

Chapter 37

Spatial Search



Liping Di and Eugene G. Yu

Abstract Urban studies concern the evolution of spatial structure in cities, where information is often tied to location. The discovery of information is in a high-dimensional space based on spatial and temporal dimensions, where the spatial relationships of components play roles in studying urban evolution. Spatial search in urban studies has to deal with diverse aspects of data structures (structured versus unstructured), data spatial context (implicit versus explicit), data spatial relationships (containment versus intersection), data volume (large volume versus large variety), spatial search speed (speed against different requirements), and spatial search accuracy (exactness versus relevance). This chapter reviews the technology in mining and extracting spatial information into urban geographic information systems, spatially indexing the urban information for effective spatially aware search, spatial relationships and their search algorithms, improving spatial relevance with different spatial similarity measures and algorithms, and open standards and interoperability in spatial search in the Web environment. Emerging technologies for spatial search in urban studies are also reviewed. Applications of spatial search in urban studies are exemplified and evaluated.

37.1 Spatial Search in the Context of Urban Studies

Urban studies is a transdisciplinary field that encompasses different academic fields, including urban geography, urban sociology, urban economics, urban housing and neighborhood development, urban environmental studies, urban governance, politics and administration, urban planning, design, and architecture (Bowen et al. 2010; Harris and Smith 2011). Search is ubiquitous in these focused research areas (Ballatore et al. 2016). In its most general form, spatial search is the search for information in a spatial and temporal context (Miller 1992). The introduction of the spatial dimension in the search problem can be viewed from two perspectives: one is as part of the information sought (i.e. the search for a place) and the other is as

L. Di (✉) · E. G. Yu

Center for Spatial Information Science and Systems, George Mason University, Fairfax, USA
e-mail: ldi@gmu.edu

© The Author(s) 2021

W. Shi et al. (eds.), *Urban Informatics*, The Urban Book Series,
https://doi.org/10.1007/978-981-15-8983-6_37

683

the context in which the search is carried out (e.g. the network of roads to be routed through with an optimal route; Miller 1992).

Spatial search in urban studies carries different connotations depending on the root subject and the application. In the context of technology and geoinformatics, spatial search includes spaceless point search, range search, k -nearest neighbor search, and aggregated spatial search (e.g. total area or total count). In economics and sociology, spatial search can be seen as a decision problem and behavior. The spatial search problem is formatted as a connected graph with physical dimensions (e.g. two-dimensional space). The spatial search problem can vary with options (e.g. perfect knowledge with fixed sample set, online without recall, online with recall, with imperfect information). In the environment of linked open data (LOD), spatial search can be described as a process of identifying the place (converting into geographic information), modeling the spatial dimensions, indexing spatially for improved performance or heuristic results, formulating the search problem, and searching for results in constrained cases.

Spatial search in urban studies involves the following components to manage and maintain a spatial information system:

- Geocoding: a process to parse and extract spatial references from a query request.
- Spatial indexing: a process to improve the performance of spatial information retrieval.
- Spatial search algorithms: a set of algorithms to achieve the efficient and effective discovery of spatial information for different applications.
- Catalog and federated catalog: a system to manage spatial metadata.

The chapter is organized as follows. The next section reviews the geocoding process. Information about popular geocoding approaches and tools is introduced in this section. This is followed by a review of the approaches and data structures used in indexing the spatial information. The third section describes the spatial search problem as expressed in computer algorithms, while the fourth section reviews the cataloging strategies of spatial data and their approaches in distributed environments. The final section briefly touches on some of the recent advances and research directions in spatial search.

37.2 Geocoding

In urban studies, place names and street addresses are commonly used in referencing data geospatially (Dueker 1974). Geocoding is the step to relate location to descriptive text or place names. In early literature, it was termed place naming (Dueker 1974; Tobler 1972). In urban areas, geocoding can be efficiently referenced using different approaches for different datasets. Street geocoding, parcel geocoding, and address-point geocoding are three of the commonly used approaches in geocoding to associate an address with spatial coordinates (Zandbergen 2008; Owusu et al. 2017). As more and more types of geocode have emerged, the levels of detail can

be associated with geocodes at different granularities. Table 37.1 shows the major generations of geocoding technologies along with major software or services for the corresponding generation. Geocoding has evolved along with the development of geographic information systems (GIS). At the beginning of GIS development, in the 1960s, the simplest geocoding schemes and systems became available. Geocoded area units could be matched to a representative point. Because these geocodes (e.g. demographic information, economic metrics) can associate with many attributes, they can be used effectively as base areal units for analyzing spatial differentiation in urban areas.

In the Web environment or connected applications, the approach is to use the API provided by geocoding services. All these services support both geocoding and reverse geocoding. The responses of these APIs are mostly in JSON, which can be easily incorporated and used by JavaScript in the Web environment (Table 37.2).

