Exploiting Location Sensitive System for Query Recommendation

Rohini Gadge[1], Manisha Patil[2]

P.G. Student, Department of Computer Engineering, SKN COE, Savitribai Phule University of Pune, Pune, India [1]

Asst. Professor, Department of Computer Engineering, SKN COE, Savitribai Phule University of Pune, Pune, India [2]

ABSTRACT: Keyword suggestion in web search is a very important featureto be considered in today's growing world. It helps user toaccess the information without any prior knowledge of how toexpress in queries. The main concept of query suggestion isused to retrieve documents from the related server byconsuming less time. Platform is provided by search enginesfor users to describe their information need more precisely by using query recommendation. Previously there has been lot ofwork done for retrieving relevant data of users to meet theirinformation need and improving performance of searchengines. This paper reviews and compares different available methods in query log processing for information retrieval.Then conclude that Existing keyword suggestion techniques are not considering the locations of the users and the queryresults which serves as a drawback of existing systems. Thespatial factor is not considered to retrieve result. The approachbased on location aware keyword query suggestion is better tounderstand user's interaction process with search engines tofind the appropriate information need.For this we have fragmented database according to the location and then used heap sort and SHA-256 algorithm to provide query suggestion.it help user to find appropriate query considering location of user and also system overhead will be less.

KEYWORDS: Location Based Search,keyword Search,Query Suggestion,Fragmentation,Spital Databases,context aware

I. INTRODUCTION

Query suggestion plays very vital role in enhancing performance of web browser. As web browser pays service to billion of user per day.and it will work on million of document for every time to provide service to user. It takes high processing and computational power and require time for processing. It is very essential for web browser to personalize his search for providing better query suggestion.In an recent survey it is proved that 4 among 5 people user search for the local shops,hotel,hospitals,bus stop etc.sothere is emerging need for providing the location based service to user. In location based search ,the result must be related to the topic of query but also they must be geographically related to the location which is associated ith the query.if user provide location manually then it may happen that location are ambiguous.this ambiguity refer to two or more location names.When ambiguious geographic name is submitted,it is a challenging and hard task for location based search to guess the user's\ intention.so in our system we are taking location though ip adreess domain so that any kind of ambiguity will not be created. Through location based searching users time can be saved.And overall performance of the system is increased,as number of document will be limited.In our proposed system location based query suggestion is provided.

1.1 Location based Search:

   In an recent survey it is proved that 4 among 5 people user search for the local shops,hotel,hospitals,bus stop etc.so there is emerging need for providing the location based service to user. In location based search ,the result must be related to the topic of query but also they must be geographically related to the location which is associated ith the query.if user provide location manually then it may happen that location are ambiguous.this ambiguity refer to two or more location names.When ambiguious geographic name is submitted,it is a challenging and hard task for location based search to guess the user's intention.so in our system we are taking location though ip adreess domain so that any kind of ambiguity will not be created. Through location based searching users time can be saved.And overall

performance of the system is increased,as number of document will be limited.In our proposed system location based query suggestion is provided.
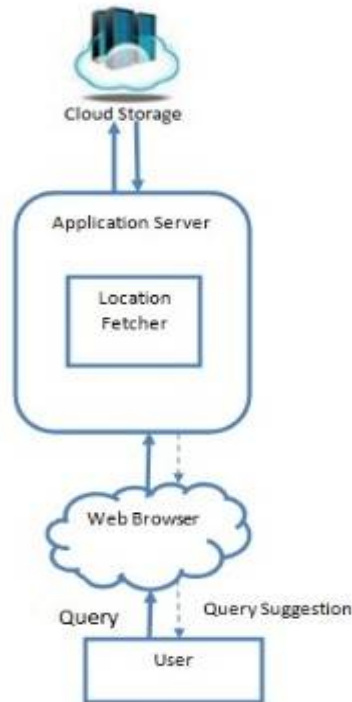

Fig 1: Proposed System Architecture

Fig 1 shows architecture of proposed system. It consist ofmain part location fetcher, cloud storage and query suggestion.User will open URL through web browser and at the sametime location fetcher will track the location of user throughIp address domain and get latitude and longitude of currentlocation. Location fetcher will send this location to the application server. The main drawback of current search engineis that they donot fragment data according to the locationhence user have to search from bigger database. If database are properly fragmented according to the location then overallprocessing time of searching will be less.In cloud storage wehave fragmented data in cloud to multiple small location basedserver.because of this user now have to search from limiteddocument.we will apply hep sort and SHA-256 algorithm to provide querysuggestion for users query.When user will fire his first keywordfrom that keyword onwards application server will provide thequery suggestion by processing database. User will select thedesired query among the provided query suggestion.

