

A QUERY WEIGHTED KEYWORD DOCUMENT GRAPH USING LOCATION BASED ON PROXIMITY

N. RAMYA

M. Tech

ramya.narayandas@gmail.com

MS A SOUJANYA

Assistant professor, CSE

annaramsoujanya@gmail.com

YERRAGUDIPADU

SUBBARAYUDU

Assistant professor, CSE

subbu.iare@gmail.com

ABSTRACT

Generally keyword generation will help to users find relevant keyword pairs for searching data to get more relevant data, simply it helps user to access the information without any prior knowledge of how to express in queries. Previously there has been a lot of work done for retrieving relevant data of users to meet their information need and improving performance of search engines. Current keyword generation process only depends on only keywords click analytics (Ex: Google), not considering the spatial location of user and documents. Taking both spatial and keyword requirements into account, the goal of a spatial keyword query is to efficiently find results that satisfy all the conditions of a search. To make our framework scalable, we propose a partition-based approach that outperforms the baseline algorithm by up to an order of magnitude. The appropriateness of our framework and the performance of the algorithms are evaluated using real data.

Keywords— Query suggestion, spatial databases, weighted keyword-document graph, document proximity

1. INTRODUCTION DATA MINING

Data mining is known to be a process of extracting usable data from a wide range of any raw data. Mining of data (once in a while called data digging or information gathering) is the way towards analyzing the data from modifying of different options and injecting out it into very useful information and very much useful fabricates salary, cuts costs, or both. Analyzing the data is all about writing computer programs or programming code is

one of various methodical mechanical assemblies for dismembering the actual data. This helps to incorporate customers to investigate data from the through variety of objectives or targets, well arrange it, and arrange the pairings perceived. Fact to be told, mining of data is the actual path towards finding associations or scenarios among various fields in many social databases.

2. LITERATURE SURVEY:

2.1 survey on location-aware instant search

Authors: - R. Zhong, J. Fan, G. Li, K.-L. Tan, and L. Zhou-Location-Based Services (LBS) have been broadly acknowledged by portable clients as of late. Existing LBS-based frameworks expect clients to sort in total catchphrases. Be that as it may, for portable clients, it is somewhat hard to sort in entire watchwords on cell phones. To lighten this issue, in this paper we consider the location-mindful moment look issue, which returns clients location-mindful answers as clients sort in questions letter by letter. The primary test is to accomplish high intuitive speed. To address this test, in this paper we propose a novel list structure, prefix district tree (called PR-Tree), to proficiently bolster location-mindful moment seek. PR-Tree is a tree-based file structure which flawlessly incorporates the literary portrayal and spatial data to list the spatial information. Utilizing

the PR Tree, we create proficient calculations to help single prefix inquiries and multi-watchword questions. Tests demonstrate that our technique accomplishes superior and fundamentally beats best in class strategies.

2.2 Survey on Location-aware tag recommendations for flickr:

Authors: - I. Miliou and A. Vlachou

Flickr is one of the biggest online picture accumulations, where shared photographs are normally explained with tags. The tagging procedure conquers any hindrance between visual substance and catchphrase look by giving an important printed depiction of the tagged question. Be that as it may, the undertaking of tagging is bulky, in this manner tag proposal is normally used to recommend significant tags to the client. Aside from literary tagging based on catchphrases, an expanding pattern of geotagging has been as of late seen, as saw by the expanded number of geotagged photographs in Flickr. Despite the fact that there exist diverse strategies for tag suggestion of photographs, the pick-up of utilizing spatial and literary data keeping in mind the end goal to prescribe more important tags to clients has not been examined yet. In this paper, we propose novel location-aware tag suggestion techniques and exhibit the adequacy of our proposed strategies.

2.3. Survey on spatial keyword query processing: an experimental evaluation

Authors: - L. Chen, G. Cong, C. S. Jensen, and D. Wu

Geo-textual files assume a critical part in spatial keyword querying. The current geo-textual lists have not been thought about efficiently under the same trial structure. This makes it hard to figure out which ordering method best backings particular usefulness. We give an overall study of 12 state-of-the-workmanship geo-textual files. We propose a benchmark that empowers the correlation of the spatial-keyword question execution. We additionally give an account of the discoveries acquired while applying the benchmark to the lists, along these lines revealing new bits of knowledge that may control file determination and also additionally investigate.

