, reduced by the terms found in the irrelevant documents or reweighted using relevant or irrelevant documents.

RF can be positive, negative or both. Positive RF only brings relevant documents into play and negative RF makes only use of irrelevant documents; any effective RF algorithms includes a "positive" component. Although positive feedback is a well established technique by now, negative feedback is still problematic and requires further investigation, yet some proposals have already been made such as grouping irrelevant documents before using them for reducing the query [3]

PROPOSED SYSTEM

It is designed to compute the new query vector using a linear combination of the original vectors, the relevant document vectors and the non-relevant document vectors, where the labels of relevance are collected in a training set.

Quantum probability is the theory of probability developed within Quantum Mechanics (QM). In QM, a probability space can be represented as vectors, matrices and operators between them. A tutorial would be out of the scope of this paper, therefore we provide the information instrumental to understanding the rest of this paper.

Detection consists of identifying the information concealed in the data which are transmitted by the source placed on one side, through a channel to the detector placed on the other side. The data are only a representation of the "true" information that one side wants to transmit [4].

IMPLEMETATION TECHNOLOGIES USED:

HTML:

HTML (hyper text markup language) is a language used to create hyper text documents that have hyper links embedded in them. It consists of tags embedded in the text of a document with HTML. We can build web pages or web documents. It is basically a formatting language and not a programming language. The browser reading the document interprets mark up tags to help format the document for subsequent display to a reader. HTML is a language for describing structured documents. HTML is a platform independent. WWW (World Wide Web) pages are written using HTML [5]. HTML tags control in part the representation of the WWW page when view with web browser. The browser interprets HTML tags in the web document and displays it. Different browsers show data differently. Examples of browsers used to be web pages include:

- Netscape
- Internet Explorer

Hypertext Markup Language (HTML) is the standard markup language for creating web pages and web applications. With Cascading Style Sheets (CSS) and JavaScript, it forms a triad of cornerstone technologies for the World Wide Web. Web browsers receive HTML documents from a web server or from local storage and render them into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document [6].

HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects, such as interactive forms, may be embedded into the rendered page. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. HTML elements are delineated by tags, written using angle brackets. Tags such as <img /> and <input /> introduce content into the page directly. Others such as ... surround and provide information about document text and may include other tags as sub-

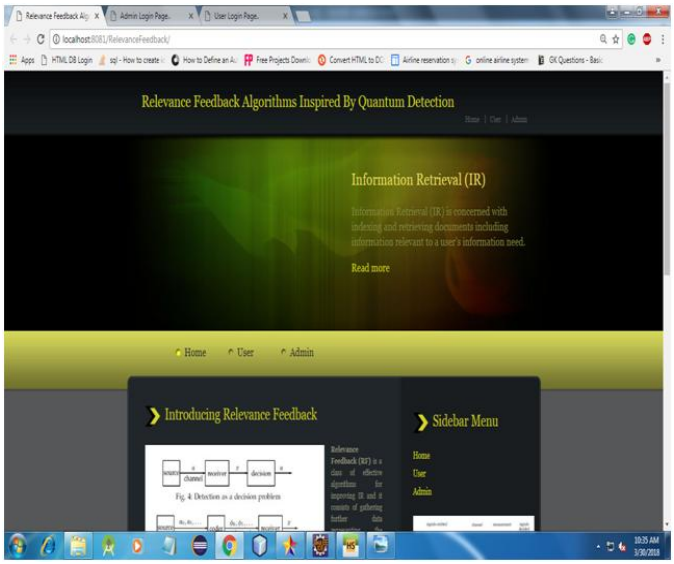
Volume No:4, Issue No:6 (November-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

elements. Browsers do not display the HTML tags, but use them to interpret the content of the page.

