Querying Geo-Textual Data: Spatial Keyword Queries and Beyond

Gao Cong
Nanyang Technological University
Singapore
gaocong@ntu.edu.sg

Christian S. Jensen
Aalborg University
Denmark
csj@cs.aau.dk

ABSTRACT

Over the past decade, we have moved from a predominantly desktop based web to a predominantly mobile web, where users most often access the web from mobile devices such as smartphones. In addition, we are witnessing a proliferation of geo-located, textual web content. Motivated in part by these developments, the research community has been hard at work enabling the efficient computation of a variety of query functionality on geo-textual data, yielding a sizable body of literature on the querying of geo-textual data.

With a focus on different types of keyword-based queries on geo-textual data, the tutorial also explores topics such as continuous queries on streaming geo-textual data, queries that retrieve attractive regions of geo-textual objects, and queries that extract properties, e.g., topics and top-k frequent words, of the objects in regions. The tutorial is designed to offer an overview of the problems addressed in this body of literature and offers an overview of pertinent concepts and techniques. In addition, the tutorial suggests open problems and new research direction.

1. INTRODUCTION

In part due to the proliferation of GPS-equipped mobile devices, notably smartphones, massive volumes of geo-located, or geo-tagged, text content is becoming available on the Web. Examples of such content include points of interest (POIs) with descriptive text, geo-tagged micro-blog posts (e.g., tweets), geo-tagged photos with text tags (e.g., as found at Flickr and Instagram), check-ins from location-based social networks (e.g., FourSquare), geo-tagged news, and geo-tagged web pages.

We refer to such data as geo-textual, or spatio-textual, data. As stated, massive volumes of such data are available. For example, Foursquare hosts over 87 million locations around the world with over 8 billion check-ins[1]. Further, new geo-textual data is being generated, and the volumes of geo-textual data are expected to grow at an accelerated pace as mobile devices continue to proliferate.

Geo-textual data can be divided into (i) streaming geo-textual data that arrives at a high rate, exemplified by geo-tagged tweets and (ii) static geo-textual data that is relatively stable, exemplified by collections of POIs.

The development described above also motivates increased research on the management of geo-textual data in the data management, data mining, and information retrieval communities. On the one hand, compared with traditional spatial data, the textual component greatly enriches the data. On the other hand, the spatial component of geo-textual data also adds a new and semantically rich aspect to textual data.

The tutorial first covers keyword-based querying of geo-textual data and then covers related functionality. A brief overview of the scope of the tutorial follows.

Spatial keyword queries. Many types of spatial database queries have been revisited for geo-textual data; and keyword queries have also been revisited in the context of geo-textual data. The resulting studies give prominence to spatial keyword queries that combine spatial functionality (e.g., range and nearest neighbor queries) with keyword queries. A typical spatial keyword query finds the objects that best match the location and keywords in the query. To address different use cases, many types of spatial keyword queries and accompanying indexing and query processing techniques have been proposed.

Querying geo-textual streams. Geo-textual data may arrive at a high rate (e.g., geo-tagged tweets or photos) and can then be modeled as a data stream. In such streaming settings, continuous queries are of particular interest. Here, users may want to be notified when interesting geo-textual objects arrive. For example, a user may want to be notified when tweets arrive that contain the term "flu" and are posted from within 5 km of the user's home.

Exploratory search and mining. Exploratory search [41] helps users search, navigate, and discover new facts and has grown in prominence. Here, querying and browsing are typically combined to enable investigation and foster learning. Geo-textual data contains both structured and unstructured data and can be readily presented on a map. One approach to enabling exploratory search on geo-textual data is to conduct search or mining of geo-textual data in a user-specified region interactively, such as finding the top-k most frequent terms in a region. Another approach is to interactively find regions with particular, user-specified properties.

[1]https://foursquare.com/about accessed January 2016.

SIGMOD'16, June 26-July 01, 2016, San Francisco, CA, USA
© 2016 ACM. ISBN 978-1-4503-3531-7/16/06...$15.00
DOI: http://dx.doi.org/10.1145/2882903.2912572

In the reminder, we review the topics covered and also cover future directions.

2. SPATIAL KEYWORD QUERIES

We briefly review different types of spatial keyword queries and indexing techniques.

