Ranked Spatial-keyword Search over Web-accessible Geotagged Data: State of the Art

Akrivi Vlachou

Abstract

Search engines, such as Google and Yahoo!, provide efficient retrieval and ranking of web pages based on queries consisting of a set of given keywords. Recent studies show that 20% of all Web queries also have location constraints, i.e., also refer to the location of a geotagged web page. An increasing number of applications support location-based keyword search, including Google Maps, Bing Maps, Yahoo! Local, and Yelp. Such applications depict points of interest on the map and combine their location with the keywords provided by the associated document(s). The posed queries consist of two conditions: a set of keywords and a spatial location. The goal is to find points of interest with these keywords close to the location. We refer to such a query as spatial-keyword query. Moreover, mobile devices nowadays are enhanced with built-in GPS receivers, which permits applications (such as search engines or yellow page services) to acquire the location of the user implicitly, and provide location-based services. For instance, Google Mobile App provides a simple search service for smartphones where the location of the user is automatically captured and employed to retrieve results relevant to her current location. As an example, a search for "pizza" results in a list of pizza restaurants nearby the user. Given the popularity of spatial-keyword queries and their wide applicability in practical scenarios, it is critical to (i) establish mechanisms for efficient processing of spatial-keyword queries, and (ii) support more expressive query formulation by means of novel

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

query types. Although studies on both keyword search and spatial queries do exist, the problem of combining the search capabilities of both simultaneously has received little attention.

1 Motivation and Problem Statement

The advent of the World Wide Web in conjunction with efficient search engines like Google and Yahoo! has made an enormous amount of information easily accessible to everybody. Search engines provide efficient retrieval and ranking of web pages based on queries consisting of a set of given keywords. Points of interest such as hotels, restaurants and tourist attractions are typically associated with a document (web page) that is geotagged, which enables the retrieval of the location of the point of interest from the web page using common information extraction techniques. An increasing number of applications support location-based keyword search, including Google Maps, Bing Maps, Yahoo! Local, and Yelp. Such applications depict points of interest on the map and combine their location with the keywords provided by the associated document(s). The posed queries consist of two conditions: a set of keywords and a spatial location. The goal is to find points of interest with these keywords close to the location. We refer to such a query as spatial-keyword query. Recent studies show that 20% of all Web queries also have location constraints [9], i.e., also refer to the location of a geotagged web page.

Moreover, mobile devices nowadays are enhanced with built-in GPS receivers, which permits applications (such as search engines or yellow page services) to acquire the location of the user implicitly, and provide location-based services. For instance, Google Mobile App provides a simple search service for smartphones where the location of the user is automatically captured and employed to retrieve results relevant to her current location. As an example, a search for "pizza" results in a list of pizza restaurants nearby the user. Twitter is another example where postings can also be geotagged, and combined with the increasing use of mobile phones for twittering the amount of information associated with locations also increases. As the num-

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

ber of mobile users increases rapidly, the need of such location-based services is expected to increase as well. More importantly, to accommodate more complex information needs of mobile users, more advanced query types are required to capture the users' intent. For example, a mobile user may be interested in retrieving cheap restaurants located in her close vicinity that serve Chinese food and are highly ranked, according to an independent ratings provider such as tripadvisor.com. The challenge is to provide to the user advanced location-based queries that rank the relevant points of interest according to different user-defined preferences. Furthermore, queries that relate to a mobile user may also involve temporal information capturing her movement.

Other applications that provide spatial-keyword search on their data include image search engines. A typical example is Flickr (http://www.flickr.com/map), where users manually annotate images using tags that provide a description of the contents and the location the image was captured. Exploiting this information, Flickr is able to support queries based on the location of images. On the other hand, modern digital cameras have integrated GPS facilities that automatically provide the exact location the photo was taken with higher accuracy than a user can provide. Even though the exact location may be known, the user still adds keywords that describe the contents of the photo, for example "sunset". This motivates even further the need of supporting spatial-keyword queries and highlights their importance in real-life applications.

Given the popularity of spatial-keyword queries and their wide applicability in practical scenarios, it is critical to (i) establish mechanisms for efficient processing of spatial-keyword queries, and (ii) support more expressive query formulation by means of novel query types. Although studies on both keyword search and spatial queries do exist, the problem of combining the search capabilities of both simultaneously has received little attention. Some existing research challenges are:

- Advanced and expressive querying mechanisms for points of interest that combine spatial information and textual relevance. To this end,

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

various novel query types complying with the paradigm of spatial-keyword search need to be introduced to cover a wide variety of user information needs.

- Novel indexing structures capable to support complex spatial-keyword queries effectively, by means of harnessing the merits of spatial data structures and text indexes.

- Efficient query processing algorithms that drastically prune the search space, capitalizing on the available indexes, and enabling ranked retrieval of results.