II. RELATED WORK

Authors in [1] propose a novel context-aware query suggestion approach which is in two steps. In the offlinemodel-learning step, to address data sparseness, click-through bipartite is clustered in order to summarize queries into concepts. In this approach queries are suggested to the user in a context-aware manner.

Authors in [2] propose a novel query suggestion algorithm based on ranking queries with the hitting time on a large scalebipartite graph. This method captures the semantic consistency between the suggested query and the query givenby user. Experiments show time is effective to generate semantically consistent query suggestions. The proposedalgorithm and its variations can successfully execute hugequeries, accommodating query suggestion.

Author [3] introduced novel, domain-independent and privacypreserving methods for enhancing MF models by expandingthe user-item matrix and by imputation of the user-itemmatrix, using browsing logs and search query logs. Theyintroduced two approaches to enhancing user modeling usingthese data. Authors show that CF systems can be enhancedusing Internet browsing data and search engine query logs,both represent a rich profile of individuals' interests.

Author [4] proposed a new query suggestion paradigm, QuerySuggestion with Diversification and Personalization thateffectively integrate diversification and personalization intoone unified framework. In the QS-DP, the suggested queriesare successfully diversified to cover different facets of theinput query and the ranking of the suggested queries arepersonalized to ensure that the top ones that align with auser's personal preferences. They propose a newrepresentation for query log. The proposed multi bipartitegraph representation comprehensively captures different kindsof relations between search queries in query log. Based on themulti-bipartite-graph representation, they design twostrategies to identify the most relevant suggestion candidate.

Author [5] proposed a method that computes likeness amongqueries based on "Query- Clicked Sequence" model. Thismodel counts weight of clicked document term by density ofdocuments containing this term on clicked sequence, andcomputation. Based on the characteristics of differentconcentration on relevant and irrelevant documents occurringon clicked document sequence, this paper proposed a querysimilarity computing method based on irrelevant feedbackanalysis, and recommended queries based on this method

Author [6] developed the Query system for personalizedquery recommendations. Query monitors the user'squerying behavior and finds matching patterns in the system'squery log, identifying same kind of users. These queries areused to recommend queries which user may find helpful. Theyexplore the use of latent factor models when, instead ofratings, the input consists of database-query log data. Andexplored how latent factor models, and in particular matrixfactorization using ALS, affect the quality of therecommendations and computational efficiency of theirframework. Such techniques have become very popular intraditional rating-based recommender systems, and in thiswork authors verified that they capture latent similaritiesbetween users and "items" even when the input is not explicit.

Author [7] proposed time aware structured query suggestionwhich clustered query suggestion along timeline so the usercan narrow down his search from a temporal point of view.When the suggested query is clicked the method presents webpages from query-URL bipartite graph. After ranking thoseaccording to click count within a particular time period thismethod helping user to access relevant web pages. It free theusers from burden of entering a specific time constraint withquery, this method can be used in the context of real usersearch tasks.

Author [8] Explained a web recommender approach based onlearning from web logs it recommends user a list of pages thatare relevant to the users proposed query by comparing withhistoric pattern and also rerank the result pages. This systemproves to be efficient as the pages desired by the users are onthe top in the result list and this method reduces the searchtime of the user. In this the recommendation is based on thefeedback of users and web log analysis.

Author [9] proposed a snippet based method to facilitateusers with query recommendations. The concepts related tothe users information needs are suggested to the users tosatisfy their current information need, extracted the conceptsfrom the web snippet. Authors proposed two weight functionsto measure the relevance between query and concept. Relatedconcepts with different meaning are selected andrecommended as query suggestions to the users.

Author [10] presented an approach based on the users searchbehaviour. Their suggested query recommendation frameworkfollows the fact that if user clicks certain result returned bysearch engine then if does not necessarily mean that the useris interested in that result but if probably reflects that the useris instead interested in the snippets of the result. This isbecause that up to that time the user clicks certain result justby viewing the snippet, the resultant document has not shownto user by that time.

