

## A Survey: Improving Efficiency of Mobile Search Engine Using Personalization

Miss. Puja N. Vilayatkar, Prof. R. R. Shelke

H.V.P.Mandal's College of Engineering & Technology Assistant Professor, Department of Computer Science & Engineering  
Amravati, Maharashtra H.V.P.Mandal's College of Engineering & Technology  
ME Second year computer science & Engg. Amravati, Maharashtra  
*poojavilayatkar@gmail.com, rajeshrshelke@rediffmail.com*

**Abstract:-** Web databases produce inquiry result pages in perspective of a customer's request. The objective of proposed structure is to think composed data which are the pages containing courses of action of data records from a social occasion of pages from different web data bases and modify them in one design, so customer can get more noteworthy data. Hence removing the data from these request result pages is fundamental for a few applications, for instance, data blend, which need to organize with different web databases. For this, data extraction and course of action methodology are proposed. For extraction, CTVS that merges both mark and regard likeness techniques are used to remove the data from different web databases. For Alignment, re-situating schedules are proposed which uses semantic equivalence to upgrade the way of rundown things. Bring the top N results returned through web searcher, and use semantic comparability's between the candidate and the request to re-rank the results. To begin with follower the situating position to a centrality scores for each candidate. By then combine the semantic closeness score with this early on criticalness score finally get the new positions. Using the hugeness score for each site page structure make sense of the relevance of data. Finally alter the data in dropping solicitation from that score.

**Keywords -** Data extraction, data record alignment, information integration, CTVS

\*\*\*\*\*

### I. Introduction

Online databases incorporate the significant web. Differentiated and webpage pages in the surface web, which can be got to by an unprecedented URL, pages in the significant web are capably made in light of a customer inquiry submitted through the request interface of a web database. In the wake of tolerating a customer's question, a web database gives back the noteworthy data, either sorted out or semi organized, encoded in HTML pages.

Various web applications, for instance, Meta questioning, data mix and examination shopping, need the data from different web databases. For these applications to further utilize the data embedded in HTML pages, modified data ex-balance is indispensable. Exactly when the data are uprooted and dealt with in a sorted out route, for instance, tables, would they have the capacity to be considered and gathered. Consequently, correct data extraction is key for these applications to perform precisely.

The objective of this endeavor is to focus data from various web data bases and change them in one association. Where anyone fires an inquiry for they get a result from one particular database and it should be obliged one. In any case, if data begin from diverse web databases, then it contains more results as com-pared to single database. The advantage of using different web databases is that we get more critical data .For this we used two databases Google and Bing. With the presence of information development, a customer has the

limit get applicable information from the World Wide Web, which contains a colossal measure of information, basically and quickly by entering request questions. On account of information and pass on it direct to the customer.

### II. Literature Survey:

Web database extraction has become much thought from the Database and Information Extraction research domains starting late due to the volume and nature of significant web data. As the returned data for an inquiry are embedded in HTML pages, the investigation has focused on the most capable technique to think this data.

UllasNambiar and SubbaraoKambhampati [2], dispersed their paper "Giving Ranked Relevant Results to Web Database Queries" in which they proposed to give situated responses to customer request by perceiving a game plan of inquiries from the request log whose answers are critical to the given customer request. They use an information recuperation based approach to manage find the closeness among request and utilize it to perceive appropriate results. The procedure can be completed without impacting the internals of a database in this way showing it could be adequately executed over any present Web databases. In any case, the work focuses just on giving situated responses to request over a singular database association and there is degrees for making approach for join questions over various relations.

R. R. Shelke, Dr. R. V. Dharaskar, Dr. V. M. Thakare [4], studied A mobile computer is effectively any computing device not constrained in its location to a desktop or data Centre. In recent years the variety of mobile computing devices available has rapidly increased. In doing so, it has also turned from theory to reality a trend for ubiquitous computing, whereby computers are all around us in the world, enabling access to digital content anytime, anyplace and anywhere.

V.kalyan Deepak and N.V.Rajeesh Kumar [3], systematically examined the issue of customized annotation approach in the paper "Recuperate Records from Web Database Using Data Alignment" which has disseminated in 2014, that first modifies the data units on a result page into assorted get-togethers such that the data in the same social occasion have the same semantic. By then, for each social affair, elucidate it from different edges and aggregate the particular annotations to foresee a last annotation name for it. They reason that exact plan is essential to fulfilling far reaching and correct annotation.