To illustrate a sample of the advanced query types that we target, consider the following scenario of a web-based map service that provides information about points of interest (hotels, restaurants, etc.) in the user's close vicinity. For example, a restaurant may be annotated with keywords such as: "restaurant", "Chinese", "live music", but also information about the quality of service (such as "cheap"), or ratings extracted from an independent provider such as tripadvisor.com. Furthermore, such keywords may be extracted from web pages. Typical queries that are not currently supported by web applications (such as Google maps) include:

1. "Which are the best (highest rated) restaurants serving Chinese food that are at most 5km from my current location?"

2. "Which are the cheapest hotels with wireless internet, located nearby the most famous tourist attraction in Paris?"

3. "Which are the top-3 hotels that have the best combination of restaurants and bars in their close vicinity?"

4. "Given my current route from Paris to Lyon, which is the best restaurant for dinner?"

In these queries, the words "restaurant", "hotel", "Chinese", "wireless internet", "tourist attraction" are indicative keywords that may appear in the annotations of points of interest.

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

2 Spatial Databases

Emerging applications require advanced query processing primitives that go beyond exact match queries. Such applications often need to handle multi-dimensional data, whether these dimensions are related to specific attributes of the data objects or are the result of advanced feature extraction algorithms. Querying multidimensional data is challenging even in a centralized domain. In this section, we introduce fundamental query types that are commonly used in processing multidimensional data. We first discuss similarity search based on range queries and highlight their relation to nearest neighbor queries. Top-k queries, that rank data objects based on some scoring function are also discussed. We conclude our exposition with the recently introduced skyline query, as a generalization of ranking using many different, and often conflicting criteria. We discuss skyline computation over subspaces of the data domain and its relationship to top-k queries.

2.1 Multidimensional Data Model

During the last decades, an increasing number of applications, such as medical imaging, molecular biology, multimedia and scientific databases, have emerged where a large amount of high-dimensional data points needs to be processed. In addition to exact match queries, emerging applications call for advanced query types. For instance, queries like "which data objects are most similar to a query object" or "which data objects are the best trade-off between different object's features" need to be support.

Let us assume a data collection S of n objects represented as points in a d-dimensional (feature) space D characterized by dimensions $\{d_1, ..., d_d\}$. A data object is treated as a point p in D defined via a set of coordinate values in that space: $p = \{p_1, ..., p_n\}$. Each coordinate value p_i may represent an attribute of the object that is of interest to the application, or, it may be the result of a scoring function that evaluates certain features of the object. In what follows, we assume that the points' coordinates are numerical non-negative values that depict certain features of database objects.

A two-dimensional example is shown in Figure 1. In the figure, a database

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

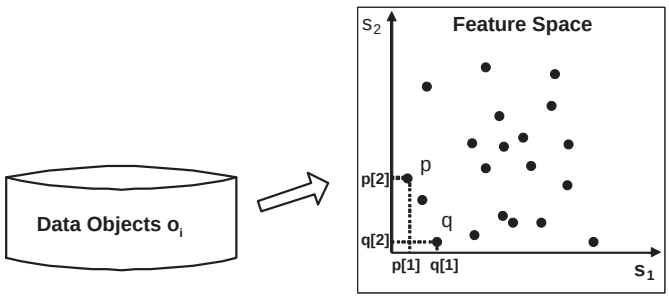


Figure 1: Feature space

of objects o_j is depicted along with a representation of the objects as points in a two-dimensional space. The coordinates of each object are calculated via two scoring functions $s_1()$ and $s_2()$. Thus, object o_j is mapped to $p{=}(p_1, p_2){=}(s_1(o_j), s_2(o_j))$. In the following, we do not distinguish between inherent attributes of the object and extracted features. We prefer to refer to its multidimensional representation and, we use the terms object and data point interchangeably.

Advanced query primitives have emerged in order to allow efficient processing of objects depicted in a high-dimensional space. Examples include similarity search based on range and nearest neighbor queries, top-k queries and the skyline operator. In the following, we shortly describe these query types.

