Context Based Search Engine for Personal Information Management

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

[1,2]*Department of M.E.(C.S.E.),*
T.P.C.T.'s College Of Engineering, Osmanabad, India.

Abstract— **In this paper we are implementing a context based search engine for managing the personal Information of a user. Personal Information management includes finding the users previously search web history.PIM includes memory management unit as a context memory for managing the context information of a user.**

Keywords—**context instances , context annotation ,PIM.**

I. INTRODUCTION

In this paper we are trying to implement the search engine which is based on the personal Information Management Unit of a user searched web pages. Many a times a user search a web page or for finding a information which is currently needed to him and close that web page .

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.

A context based search engines includes finding of information based on its context. Context is nothing but any information that can characterize and is relevant to the interaction between a user and an application .

Diversity of methods have been implemented for web information for access and reuse.Number of techniques such as bookmarks, History retrieval, 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.

II. RELATED WORK

Any information that we have searched may be analysed by any web search. One of the basic theme 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 techniques 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.

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].

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 consist of multiple levels of association and algorithms for collecting the association information and providing answers to real-time queries.

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 facilitate the user to search old information and answers to conflicting information goals.

III.METHODOLOGY

A. Personal Information Management Unit

Context memory model is useful for context based information refinding [01].Context memory is divided into two units: Short Term context Memory and Long term Context Memory. Short Term Context Memory is limited in capacity and lasts for few days. Long Term Context Memory is unlimited in capacity and lasts for as short as few days or as long as decades.

Long Term Context Memory consist of Two units Permanent and Evolving. Permanent unit has life long accessing experience and evolving unit will decay later.

Same methodology we will use for the implementation purpose that number of pages revisited frequencies and access logs will help us to determine the exact web page that the user need.

Number of circumstances under which we are accessing user's information is nothing but the access context. Context information may be internal or external. Such as user name, activity, agenda etc. or name of the author ,publishing date ,time etc.

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. Creation Of Context Instances

Context is nothing but any information that can characterize and is relevant to the interaction between a user and an application[1].Techniques that are presented for refinding of information is explained in[1].In this techniques Personal Information Management System is used for the context analysis. Context for the Metadata is presented in[2] in which properties of data contents is incorporated and indexed for personal information retrieval.

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.

C. Clustering Of Context Instances

Clustering of context instances can be obtained by combining all the searched web pages starting keyword or the related word referencing to that specific search. Each and every cluster is associated with its searched web pages. A presented in[1]cluster of all the attributes is implemented for the association of context instances. Similarity measure of attributes related to the context instances can also be calculated. The measure of similarity outputs the result.

D. Association of Clustered Context Attributes

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.

The responsibility of context manager is to identify the closely matched result and return the linked web pages via file path and the URL stored in the personal linkage repository as shown in the overall architecture in Fig 1 the database is retrived with the most searched pages.

E. Algorithm For Finding the searched web pages by context

Algorithm 1. Context clustering and association based information accessing.

Input: Context attribute and related searched web pages and a query Q.

Output :A ranked list of web pages that matches with the keyword query Q.

1:Create keywords as the context attribute $K_1, K_2, K_3, \ldots K_n$.

2:Store related web pages title and URL of searched web pages $W_1, W_2, W_3, \ldots W_n$.

3.Clustering of $W_{11}, W_{12}, W_{13}, \ldots W_{1n}$ with K_1 and so on.

4.Association of $K_1, K_2, K_3, \ldots K_n$ with $W_1, W_2, W_3, \ldots W_n$.

5.Rankedf List $L_1, L_2, L_3, \ldots L_n$ of searched web pages for query Q.