3.METHODOLOGY:

The Administrator principle client of our application, he gather spatial database and play out our proposed technique PA for clients. It is essential that the articles many databases like spatial (restaurants/Hotels) are connected with keyword(s) to show their associations/organizations/keys.

rid	R name	About	keywords	latitude	Longitude	city	Restaurar
1	Surya Hotel	Welcome to best Non Veg Hotel	chicken biryani, spagati, noodles	17.44194651690908	78.4542955122451	Hyderabad	500016
2	Bavarchi Restaur	This Non and Veg Hotel	spagati, noodles	17.47068563726902	78.4223664961318	Hyderabad	500018
3	Hyderabad Dhaba	Its Completly Hyderabad Specia	spagati, chicken biryani	17.46446330629862	78.4282029829482	Hyderabad	500018
4	Blue Fox	Luxury Hotel for a	chicken biryani, noodles, veg soup, biryani	17.43695150672265	78.4450257978896	Hyderabad	500082
5	Ruchi Resturent	Pure veg Hotel	spagati, veg soup	17.44366607892473	78.4774697981338	Hyderabad	500003

6	Kruthinga Resturent	Hotel for all Veg ar Non Veg	spagati, veg biryar noodles	17.46970317809523	78.4275163374404	Hyderabad	500018
---	---------------------	------------------------------	-----------------------------	-------------------	------------------	-----------	--------

Document Table Data

The client enrolls themselves by giving their legitimate subtle elements, for example, email id, client name, telephone number, and so on. Client will seek in our application utilizing arbitrary keyword and he will get agreeing comes about. After he will get a few catchphrases proposals in light of our PA calculation. Those outcomes imply those watchwords are identified with question and close to his separation.

Customer in our application information proprietor of our application, he/she present the spatial database of possess business information with geo properties.

4. ALGORITHM

The Partition Based Algorithm the testing and associated results are carried out. This research project expects to have following outcomes by the end of the project.

- User will search in our application using random keywords, and he will get according results.
- After collect documents according user location.
- Perform weight keyword calculations.
- Find keywords suggestions based on our PA algorithm.
- Those results mean those keywords are related to query and near to his distance.

PARTITION BASED ALGORITHM

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

Output: Keywords suggestions K.

Initialization:

- i. User current location, defined $\lambda_u = x_u, y_u$
- ii. Let Documents $D = \{ d_1, d_2, \dots, d_n \}$
- iii. Priority Queue $Q \leftarrow 0$, Active Ink $AINK \leftarrow 0$, Distribution Ratio $distratio = 1$

```

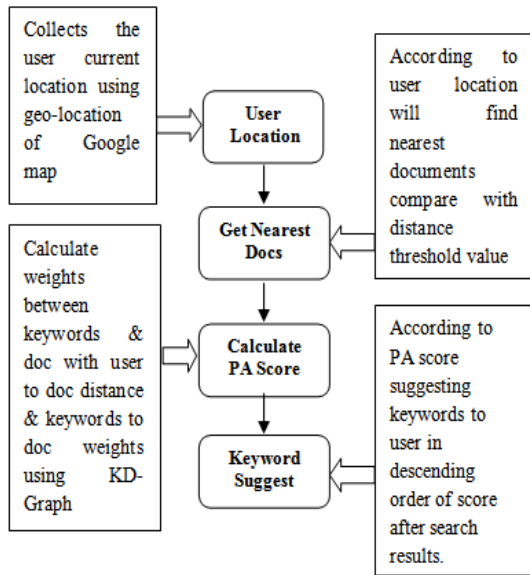
for each D
calculate distance  $dist(\lambda_u, D(i))$ 
  if ( $dist \leq 1$ )
    According to Euclidian Distance
    calculate Score ( $\acute{w}$ );
    Keywords to Document
     $w(e)$  weight of edge. According to KD Graph
     $\beta = 0.5$  normalized to take values in [0, 1]
    Documents to Keywords
     $w(e)$  weight of edge. According to KD Graph
     $\beta = 0.5$  normalized to take values in [0, 1]

 $\acute{w}(e') = \beta * w(e') + (1 - \beta) * (1 - mindist(\lambda_u, D_i))$ 
)

else break;
end for;
distratio = distratio - 0.5;
while Q != 0
for each node v connected top in G do
  Distribute distratio according to  $\acute{w}(e)$ ;
   $AINK \leftarrow AINK - top. aink$ ;
  Distribute distratio according to
 $\acute{w}(e')$ ;end for;
end while;
return top G;

```

5. FLOW OF EXECUTION:

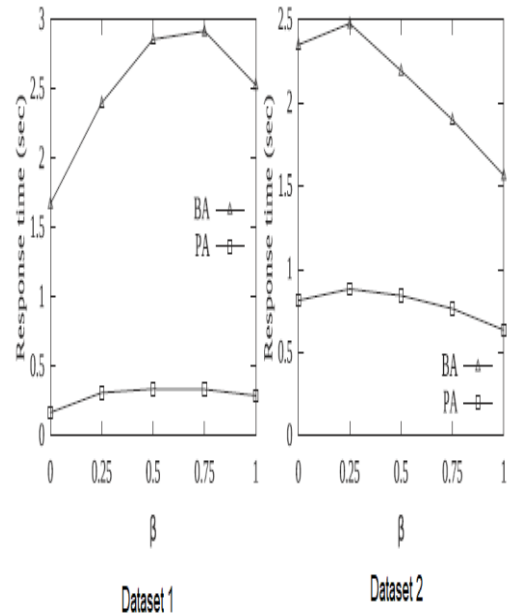


We design the first ever Location-aware Keyword query Suggestion framework, for suggestions relevant to the user's information needs that also retrieve relevant documents close to the query issuer's location to get Keyword suggestions by PA algorithm.