2.1 Standard Queries

Standard queries generalize fundamental queries from spatial databases and information retrieval. In spatial databases, the arguably most fundamental queries are range and k nearest neighbor queries. In information retrieval, queries may be Boolean, returning results that contain query keywords, or ranking-based, returning the k objects that are most similar to query keywords according to some text similarity function.

The content that is queried is a set $\mathcal{D}$ of objects. An object $p \in \mathcal{D}$ has two attributes: $\langle \lambda, \psi \rangle$, where λ encodes a geo-location and ψ is a text value. These objects are digital, are often assumed to be available on the web, and typically represent a physical, geo-located entity that may be of potential interest to a user. The objects are also called places, POIs, geo-textual objects, or web objects. Standard queries return a set or ranked list of objects from $\mathcal{D}$.

We cover several types of standard queries. Let ρ be a spatial region, λ be a point location, ψ be a set of keywords, and k be the number of objects to return.

1. A *Boolean range query* $q = \langle \rho, \psi \rangle$ returns all objects in $\mathcal{D}$ that are located in region ρ and that contain all the keywords in ψ.

2. A *Boolean kNN query* $q = \langle \lambda, \psi, k \rangle$ returns up to k objects from $\mathcal{D}$, each of which contains all the keywords in ψ, ranked in increasing spatial distance from λ.

3. A *top-k range query* $q = \langle \rho, \psi, k \rangle$ returns up to k objects from $\mathcal{D}$ that are located in the query region ρ, now ranked according to their text relevance to ψ.

4. Finally, a *top-k kNN query* $q = \langle \lambda, \psi, k \rangle$ retrieves k objects from $\mathcal{D}$, ranked according to a score that takes into consideration both spatial proximity and text relevance.

We name the above queries by following the work [4]; the queries may be named differently in other studies [10, 13, 19, 21, 27, 29, 43, 50, 61, 65, 67].

2.2 Beyond Single Object Result Granularity

In some scenarios, a user's needs are not met best by a query result that returns a set or ranked list of geo-textual objects from $\mathcal{D}$, each of which aims to satisfy the user's needs. Instead, the user's needs may be better satisfied by an aggregation of several objects that are near each other, meaning that the possible answers are subsets of $\mathcal{D}$, sets of subsets of $\mathcal{D}$, or ranked lists of subsets of $\mathcal{D}$.

For example, several objects may combine to collectively meet the user's needs, while no individual object meets the needs well. Consider a query with keywords "hotel, pub, beach." Perhaps no nearby single place is a good match for this combination of keywords. Instead, a group of three places that are close to each other and are close to the query location may combine to provide a result that meets the user's needs better than any single object.

The *m-closest keywords* (mCK) query [26, 63, 64] retrieves a set of objects with text descriptions that combine to contain a set of m query keywords, while also minimizing the maximum distance between any two objects in the result set. Apart from retrieving a set of objects satisfying the user needs, the mCK query can also be used for geo-tagging a document or a photo with textual tags.

The *collective keyword* query [6, 7, 37] further generalizes the mCK query. A query q takes a location λ and a set of keywords ψ as arguments. Its search space is all subsets of the set of objects $\mathcal{D}$, and it returns a set of objects such that (i) the textual descriptions of the objects collectively cover ψ, (ii) the result objects are all close to λ, and (iii) the result objects are close to each other.

The *top-k groups* query [2, 47] aims to support users who wish to explore different objects. For example, a user may want to explore different restaurants before deciding where to have dinner. Given a query location and query keywords, the query retrieves a ranked list of k groups, or sets, of objects that score the best according to a ranking function that takes into account the diameter of the group, the group's distance to the query location, and the relevance of the group's objects to the query keywords. Unlike in the collective keyword query, where result objects combine to be an answer, each object in a result group is a possible answer.

The *keyword-aware route planning* query [5, 60] retrieves a route that covers all query keywords and perhaps satisfies a certain constraint (e.g., distance), while optimizing an objective, e.g., the popularity of the route.

2.3 Other Queries

Generalizations of many other types of spatial queries are revisited in the context of geo-textual data, including the following.

1. *Moving continuous queries* [54, 57–59]. This type of query enables a mobile user (e.g., a pedestrian or driver) to be continuously aware of the k geo-textual objects that best match a query with respect to location and text relevancy. In addition to static POIs, the objects may model users that play the role of service providers and thus offer moving services, with textual descriptions, to other users.