IV. IMPLEMENTATION

A. Context Based Search Engine Architecture

The Overall architecture of the Context Based Search Engine is shown in the following fig.1.In this architecture the first step is to create context instances follows with keyword allocation for the related context instances for the ease of searching web pages. Clustering and associations are performed on the context attributes and context instances respectively
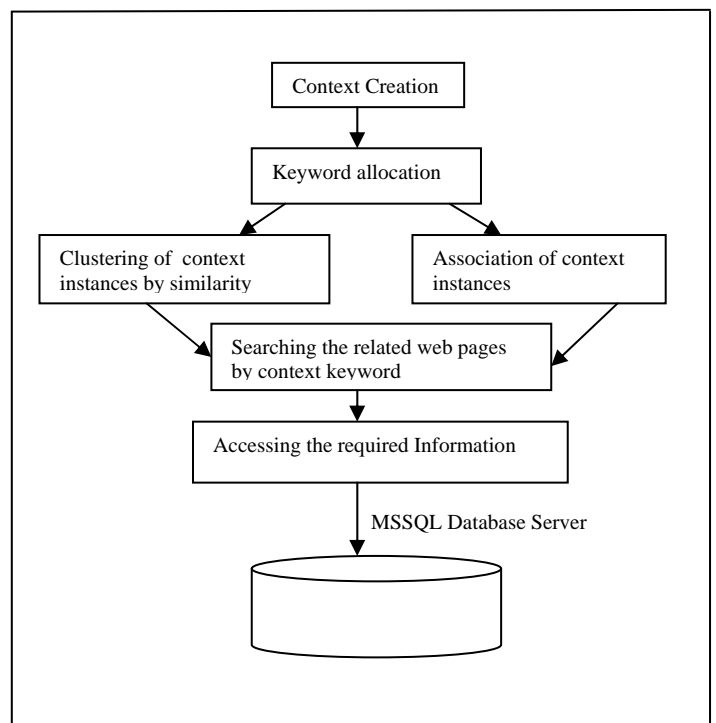

Fig1.The overall architecture of Context Based Search Engine

B. User Interaction for Information Refinding

In the implementation of Context based search engine user is interacted with the window for searching out the previously searched web pages by a search bar as shown in the following Fig 2.
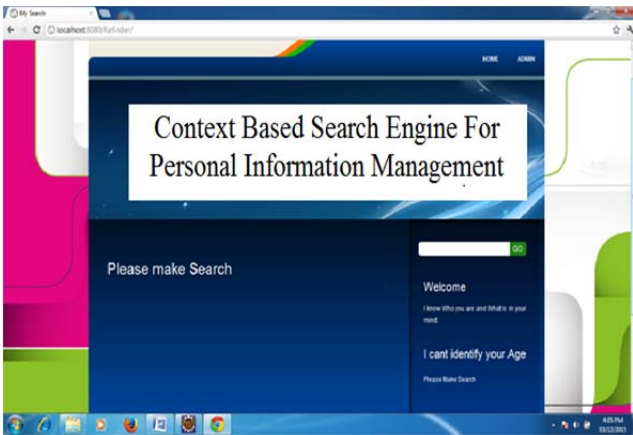
Fig 2.Context Based Search Engine Window.

One of the administrator platform is there to manage the URL of context memory unit, so that the information is saved in the preferred database as shown in the following Fig3.
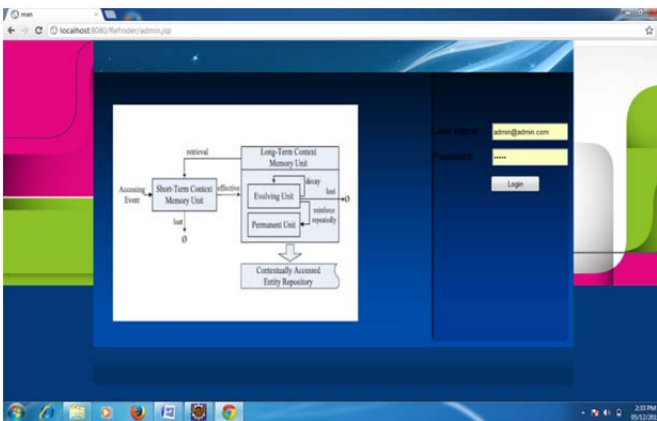

Fig 3.Login Window for admin to manage information.

New context attribute or the keyword can be inserted by the admin for managing the context attribute generation as presented in Fig4.
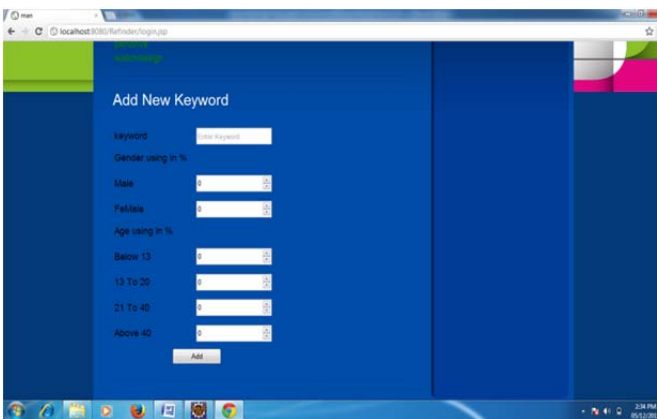