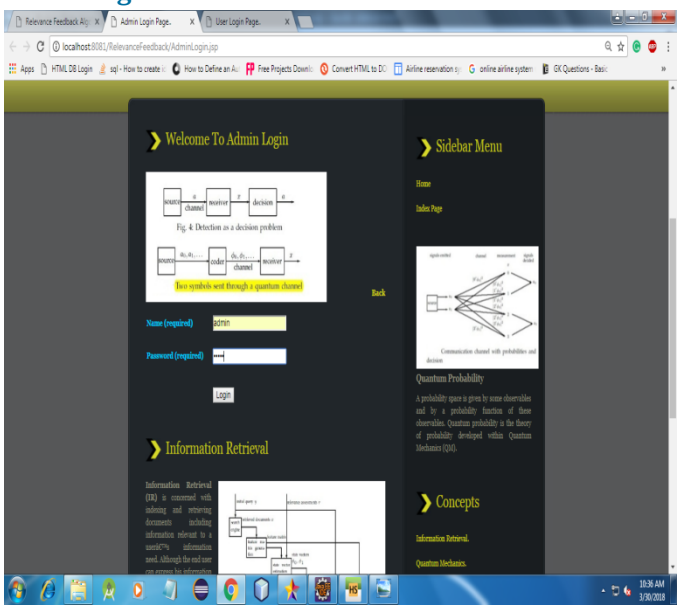
HTML can embed programs written in a scripting language such as JavaScript which affect the behavior and content of web pages. Inclusion of CSS defines the look and layout of content [7-12].