Authors et al [7], introduces review Mobile Computing is "taking a computer and all necessary files and software out into the field". There are several different dimensions under which mobile computers can be defined: in terms of physical dimensions; in terms of how devices may be hosted; in terms of when the mobility occurs; in terms of how devices are networked; in terms of the type of computing that is performed.

Maker SureshKumar.T, Sivaranjani.S and Dr.Shanthi.N [6], review extraction mechanical assemblies and consider their execution estimations for both touching and non-flanking pages consolidated in paper "A Survey of Tools for Extracting and Aligning the Data in Web" in walk 2014.

Table 1 Data Extraction Method Summarization

Tools	Nested Structure processing	Single Result Page	Non-contiguous Data Result
CTVS	Yes	Yes	Yes
DeLa	Yes	Yes	No
Viper	No	Yes	No

CTVS achieves higher precision than the present systems. Notwithstanding the way that DeLa and ViNT perform well on data extraction, they disregard to cover the settled structure planning when in doubt while CTVS has the limit cover the settled sorted out pages in web. Snake

has extraordinary execution comes to fruition anyway it doesn't handle non-coterminous pages.

Bincy S Kalloor, Shiji C.G [8], proposed customized multi-annotator approach in "A Survey on Data Annotation for Web Databases" in September 2014. In this paper display a customized annotation approach, first alters the data units on a result page into unmistakable social affairs, such that the data in the same get-together have the same which implies. By then for each social affair remark it from particular component and total the assorted annotations to expect a last annotation mark. Priyanka C. Ghegade, Prof. VinodWadane [17], describes personalized web search(PWS) is one of the active ongoing research field that related to the retrieval of the relevant web page results based on the user interest and preferences. ARUNPANDI.V, SUNDARAMOORHTY.S, VARADHARAJAN.E [18], proposed the personalization involve in collecting the user interests implicitly or explicitly called as user profile. Collecting the user interests in explicit manner is not so easy. Because the users are not interested to provide their interest to the server. So we go for other methods to create user profile in implicit manner.

Weifeng Su, Jiying Wang, Frederick H. Lochovsky [10], were accessible a novel data extraction and course of action technique called CTVS in "Uniting Tag and Value Similarity for Data Extraction and Alignment" in July 2012, that joins both mark and regard comparability. CTVS normally expels data from inquiry result pages by first recognizing and dividing the request result records (QRRs) in the inquiry result pages and a short time later conforming the isolated QRRs into a table. They prescribed that technique to handle the circumstance when the QRRs are not flanking and for dealing with any settled structure that may exist in the QRRs. In like manner arrangement a record course of action estimation that changes the qualities in a record, first pairwise and a short time later exhaustively, by uniting the tag and data regards similarity information. Neha Dewangan, Rugraj [14], were proposed a client side privacy protection framework called UPS i.e. User customizable Privacy preserving Search is presented in the paper. Any PWS can adapt UPS for creating user profile in hierarchical taxonomy.

Greeshma A S., Lekshmy P. L. [15], were provides a review on personalized web search and the related security concepts. The PWS techniques are developed remarkably in the last decades. A variety of techniques have emerged to increase search effectiveness and to protect privacy using multiple algorithms. Different methods conclude that privacy preservation is not handled well. R.Aravindhan, Mano Chitra.M [16], describes the comparison and analysis

between various methods involved in developing ontology based search engines. It also illustrates that there are many techniques or patterns that can be followed for getting an efficient result. The kind of comparison that has been done actually reflects that the efficiency of search engine differs from each method. From the above comparison, it makes clear that the usage of ontology based search engine will provide accurate results depending on the literal meaning of the query and the semantic search engines will produce results based on the query logic.

HimaniArya, JaytrilokChoudhary, Deepak Singh Tomar[19], proves personalization of web search is a necessity now-a-days to reveal user preferences in search results. In this paper, a survey of personalization has been given. The maximum number of web personalization methods are based on web page content analysis and textual similarity with the user preferences. Some approaches uses ontological user profile for personalization. In all the approaches, it is unclear whether personalization is working efficiently and effectively for all users in different search context. K. Mala, M. A. Hema [20], describes the users are given with predefined queries and topical interests, users have to synthesize the needed information from the given queries and topical interests and conduct the searches correspondingly. Ideally, a large scale user study should be conducted in which PWSE is subjected to real life use, users' behaviors are monitored transparently and satisfaction of the users is analyzed and compared with other systems, but a large scale is implemented.