6. EXPERIMENTAL ANALYSIS:

X- Shows the response time when varying Y-axis. Both BA and PA are bell-shaped. When β approaches 0 or 1, only the documents that are either close to the query location or have high textual relevance to the user supplied keywords are involved in the computation. Algorithm PA outperforms BA for all values

of Y-axis by a wide margin



7. FUTURE WORK:

There are two mechanisms which can be used in the future. The more highly keyword-document graph is the store data and searching based on keyword PA algorithm will increase the application performance and data accuracy. In future, we apply the concept of keyword suggestion in instant search results.

8. CONCLUSION:

In the proposed method, an LKS framework providing keyword suggestion that are relevant to the user information needs and at the same time can retrieve relevant documents near the user location. A baseline algorithm will come out from calculation BCA is acquainted with tackle the issue.

The admin will make a segment based calculation which processes the scores of the applicant watchword inquiries at the segment level and uses a languid component to significantly diminish the computational cost. Exact investigations are led to contemplate of adequacy related to system and the execution of the proposed calculations. The outcome

demonstrates that the system can offer helpful recommendations and that PA beats the pattern calculation fundamentally. Later on, we intend to additionally think about the adequacy of the structure by gathering more information and outlining a benchmark. What's more, the main thing will be accessibility of information.

9. REFERENCES

- [1]. R. Baeza-Yates, C. Hurtado, and M. Mendoza, "Query recommendation using query logs in search engines," in *Proc. Int. Conf. Current Trends Database Technol.*, 2004, pp. 588–596.
- [2]. D. Beeferman and A. Berger, "Agglomerative clustering of a search engine query log," in *Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2000, pp. 407–416.
- [3]. 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.
- [4]. N. Craswell and M. Szummer, "Random walks on the click graph," in *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2007, pp. 239–246.
- [5]. Q. Mei, D. Zhou, and K. Church, "Query suggestion using hitting time," in *Proc. 17th ACM Conf. Inf. Knowl. Manage.*, 2008, pp. 469–478.
- [6]. 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.
- [7]. T. Miyanishi and T. Sakai, "Time-aware structured query suggestion," in *Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2013, pp. 809–812.
- [8]. A. Anagnostopoulos, L. Becchetti, C. Castillo, and A. Gionis, "An optimization framework for query recommendation," in *Proc. ACM Int. Conf. Web Search Data Mining*, 2010, pp. 161–170.
- [9]. P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna, "The query-flow graph: Model and applications," in *Proc. 17th ACM Conf. Inf. Knowl. Manage.*, 2008, pp. 609–618.
- [10]. Y. Song, D. Zhou, and L.-w. He, "Query suggestion by constructing term-transition graphs," in *Proc. 5th ACM Int. Conf. Web Search Data Mining*, 2012, pp. 353–362.
- [11]. L. Li, G. Xu, Z. Yang, P. Dolog, Y. Zhang, and M. Kitsuregawa, "An efficient approach to suggesting topically related web queries using hidden topic model," *World Wide Web*, vol. 16, pp. 273–297, 2013.
- [12]. D. Wu, M. L. Yiu, and C. S. Jensen, "Moving spatial keyword queries: Formulation, methods, and analysis," *ACM Trans. Database Syst.*, vol. 38, no. 1, pp. 7:1–7:47, 2013.
- [13]. D. Wu, G. Cong, and C. S. Jensen, "A framework for efficient spatial web object retrieval," *VLDB J.*, vol. 21, no. 6, pp. 797–822, 2012.
- [14]. J. Fan, G. Li, L. Zhou, S. Chen, and J. Hu, "SEAL: Spatio-textual similarity search," *Proc. VLDB Endowment*, vol. 5, no. 9, pp. 824–835, 2012.
- [14]. P. Bouros, S. Ge, and N. Mamoulis, "Spatio-textual similarity joins," *Proc. VLDB Endowment*, vol. 6, no. 1, pp. 1–12, 2012.
- [15]. Y. Lu, J. Lu, G. Cong, W. Wu, and C. Shahabi, "Efficient algorithms and cost models for reverse spatial-keyword k -nearest neighbor search," *ACM Trans. Database Syst.*, vol. 39, no. 2, pp. 13:1–13:46, 2014.
- [16]. S. Basu Roy and K. Chakrabarti, "Location-aware type ahead search on spatial databases: Semantics and efficiency," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2011, pp. 361–372.
- [17]. R. Zhong, J. Fan, G. Li, K.-L. Tan, and L. Zhou, "Location-aware instant search," in *Proc. 21st ACM Conf. Inf. Knowl. Manage.*, 2012, pp. 385–394.
- [18]. I. Miliou and A. Vlachou, "Location-aware tag recommendations for flickr," in *Proc. 25th Int. Conf. Database Expert Syst. Appl.*, 2014, pp. 97–104.