TOPk-VO: MOVING TOP-K SPATIAL KEYWORD QUERIES WITH VERIFICATION OBJECT

Mohamed Jamshad K, Assistant Professor of Computer Science, EMEA College of Arts & Science, Kondotty, Malappuram(dist.), Kerala

Abstract

Conventional spatial searches, such as range search and closest neighbor retrieval, require just geometric property constraints on objects. Many contemporary applications need new types of queries that seek objects that fulfill both a spatial criterion and a predicate on their related textual. Instead of evaluating all of the restaurants, a nearest neighbor question might ask for the restaurant that is the closest among those whose menus include "steak, pasta, and brandy" all at the same time. Currently, the best answer to such queries is based on the IR (information retrieval)-tree, which has a few flaws that significantly affect its efficiency, as shown in this research. Using efficient Moving Top-k Spatial Keyword (MkSK) searches, the proposed method named as TOPk-VO for efficient spatial keyword queries. An MkSK query that includes location in the query allows a client to be aware of the spatial web objects that best fit a query in terms of location and text relevancy. For validating MkSK requests, the system creates a verification object (VO). The VO is sent to the customer and is immediately updated on the website. The query will obtain the relevant latitude and longitude information from the data owner using MkSK query processing and point out the values in Google map via the VO along with the client request. This system additionally employs authentication data structures such as MIR-tree and MIR*-tree to ensure that the client obtains the correct top-k result even while its geographical position changes constantly.

Keywords: MkSK, VO, Spatial, Keyword Query, MIR, MIR*

I. INTRODUCTION

Data mining is the computer process of finding patterns in massive data sets using techniques from statistics and database systems [1].

A spatial keyword query combines location and text search by accepting as inputs a location and a list of keywords. Spatial web objects may be places of attraction (for example, restaurants) with an online presence and written descriptions. The MkSK query takes into account a query's continuously changing position and keeps a mobile client informed of the top-k spatial web items that best match a query in terms of location and text relevance [2]. A mobile client, for example, may initiate a "cafe" query to be informed about nearby coffee shops. While a client is travelling, the MkSK query guarantees that the client always receives an up-to-date answer. The client may ignore an undesirable outcome and continue until a more desirable result arises. A straightforward response to the MkSK question is to periodically activate an existing snapshot of the spatial keyword query processing technique [3]. However, even if snapshot queries are run often, which is inefficient and wasteful given that subsequent responses are likely to be very similar, there is no guarantee that the user will always get the right, most recent result. Another possibility is to use a buffering technique developed for processing spatial k-Nearest Neighbor (kNN) queries to top-k spatial keyword search [4]. Given a kNN query, the function returns the k nearest neighbors and uses them to create a buffer region that is stable as long as the user walks inside it. Using the k objects, the kNN result may be computed [14]. It is unclear, however, if the technique can be extended to MkSK queries that include both text relevance and geographical distance. A safe zone-based approach to MkSK query processing provides the client with a safe zone in addition to the query answer. The safe zone is a region that includes the user's location and maintains a consistent top-k result. Due to the fact that the customer only has to request a new result when leaving the safe zone, the safe zone-based approach significantly reduces communication between clients and Service Providers (SPs), as well as computation expenses [15]. Consider the MkSK query q, which contains the search term "vanilla coffee." The text relevance of objects p1 and p2 to q is 2 and 1, respectively, when the number of matching keywords is used to determine relevance. The SP returns object p2 as the top-ranked result and the grey circle as p2's safe zone, indicating that object p2 is the top-ranked result as long as the client remains inside the grey circle. The curved path reflects the client's movement. When the client crosses the boundary of the grey circle (at q0), it communicates its updated position to the SP, which computes and returns a new top-1 result p1, as well as the white region as a safe zone. Completeness ensures that no legitimate result object is missing from a query result. Authentication techniques have been created for a variety of types of queries, including relational inquiries, sliding window queries, spatial queries, text similarity searches, shortest route searches, moving kNN queries, moving range queries, and sub graph searches. Among the improvements are a data structure for MkSK query authentication and the MIR-tree, which requires less processing and transmission [16, 17]. The MIR-tree may be used to answer a broad variety of spatial keyword query variations. To verify the top-k results and safe zones of MkSK searches, a Verification Object (VO) is generated. Algorithms are being developed for generating VO and verifying top-k results and safe zones using VO. An enhanced data structure (MIR*-tree) is proposed to significantly reduce transmission costs. The idea of the MIR*-tree is applicable to any tree-structured ADS, such as the MIR-tree, in which each node contains several items [18]. To assess performance, a comprehensive experimental investigation on real-world data is carried out.