III. PROPOSED ALGORITHM

Heap Sort Algorithm:

(**1**) Build a max heap from the input data.

(**2**) At this point, the largest item is stored at the root of the heap. Replace it with the last item of the heap followed by reducing the size of heap by 1. Finally, heap if y the root of tree.

(**3**) Repeat above steps while size of heap is greater than 1.

SHA 256 Algorithm:-

In SHA-256 massage to be hashed first.

(1) Padded with its length in such a way that the result is a multiple of 512 bits long, and then

(2) Parsed into 512-bit message blocks $M^{(1)}, M^{(2)}, \ldots, M^{(N)}$.

The message blocks are processed one at a time: Beginning with a fixed initial hash value $H^{(0)}$, sequentially compute

$$H^{(i)} = H^{(i-1)} + C_M^{(i)}(H^{(i-1)}),$$

Where C is the Sha-256 compression function and + means word-wise mod 2^{32} addition $H^{(N)}$ is the hash of M.

IV. PSEUDO CODE

Step 1: Input the dataset to the application.
Step 2: fetch the location through latitude and longitude.
Step 3: Fragment the database according to location of user.
Step 4: Compute the hash value of each hash function.
Step 5: Generate counter for hash value in hash table
Step 6: Apply heap sort algorithm.
Step 7: Perform query suggestion on basis of user's location
Step 8: Provide the query suggestion.
Step 9: go to step 2.
Step 10: End.

V. PERFORMANCE RESULT AND ANALYSIS

Experimental evaluation result to compare the proposed system with the existing system for evaluating the performance. System is built using Csharp on Windows platform. The evaluation results are carried out by considering the no of input query. The parameters considered for evaluation are response time with respect to no of input queries By observing the above results can be conclude that the proposed system gives better results as compared to the existing system in terms of various parameters calculated.

Fig 1. Shows that the overall processing time of proposed system with repect to no of record and it provides us no of keystroke that are required for this process. Fig 2. Shows the performance of the system in existing system performance was so low because of database were not fragmented according to location and there is improper management of database in proposed work we eliminate the process of document search. In fig. 3 shows scalability of the system. Existing system was not that much scalable on large dataset proposed system overcome this by fragmenting database according to location of user everytime.
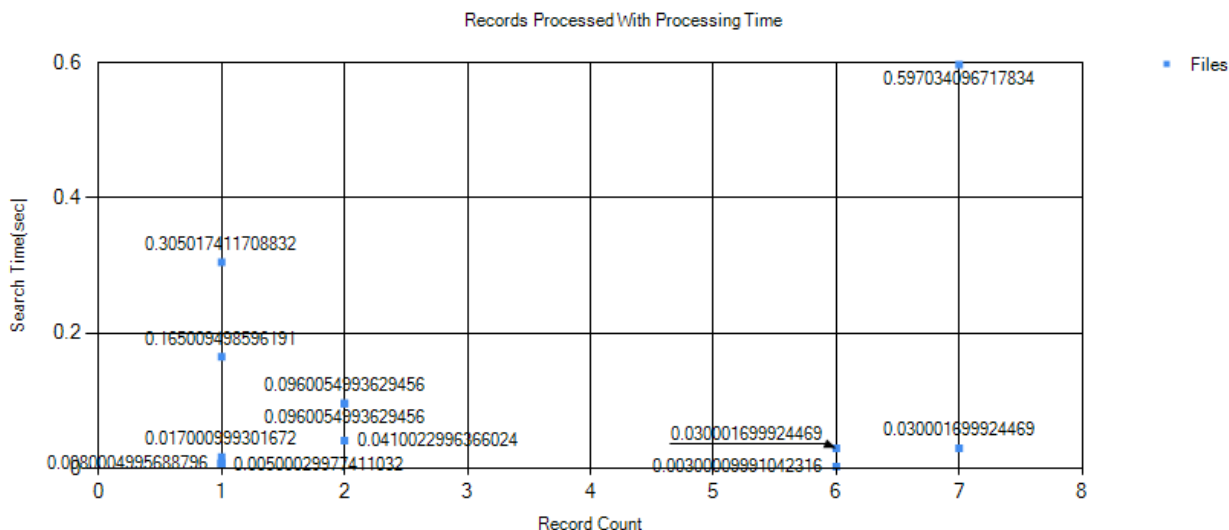
Fig.1. No of record processed with processing time of proposed system
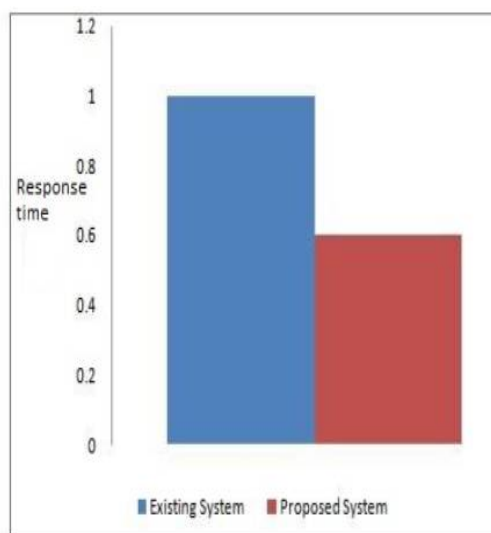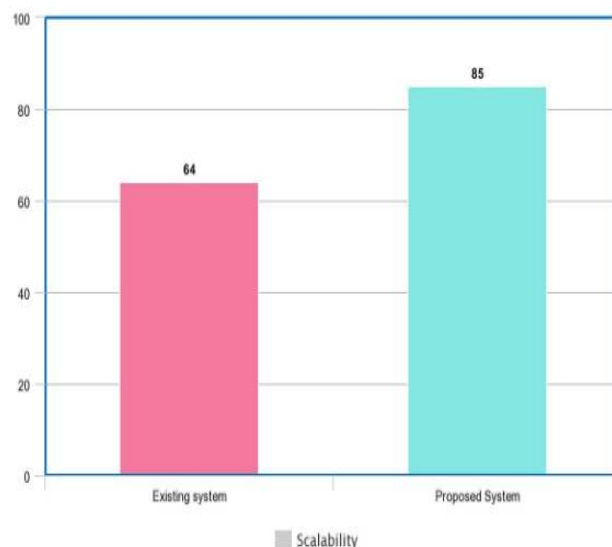