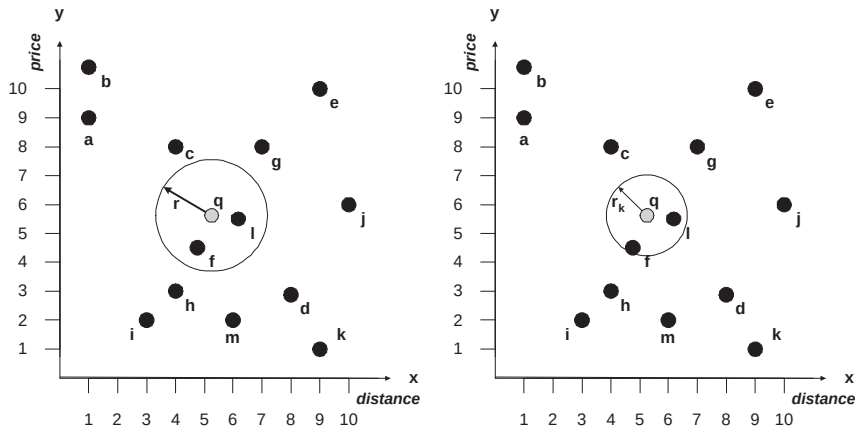
2.2 Similarity Search in Metric Spaces

Several applications, such as multimedia databases [24], employ feature transformation, which projects important features or properties of data objects into a high-dimensional space. Subsequent processing in that space often requires support for similarity search in order to retrieve similar objects. Similarity search in metric spaces focuses on supporting queries, whose purpose is to retrieve objects which are similar to a query point, when a metric distance function *dist* is used to measure the objects' (dis)similarity.

More formally, a metric space is a pair $M = (D, dist)$, where D is a domain of feature values and *dist* is a distance function with the following properties:

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.



(a) Range query (b) k-nearest neighbor search

Figure 2: Similarity search examples

1. $dist(p, q) \geq 0$ (non negativity),
2. $dist(p, q) = 0$ iff p=q (identity of indiscernibles),
3. $dist(p, q) = dist(q, p)$ (symmetry),
4. $dist(p, q) \leq dist(p, o) + dist(o, q)$ (triangle inequality).

A smaller distance between two objects is used by the application to indicate higher degree of similarity among them. Similarity is symmetric, however, because of the triangle inequality, an object may be similar to two dissimilar objects.

Similarity search in metric spaces involves, two different types of queries, namely range and nearest neighbor queries.

2.2.1 Range Query

Range queries are specified by a query object q and a range (radius) value r. The result set of the query is defined to contain all the objects o from the dataset that have a distance to the query object q that is less than or equal to radius r:

Definition 1. ***Range query*** $R(q, r)$***:*** *Given a query object* q *and a radius* r*, a point* $p \in S$ *belongs to the result set* R_q^r *of the range query iff* $dist(q, p) \leq r$.

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

A range query $R(q, r)$ can be interpreted as "retrieve all objects that are within distance r to q". Figure 2(a) depicts a range query defined by query point q and a radius r. The result set of this query contains data points f and l.

2.2.2 k-Nearest Neighbor Query

A drawback of range queries is that the cardinality of the result set is not known in advance, but can be anything between zero and the size of the database. Consequently, selection of an inappropriate value for the query range may lead to very few or too many query results. In the first case, a new range query has to be posed with a larger range, which leads to redundant processing cost. In the second case, more than necessary objects are retrieved, which again leads to increased processing cost. In practice, a good selection of a radius value r is difficult to obtain as it requires knowledge of the underlying distribution of data in the projected feature space.

The k-nearest neighbor (k-NN) [22, 13] query overcomes this problem by giving the user the ability to specify the size k of the answer set. This query type does not require a user to provide a query range and is therefore far easier to use than the similarity range query. The k-nearest neighbor query returns the k most similar (to a query point q) data points from the dataset and is defined as follows:

Definition 2. ***k-nearest neighbor query*** $NN_k(q)$***:*** *Given a query object q and a positive integer k, the result set NN_q^k of the k-nearest neighbor, is a set such that $NN_q^k \subseteq S$, $|NN_q^k| = k$ and $\forall u, v : u \in NN_q^k, v \in S - NN_q^k$ it holds that $dist(q, u) \leq dist(q, v)$.*

In the definition we assumed that the dataset contains more than k points ($n = |S| \geq k$), which is typically the case. Otherwise, the result of the query is, trivially, the set of all objects in S. Moreover, the k data objects with the smallest distance may not be unique. When more than one objects have the same distance to the query point, one or more of them may be chosen randomly in order to produce a result set containing exactly k points.

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

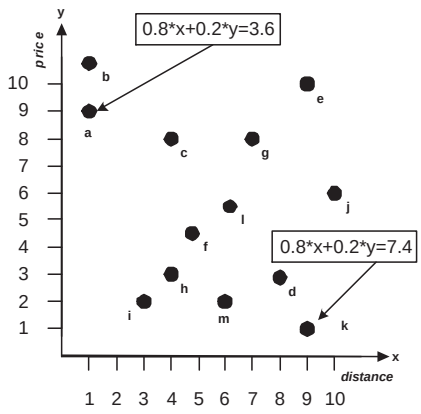


Figure 3: Top-k example

Intuitively, a k-nearest neighbor query states "retrieve the k objects in S which are closest in distance to a given object". Figure 2(b) illustrates the works of the nearest neighbor query. Given the query point q the figure depicts the results of the 2-nearest neighbor search (k=2), namely points f and l.

Given a query object q, a k-nearest neighbor query is equivalent to a range query specified by query point q and a radius equal to the distance of the k-th neighbor.