SCREENSHOTS

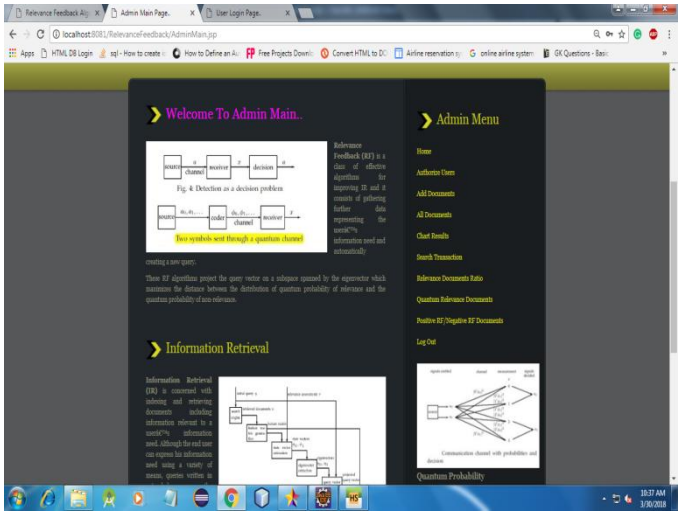
Home page



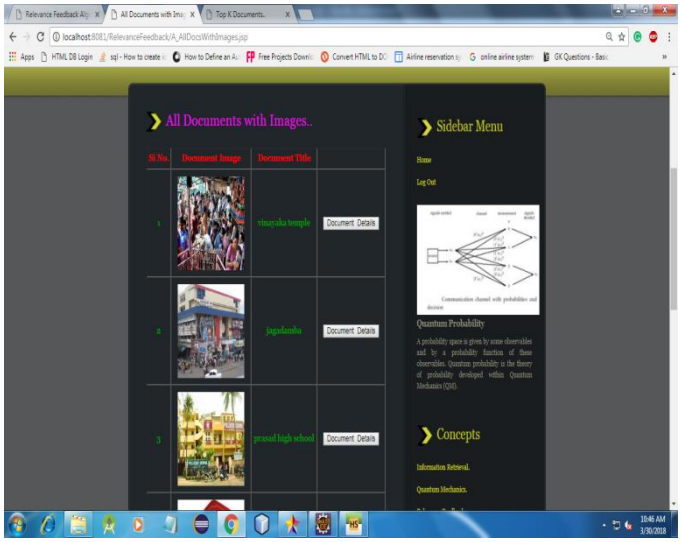
Admin login

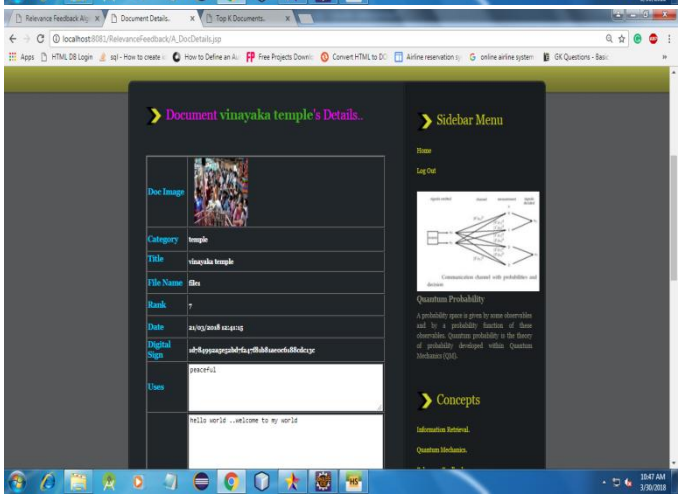


Admin main page



All documents





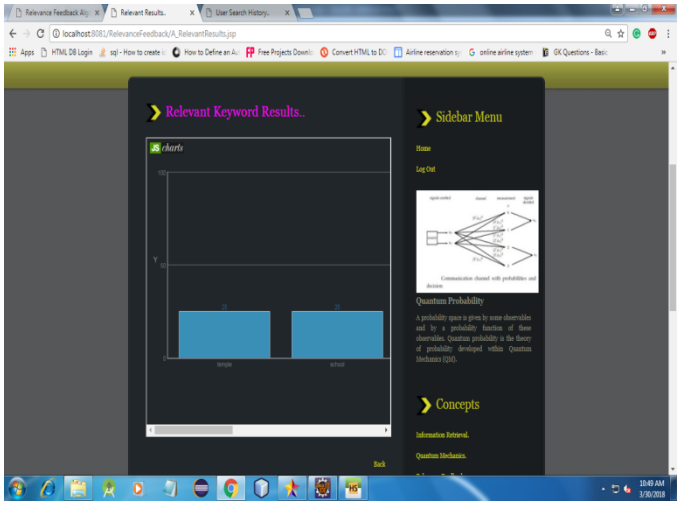
Volume No:4, Issue No:6 (November-2018) ISSN No : 2454-423X (Online)

International Journal of Research in Advanced Computer Science Engineering

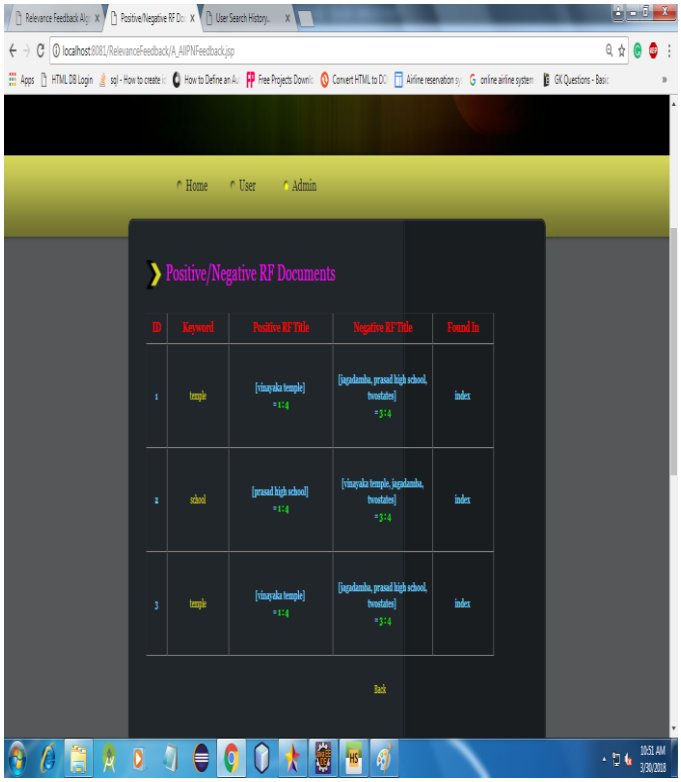
A Peer Reviewed Open Access International Journal