Fig 2.Comapritive result for response time

Fig 3.Scalability Evaluation

VI. CONCLUSION

User satisfaction is very important for information retrieval by providing alternate query through query suggestion. Query suggestion is best method for helping users to satisfy the users information need by suggesting alternating queries related to current query.In our system we are providing keyword query suggestion from first character,so no of hits required are less.This will also reduce system overhead.

REFERENCES

1. H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, *Context-aware query suggestion by mining click-through and session data,* in Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2008, pp. 875–883.
2. Mei, Qiaozhu, Dengyong Zhou, and Kenneth Church. *Query suggestion using hitting time.* Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008
3. Royi Ronen and et.al, *Recommendations Meet Web Browsing: Enhancing Collaborative Filtering using Internet Browsing Logs*, IEEE 32nd International Conference on Data Engineering (ICDE), pp 1230 – 1238, 2016
4. Di Jiang, Kenneth Wai-Ting Leung, Lingxiao Yang, Wilfred Ng, *Query suggestion with diversification and personalization*, Knowledge-Based Systems. Volume 89, Pages 553–568, 2015.
5. Bo Zhang, Bin Zhang, Shubo Zhang, Chao Ma, "Query Recommendation Based on Irrelevant Feedback Analysis",8th International Conference on Bio Medical Engineering and Informatics, pp 644-648, 2015.
6. Magdalini Eirinaki,*QueRIE reloaded: Using matrix factorization to improve database query recommendations*, IEEE international conference on Big Data, pp 1500 – 1508, 2015.
7. T. Miyanishi and T. Sakai, *Time-aware structured query suggestion,* in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval,pp. 809–812, 2013.
8. R. Bhushan and R. Nath. 2012. *Recommendation of optimized web pages to users using Web Log mining techniques*. Advance Computing Conference (IACC), 2013 IEEE 3rd International. IEEE, 2012.
9. Goyal, Poonam, and N. Mehala. 2011. *Concept based query recommendation*. Proceedings of the Ninth AustralasianDataMining Conference-Volume 121. Australian Computer Society. Inc.
10. Y. Liu, Junwei Miao, Min Zhang, Shaoping Ma, and Liyun Ru. 2011. *How do users describe their information need: Query recommendation based on snippet click model*. Expert Systems with Applications 38, no. 11 (2011):13847-13856.\