IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Implementation of Dynamic Keyword Query Suggestions on Geo Location using Document Proximity

P Sujini [1*,] D.N. Vasundhara[2]

[1*,2]Department of CSE, VNR Vignana Jyothi Institute of Engineering and Technology, JNTUH, Hyderabad, India.

[*]*Corresponding Author:* sujinimtech2016@gmail.com

Available online at: www.ijcseonline.org

Accepted: 19/Jul/2018, Published: 31/Jul/2018

Abstract- A search engine is an online tool for searching any information on the World Wide Web like documents, business services, images, videos and so on. Many tools were there for helping the search engine helping the search engine like page ranking algorithms, voice search, image mining and keywords suggestions. From all this tools keyword suggestions will help online users to retrieve needed information and Express the queries without any background knowledge. Many tools were proposed to enhance the keyword suggestions like click-through, random walk method etc. But those are not satisfied user requirements in the modern world. In this paper proposing a novel structure of Keyword recommendation using location proximity which gives useful suggestion on geo based location. From the baseline algorithms, we are enhancing the partition based algorithm to achieve keyword suggestions. In addition, we enhance instant search keywords by using location proximity. Our proposed algorithm achieves better performance compared with existing algorithms in time ratio.

Keywords: Baseline algorithm, partition based algorithm, location proximity, keyword suggestions.

I. INTRODUCTION

In our modern world, we are very much habituated with search engine by using mobiles, laptops, tabs and etc. generally to retrieve anything from search engine we need to type a query. Query is a set of words submitted by the web user by typing some characters in their machines. By this query search engine will return the results from the web data. So for using search engines and to get proper results from the search engine query formulation plays crucial role in net surfing. But keyword formulation is very difficult if user doesn't have any prior knowledge of the data and most of the cases user doesn't know about the query formulation and doesn't spend much time to form a query words. According to these situations user needs web search engine keyword suggestions.

Keyword suggestions are two types in any search engine. First one, while typing the keywords user will get suggestions according to character by character of user query. This type is popularly named as instant keyword suggestions. Second one after submitting the query user will get results, after the results user will get keyword suggestions. This type popularly named fuzzy keyword suggestions. In this paper we are discussing and enhancing of the fuzzy keyword suggestions.

Most of popular web search engines like Google, Bing and Yahoo providing appropriate keyword suggestions to online users. But main problem of these search engines not considering the local attributes or local services. In this paper we are proposing a concept called keyword suggestions using location proximity. In this concept we are considering Keyword Document (KD) score [2] and location of keywords. For this we are proposing partition based on local keywords add user location. We are constructing keyword suggestions at two levels, instant and fuzzy keyword suggestions. Existing baseline algorithms models of bookmark coloring algorithm BCA [3][4][5] which are used calculate the KD means keyword document score. in proposed suggestions VR enhancing the baseline algorithms [1] with partition based algorithms for both in instant and fuzzy keyword suggestions.

II. RELATED WORK

Doug Beeferman and Adam Berger[1] describe that http links and protocols in commercial search engines can able to crawl the data of user's information like name, IP address and Mac Address of user's machine which are can get and browser information like name of the browse, provider, date and time and so on. Huanhuan Cao, Daxin Jiang and Jian Pei [3] describes that to depict the comparative queries into a concept, we need to extent the comparison between queries. When a user searches for a query to a web search engine, a set of URLs will be returned as the answers.

Baeza-Yates, C. Hurtado, and M. Mendoza[2] proposed a method that, suggests a list of similar queries upon query submitted to a web portal. The related queries are based on

the previously provided queries, and that may be issued by the user to the search engine to redirect or tune the search process. Qiaozhu Mei, et.al [16] present that our methodology utilizing hitting time on a bipartite diagram can be effortlessly changed in accordance with prompt redid question proposals. Normally, when we ken the character of the utilizer (e.g., his IP address), we ought to revive our discernment about the information goal of this request. One may verbalize that a direct system is to build up the bipartite outline only predicated on the verifiable setting of that utilizer. Regardless, that could witho1ut much of a stretch fall into the pickle of data lack. The fundamental treatment additionally loses the possibility of utilizing customary insight. If an utilizer starting at now kens what question to utilize (e.g., picking up from his history), it isn't pellucid how much request proposition could benefit.