II. BACKGROUND STUDY

The data outsourcing concept was originally suggested in [4], in which a data owner outsources its data to a third-party service provider who is responsible for responding to data requests from the data owner or other users. In general, data outsourcing raises two security concerns: data privacy and query integrity [5]. To maintain data privacy, the data owner must outsource encrypted data to a service provider, and effective ways for accessing encrypted data must be developed. The pioneering work [4] proposed that SQL queries be performed over encrypted data through bucketization. Their strategy is to delegate as much of the query execution to the service providers as feasible without decrypting the data. The client is responsible for decryption and the remainder of query processing. Numerous researches on performing various queries on encrypted data have been published since then. Another area of research has been devoted to ensuring query integrity, that is, that a query answer is generated from the outsourced data (the authenticity criteria) and contains all of the data satisfying the query (the correctness requirement). Query authentication was originally discussed in the literature of cryptography. Numerous types of questions were examined, including range queries, kNN queries, shortest-path searches, and spatial skyline queries. Another focus of the study was to safeguard data privacy against unreliable service providers. A common practice is to encrypt the dataset prior to outsourcing it to a third-party service provider, and several methods have been proposed to enable rapid query processing over encrypted data. The first study focuses on both one-dimensional and multi-dimensional range queries. Recent work has concentrated on secure ranked keyword search, access control, and circular range queries over encrypted data. The Merkle Hash Tree (MH-tree) [6] is a main-memory binary tree used to organize hash data hierarchically. After building the tree, the data owner signs the hash value stored at the root of the MH-tree using a public key digital signature technique. To authenticate one-dimensional range queries, Devanbu et al. [7] organize the database entries according to the query attribute and index them using an MH-tree. To validate multidimensional range queries in [7], a combination of the MH-tree and the range search tree [8] is utilized. Martel et al. [9] extend the concept of the MH-tree to arbitrary search DAGs, such as dictionaries, tries, and optimal range search trees. ADSs for graph and geometric searches are proposed by Goodrich et al. [10]. On the other hand, these techniques are mostly theoretical in nature and focus on primary memory. The VB-tree [11] is the first disk-based ADS to be described in the Database literature, and it confirms the validity but not the completeness of 1D range results. A subsequent signature chaining technique [12, 13] verifies both the soundness and completeness of the signature.

a) PROBLEM IDENTIFICATION

This advancement makes it possible to outsource spatial keyword searching. In contrast to conventional client-server architectures, outsourcing query capabilities to a service provider (SP) by an organization that want to deliver a service, referred to as a data owner (DO), provides a variety of advantages, including elastic scalability, cost savings, and high availability.

Additionally, data owners are not obliged to invest in high-end hardware or hire technical staff to operate the service. Additionally, outsourcing is anticipated to improve availability. Additionally, outsourcing may help reduce network latency associated with the transmission of query results.

III. TOPk-VO MODEL

The authentication of top-k spatial keyword searches that is moving. Additionally, this suggested approach may be utilized to verify static top-k spatial keyword searches (by validating just the top-k result), which is a sub-task of MKSK query authentication. To evaluate the proposed system's performance, this paper undertakes extensive experimental study utilizing real-world data. This paper introduced a new authenticated data structure, the MIR-tree, to efficiently authenticate moving top-k spatial keyword query results. It is designed to serve as a validation object for MkSK searches' top-k results and safe zones.