Observation 1. *Given a query object q, let r_k be the distance of the k-th nearest neighbor p, i.e., $r_k = dist(q,p)$ then $\forall r \in NN_q^k$ it holds that $r \in R_q^{r_k}$.*

Of course the range query may return a few additional objects whose distance from q is exactly r_k. Observation 1 expresses that any nearest-neighbor query can be transformed and have its results computed via a range query, if the distance to the k-th nearest neighbor was known a-priori.

2.3 Top-k Query

For decades, top-k queries were mainly studied in the information retrieval research field, aiming at ranking text documents according to some query terms, both efficiently and effectively. More recently [5, 14] the data management community has realized the benefits of top-k queries in database

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

systems and several efficient algorithms for their evaluation have emerged. Top-k queries on multidimensional datasets compute the k most interesting results with regards to a monotone score aggregation function, such as weighted aggregation, applied on the attribute values.

Definition 3. *An aggregation function f is increasingly monotone, if $\forall p, p' \in S$ with $p_i \leq p'_i$, $\forall i$, then $f(p) = f(p_1, ..., p_d) \leq f(p'_1, ..., p'_d) = f(p')$.*

The property of increasing monotonicity means that whenever the score of all dimensions of the point p is at least as good as that of another point p', then we expect that the overall score of p is as least good as p'. The result of a top-k query is the ranked list of the k objects with lowest *score* values. As in the case of k nearest neighbor queries, when the database consists of fewer than k points, the result contains the whole dataset.

Definition 4. ***Top-k query:*** *Given a positive integer k, the result set TOP_k of the top-k, is a set such that $TOP_k \subseteq S$, $|top_k| = k$ and $\forall u, v : u \in TOP_k, v \in S - TOP_k$ it holds that $f(u) \leq f(v)$, assuming that minimum values are preferable.*

A special case of monotone functions is the weighted sum function, also called linear. Each feature (dimension) d_j has an associated query-dependent weight w_j indicating the dimension's relative importance for the query. The aggregated score for object p is defined as a weighted sum of the individual scores: $score(p) = \sum_{j=1}^{d} w_j \times p_j$, where $w_j \geq 0$ $(1 \leq j \leq d)$ and $\exists j$ such that $w_j > 0$. If some weights are set equal to zero, then a top-k query refers to only to a subset of the available features. The weights indicate the user's preferences and influence the ordering of the data objects and therefore the top-k result set. For example, consider the dataset depicted in Figure 3. By assigning a high weight to values of dimension x (distance), point a is the top-1 object, while if a low weight is used, point k becomes the top-1 object.

A top-k query takes two parameters: a user specified monotone function f and the number of requested objects k. Notice that both the scoring function and the parameter k may differ for each query and we are interested in retrieving the k objects with the best (minimum) values of the scoring function. In the special case of the weighted sum, the user specifies the weighting

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

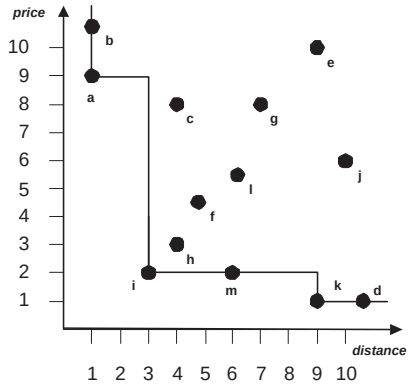


Figure 4: Skyline example

of each feature, i.e., how important this feature is based on his preferences and therefore, a top-k query takes two parameters: a d-dimensional vector $w = \{w_1, \ldots, w_d\}$ and the number of requested objects k.

2.4 Skyline Operator

Skyline queries [2] have attracted much attention recently, since they help users to make intelligent decisions over complex data, where many conflicting criteria are considered. Let us assume for example a database containing information about hotels. Each tuple of the database is represented as a point in a data space consisting of numerous dimensions. In our example, the y-dimension represents the price of a room, whereas the x-dimension captures the distance of the hotel to a point of interest such as the beach (Figure 4). According to the dominance definition, a hotel dominates another hotel because it is cheaper and closer to the beach. Thus, the skyline points, in the example points a, i and k, are the best possible trade-offs between price and distance from the beach.

In the following, we define the skyline and subspace skyline queries and point out their relation to top-k queries.

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

2.4.1 Skyline queries

Definition 5. ***Skyline:*** *A point $p \in S$ is said to* dominate *another point $q \in S$, denoted as $p \prec q$, if (1) on every dimension $d_i \in D$, $p_i \leq q_i$; and (2) on at least one dimension $d_j \in D$, $p_j < q_j$. The* skyline *is a set of points $SKY \subseteq S$ which are not dominated by any other point. The points in SKY are called skyline points.*

Without loss of generality, we assume that skylines are computed with respect to min conditions on all dimensions and that all values are non-negative.