Fig. 4.Keyword annotation as context attribute

The searched web pages URL can be managed by content and title of searched web pages as in Fig 5.
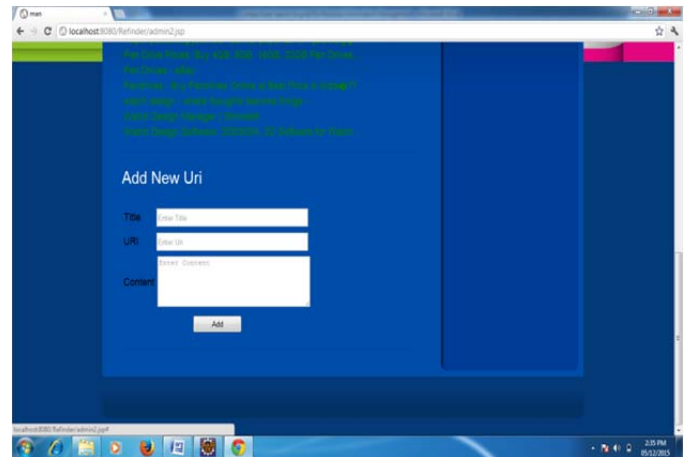

Fig 5.Managing URL of searched web pages.

When user wants to search a particular information of previously searched web pages one has to enter the related context of that information or the context related to that information. Fig 6 shows the search page for the user query.
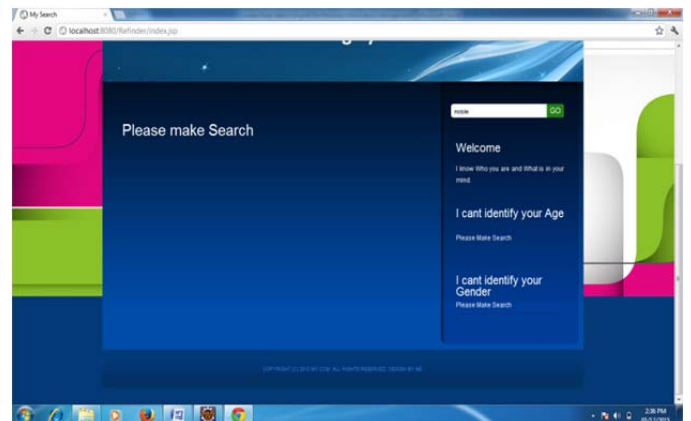

Fig 6.User Search for searched pages.

After entering user query a ranked list of searched pages that being matched by the context similarity measure will be in the output window as in Fig 7.
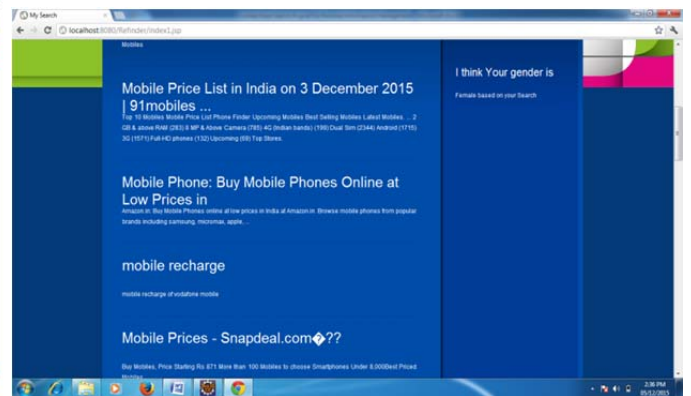

Fig 7.Ranked List Of Searched Web pages.

When user enters a qury or the related context to the search engine output is the ranked list of searched web pages as implemented in the above methodology.

V. EVALUATION

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 gives the best answer to find the quality of the system.

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

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

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

The precision,recall and F-measuire for the present system are(74.38,98.09 and 84.60 percent).

VI. 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.

REFERENCES

[1] Tangjian Deng,Liang Zhao,Hao Wang,Qingwei Liu and Ling Feng"Refinder:A Context Based Information Refinding System"IEEE transaction an 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 ReFinding," 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.KeranaHanirex, Dr.k.P.K.Kaliyamurthie "A Comparative Study Of Context-Based Information Refinding" An international lournal 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 inPsychological 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.