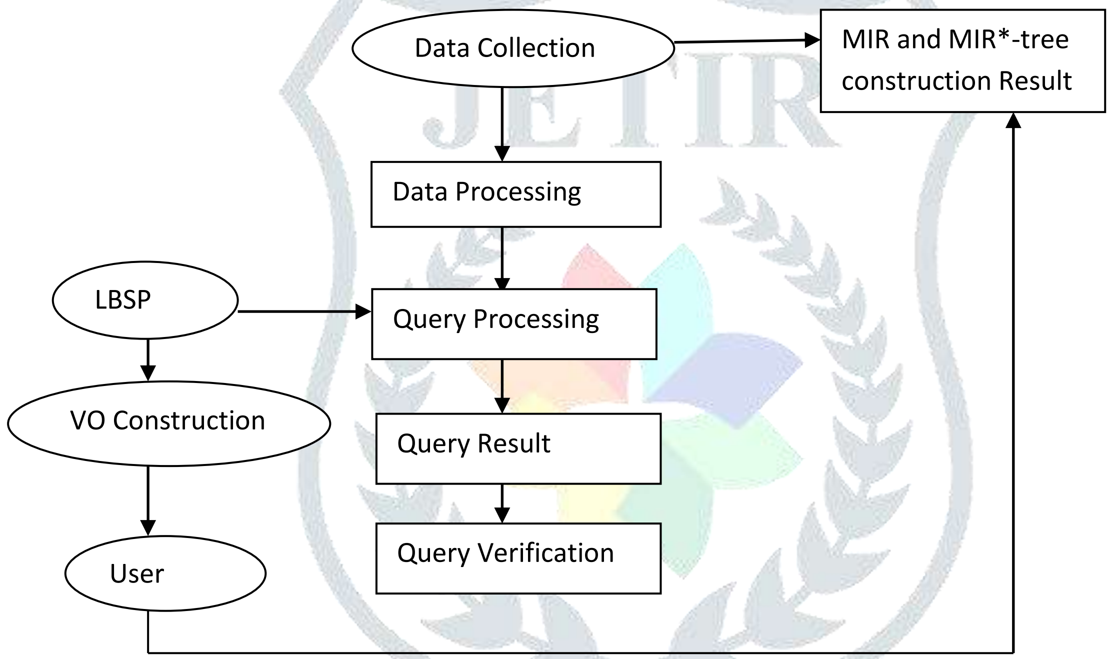

Figure 1: Proposed Architecture Diagram

1. Data Outsourcing

In this Module, Initially the data set consists of Hotel name, addresses and the location in the form of latitude and longitude. And also the tourism places are maintained.

2. Location Tracking

In this module, track the user's location like latitude and longitude values. The User can find his/her current location and views all the hotels in map by entering the location.

3. Authentication Framework:

The framework is divided into two stages, the startup phase and the query processing and authentication phase. The DO obtains a private key from a key distribution centre during the startup phase. The client obtains a public key and signatures from the key distribution centre.

4. **Query Processing**

The client initiates the query processing and authentication step by issuing an MkSK query. When the SP receives the query, it computes the top-k result. Spatial keyword query processing is also gaining popularity in the academic community, with a variety of methods suggested for effectively processing spatial keyword searches.

5. **MIR & Verification Object**

Based on the IRtree, the Merkle-IR-tree (MIR-tree) is created by embedding a sequence of digests in each node of the tree. Each item in the IR-tree summarises the geographic distances and text relevance of the elements in its child node. This index allows for the efficient processing of spatial keyword searches since it may reduce the search space based on spatial proximity and text relevance at the same time.

The MIR-tree has the advantageous feature of requiring just a small number of objects and MIR-tree entries to be put into the VO.