### A. Personalized Web Search

The search web field involves in the information retrieval from the structure of hyperlinked web pages like Google. This kind of engines having the following problems. (1) Allocated weight for Web page, and (2) Hyperlinked Web pages may have related contents that are not considered. In [5] and [6] the use of personalized PageRank to enable personalized Web searches was suggested as a modification of the global PageRank algorithm, which computes a universal notion of importance of a Web page. Experiments in this work concluded that the use of personalized PageRank scores can improve a Web search. However, browsing patterns, bookmarks, and so on were not considered for the experiment.

### B. Personalized web site

Personalized Web site is constructed using the contents that present in the web pages, the structure of the contents, the link topologies that are used in web pages. Then [9], a search process involving the link personalization and content personalization are the types in which the web site personalization takes place. The link personalization deals

with the site URLs and the links given in those web pages. The content personalization involves in the content analyzing.

### C. Recommender system

It has become increasingly difficult to search for useful information on the Web because the amount of information on the Web continues to grow. From [11] and [13] we get the feeling of being overwhelmed by the number of choices. This is known as "information overload." An approaches to reduce this overload, recommender systems have emerged in domains such as E-commerce, digital libraries, and knowledge management. These systems provide personalized suggestions based on user preferences. This system collects feedback from the users in the form of finding the similarities between the contents, similarity between user profiles and rating the items in a particular domain.

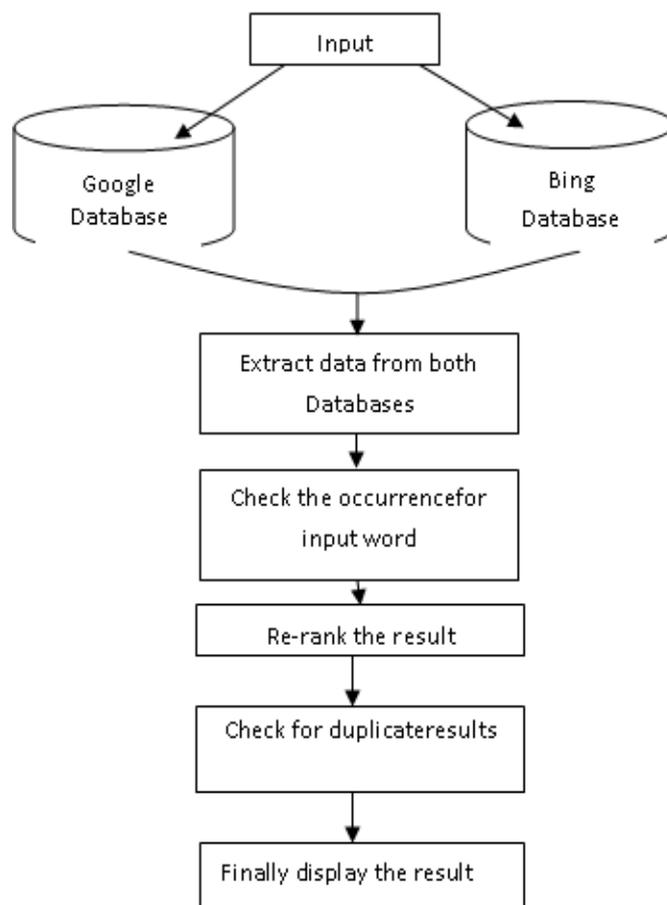


Figure: General architecture of system

### III. Discussion

Search engine plays a major important way to search the applicable information from the web. Though, the investigate results acquire may not forever be cooperative to

the user, as investigate engine fail to be familiar with the user purpose behind the query. Anuradha R. Kale, Prof V.T.Gaikwaid, Prof H.N.Datir [1], describes the data extraction and alignment for multiple web databases. UllasNambiar and SubbaraoKambhampati [2], dispersed their paper "Giving Ranked Relevant Results to Web Database Queries" in which they proposed to give situated responses to customer request by perceiving a game plan of inquiries from the request log whose answers are critical to the given customer request. V.kalyan Deepak and N.V.Rajeesh Kumar [3], systematically examined the issue of customized annotation approach in the paper "Recuperate Records from Web Database Using Data Alignment". R. R. Shelke, Dr. R. V. Dharaskar, Dr. V. M. Thakare [4], studied A mobile computer is effectively any computing device not constrained in its location to a desktop or data Centre. Prasad B. Dhore, Rajesh B. Singh [5] and Suresh Kumar. T, Sivaranjani. S and Dr. Shanthi. N, [6] proposed the use of personalized PageRank to enable personalized Web searches was suggested as a modification of the global PageRank algorithm, which computes a universal notion of importance of a Web page.