The cardinality of the skyline set SKY depends on the data distribution, the dimensionality and the cardinality of the dataset. It has been shown [6, 11] that the expected number of skyline points is $\Theta(\ln^{d-1} n/(d-1)!)$ for a random dataset. The result suggests that the skyline cardinality increases with the dataset dimensionality. The intuition is that as the number of dimensions increases, it is more likely for any point p that there exists another point q, where p and q are better than each other in different subsets of dimensions. In other words, the probability of one point dominating another point in the full space is decreasing as the dimensionality increases. Therefore, the cardinality of the skyline set increases rapidly with the dimensionality of the dataset.

2.4.2 Subspace skyline queries

Applications often provide numerous candidate attributes that they can use for data analysis. In our running example, the hotel database could contain numerous other attributes, such as the number of rooms, the age of the hotel, the size of room, the star rating, etc. The notion of skyline can be extended to subspaces, where given a set of d-dimensional objects, a subspace skyline query only refers to a user-defined subset of attributes. Each non-empty subset U of D ($U \subseteq D$) is referred to as a *subspace* of D. The data space D is also referred as full space of the dataset S.

Definition 6. ***Subspace Skyline:*** *A point $p \in S$ is said to* dominate *another point $q \in S$ on subspace $U \subseteq D$, denoted as $p \prec_U q$, if (1) on every*

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

dimension $d_i \in U$, $p_i \leq q_i$; *and (2) on at least one dimension* $d_j \in U$, $p_j < q_j$. *The* skyline *of a subspace* $U \subseteq D$ *is a set of points* $SKY_U \subseteq S$ *which are not dominated by any other point on subspace* U. *The points in* SKY_U *are called* skyline points *on subspace* U.

Consider for example the dataset depicted in Figure 4. The skyline points are $SKY = \{a, i, k\}$, while for the (non-empty) subspace $U = \{x\}$ the skyline points on U are $SKY_U = \{a, b\}$. Notice that point b is a skyline point on the subspace $\{x\}$ but it is dominated by point a in the full space $\{x, y\}$.

Observation 2. *A skyline point* $p \in SKY_U$ *on a subspace* $U \subseteq D$ *is either a skyline point on* D, *or is dominated on* D *by another skyline point* $q \in SKY_U$, *for which* $p_i = q_i$, $\forall i : d_i \in U$.

2.4.3 Relation to top-k queries

Skyline queries relate to top-k queries, and can be used to discard points that cannot belong to the top-k result set.

Observation 3. *The top-*1 *object for any increasingly monotone aggregation function belongs to the skyline set.*

Proof: Consider a point q that does not belong to the skyline, but it is the top-1 for a query defined by an increasingly monotone function f. Then, there exists another point p that dominates q, i.e., on each dimension $d_i \in D$, $p_i \leq q_i$; and on at least one dimension $d_j \in D$, $p_j < q_j$, hence $f(p) < f(q)$, and since f is increasingly monotone this leads to a contradiction, because q is the top-1, i.e., $f(q) < f(p)$. Thus, the top-1 object for any increasingly monotone function belongs to the skyline.

For example, consider the dataset depicted in Figure 3. By assigning a high weight to the score of attribute x, point a is the top-1 object, while if a low weight is used, point k becomes the top-1 object. Both a and k belong to the skyline set. This observation can be adopted for efficient top-k evaluation, by using the notion of the k-skyband operator [18]. The result set of any top-k query is a subset of the k'-skyband set, with $k \leq k'$.

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

3 Text Retrieval

A document d is represented by a vector. Each dimension corresponds to a distinct term in the document. The value $d(t)$ of a term t in the vector is computed by a language model [19]. A commonly used model is described by the following equation:

$$d(t) = (1 - \lambda)\frac{tf(t,d)}{|d|} + \lambda\frac{tf(t,C)}{|C|}$$

where $tf(t,d)$ is the term frequency (number of occurrences) of term t in document d, $tf(t,C)$ is the term frequency of term t in the entire document collection C, and λ is the smoothing parameter.

Definition 7. *A **keyword query** is defined by a set of keywords $w_1, \ldots, w_m$. The result is a list of objects ordered by the relevance of their textual descriptions to the query keywords, as measured by an IR ranking function [19].*

Definition 8. *A special case is the **Boolean keyword query** which returns the set of all objects whose text document contains all of $w_1, \ldots, w_m$.*

Approximate string retrieval has been also studied in the related literature. The main goal is to find strings that match the given keyword approximately and is important in the case that users have spelling errors when they type the queries. Also, approximate string retrieval is necessary for type ahead search. In order to measure the similarity of two string usually the edit distance is applied. The edit distance of two strings is defined as the minimum number of changes in spelling (insertion, deletion, or substitution) required to change one string into another. In following we define three commonly used predicates used for approximated string retrieval.