The VO size is essential in order to decrease client/server communication costs and client-side result verification time. This paper next suggest an improved ADS, the MIR*-tree, which allows us to reduce the VO size. The digests in the MIR*-tree are calculated differently from those in the MIR-tree.

When the user searched for a place, the verification object was delivered.

6. **Top K result**

The problem of validating a top-k result entails creating a small VO with minimal communication costs and a fast authentication period. The client retrieves and authenticates the top-k result from the VO.

Retrieve the top-k items with the best ranking scores, as determined by their distance from the query location and the relevancy of their text description to the query keywords chicken, mutton, and vegetables.

7. **VO CONSTRUCTION ALGORITHM**

The Authenticated Data Structure and MIR-tree both have the benefit of needing just a limited amount of objects and MIR-tree entries to be stored in the VO. A Verification Set (VS) comprised of all objects having a rank less than rank [3]. The VO is calculated via a depth-first traversal of the MIR-tree under the following conditions: I If the ranking score of a non-leaf entry e is greater than the ranking score of rank k, the VO entry for e is constructed and added to the VO, but its subtree is not visited. (ii) All items included inside each visited leaf node are added to the VO. When the VO is constructed, a traversal string str is generated to track the MIR-tree search. It contains the identifiers for the tree entries and objects added to the VO, as well as the special letters '12' and '_' used to denote the scope of a node. The traversal string is needed to avoid duplication of entries in the VO, since the VOs for the top-k result and the safe zone may otherwise have identical entries. Finally, the client is supplied with the VO and traversal string str.

IV. RESULTS AND DISCUSSION

The Proposed Topk-VO model has implemented by using Asp.Net Programming language and Google API used for MAP generation.