R. R. Shelke, Dr. R. V. Dharaskar, Dr. V. M. Thakare [7] and Bincy S Kalloor, Shiji C.G [8], introduces review Mobile Computing is "taking a computer and all necessary files and software out into the field". Weifeng Su, Jiyang Wang, Frederick H. Lochovsky [10] and Y. Zhai and B. Liu [11], were accessible a novel data extraction and course of action technique called CTVS in "Uniting Tag and Value Similarity for Data Extraction and Alignment". Deepika. J. [13] and Neha Dewangan, Rugraj [14], describes supporting privacy protection in personalized web search- a review and Non-Duplicate Data Extraction in Web Databases by Combining Tag and Value Similarity. Greeshma A S., Lekshmy P. L. [15] and R. Aravindhan, Mano Chitra. M [16], were provides a review on personalized web search and the related security concepts. The PWS techniques are developed remarkably in the last decades. A variety of techniques have emerged to increase search effectiveness and to protect privacy using multiple algorithms. Priyanka C. Ghegade, Prof. Vinod Wadane [17] and ARUNPANDI. V, SUNDARAMOORHTY. S, VARADHARAJAN. E [18], describes personalized web search (PWS) is one of the active ongoing research field that related to the retrieval of the relevant web page results based on the user interest and preferences. Himani Arya, Jaytrilok Choudhary, Deepak Singh Tomar [19] and K. Mala, M. A. Hema [20], proves personalization of web search is a necessity now-a-days to reveal user preferences in search results. In this paper, a survey of personalization has been given. The maximum number of web

personalization methods are based on web page content analysis and textual similarity with the user preferences.

#### IV. Personalized web search

The personalized web search is takes place in three ways: (1) client side personalization, (2) Server side personalization and (3) client-server cooperative personalization.

##### A. Server side personalization

Level II privacy protection can be achieved. But when the search engine uses the user login ID to collect user information, [3] this method will not achieve Level II privacy protection; when the search engine only uses the IP address to aggregate the user information, this method works. Sometimes, search engines group users randomly or according to some criteria before they release the search engine logs. Then we will also have Level II privacy protection to those third parties which receive the search engine logs. It is impossible to implement Level III or Level IV privacy protection if personalization is done on the server side.

##### B. Client side personalization

A client-side personalized search agent can do query expansion to generate a new query before sending the query to the search engine. The sensitive contextual information is generally not a major concern since it is strictly stored and used on the client side and the overhead in computation and storage for personalization can be distributed among the clients. [5] A main drawback of personalization on the client side is that the personalization algorithm cannot use some knowledge that is only available on the server side (e.g., PageRank score of a result document).

##### C. Client-server cooperative personalization

The user profile is still stored on the client side, but the server also participates in personalization. [2] When a query is given to the search engine then the client extracts contextual information from the user profile. The combination of extracted information from the profile and query is sends it to the search engine. The search engine then does personalization with the received context. The contextual information sent to the server specifies the user's search preferences (e.g., query expansion terms, topic weight vector). This architecture provides the same level of privacy protection as server-side personalization. However, the personally identifiable information collectable on the server side is less than in the case of pure server side personalization.

## V. Conclusion

Personalized web search(PWS) is one of the active ongoing research fields that related to the retrieval of the relevant web page results based on the user interest and preferences. This paper aims to discovering the number of diverse user search goals for a query and depicting each goal with some keywords automatically. The kind of comparison that has been done actually reflects that the efficiency of search engine differs from each method. From the above comparison, it makes clear that the usage of ontology based search engine will provide accurate results depending on the literal meaning of the query and the semantic search engines will produce results based on the query logic.