www.ijracse.com

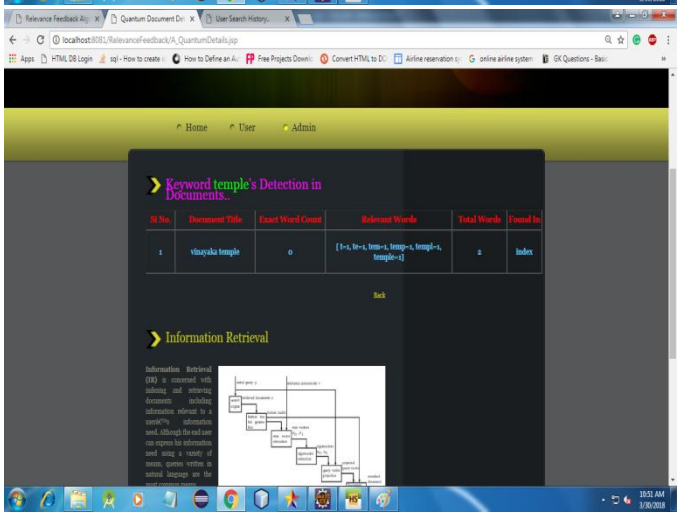
Relevant words



Positive/ Negative RF



Documents ratio



User registration

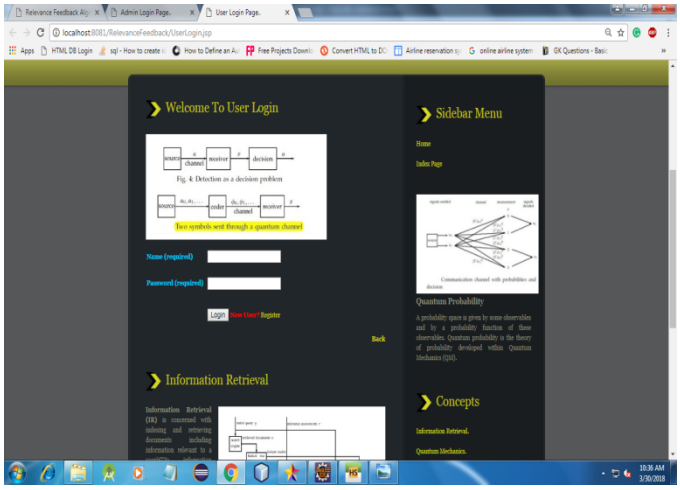
Volume No:4, Issue No:6 (November-2018) ISSN No : 2454-423X (Online)

International Journal of Research in Advanced Computer Science Engineering

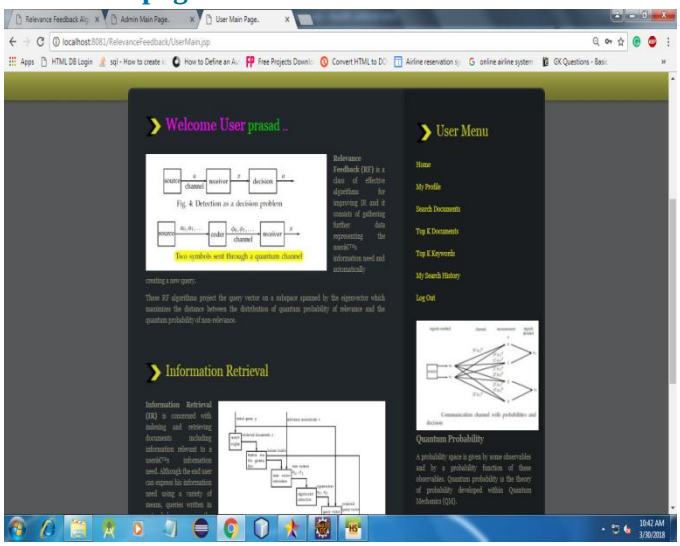
A Peer Reviewed Open Access International Journal

www.ijracse.com

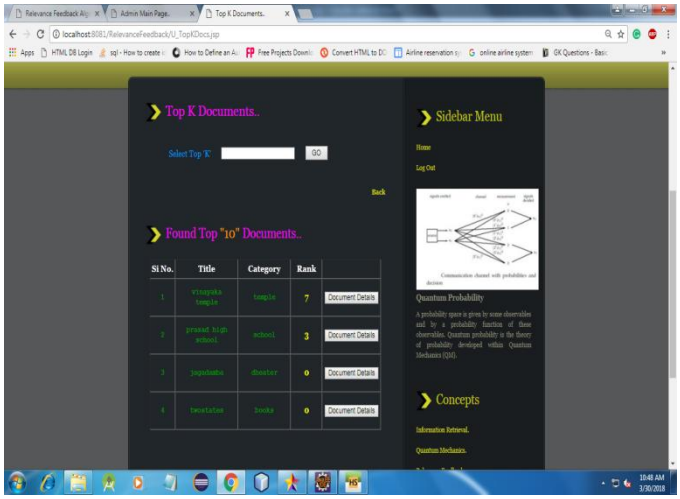
User login



User home page



Top K documents



Top K keywords

Sl No.	Keyword	Rank
1	books	7
2	temple	7
3	vizag	5
4	school	4
5	file1	1
6	mumbai	1
7	temple	1
8	properties props = system.getproperties(); props.put(	1