Figure 2: Home Page for Proposed System

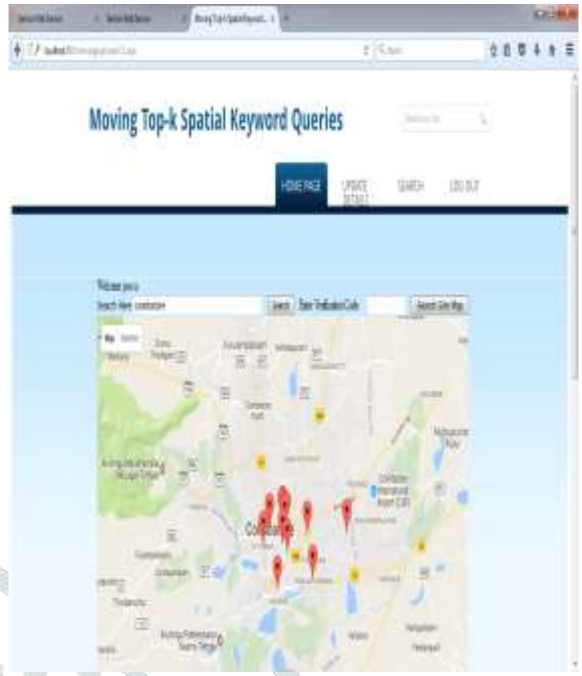

Figure 3: Location Map Marker is displayed from user side

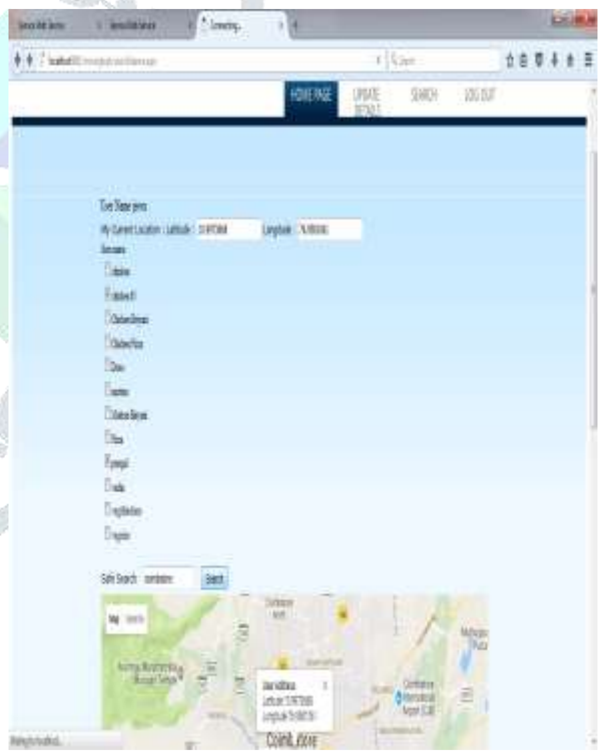

Figure 4: User searching for particular foods

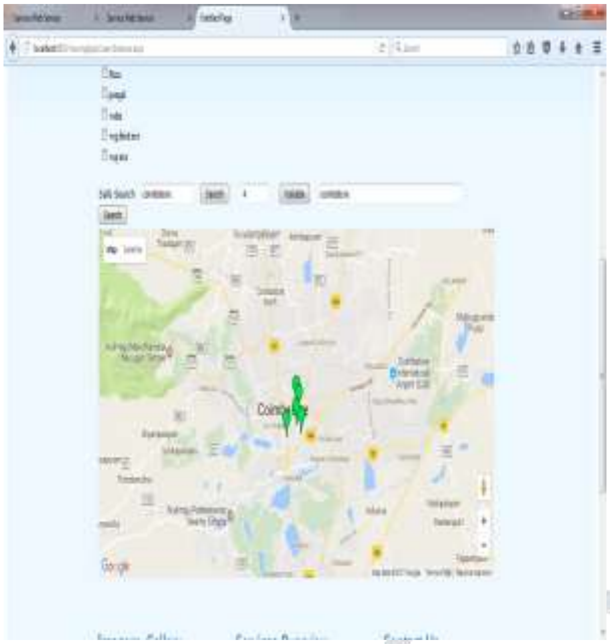

Figure 5: Hotel markers are displayed from user search query

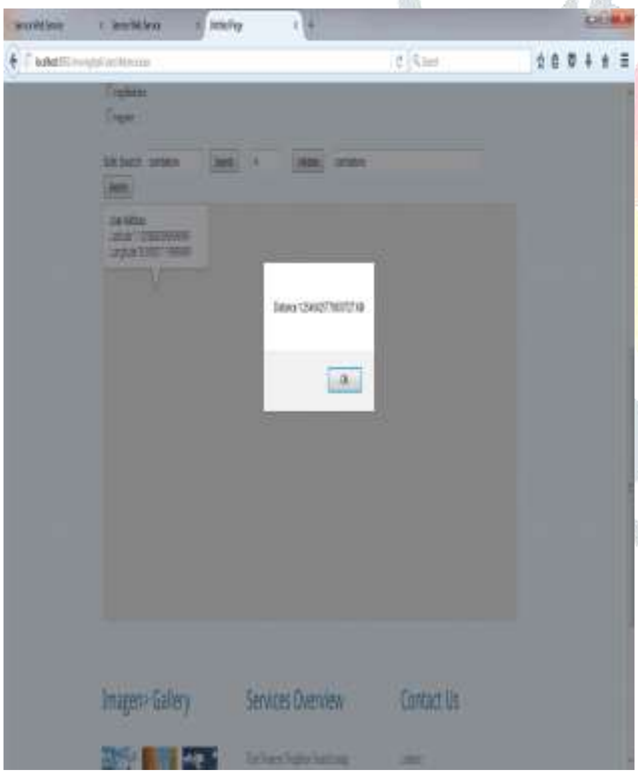

Figure 6: The distance calculated from current location to destination

V. CONCLUSION

The proposed TOPk-VO model achieved the best search result from spatial keyword queries. There have been many applications that have requested a search engine that can effectively handle new kinds of spatial queries that are combined with keyword search. Existing solutions to comparable queries either use an inordinate amount of space or are incapable of responding in real time. Spatial top-k searches are an invaluable resource for location-aware applications. This paper develops a new technique for processing top-k spatial searches. The top-k results and safe zones for MkSK queries are created. Algorithms for generating and distributing verification objects are being developed for verifying top-k results and safe zones. The MIR*-tree is proposed as an enhancement to the MIR-tree in order to reduce communication costs. Extensive empirical testing on real-world data sets demonstrates that the techniques outperform two existing baseline algorithms by orders of magnitude.

VI. REFERENCES

[1] R R. Zhang, Y. Zhang, and C. Zhang, “Secure Top-k Query Processing via Untrusted Location-Based Service Providers,” Proc. IEEE INFOCOM ’12, Mar. 2012.

[2] H. Hacigumus, S. Mehrotra, and B. Iyer, “Providing Database as a Service,” Proc. IEEE 18th Int’l Conf Data Eng. (ICDE), Feb. 2002.

[3] W.-S. Ku, L. Hu, C. Shahabi, and H. Wang, “Query Integrity Assurance of Location-Based Services Accessing Outsourced Spatial Databases,” Proc. Int’l Symp. Advances in Spatial and Temporal Databases, July 2009.