2. *Why-not queries* [14, 15]. When issuing a top-k NN query, a user may expect some known object to be in the result of the query. Should such an object be missing from the result, a why-not query offers an explanation of why the expected object is missing and suggests a similar query with revised parameters that includes the missing object in its result.

3. *Reverse kNN queries* [18, 38]. Different from a traditional user, a business may be interested in finding users or business objects that have a query object in their lists of top-k objects ranked by a ranking function that takes both spatial proximity and text relevance into consideration. This query is called a reverse spatial and textual kNN query.

4. *Similarity join queries* [3, 22]. Given a set of geo-textual objects, a distance threshold, and a text similarity threshold, the spatio-textual similarity join query

finds all pairs of objects such that the distance between each pair is smaller than the distance threshold and their text similarity exceeds a threshold.

2.4 Querying in Road Networks

The proposals reviewed above generally assume a Euclidean space setting. Other proposals (e.g., [8, 44, 62]) on efficiently processing of spatial keyword queries assume a spatial-network setting and define spatial distance as road network distance, which is more computationally expensive to compute than Euclidean distance. A distributed solution [39] is proposed to process Boolean range queries and spatial group keyword queries on road networks.

2.5 Indexing Techniques

A common challenge in answering the kinds of spatial keyword queries covered above is to develop index structures and algorithms that enable efficient query processing. The tutorial does not cover algorithms, but it offers a brief coverage of indexing techniques for geo-textual data and related techniques used in open sources/commercial systems.

Existing spatial-keyword indexing techniques usually combine a spatial index and a text index so that both textual and spatial information can be utilized to prune the search space when processing spatial keyword queries. The existing indices can be categorized according to the spatial index they utilize: (i) R-tree based indices [21, 27, 29, 43, 55, 56, 67]; (ii) grid or quad-tree based indices [20, 34, 50, 61, 65]; and (iii) space filling curve based indices [16, 19]. Using the text index employed, indexing techniques can also be classified as inverted file based (e.g., [67]) and signature-file based (e.g., [29]). In addition, some techniques (e.g., [67]) loosely combine a spatial and a text index, while other others integrate them tightly, resulting in hybrid index structures (e.g., [29]), and yet another approach captures an object's spatio-textual part as a compact bit string that can be indexed using a standard index [46] such as a B-tree.

3. QUERYING GEO-TEXTUAL STREAMS

Streaming geo-textual data, exemplified by streams of geo-tagged microblog posts, is an increasingly available and thus important type of data that is attracting increasing interest. Most of existing research on querying geo-textual data streams aims to develop efficient solutions to handling a large number of spatial-keyword subscription, or continuous, queries over geo-textual streams. We cover recent studies on querying geo-textual data streams.

Boolean subscription queries. Several proposals [11, 35, 53] consider Boolean subscription queries over streaming data, where both spatial and keyword conditions serve as Boolean filters. These proposals adopt slightly different settings.

For example, the keyword condition in one study [11] supports both Boolean AND and Boolean OR semantics, while the other studies [35, 53] focus on Boolean AND. The geo-textual objects in one study [35] can be associated with regions, while the geo-textual objects in the other studies have point locations. These proposals present solutions to indexing and grouping subscription queries such that a group of subscriptions can be processed together over the geo-textual data stream, rather than being processed individually, which will be computationally expensive. In addition to being able to efficiently process a large number of subscription queries over geo-textual data streams where geo-textual objects arrive at a high rate, such solutions should be capable of efficiently handling the arrival of new subscriptions and the expiry or discontinuation of existing subscriptions.

Similarity based subscription queries. Instead of employing Boolean conditions in subscription queries, several recent studies [12, 28] employ similarity notions in subscription conditions. In one study [28], a subscription condition can be defined on the spatial-keyword similarity between an incoming geo-textual object and a subscription. Thus, if the similarity exceeds a threshold, the condition is satisfied, and the object is emitted as a result of the subscription.

In contrast, another study [12] aims at maintaining the top-k best objects for each subscription over a stream of geo-textual objects, where the ranking score for each object is computed by considering text relevance, spatial proximity, and the freshness of the object. In yet another study [25], a different type of subscription query is proposed, where the user of each subscription query moves.

Other queries There exists little work on supporting one-time queries on geo-textual data streams. A system [40] is built to support querying geo-textual data within given spatial and temporal ranges.

