Performance Measurement of Refinding Queries in Context Based Search

Sayyad F S[1], Prof. P.P. Kalyankar [2]

Lecturer, Department of CSE, COE, Osmanabad, India [1]

Professor, Department of CSE, COE, Osmanabad, India [2]

Abstract: In this paper we are measuring the performance of refinding queries that contains the information retrieval for Personal Information Management System. When user needs the already searched web pages repeatedly then this search engine provides the user interface for saving the title and the web page URL for reference .In this keyword and related web pages are stored we can refind needed. Also search engine contains the association and clustering of context instances.

Keywords: context instances, Personal Information Management, refinding queries, SVM.

I. INTRODUCTION

In this paper we are trying to measure the performance of refinding queries for the personal Information Management fundamentals. The implemented methodology gives one search bar for finding the personal information of a user which is previously searched and gives one database to save the searched web pages.

Numbers of methods have been implemented for information retrieval. Diversity of methods has been implemented for web information for access and reuse. Number of techniques such as bookmarks, History retrieval, and search engines and so on are present in the proposed system. In the existing search engine we are trying to implement the context based search for finding the personal information management. When user enters a query if it is similar to a previous query it obtains the current result from its cache. In this search engine we can also use textual information for the particular web search such as time of accessing, name of the author, publishing date etc. for example. The user needs to store the title and the URL of the searched web page in the database. The association among the keyword and the searched web pages gives the best result for the information retrieval.

II. RELATED WORK

Any information that we have searched may be analysed by any web search. One of the basic themes is that information in the Personal Information Management [1] can be analysed in the context only.

While great efforts have been made to find the contextual information of any user like session, place, field of work. Context is nothing but any information that can characterize and is relevant to the interaction between a user and an application.Techniques that are presented for refinding of information is explained in [1]. In this technique Personal Information Management System is used for the context analysis. Context for the Metadata is presented in [12] in which properties of data contents is incorporated and indexed for personal information retrieval.

Soules and Ganger [22] developed a file search engine combining content - based search with temporal relationships between files gathered from user's file operations. The main aim of this search engine is to find the refinding queries occurrence and quality of the search engine. A more general way for the analysis of context creation for the system demonstration such as activity, time and place are used for the refinding of the user searched information [1].

Many of the techniques related with the data mining and information retrieval contains the theme for refinding queries. The implemented search engine contains context creation and the web page linking to the related context Attribute. One of the technique used in the[17] uses Information Finding by association. The system interface is such a designed model for the information finding by association. It consists of multiple levels of association and algorithms for collecting the association information and providing answers to real-time queries.

We are also measuring the performance for the number of context instances and the number of dimensions used for the context instances .A comparative study of all the refinding methodologies is presented in [16] gives the basic idea of all the information finding systems for the analysis of all the methods.

Information or the history finding of the user by using the context analysis is one of the part of information refinding but information refinding is not the just information finding is explained in the[12].

How to improve a recall search of the user history is implemented in the [3].That system provides the recall capability of the user for the search history.

One of the technology used for the search engine is explained in the [14] for researching of the web pages by using the finding and refinding of the search pages. This technique is useful for the web history searching fundamentals. This system facilitates the user to search old information and answers to conflicting information goals.

Presented technology also contains search engine for information retrieval and information management with the help of context instances their hierarchy and context creation for system manipulation.

III. METHODOLOGY

A. Synthetic Data Generation

Data required for gathering of information is related with the daily user's information refinding on the web. Many times user needs the same information repeatedly for the web management or daily information management. For such work the presented system contains context instances creation and linking of web pages.

Context instances are managed by the values we are storing in the databases. In the implementation technique we are using the context based search engine that takes keyword as a context and association of all the attributes related to that context. The search engine is nothing but the one of the form of personal information management system.

B. Linking of web pages

After creating the context attribute we link the repeatedly searched web pages to the keyword. Context instances are the objects that are used for accessing information of a particular user. In the presented search engine the keywords are used as a context attribute of a user for finding the previously searched information. Association of clustered context can be done after finding the measure of similarity between the context attributes.. For information accessing the related keyword is used in the search bar and the searched web pages are listed as a ranked list of output pages. The information can be retrieved from the specific database used in the implementation. The output is nothing but the previously searched web pages by the user for refinding of accessed information. So the system is completely implemented for Personal Information Management for the searched web pages.

C. Generation of Refinding queries

A group of user is used in the search engine for the performance measurement of the refinding queries. That group is formed by the range of age values and the gender of the user that finds the related information. In generating refinding request the user is repeatedly searching the information so the system needs to understand the refinding request parameters.

D. Context Memory Snapshot

Context memory snapshot gives the short interval of context memory for the refinding request. Long Term Context Memory consist of Two units Permanent and Evolving. Permanent unit has lifelong accessing experience and evolving unit will decay later. In performance measurement of refinding request on age group is making the search and system can evaluate the percentage of accessing the gender and age group of the user. In this way we are calculating the used percentage of gender searching repeatedly the same information. At the same time system also calculates the age wise percentage of same query searched by the user in different age group. Like below 13,13-20,21-40 and above 40 and so on. The % wise

distribution gives the result for accessing information and that will help for the calculation user precisiohn and recall parameters.