III. METHODOLOGY

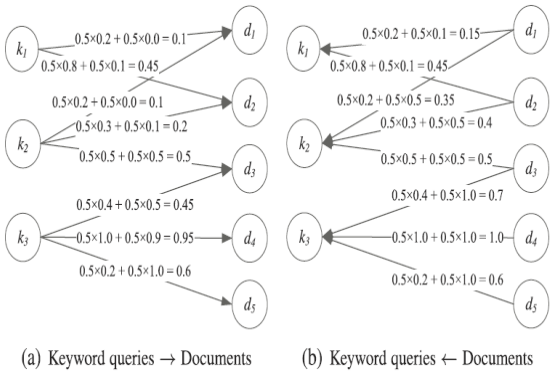
Keyword suggestions are a concept of analyze of web data and form keywords which are easily retrieve web data for the user. Web data is a combination of multiple data sources like Meta keywords, descriptions, titles, images, locations etc. suggestions are two types in any search engine. First one, while typing the keywords user will get suggestions according to character by character of user query. This type is popularly named as instant keyword suggestions. Second one after submitting the query user will get results, after the results user will get keyword suggestions. This type popularly named fuzzy keyword suggestions. To achieve two types of Keyword suggestions we proposing two algorithms of partition based using location proximity.

Partition Based Algorithm

This algorithm we are proposing for getting the keyword suggestions to the users while searching by the consideration of user location and document locations. Here documents nothing but business services locations. Partition based algorithm is a enhanced model of bookmark coloring algorithm. In bookmark coloring algorithm we can distribute some colors to neighbor nodes. Same as BCA in partition based algorithm with distribute value according to the keyword and document weights.

Weight adjustment

In partition based algorithm whichever documents participating in ink distribution we call it as active ink nodes (AINK). weight distribution of AINK is two types. One is keywords to document and another one is document to keywords.



(a) Keyword queries → Documents (b) Keyword queries ← Documents

Fig 1. Weight adjustment

Keywords to Document

According to figure 1 we are distributed Ink from keywords to documents based on following equation
w(e) weight of edge. According to KD Graph
β=0.5 normalized to take values in [0,1]

$$\acute{w}(e)=\beta*w(e)+(1-\beta)*(1-\text{dist}(\lambda_u, di)) \quad (1)$$

Document to Keywords

According to figure 1 we are distributing income from documents to keywords based on following equation
w(e) weight of edge. According to KD Graph
β=0.5 normalized to take values in [0,1]

$$\acute{w}(e`) = \beta*w(e^{l})+(1-\beta)*(1-\text{mindist}(\lambda_u, Di)) \quad (2)$$

Here in algorithm 1 we are performing instant keyword suggestions, we are collecting the keywords of documents of user's City and we are searching the keywords character by character which are matched with the keywords application will return the matched keywords of documents.

Algorithm1:

Input: User location (x, y), Documents D

Output: Keyword suggestions K.

Initialization:

(i) User current city, defined λ_u

(ii) Let Documents D= {d_1, d_2, …., dn} ∈ λ_u

```
for each Dλu
        for each character of query
                find relevant d.
                K=K+d;
        end for
end for
return K;
```


In algorithm 2, we are performing fuzzy key word suggestions; here we are distributing the Ink in between AINK according to weight adjustment. According to the equation 1 and 2 we are distributing the distribution ratio between AINK.First we collect the user location and documents which are near to users (near to 1 kilometer). We are using Euclidian Distance to calculate the distance between user and documents. We are assuming the distribution ratio as 1.

Algorithm2:

Input: User location (x, y), Documents D

Output: Keyword suggestions K.

Initialization:

User current city, defined $\lambda_u = x_u y_u$

Let Documents D={$d_1, d_2, \ldots . dn$}

Priority Queue Q ←0, Active Ink AINK← 0, Distribution Ratio=1

for each D

calculate distance dist(λ_u, D(i))

if dist<=1; According to Euclidian Distance

calculate Score($\dot{w}$);

keywords to Documents

w(e) weight of edge. According to KD Graph

β=0.5 normalized to take values in [0,1]

$\dot{w}(e) = \beta * w(e) + (1-\beta) * (1 - \text{dist}(\lambda_u, di))$;

Document to Keywords

w(e) weight of edge. According to KD Graph

β=0.5 normalized to take values in [0,1]

$\dot{w}(e`) = \beta * w(e^l) + (1-\beta) * (1 - \text{mindist}(\lambda_u, Di))$;

else break;

end for;

distratio=distratio-0.5;

while Q!=0

foreach nodev connected topin G do

Distribute distratio according to $\dot{w}(e)$;

AINK← AINK-top.aink;

Distribute distratio according to $\dot{w}(e`)$

end for;

end while;

return top G;

IV. EXPERIMENTAL RESULTS

In this section we can see the results of our Enhanced Partition based algorithm. We experiment the proposed architecture and application in i5 processor with 4 GB RAM in windows operating system. For locations and related maps, we use Google Maps services of JavaScript API.