Definition 9. ***Prefix match*** *A string s is a prefix match of t, if s is a prefix of t.*

Definition 10. ***Substring match*** *A string s is a substring match of t, if s is a substring of t.*

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

Symbols	Description
O	Database
$\|O\|$	Cardinality
o_i	Data object
p_i	Multi-dimensional data point of o_i
d_i	Textual description of o_i
Q	Spatial keyword query
k	Number of requested results
q	Multi-dimensional query point of Q
$\{w_1, \ldots, w_m\}$	Set of keywords of Q
f	Ranking function that combines the spatial and textual relevance

Table 1: Overview of symbols.

Definition 11. ***Approximate match*** *Given a threshold τ, a string s is an approximation of t if the edit distance between s and t is less than τ.*

4 Spatial-Keyword Search

In this report, we assume a database O that stores $|O|$ geo-tagged data objects o_i. Each object o_i relates to a multi-dimensional data point p_i and document d_i. The point p_i is a location descriptor in the multidimensional space, while the document d_i captures the textual description of objects o_i

Figure 5 depicts an example of a database storing information of hotels. In this example [10], each object o_i is associated with a 2-dimensional points p_i that consists of the latitude and the longitude of the hotel and describes its location. The textual information d_i of hotel o_i is the is the concatenation of the *name* and *amenities* attributes.

As mentioned before a nearest neighbor query searches through the multidimensional space to find the k nearest objects to the specified query point q. Then, the spatial objects are ranked by distance such that an object closer to q has a higher rank. Keyword search enables retrieving data objects based on textual information. Spatial-keyword search combines both approaches allowing users to retrieve data objects based on their spatial and textual

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

	Name	Latitude	Longitude	Amenities
H_1	Hotel A	25.4	-80.1	tennis court, gift shop, spa, Internet
H_2	Hotel B	47.3	-122.2	wireless Internet, pool, golf course
H_3	Hotel C	35.5	139.4	spa, continental suites, pool
H_4	Hotel D	39.5	116.2	sauna, pool, conference rooms
H_5	Hotel E	51.3	-0.5	dry cleaning, free lunch, pets
H_6	Hotel F	40.4	-73.5	safe box, concierge, internet, pets
H_7	Hotel G	-33.2	-70.4	Internet, airport transportation, pool
H_8	Hotel H	-41.1	174.4	wake up service, no pets, pool

Figure 5: Example of sample database of hotels [10]

information.

Definition 12. *A **top-k spatial keyword query** Q is a combination of a top-k spatial query and a keyword query. In particular, Q is defined by a number k of requested results, a multidimensional point q, a set of keywords $\{w_1, \ldots, w_m\}$, and a ranking function that combines the spatial and textual relevance to an overall score. The result of Q is a list of the top-k objects ranked according to the ranking function.*

In [12], the spatial-keyword query is defined as a combination of range queries and boolean keyword search. We refer to this query type as *boolean-range spatial keyword queries.*

Definition 13. *A* boolean-range spatial keyword queries *[12] is defined by a spatial part specified as a minimum bounding rectangle (MBR) and a set of keywords. This query applies the AND semantics and is defined as the one in which all the keywords are required to be present in the retrieved records and the locations of all retrieved records should fall in the given MBR.*

Definition 14. *A special case is the **distance-first top-k spatial keyword query** [10], which returns a ranked list of the k objects that contain all of $w_1, \ldots, w_m$ and are closest to q. That is, distance-first top-k spatial keyword query is a combination of a top-k spatial query and a Boolean keyword query.*

In the example depicted in Figure 5, assuming the query point $q = [30.5, 100.0]$, the hotel $H4$ is ranked first if only the spatial information is

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

considered. Considering a boolean keyword query with keywords $\{internet, pool\}$, the result set contains two hotels namely $H2$ and $H7$. The result of a distance-first top-k spatial query with $k = 2$, $q = [30.5, 100.0]$ and $\{w_i\} = \{internet, pool\}$ is the ranked list $\{H7, H2\}$.

5 Indexing techniques

The problem of indexing location information as well as text information was studied in [30]. The main challenge in combing spatial and textual indexes is that the location information is two-dimensional and in Euclidean space, while the the index of conventional text search is set-oriented. The authors proposed to use a hybrid index structure, which integrates inverted files and $R*$-trees, to handle both textual and location aware queries. Three different combining schemes were studied: (1) inverted file and R*-tree double index, (2) first inverted file then R*-tree, (3) first R*-tree then inverted file.

In [12], the authors define and study boolean-range keyword queries. The authors study the performance bottlenecks of different indexing mechanisms and develop a novel indexing structure called $KR*$-tree that captures the joint distribution of keywords in space and significantly improves performance over other index structures.

In [10], the problem of top-k spatial keyword search is studied. The authors focus on boolean keyword search leading to the definition of *distance-first top-k spatial keyword query*. The IR^2-Tree is proposed, which is an efficient indexing structure that stores spatial and textual information for a set of objects. The IR^2-Tree is a combination of an R-Tree and signature files and each node of an IR^2-Tree contains both spatial and keyword information. The spatial information is summarized by using minimum bounding boxes, while the textual information is stored by using signatures. The authors propose an efficient incremental algorithm is presented to answer top-k spatial keyword queries using the IR^2-Tree and show that their approach performs efficiently.