E .Algorithm for refinding queries calculation
Input: A keyword to search engine.
Output: A list of search web pages already searched.
Algorithm:
1. Create keyword for the web pages with the specific search.
2. Save web page URL, content and title in a search engine window.
2. For each search save the pages with the related keyword.
3. Enter the keyword for search.
4. Get the link pages of searched pages.
5. Search the next pages with related context.

IV. EVALUATION

In Evaluation we are measuring the performance of refinding queries based on the gender and the age group that classify the refinding queries in different groups.
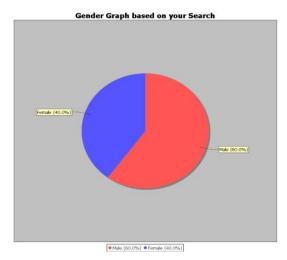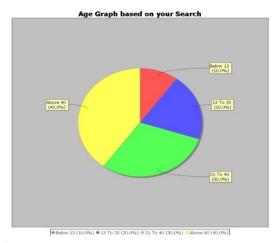

Fig1.1 Gender graph based on the search

In all the refinding queries the group is classified into different ranges as shown in the above Fig 1.1. The classification gives the percentage of gender that repeatedly accessing the same information for the web page search.

Fig1.2 Age graph based on the query to be searched.

The classification by using the age group is also shown in Fig 1.2 that gives the age range of users that are repeatedly accessing the information.

In refinding queries number of methods have been implemented that uses different algorithms for the presented implementation.

Number of techniques are present now a days for the re-finding of information that was previously searched because the search methods are same for the user as inserting the just keyword or the related word of search at these instances context based web searching is useful.

With the technique used above the context similarity calculation is presented as a part of web history search engine methodologies. The basic theme of the presented search engine is to manage Personal Information Management System of a user searched web pages. As the human memory is decaying. its reinforcement is necessary. Above implemented technique is well suited for finding the previously searched information of a user as a part of Personal Information Management.

The graph for the search performance with the help of different searched methods is shown in Fig 1.3.The methods such as SVM (Searched Vector Method),Bayesian algorithm, Decision Tree algorithm is also implemented with this search.
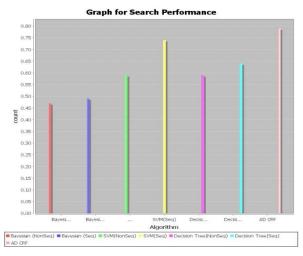

Fig1.4. Graph for search performance with different algorithms

Context based search engine is implemented in Java running on Windows & and more with JDK_1.7 and Apache Tomcat Server v6.0 on eclipse-jee-luna-R-win32 as IDE. The matched results are calculated by using the precision and recall parameters for the information refinding.

The precision and recall values give the best answer to find the quality of the system

.

$$\text{Precision} = \frac{\text{No.of true result being matched}}{\text{No.of matched result}}$$

$$\text{Recall} = \frac{\text{No.of true result being matched}}{\text{No.of true result}}$$

$$\text{F-measure} = 2 \cdot \frac{\text{Precisio.Recall}}{\text{Precision+Recall}}$$

The graph for the precision and recall parameters is shown in the following Fig 1.5.Above formula is used for finding the precision and recall parameters of the related search. After calculating precision and recall parameters we can calculate the F-measure of the search.
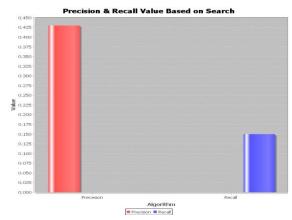

Fig 1.5 Graph of precision and recall parameters

V. APPLICATIONS

Number of techniques are present now a days for the re-finding of information that was previously searched because the search methods are same for the user as inserting the just keyword or the related word of search at these instances context based web searching is useful.

With the technique used above the context similarity calculation is presented as a part of web history search engine methodologies. The basic theme of the presented search engine is to manage Personal Information Management System of a user searched web pages. As the human memory is decaying. its reinforcement is necessary. Above implemented technique is well suited for finding the previously searched information of a user as a part of Personal Information Management. Number of methods using clustering and association of context attributes, so we can easily form the user characteristics. Decisions can be made for the user history, user behavior, accessed information, access context etc.

VI. CONCLUSION

We have used the basic fundamentals for finding the similarity of context instances and their attributes. Selection criteria are dependent on user's choice. Using any web history search engine context analysis and context annotation is the basic fundamental step. In nowadays peoples are searching huge amount of information on the web and saving to there system or just seen it and forgot for a small amount of time. Numbers of search engines are there for finding relevant information by a single search. But sometimes we doesn't get the exact web page for the single search in this situation also context based search techniques are useful. While using context as the way of search methodologies context attributes are grouped together to form the context instance and then after clustering and association can be performed as the basis of search engines. Hence we can use this search engine in Personal Information Management.

VII. FUTURE SCOPE

Context degradation is one of the important part as we are dealing with the context information finding. Context attributes value degrades along a hierarchy from specific to general. As human memory is also degraded as the age goes by, same like the unnecessary information present in the context memory snapshot is also to be removed for inserting new values of context attributes and context instances.