4. EXPLORATORY SEARCH AND MINING

We proceed to outline studies on exploring geo-textual data. We group them into two categories.

Region search. Given a collection of geo-textual objects, the region search problem is to identify a region, that satisfies a user-specified condition (e.g., involving the size and shape of the region) and that maximizes some aggregate score (e.g., a SUM function) of the objects inside it, for user exploration. Based on a returned result, users may want to modify the query parameters and issue a new region search query.

Given a set $\mathcal{D}$ of geo-textual objects and a rectangle r of a given size (length and width), the *Maximizing Range Sum* (MaxRS) problem is to find a location for r that maximizes the sum of the scores of all the objects covered by r. The problem was first studied in the computational geometry community. Imai et al. [30] propose an $O(n \log n)$ optimal algorithm, where n is the number of objects in $\mathcal{D}$; Nandy and Bhattacharya [42] propose a line-sweeping-based algorithm with the same complexity. An external memory algorithm for the MaxRS problem has also been proposed [17], as has an approximate algorithm for the problem [49]. In recent work [23], the aggregate score functions in the region search problem are defined to be submodular monotone set functions, rather than SUM as used in MaxRS query. These existing studies on region search consider a collection of static geo-textual objects.

The problem of finding a region that does not exceed a given size and that contains objects with the maximum sum of scores is also studied in a road network setting [9]. Here, vertices are created for all geo-textual objects, and weights are associated with vertices that denote the relevance to query keywords of an object. A region is a connected subgraph, and subgraph size is formalized as the sum of the lengths of the edges in in the subgraph. The goal is then to find a qualifying subgraph where the sum of the vertex weights is maximal.

The above studies do not aim at supporting interactive exploration. In contrast, Alexander et al. [31] propose semantic window to study the region search problem for interactive data exploration of multidimensional data, in which a user explores a data space by posing a number of queries that find rectangular regions of interest.

Region exploration. The problem of region exploration is to explore and discover properties of user-specified regions. Based on the returned results, users may interactively specify different regions and pose a different region exploration queries. Region exploration over geo-textual data is a relatively new research topic. Given a user-specified region, one study [48] considers the problem of retrieving the top-k frequent words over the geo-textual data stream for the region. Another study considers the problem of selectivity estimation [52] for a user-specified region. Based on pre-computed probabilistic topic models for each grid cell, a recent study [66] proposes an approach to efficiently discover topics for a user-specified region. Additionally, event detection [1] is performed for a given region over text streams [51]. Feng et al. [24] propose a system called StreamCube designed to enable exploration of events over the spatio-temporal Twitter stream by clustering hashtags. Sankaranarayanan et al. [45] develop a news processing system, called TwitterStand, to continuously acquire breaking news from tweets. In TwitterStand, tweets falling into a specified region are clustered, and each cluster of tweets is associated with a set of geographical locations.

5. FUTURE DIRECTIONS

While good progress has already been made, research on geo-textual data has just begun, and there are many opportunities for continued research. Here, we discuss some of the possible directions.

Effective ranking and user evaluation. Numerous signals are in play when search engines perform ranking of web pages, such as the quality of a web page, click through, and diversity. It is natural to consider whether the signals used for conventional web-page ranking can be useful for the ranking of geo-textual data. Further, geo-textual data contains new features that can be considered for ranking, such as the popularity or rating of a POI.

Reliable evaluation of ranking proposals for geo-textual objects is essential in order to make progress. There is often no mathematical definition of the right, or best, result. Rather, the utility of a result is user dependent, and we need to determine how useful users will find a result. It is thus challenging to establish a reliable ground truth for the results of ranking queries, and user evaluation is an important ingredient when attempting to evaluate the effectiveness of a proposal for the ranking of geo-textual objects. Being able to assess how good ranking functions are will enable better and more complex ranking function [32]. This may in turn call for new indexing and query processing techniques.

Personalized spatial keyword queries. The existing work on spatial keyword queries does not consider personalization. On the other hand, work on personalized location recommendation [36] does not consider spatial keyword search, but instead aims to understand users' topical interests and mobility preference from users' historical geo-textual data. It is of interest to attempt to bridge this gap, thus enabling personalized search.