Two different [9, 16] indexing approaches have been proposed that employ a hybrid index that augments the nodes of an R-tree with inverted indexes.

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

The inverted index at each node refers to a pseudo-document that represents all the objects under the node. During query processing, the index is exploit based on the spatial information and in order to verify that a node is relevant to query keywords, the inverted index at each node is retrieved. The score that derives from the vector that represents the pseudo-document of a node is an upper bound for the textual relevance of any document index by this node. Therefore, the nodes can be rank and accessed based on their spatial and textual relevance and a top-k query can be processed efficiently. Another hybrid indexing structure that combines the R*-tree and bitmap indexing to process the spatial-keyword query was proposed in [29].

The *Spatial Inverted Index (S2I)* was proposed in [20] for processing top-k spatial keyword queries more efficiently. The S2I index maps each keyword to a distinct aggregated R-tree that stores the objects with the given term. The aggregated R-tree stores the latitude and longitude of the objects, and maintains an aggregated value that represents the maximum term impact of the objects under the node. In fact, the aggregated R-tree is employed only when the number of objects exceeds a given threshold. As long as the threshold is not exceeded, the objects are stored in a file, one block per term.

The authors in [7, 8] notice that there is significantly more textual than spatial data in current search engines and claim that it is important to focus on the textual aspect of the problem. In order to verify experimentally this claim comparing three basic indexing approaches are compared: 1) an multi-dimensional index in which an inverted index is maintained at each leaf, indexing all documents within the leaf's MBR, 2) R*-Tree that assumes an oracle that can prune unproductive subtrees, and thus supersedes many of the optimizations in the literature and 3)a brute-force Text-First baseline first determines all textually relevant documents using a state-of-the-art inverted index implementation. The experimental evaluation shows that state-of-the-art query processing techniques for text data are quite efficient. Thus, it is important to consider approaches that preserve and exploit these techniques.

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

6 Advanced query types

Reverse Spatial Textual k Nearest Neighbor (RSTkNN) search is defined in [17]. The aim of RSTkNN queries is to find the objects that take the query object as one of their k most spatial-textual similar objects. The authors design a hybrid index tree called IUR-tree (Intersection-Union R-Tree) and employ a branch-and-bound search algorithm to process RSTkNN queries efficiently.

The problem of retrieving a group of spatial web objects such that the groups keywords cover the query's keywords and such that the objects are the nearest to the query location and have the lowest inter-object distances was studied in [4]. This problem is NP-complete and can be solved efficiently by approximate algorithms.

In [3] the concept of prestige-based relevance to capture both the textual relevance of an object to a query and the effects of nearby objects is proposed. Therefore, a new type of query, the Location-aware top-k Prestige-based Text retrieval (LkPT) query, is proposed that retrieves the top-k spatial web objects ranked according to both prestige-based relevance and location proximity.

A top-k spatial keyword query returns the k best objects ranked in terms of both distance to the query location and textual relevance to the query keywords. Most of the approaches assume Euclidean space. However, for most real applications, the distance between the objects and query location is constrained by a road network (shortest path). In [21], the problem of processing top-k spatial keyword queries on road networks (where the distance between the query location and the spatial object is the shortest path) is studied. Wu et al. [27] study the problem of keeping the result set of spatial-keyword queries up-to-date, while the user is moving on a road network.

Moreover, [26] joint processing of multiple top-k spatial keyword queries, which is important for the query response time during high query loads. Finally, effective caching of shortest paths for location-based services is studied in [25]

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

7 Approximate string search

Approximate-keyword queries are studied in [1]. An index structure called LBAKtree that combines approximate indexes with a tree-based spatial index is proposed, so that location-based approximate-keyword queries can be processed efficiently. In [28], the MHR-tree is designed for approximate string search in spatial databases. The MHR-tree is based on the R-tree augmented with the min-wise signature and the linear hashing technique.

Users that use mobile devices often use services like yellow page to to find businesses using keywords near their current location. Typing the entire query is cumbersome and prone to errors, especially from mobile phones. In [23], the authors address this problem by introducing type-ahead search functionality on spatial databases. Like keyword search on spatial data, type-ahead search needs to be location-aware, i.e., with every letter being typed, it needs to return spatial objects whose names (or descriptions) are valid completions of the query string typed so far, and which rank highest in terms of proximity to the user's location and other static scores. An efficient method for *location-aware type-ahead search* is proposed by using a formal model for query processing cost and developing techniques that optimize that cost.

A similar problem is studied in [15]. The textual information is a set of keywords and the location-based instant search returns those documents that contain all keywords and have a keyword that satisfy the prefix relation with the keyword that is currently typed. Furthermore, the approach proposed in [15] support both range queries and nearest-neighbor queries for the spatial part of the query.