Another way of improving the search engines methodology is that context can be annotated automatically. This technique can be used for the better analysis of inserting and deleting the contexts for the search engines.

As context degradation gives the new context in the web searching they also needs to go with similarity measure findings of context attributes and context instances.

REFERENCES

[1] Tangjian Deng, Liang Zhao, Hao Wang, Qingwei Liu and Ling Feng" Refinder: A Context Based Information Refinding System" IEEE transaction a Knowledge and Data Engineering, Vol.25, NO.9, September 2013.
[2] M. Fuller, L. Kelly, and G.J.F. Jones, "Applying Contextual Memory Cues for Retrieval from Personal Information Archives, "Proc. Personal Information Management Workshop at CHI, 2008.
[3] J. Hailpern, N. Jitkoff, A. Warr, R. Karahalios, K. Sesek, and N. Shkrob, "YouPivot: Improving Recall with Contextual Search," Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI), 2011.
[4] S.K. Tyler and J. Teevan, "Large Scale Query Log Analysis of Re-Finding," Proc. Third ACM Int'l Conf. Web Search and Data Mining (WSDM), 2010.
[5] C.A.N. Soules and G.R. Ganger, "Connections: Using Context to Enhance File Search," Proc. 20th ACM Symp. Operating Systems Principles (SOSP), 2005.
[6] Y. Chen and G. Jones, "Integrating Memory Context into Personal Information Re-Finding," Proc. Second Symp. Future Directions in Information Access, 2008.
[7] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D.C.Robbins, "Stuff I've Seen: A System for Personal Information Retrieval and Re-Use," Proc. 26th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2003.
[8] S. Won, J. Jin, and J. Hong, "Contextual Web History: Using Visual and Contextual Cues to Improve Web Browser History," Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI), 2009.
[9] D. Morris, M.R. Morris, and G. Venolia, "Searchbar: A Search-Centric Web History for Task Resumption and Information Re-Finding," Proc. SIGCHI Conf. Human Factors in Computing Systems(CHI), 2008.
[10] B. MacKay, M. Kellar, and C. Watters, "An Evaluation of Land-marks for Re-Finding Information on the Web," Proc. Extended Abstractson HumanFactors in ComputingSystems(CHI'05 EA), 2005.
[11] L. Feng, P.M. Apers, and W. Jonker, "Towards Context-Aware Data Management for Ambient Intelligence," Proc. 15th Int'l Conf. Database and Expert Systems Applications (DEXA), 2004.
[12] R. Capra, M. Pinney, and M.A. Perez-Quinones, "Refinding Is Not\ Finding Again," technical report, Aug. 2005.
[13] L. Tauscher and S. Greenberg, "How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems," Int'l J. Human Computer Studies, vol. 47, pp. 97-137, 1997.
[14] J. Teevan, "The Re:Search Engine: Simultaneous Support for Finding and Re-Finding," Proc. 20th Ann. ACM Symp. User Interface Software and Technology (UIST), 2007.
[15] E. Tulving, "What is Episodic Memory?" Current Directions in Psychological Science, vol. 2, no. 3, pp. 67-70, 1993.
[16] A.P. Nivethith,, D. Kerana Hanirex, Dr. k. P.K. Kaliyamurthie "A Comparative Study Of Context-Based Information Refinding" An international journal of advanced computer technology, 2004.
[17] Due Horng Chau, Brad Myers and Andrew Faulring" Feldspar: A System for Finding Information by Association" CHI 2008, Florence, Italy.
[18] Google Web History, http://www.google.com/history, 2013.
[19] Prathyusha Kanakam1, S. Mahaboob Hussain2, Dr. Sumit Gupt3, Dr. D. Surya Narayana4 "An Analysis of Exploring Information from Search Engines in Semantic Manner".
[20] Steve Lawrence and C.Lee Giles" Context and Page Analysis for Improved Web Serach" NEC Research Institute.
[21] J.Z. Tsien, The Memory. Scientific Am., 2007.
[22] E. Tulving, Elements of Episodic Memory. Oxford Univ. Press, 1983.
[23] E. Tulving, "What is Episodic Memory?" Current Directions in Psychological Science, vol. 2, no. 3, pp. 67-70, 1993.
[24] M. Mayer, "Web History Tools and Revisitation Support: A Survey of Existing Approaches and Directions," Foundations and Trends in Human-Computer Interaction, vol. 2, no. 3, pp. 173-278, 2009.

BIOGRAPHIES

Ms. Sayyad Farhina Anjum Shabbir Ahmed is a Student Of ME(CSE)at TPCT's College Of Engineering Osmanabad.Her research interest include in a Data Mining and Knowledge Discovery. She was the University Ranker of Dr.BAM University.

Prof. P. P. Kalyankar is a Assistant Professor Of Computer Science And Engineering Department at TPCT's College Of Engineering Osmanabad.His research interest include in a computerscience. Membership: LIFE MEMBER OF LMISTE AMIE. Working as a HOD (MCA).He is also recognized as a PG Teacher for ME and MCA Classes.