Querying and mining geo-textual data streams. It is an open problem to effectively and efficiently support continuous queries on geo-textual streams. For example, instead of receiving individual tweets from a stream, users may want to be notified in real time of relevant trending events or even of casual relationships among events. Furthermore, high velocity geo-textual data streams call for distributed solutions. Additionally, geo-textual data streams can be integrated with static geo-textual data, such as POIs. By bridging dynamic geo-textual data streams and static geo-textual data, exciting opportunities for data analytics emerge.

Exploratory search and mining. Being a new topic for geo-textual data, many research problems remain open within the scope of this topic. For example, what are interesting exploratory search and mining tasks on geo-textual data? What factors should be considered in such tasks, e.g., diversity [33]? How can the tasks be performed efficiently? How can interactive exploratory search and mining be performed efficiently?

6. REFERENCES

[1] H. Becker, M. Naaman, and L. Gravano. Learning similarity metrics for event identification in social media. In *WSDM*, pages 291–300, 2010.

[2] K. S. Bøgh, A. Skovsgaard, and C. S. Jensen. GroupFinder: A new approach to top-k point-of-interest group retrieval. *PVLDB*, 6(12):1226–1229, 2013.

[3] P. Bouros, S. Ge, and N. Mamoulis. Spatio-textual similarity joins. *PVLDB*, 6(1):1–12, 2012.

[4] X. Cao, L. Chen, G. Cong, C. S. Jensen, Q. Qu, A. Skovsgaard, D. Wu, and M. L. Yiu. Spatial keyword querying. In *ER*, pages 16–29, 2012.

[5] X. Cao, L. Chen, G. Cong, and X. Xiao. Keyword-aware optimal route search. *PVLDB*, 5(11):1136–1147, 2012.

[6] X. Cao, G. Cong, T. Guo, C. S. Jensen, and B. C. Ooi. Efficient processing of spatial group keyword queries. *ACM Trans. Database Syst.*, 40(2):13, 2015.

[7] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi. Collective spatial keyword querying. In *SIGMOD*, pages 373–384, 2011.

[8] X. Cao, G. Cong, C. S. Jensen, and M. L. Yiu. Retrieving regions of interest for user exploration. *PVLDB*, 7(9):733–744, 2014.

[9] X. Cao, G. Cong, C. S. Jensen, and M. L. Yiu. Retrieving regions of interest for user exploration. *PVLDB*, 7(9):733–744, 2014.

[10] A. Cary, O. Wolfson, and N. Rishe. Efficient and scalable method for processing top-k spatial Boolean queries. In *SSDBM*, pages 87–95, 2010.

[11] L. Chen, G. Cong, and X. Cao. An efficient query indexing mechanism for filtering geo-textual data. In *SIGMOD*, pages 749–760, 2013.

[12] L. Chen, G. Cong, X. Cao, and K. Tan. Temporal spatial-keyword top-k publish/subscribe. In *ICDE*, pages 255–266, 2015.

[13] L. Chen, G. Cong, C. S. Jensen, and D. Wu. Spatial keyword query processing: An experimental evaluation. *PVLDB*, 6(3):217–228, 2013.

[14] L. Chen, X. Lin, H. Hu, C. S. Jensen, and J. Xu. Answering why-not questions on spatial keyword top-k queries. In *ICDE*, pages 279–290, 2015.