CONCLUSION AND FUTURE ENHANCEMENT

In this paper, a class of RF algorithms inspired by quantum detection has been proposed to re-weight query terms by projecting the query vector on the subspace represented by the eigenvector which is the optimal solution to the problem of finding the maximal distance between two quantum probability distributions. RF is then viewed as a signal detection technique – relevance is the document state to be detected and the queries are the detectors. First, the documents retrieved by an IR system to answer the original query are used to extract a feature matrix. Second, some relevance assessments are obtained according to whether RF is explicit or pseudo. The quantum probability distributions can be estimated and the optimal solution of a distance between two quantum probability distributions can be calculated. The eigenvector that results from this optimization problem can be utilized to project the query vector. Third, the retrieved documents can be re-ranked to answer the modified query.

The query term re-weighting is different from the re-weighting performed by the classical RF algorithms since each query term variation depends on the other

Volume No:4, Issue No:6 (November-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

query term variations, thus capturing a kind of term dependence which is not captured by other RF algorithms. Our approach has low complexity and can be used in reality. For each query, the running time of the first document retrieval depends on the number of query terms as usual. The construction of the feature matrix depends on the number of retrieved documents used to estimate the probability distributions – our experiments showed that a few dozen documents can be sufficient.

REFERENCES

[1] A. Aji, Y. Wang, E. Agichtein, and E. Gabrilovich, "Using the past to score the present: Extending term weighting models through revision history analysis," in Proc. 19th ACM Conf. Inf. Knowl. Manage., 2010, pp. 629–638.

[2] R. Blanco and P. Boldi, "Extending BM25 with multiple query operators," in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2012, pp. 921–930.

[3] C. Carpineto and G. Romano, "A survey of automatic query expansion in information retrieval," ACM Comput.Surv., vol. 44, no. 1, pp. 1–50, Jan. 2012.

[4] K. Collins-Thompson, P. N. Bennett, R. W. White, S. de la Chica, and D. Sontag, "Personalizing web search results by reading level," in Proc. 20th ACM Int. Conf. Inf. Knowl. Manage., 2011, pp. 403–412.

[5] W. Croft, D. Metzler, and T. Strohman, Search Engines: Information Retrieval in Practice. Reading, MA, USA: Addison-Wesley, 2009.

[6] I. Frommholz, B. Larsen, B. Piwowarski, M. Lalmas, P. Ingwersen, and K. van Rijsbergen, "Supporting polyrepresentation in a quantum-inspired geometrical retrieval framework," in Proc. 3rd Symp. Inf. Interaction Context ins, 2010, pp. 115–124.

[7] I. Frommholz, B. Piwowarski, M. Lalmas, and K. van Rijsbergen, "Processing queries in session in a quantum-inspired IR framework," in Proc. Eur. Conf. Inf. Retrieval, 2011, pp. 751–754.

[8] R. B. Griffiths, Consistent Quantum Theory. Cambridge, U.K.: Cambridge Univ. Press, 2002.

[9] M. Gupta and M. Bendersky, "Information retrieval with verbose queries," Found. Trends Inf. Retrieval, vol. 9, nos. 3/4, pp. 91–208, 2015.

[10] D. Harman, "Relevance feedback revisited," in Proc. 15th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Copenaghen, Denamrk, 1992, pp. 1–10.

[11] B. He, J. X. Huang, and X. Zhou, "Modeling term proximity for probabilistic information retrieval models," Inf. Sci., vol. 181, no. 14, pp. 3017–3031, Jul. 2011.

[12] R. Hughes, The Structure and Interpretation of Quantum Mechanics. Cambridge, MA, USA: Harvard Univ. Press, 1989.