8 Summary

With the popularization of geotagging an increasing number of applications have flourished that provide location-based services with some primitive form of keyword search. This report overviews the existing techniques related to spatial-keyword search. Several research challenges associated with the effi-

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

cient support of spatial-keyword search are still not address by the existing approaches. Two main factors are critical for location-based Web search engines, thus determining their overall performance: the efficiency of query processing for spatial-keyword queries and the quality of the retrieved points of interest. The efficiency of query processing directly influences the query throughput, which is very important in the context of Web applications, since the aim is to serve many concurrent user requests with minimum latency. On the other hand, the quality of the result is an equally important parameter that influences the user satisfaction from the search engine. High quality of service requires combination of techniques used in keyword search (for example approximate keyword search) and spatial queries (for example advanced spatial query types).

References

[1] S. Alsubaiee, A. Behm, and C. Li. Supporting location-based approximate-keyword queries. In *GIS*, pages 61–70, 2010.

[2] S. Borzsonyi, D. Kossmann, and K. Stocker. The skyline operator. In *Proceedings of IEEE International Conference on Data Engineering (ICDE)*, pages 421–430, 2001.

[3] X. Cao, G. Cong, and C. S. Jensen. Retrieving top-k prestige-based relevant spatial web objects. *PVLDB*, 3(1):373–384, 2010.

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

[5] M. J. Carey and D. Kossmann. On saying "enough already!" in sql. In *Proceedings of International Conference on Management of Data (SIGMOD)*, pages 219–230, 1997.

[6] S. Chaudhuri, N. N. Dalvi, and R. Kaushik. Robust cardinality and cost estimation for skyline operator. In *Proceedings of IEEE International Conference on Data Engineering (ICDE)*, page 64, 2006.

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

[7] Y.-Y. Chen, T. Suel, and A. Markowetz. Efficient query processing in geographic web search engines. In *SIGMOD Conference*, pages 277–288, 2006.

[8] 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.

[9] G. Cong, C. S. Jensen, and D. Wu. Efficient retrieval of the top-k most relevant spatial web objects. *PVLDB*, 2(1):337–348, 2009.

[10] I. D. Felipe, V. Hristidis, and N. Rishe. Keyword search on spatial databases. In *ICDE*, pages 656–665, 2008.

[11] P. Godfrey. Skyline cardinality for relational processing. In *Proceedings of FoIKS*, pages 78–97, 2004.

[12] 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.

[13] G. R. Hjaltason and H. Samet. Distance browsing in spatial databases. *ACM Transactions on Database Systems (TODS)*, 24(2):265–318, 1999.

[14] I. F. Ilyas, G. Beskales, and M. A. Soliman. A survey of top-k query processing techniques in relational database systems. *ACM Computing Surveys (CSUR)*, 40(4):1–58, 2008.

[15] S. Ji and C. Li. Location-based instant search. In *SSDBM*, pages 17–36, 2011.

[16] Z. Li, K. C. K. Lee, B. Zheng, W.-C. Lee, D. L. Lee, and X. Wang. Ir-tree: An efficient index for geographic document search. *IEEE Trans. Knowl. Data Eng.*, 23(4):585–599, 2011.

[17] J. Lu, Y. Lu, and G. Cong. Reverse spatial and textual k nearest neighbor search. In *SIGMOD Conference*, pages 349–360, 2011.

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

[18] D. Papadias, Y. Tao, G. Fu, and B. Seeger. Progressive skyline computation in database systems. *ACM Transactions on Database Systems*, 30(1):41–82, 2005.

[19] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *SIGIR*, pages 275–281, 1998.

[20] 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.

[21] J. B. Rocha-Junior and K. Nørvåg. Top-k spatial keyword queries on road networks. In *EDBT*, pages 168–179, 2012.

[22] N. Roussopoulos, S. Kelley, and F. Vincent. Nearest neighbor queries. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, pages 71–79, 1995.

[23] S. B. Roy and K. Chakrabarti. Location-aware type ahead search on spatial databases: semantics and efficiency. In *SIGMOD Conference*, pages 361–372, 2011.

[24] T. Seidl and H.-P. Kriegel. Efficient user-adaptable similarity search in large multimedia databases. In *Proceedings of International Conference on Very Large Data Bases (VLDB)*, pages 506–515, 1997.

[25] J. R. Thomsen, M. L. Yiu, and C. S. Jensen. Effective caching of shortest paths for location-based services. In *SIGMOD Conference*, pages 313–324, 2012.

[26] 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.

[27] 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.

European Union
European Social Fund
OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

[28] B. Yao, F. Li, M. Hadjieleftheriou, and K. Hou. Approximate string search in spatial databases. In *ICDE*, pages 545–556, 2010.

[29] 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.

[30] 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.

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND
Co- financed by Greece and the European Union

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.