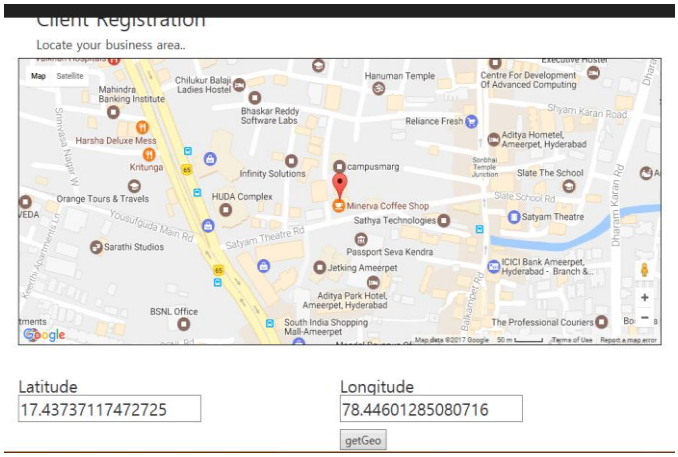


Fig2: Client Location collection with lat and lon values.

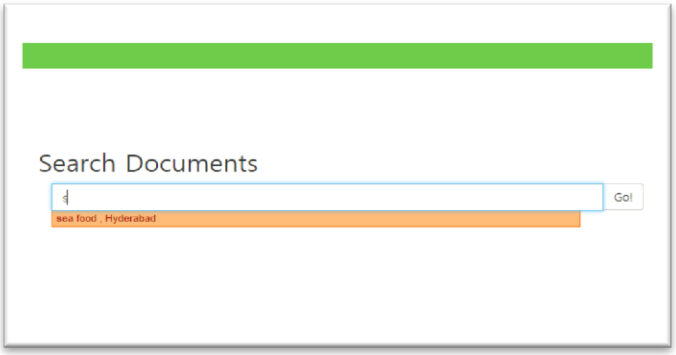


Fig 3: Instant search result.

Initial KD-Graph and Distances

Sno	Keyword	Document	Distances	KD Score
1	sea food	d7	0.281113	0.333333
2	veg restuarent	d7	0.281113	0
3	sea food	d6	0.260932	0.25
4	veg restuarent	d6	0.260932	1
5	sea food	d8	0.0715355	0

Fig 4. Collection of document with 1 km range of user location along with KD Score

Keyword Queries --> Documents

Sno	Distances	Keyword	Document	KD Score	Edge Weight w(e)
1	0.281007	sea food	d7	0.333333	0.526163
2	0.281007	veg restuarent	d7	0	0.359497
3	0.260861	sea food	d6	0.25	0.494569
4	0.260861	veg restuarent	d6	1	0.869569
5	0.0714132	sea food	d8	0	0.464293
6	0.0714132	veg snacks	d8	0.5	0.714293
7	0.0051403	lobster restaurant	d2	0.266667	0.630763
8	0.00513855	lobster restaurant	d4	1	0.997431

Fig 5: Equation 1 result. Keywords to Documents

Documents --> Keyword Queries

Sno	Distances	Keyword	Document	KD Score	Edge Weight w(e')
1	0.281007	sea food	d7	0.333333	0.664097
2	0.281007	veg restuarent	d7	0	0.497431
3	0.260861	sea food	d6	0.25	0.622431
4	0.260861	veg restuarent	d6	1	0.997431
5	0.0714132	sea food	d8	0	0.497431
6	0.0714132	veg snacks	d8	0.5	0.747431
7	0.0051403	lobster restaurant	d2	0.266667	0.630764
8	0.00513855	lobster restaurant	d4	1	0.997431

Fig 6: Equation 2 result. Documents to Keywords.

Sno	Keyword	Score	Score
1	sea food	0.06579233262136792	0.2392178795167378
2	veg restuarent	0	0.2392178795167378
3	sea food	0.06579233262136792	0.2392178795167378
4	veg restuarent	0	0.2392178795167378
5	sea food	0.06579233262136792	0.2392178795167378
6	veg snacks	0	0.2392178795167378
7	lobster restaurant	0	0.2392178795167378
8	lobster restaurant	0	0
9	sea food	0.06579233262136792	0
10	lobster restaurant	0	0
11	sea food	0.06579233262136792	0
12	lobster restaurant	0	0

Fig 7: Performing Partition based algorithm.