[15] L. Chen, X. Lin, H. Hu, C. S. Jensen, and J. Xu. Answering why-not spatial keyword top-k queries via keyword adaption. In *ICDE*, to appear, 2016.
[16] Y. Chen, T. Suel, and A. Markowetz. Efficient query processing in geographic web search engines. In *SIGMOD*, pages 277–288, 2006.
[17] D.-W. Choi, C.-W. Chung, and Y. Tao. A scalable algorithm for maximizing range sum in spatial databases. *PVLDB*, 5(11):1088–1099, 2012.
[18] F. Choudhury, J. S. Culpepper, T. Sellis, and X. Cao. Maximizing bichromatic reverse spatial and textual k nearest neighbor queries. *PVLDB*, 9(6):456–467, 2016.
[19] M. Christoforaki, J. He, C. Dimopoulos, A. Markowetz, and T. Suel. Text vs. space: efficient geo-search query processing. In *CIKM*, pages 423–432, 2011.
[20] G. Cong, H. Lu, B. C. Ooi, D. Zhang, and M. Zhang. Efficient spatial keyword search in trajectory databases. *CoRR*, abs/1205.2880, 2012.
[21] G. Cong, C. S.Jensen, and D. Wu. Efficient retrieval of the top-k most relevant spatial web objects. *PVLDB*, pages 337–348, 2009.
[22] J. Fan, G. Li, L. Zhou, S. Chen, and J. Hu. SEAL: spatio-textual similarity search. *PVLDB*, 5(9):824–835, 2012.
[23] K. Feng, G. Cong, S. S. Bhowmick, W.-C. Peng, and C. Miao. Towards best region search for data exploration. In *SIGMOD*, to appear, 2016.
[24] W. Feng, C. Zhang, W. Zhang, J. Han, J. Wang, C. Aggarwal, and J. Huang. STREAMCUBE: hierarchical spatio-temporal hashtag clustering for event exploration over the Twitter stream. In *ICDE*, pages 1561–1572, 2015.
[25] L. Guo, D. Zhang, G. Li, K. Tan, and Z. Bao. Location-aware pub/sub system: When continuous moving queries meet dynamic event streams. In *SIGMOD*, pages 843–857, 2015.
[26] T. Guo, X. Cao, and G. Cong. Efficient algorithms for answering the m-closest keywords query. In *SIGMOD*, pages 405–418, 2015.
[27] R. Hariharan, B. Hore, C. Li, and S. Mehrotra. Processing spatial-keyword (SK) queries in geographic information retrieval (GIR) systems. In *SSDBM*, page 16, 2007.
[28] H. Hu, Y. Liu, G. Li, J. Feng, and K. Tan. A location-aware publish/subscribe framework for parameterized spatio-textual subscriptions. In *ICDE*, pages 711–722, 2015.
[29] I.D.Felipe, V.Hristidis, and N.Rishe. Keyword search on spatial databases. In *ICDE*, pages 656–665, 2008.
[30] H. Imai and T. Asano. Finding the connected components and a maximum clique of an intersection graph of rectangles in the plane. *Journal of Algorithms*, 4(4):310–323, 1983.
[31] A. Kalinin, U. Cetintemel, and S. Zdonik. Interactive data exploration using semantic windows. In *SIGMOD*, pages 505–516, 2014.
[32] I. Keles, S. Šaltenis, and C. S. Jensen. Synthesis of partial rankings of points of interest using crowdsourcing. In *GIR*, pages 15:1–15:10, 2015.
[33] H. A. Khan, M. A. Sharaf, and A. Albarrak. Divide: efficient diversification for interactive data exploration. In *SSDBM*, pages 15:1–15:12, 2014.
[34] A. Khodaei, C. Shahabi, and C. Li. Hybrid indexing and seamless ranking of spatial and textual features of web documents. In *DEXA (1)*, pages 450–466, 2010.
[35] G. Li, Y. Wang, T. Wang, and J. Feng. Location-aware publish/subscribe. In *KDD*, pages 802–810, 2013.
[36] X. Li, G. Cong, X. Li, T. N. Pham, and S. Krishnaswamy. Rank-GeoFM: A ranking based geographical factorization method for point of interest recommendation. In *SIGIR*, pages 433–442, 2015.
[37] C. Long, R. C. Wong, K. Wang, and A. W. Fu. Collective spatial keyword queries: a distance owner-driven approach. In *SIGMOD*, pages 689–700, 2013.
[38] J. Lu, Y. Lu, and G. Cong. Reverse spatial and textual k nearest neighbor search. In *SIGMOD*, pages 349–360, 2011.
[39] S. Luo, Y. Luo, S. Zhou, G. Cong, and J. Guan. Distributed spatial keyword querying on road networks. In *EDBT*, pages 235–246, 2014.
[40] A. Magdy, L. Alarabi, S. Al-Harthi, M. Musleh, T. M. Ghanem, S. Ghani, and M. F. Mokbel. Taghreed: a system for querying, analyzing, and visualizing geotagged microblogs. In *SIGSPATIAL*, pages 163–172, 2014.
[41] G. Marchionini. Exploratory search: From finding to understanding. *Commun. ACM*, 49(4):41–46, Apr. 2006.
[42] S. C. Nandy and B. B. Bhattacharya. A unified algorithm for finding maximum and minimum object enclosing rectangles and cuboids. *Computers & Mathematics with Applications*, 29(8):45–61, 1995.
[43] J. B. Rocha-Junior, O. Gkorgkas, S. Jonassen, and K. Nørvåg. Efficient processing of top-k spatial keyword queries. In *SSTD*, pages 205–222, 2011.
[44] J. B. Rocha-Junior and K. Nørvåg. Top-k spatial keyword queries on road networks. In *EDBT*, pages 168–179, 2012.
[45] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In *SIGSPATIAL*, pages 42–51, 2009.
[46] A. Skovsgaard and C. S. Jensen. Top-k point of interest retrieval using standard indexes. In *SIGSPATIAL*, pages 173–182, 2014.
[47] A. Skovsgaard and C. S. Jensen. Finding top-k relevant groups of spatial web objects. *VLDB J.*, 24(4):537–555, 2015.
[48] A. Skovsgaard, D. Sidlauskas, and C. S. Jensen. Scalable top-k spatio-temporal term querying. In *ICDE*, pages 148–159, 2014.
[49] Y. Tao, X. Hu, D.-W. Choi, and C.-W. Chung. Approximate maxrs in spatial databases. *PVLDB*, 6(13):1546–1557, 2013.
[50] S. Vaid, C. B. Jones, H. Joho, and M. Sanderson. Spatio-textual indexing for geographical search on the web. In *SSTD*, pages 218–235, 2005.
[51] M. Walther and M. Kaisser. Geo-spatial event detection in the Twitter stream. In *ECIR*, pages 356–367, 2013.
[52] X. Wang, Y. Zhang, W. Zhang, X. Lin, and W. Wang.