[4] Merkle, R. A Certifi ed Digital Signature. CRYPTO, 1989.

[5] Devanbu P., Gertz M., Martel C., Stubblebine S., Authentic Data Publication over the Internet Journal of Computer Security 11(3): 291-314, 2003.

[6] Berg M., van Kreveld M., Overmar M., Schwarzkopf, Computational Geometry: Algorithms and Applications. Springer, 1997.

[7] Martel C., Nuckolls G, Devanbu P, Gertz M, Kwong, A, Stubblebine S, A General Model for Authenticated Data Structures. Algorithmica, 21-41, 2004.

[8] Goodrich M, Tamassia R, Triandopoulos N, Cohen R, Authenticated Data Structures for Graph and Geometric Searching, CT-RSA, 2003.

[9] Pang H, Tan K.. L, Authenticating Query Results in Edge Computing, ICDE, 2004.

[10] Pang H, Jain A, Ramamritham K, Tan K. L, Verifying Completeness of Relational Query Results in Data Publishing. SIGMOD, 2005.

[11] Narasimha M, Tsudik G, Authentication of Outsourced Databases Using Signature Aggregation and Chaining, DASFAA, 2006.

[12] F. Chen and A. Liu, SafeQ: Secure and Effi cient Query Processing in Sensor Networks, Proc. IEEE INFOCOM‟10, pp. 1-9, Mar. 2010.

[13] R. Zhang, J. Shi, Y. Liu, and Y. Zhang, Verifi able Fine-Grained Top-K Queries in Tiered Sensor Networks, Proc. IEEE INFOCOM‟10, Mar. 2010.

[14] R. Merkle, Protocols for Public Key Cryptosystems, Proc. IEEE Symp. Security and Privacy (S&P‟80), pp. 122- 134, Apr. 1980.

[15] B. Sheng and Q. Li, Verifi able Privacy-Preserving Range Query in Sensor Networks, Proc. IEEE INFOCOM‟08, pp. 46-50, Apr. 2008.

[16] B. Hore, S. Mehrotra, and G. Tsudik, A Privacy-Preserving Index for Range Queries, Proc. 30th Int‟l Conf. Very Large Data Bases (VLDB‟04), pp. 720-731, Aug. 2004.

[17] M. L. Yiu, Y. Lin, and K. Mouratidis. Effi cient verifi cation of shortest path search via authenticated hints. In ICDE, pages 237–248, 2010.

[18] Y. Yang, D. Papadias, S. Papadopoulos, and P. Kalnis. Authenticated join processing in outsourced databases. In SIGMOD, pages 5–18, 2009.