## VI. References

- [1] Anuradha R. Kale, Prof V.T.Gaikwaid, Prof H.N.Datir "Data Extraction and alignment for multiple web Databases" *International Journal of Scientific & Engineering Research*, Volume 4, Issue 7, July-2013 2422 ISSN 2229-5518.
- [2] UllasNambiar, SubbaraoKambhampati, "Providing Ranked Relevant Results for Web Database Queries".
- [3] V.kalyan Deepak, N.V.Rajeesh Kumar, "Retrieve Records from Web Database Using Data Alignment" (*IJCSIT International Journal of Computer Science and Information Technologies*, Vol. 5 (2), 2014, 1552-1554
- [4] R. R. Shelke ,Dr. V. M. Thakare, Dr. R . V. Dharaskar, "Study of Data Mining Approach for Mobile Computing Environment",*International Journal on Computer Science and Engineering (IJCSE)* , ISSN : 0975-3397, Vol. 4, 12 Dec 2012 , pp.1920-1923
- [5] Prasad B. Dhore, Rajesh B. singh, "Annotating Search Record from Web Databases", *International Journal of Software and Hardware Research in Engg*, ISSN No:2347-4890, Volume 2 Issue 12, December 2014
- [6] SureshKumar.T, Sivaranjani.S and Dr.Shanthi.N, "A Survey of Tools for Extracting and Aligning the Data in Web", *International Journal of Computer Science & Engineering Technology (IJCSET)*, ISSN : 2229-3345 Vol. 5 No. 03, Mar 2014
- [7] R. R. Shelke, Dr. R. V. Dharaskar, Dr. V. M. Thakare, "DATA MINING FOR MOBILE DEVICES USING WEB SERVICES", *International Conference on Industrial automation And Computing ( ICIAC - 12<sup>th</sup> & 13<sup>th</sup> April 2014)* , Jhulelal Institute of Technology, Nagpur.
- [8] Bincy S Kalloor, Shiji C.G, "A Survey on Data Annotation for Web Databases", *International Journal of Engineering and Innovative Technology (IJEIT)* ISSN: 2277-3754, Volume 4, Issue 3, September 2014
- [9] Weifeng Su, Jiying Wang, Frederick H. Lochovsky, "Combining Tag and Value Similarity for Data Extraction and Alignment" *IEEE Transactions On Knowledge And Data Engineering*, Vol. 24, No. 7, July 2012
- [10] Weifeng Su, Jiying Wang, Frederick H. Lochovsky, "Rec-ord Matching over Query Results from Multiple Web Databases" *IEEE Transactions On Knowledge And Data Engineering*, Vol. 22, No. 4, April 2010
- [11] Y. Zhai and B. Liu, "Structured Data Extraction from the WebBased on Partial Tree Alignment," *IEEE Trans. Knowledge and Data Eng.*,vol.18, no.12, pp.1614-1628, 2006.
- [12] Ruofan Wang, Shan Jiang and Yan Zhang: *Re-ranking Search Results Using Semantic Similarity*, 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)
- [13] Deepika.J, "Non-Duplicate Data Extraction in Web Databases by Combining Tag and Value Similarity", *International Journal of Advanced Information Science and Technology (IJAIST)* ISSN: 2319:2682 Vol.9, No.9, January 2013.
- [14] NehaDewangan, Rugraj, "SUPPORTING PRIVACY PROTECTION IN PERSONALIZED WEB SEARCH- A REVIEW", *International Journal of Computer Engineering and Applications*, Volume VIII, Issue II, November 14.
- [15] Greeshma A S., Lekshmy P. L., "Privacy Protection in Personalized Web Search- A Survey", *International Journal of Innovative Research in Advanced Engineering (IJIRAE)* ISSN: 23 Issue 6, Volume 2 (June 2015) [www.ijirae.com](http://www.ijirae.com)
- [16] R.Aravindhan, Mano Chitra.M, "A Review on Ontology Based Search Engine", *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 3, Issue 10, October 2014.
- [17] Priyanka C. Ghegade, Prof. VinodWadane, "A Survey of Personalized Web Search in Current Techniques", (*IJCSIT International Journal of Computer Science and Information Technologies*, Vol. 5 (6) , 2014.
- [18] ARUNPANDI.V, SUNDARAMOORHTY.S, VARADHARAJAN.E, "SURVEY ON PERSONALIZED WEB SEARCH WITH PRIVACY PROTECTION", *JOURNAL OF INTERNATIONAL ACADEMIC RESEARCH FOR MULTIDISCIPLINARY* Impact Factor 1.625, ISSN: 2320-5083, Volume 2, Issue 11, December 2014
- [19] HimaniArya, JaytrilokChoudhary, Deepak Singh Tomar, "A Survey on Techniques for Personalization of Web Search", *International Journal of Computer Applications* (0975 – 8887) Volume 94 – No. 18, May 2014
- [20] K. Mala, M. A. Hema, "An Efficient Query Processing for Personalized Web Search Engine", Volume 2, Issue 5, May 2014 *International Journal of Advance Research in Computer Science and Management Studies*.