Selectivity estimation on streaming spatio-textual data using local correlations. *PVLDB*, 8(2):101–112, 2014.

[53] X. Wang, Y. Zhang, W. Zhang, X. Lin, and W. Wang. AP-Tree: Efficiently support continuous spatial-keyword queries over stream. In *ICDE*, pages 1107–1118, 2015.

[54] D. Wu, B. Choi, J. Xu, and C. S. Jensen. Authentication of moving top-k spatial keyword queries. *IEEE Trans. Knowl. Data Eng.*, 27(4):922–935, 2015.

[55] D. Wu, G. Cong, and C. S. Jensen. A framework for efficient spatial web object retrieval. *VLDB J.*, 21(6):797–822, 2012.

[56] D. Wu, M. L. Yiu, G. Cong, and C. S. Jensen. Joint top-k spatial keyword query processing. *IEEE Trans. Knowl. Data Eng.*, 24(10):1889–1903, 2012.

[57] D. Wu, M. L. Yiu, and C. S. Jensen. Moving spatial keyword queries: Formulation, methods, and analysis. *ACM Trans. Database Syst.*, 38(1):7, 2013.

[58] D. Wu, M. L. Yiu, C. S. Jensen, and G. Cong. Efficient continuously moving top-k spatial keyword query processing. In *ICDE*, pages 541–552, 2011.

[59] X. Xie, P. Jin, M. L. Yiu, J. Du, M. Yuan, and C. S. Jensen. Enabling scalable geographic service sharing with weighted imprecise voronoi cells. *IEEE Trans. Knowl. Data Eng.*, 28(2):439–453, 2016.

[60] B. Yao, M. Tang, and F. Li. Multi-approximate-keyword routing in GIS data. In *SIGSPATIAL*, pages 201–210, 2011.

[61] C. Zhang, Y. Zhang, W. Zhang, and X. Lin. Inverted linear quadtree: Efficient top k spatial keyword search. In *ICDE*, pages 901–912, 2013.

[62] C. Zhang, Y. Zhang, W. Zhang, X. Lin, M. A. Cheema, and X. Wang. Diversified spatial keyword search on road networks. In *EDBT*, pages 367–378, 2014.

[63] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsuregawa. Keyword search in spatial databases: Towards searching by document. In *ICDE*, pages 688–699, 2009.

[64] D. Zhang, B. C. Ooi, and A. K. H. Tung. Locating mapped resources in web 2.0. In *ICDE*, pages 521–532, 2010.

[65] D. Zhang, K.-L. Tan, and A. K. H. Tung. Scalable top-k spatial keyword search. In *EDBT*, pages 359–370, 2013.

[66] K. Zhao, L. Chen, and G. Cong. Topic exploration in spatio-temporal document collections. In *SIGMOD*, to appear, 2016.

[67] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma. Hybrid index structures for location-based web search. In *CIKM*, pages 155–162, 2005.