V. Conclusion

In this paper we have proposed a framework of Keyword suggestions using location proximity. From the baseline algorithms we are enhancing the partition based algorithm to achieve keyword suggestions. In addition, we enhance instant search keywords by using location proximity. Our proposed algorithm achieves better performance compare with existing algorithms in time ratio. Our empirical results show that our proposed framework gives useful keywords suggestions and that enhanced PA outperforms significantly.

REFERENCES

[1] Shuyao Qi "Location aware keyword query suggestion based on document proximity" IEEE 10.1109/ICDE.2016.7498428

[2] H. Tong, C. Faloutsos, and J.-Y. Pan, "Fast random walk with restart and its applications," in Proc. 6th Int. Conf. Data Mining,2006, pp. 613–622.

[3] P. Berkhin, "Bookmark-coloring algorithm for personalized page rank computing," Internet Math., vol. 3, pp. 41–62, 2006.

[4] M. Gupta, A. Pathak, and S. Chakrabarti, "Fast algorithms for top k personalized page rank queries," in Proc. 17th Int. Conf. World Wide Web, 2008, pp. 1225–1226.

[5] Song and L.-W. He, "Optimal rare query suggestion with implicit user feedback," in Proc. 19th Int. Conf. World Wide Web 2010, pp. 901–910.-9.

[6] Y. Fujiwara, M. Nakatsuji, H. Shiokawa, T. Mishima, and M.Onizuka, "Efficient ad-hoc search for personalized PageRank," inProc. ACM SIGMOD Int. Conf. Manage. Data, 2013, pp. 445–456.

[7] Y. Fujiwara, M. Nakatsuji, M. Onizuka, and M. Kitsuregawa, "Fastand exact top-k search for random walk with restart," Proc. VLDBEndowment, vol. 5, no. 5, pp. 442–453, Jan. 2012.

[8] S. Cucerzan and R. W. White, "Query suggestion based on userlanding pages," in Proc. 30th Annu. Int. ACM SIGIR Conf. Res.Develop. Inf. Retrieval, 2007, pp. 875–876.

[9] R. Li, B. Kao, B. Bi, R. Cheng, and E. Lo, "DQR: A probabilistic approach to diversified query recommendation," in Proc. 21st ACM Conf. Inf. Knowl. Manage., 2012, pp. 16–25.

[10] P. Berkhin, "Bookmark-coloring algorithm for personalized PageRank computing," Internet Math., vol. 3, pp. 41–62, 2006.

[11] T. Gaasterland, "Cooperative answering through controlled query relaxation," IEEE Expert, vol. 12, no. 5, pp. 48–59, Sep. 1997

[12] H. Tong, C. Faloutsos, and J.-Y.Pan, "Fast random walk with restart and its applications," in Proc. 6th Int. Conf. Data Mining,2006, pp. 613–622.

[13] N. Lao and W. W. Cohen, "Relational retrieval using a combination of path-constrained random walks," Mach. Learn., vol. 81,no. 1, pp. 53–67, 2010.

[14] R. Li, B. Kao, B. Bi, R. Cheng, and E. Lo, "DQR: A probabilistic approach to diversified query recommendation," in Proc. 21st ACM Conf. Inf. Knowl. Manage., 2012, pp. 16–25.

[15] Y. Song and L.-W.He, "Optimal rare query suggestion with implicit user feedback," in Proc. 19th Int. Conf. World Wide Web,2010, pp. 901–910.

[16] Q. Mei, D. Zhou, K. Church,"Query suggestion using hittingtime," In CIKM, 2008, pp. 469–478.

Authors Profile

Mrs. Sujini Puttamsetty has pursuing her M. Tech (S.E) in VNRVJIET, Hyderabad. She has done her B. Tech from Audisankara college of Engineering and Technology Gudur Andhara Pradesh in 2008, with an aggregate of 70.2%. She has done her Higher school education from Nalanda Junior College,Nellore Andhara Pradesh in 2004, with an aggregate of 80%. She completed her SSC from V.R High School,Nellore Andhara Pradesh. with an aggregate of 79%. She had experience of 1.5 years in teaching field and 2.4 years of experience in IT field. She is familiar with C, Java, SQL and Web Designing and Talend data Integration (ETL tool). Her area of interests is Data Mining.

Mrs. D N Vasundhara completed her Bachelor's degree in Engineering with Computer Science and Information Technology as her Specialization. She did her post-graduation in Software Engineering. She is interested in Image Processing and Data Mining. She is working as Assistant Professor in Computer Science and Engineering Department in VNR Vignana Jyothi Institute of Engineering and Technology. She guided 12 Under Graduation projects and 4 Post Graduation projects. She had an experience of 11 years in teaching field.