A place name may evolve over time, and sometimes, a place may carry multiple alternative names. In such cases, a gazetteer (a searchable database of toponyms) is useful and may be adapted to provide specific geocoding assistance. A gazetteer also contains basic information about the place in addition to geographic coordinates. This basic information may include demographic statistics, physical features, literacy, and economic conditions. The NGA GEONet Names Server (GNS) is one of the sources

Table 37.1 Brief history of geocoding development

Generation	Geocoding technologies	Representative system or service
1960s	City block codes; street segments; representative point; Address Coding Guides (ACG) (Dueker 1974)	Automatic Location Table (AULT; Dueker 1974) Street Address Conversion System (SACS; Dueker 1974)
1970s	Dual Independent Map Encoding (DIME) (Farnsworth and Curry 1970)	Address Matching System (ADMATCH), Geographic Base File System (DIME), Computer Mapping System (GRIDS) (Farnsworth and Curry 1970)
1980s	Geographic Base File (GBF) (Davis et al. 1992)	GBF/DIME (Davis et al. 1970)
1990s	Topologically Integrated Geographic Encoding and Referencing (TIGER) (Broome and Meixler 1990)	
2000s	Commercial geocoding scheme Multilevel geocoding (Zandbergen 2008; Goldberg 2017) ADDRESS-POINT™ (Mesev 2005) Geocoded National Address File (G-NAF) (Paull 2003) Open Street Map (OSM)	Commercial software and services (Goldberg et al. 2007)
2010s	Master Address File (MAF) (Trainor 2003)	MAF/TIGER (Galdi 2005; Trainor 2005) Commercial geocoding Application Programming Interface (API) (Panasyuk et al. 2019)

Table 37.2 List of selected geocoding web services

Name	Limitation for service	Reference data	Reference or endpoint
Google geocoding API	50 requests per second; free credit \$200 each month	Google maps	https://maps.googleapis.com/maps/api/geocode/json
Bing Locations API	Maximum 2 jobs at the same time. 50 jobs/24 h. 5 data sources and 2500 entities per source	Bing maps (NAVTEQ)	https://docs.microsoft.com/en-us/bingmaps/rest-services/locations
Yahoo geocoding API	5000 queries per IP address per day	Yahoo maps (NAVTEQ)	https://local.yahooapis.com/MapsService/V1/geocode
Baidu geocoding/reverse geocoding API	One million times/day	Baidu maps	https://api.map.baidu.com/telematics/v3/geocoding https://api.map.baidu.com/telematics/v3/reverseGeocoding
Yandex geocoder API	25,000 total requests per day to the geocoder, router, and panorama service combined	Yandex maps (NAVTEQ)	https://tech.yandex.com/maps/geocoder
Gaode geocoder API	250 query requests per day (API calls)	Gaode map	https://lbs.amap.com/api/javascript-api/guide/services/geocoder
Nominatim	1 request per second	OpenStreetMap (OSM)	https://nominatim.openstreetmap.org/search https://nominatim.openstreetmap.org/reverse
Texas A&M Geoservices Geocoder	2500 queries	Combined resources	https://geoservices.tamu.edu/Services/Geocode

used in these services. These services from gazetteers have been found very useful in urban studies (Janowicz et al. 2019; Dimou and Schaffar 2009). Table 37.3 lists a few of the most widely used gazetteers for retrieving geographic dimensions or coordinates of a place name and basic information about the place. The capabilities of gazetteers in disambiguating place names and putting place in context have led to many applications in the semantic analytics of urban studies (Janowicz et al. 2019).

Table 37.3 List of selected gazetteers

Name	Coverage/reference data	Usage	Limits	URL
Getty thesaurus of geographic names	Worldwide	Find the modern analog	Not all records have geographic coordinates	https://www.getty.edu/research/tools/vocabularies/tgn/index.html
Geonames.org	Worldwide	Geocoding/reverse-geocoding	2000/hour and 30,000/day	https://geonames.org/
GEOnet names server (GNS)	Worldwide	Official spelling, map, info, OGC services	Maximum number of features 8500	https://geonames.nga.mil/gns/html/gns_services.html
JRC Fuzzy Gazetteer	Worldwide, 7 million place names (from GNS)	Spelling correction, old names	Browser-based	https://dma.jrc.it/services/fuzzyg
Falling Rain Global Gazetteer	Worldwide, directories	Easy to follow through country by country	Browser-based	https://www.fallingrain.com/world/index.html

37.3 Spatial Indexing

Spatial indexing is the process of creating an effective and efficient data structure to help in speeding up spatial queries. Spatial indexing differs from common database indexing in having spatial properties: the object is not just one value but has two or more dimensions, and the size of an object may be non-zero (that is, a line, area, or volume; Kriegel and Seeger 1988). These properties lead to spatial relationships that are more complex than simple linear relationships. Many spatial indexing schemes have been developed along with the development of computer technologies (Kriegel and Seeger 1988; Lu and Ooi 1993). The basic goal of such spatial indexing is to reduce the computation required to retrieve matched spatial objects, given a set of geometrical criteria.

To create a spatial index, it is first necessary to identify the features to be indexed. For example, in a 2D spatial world, geographic features are commonly expressed as points, lines, or areas. Points can be represented as a pair of coordinates, which can be treated as fields to be indexed in a spatial database. Most spatial indexing approaches are specially designed for points (Lu and Ooi 1993). Lines and areas cannot be represented accurately as fields fit for indexing in a spatial database without losing information. Representative features need to be either selected or extracted for complex geographic objects. The processes are analogous to feature selection and feature extraction in machine learning, statistics, and information theory. In other words, the selection of features does not change the values which can be interpreted as dimensions. For example, the minimum bounding rectangle (MBR), the two-dimensional case of the minimum bounding box, can be treated as a selected feature, since its value can be found in the array of coordinates representing the geographic object. Any selected coordinate from the represented arrays (e.g. start point, end point, or middle point) can also be selected as the basis of indexing. The process can be generalized as one of transforming a k -dimensional space to a $2k$ -dimensional space as described by Kriegel and Seeger (1988). For example, a rectangle aligned with the axes in 2D space can be defined by four coordinates. One encoding can be the corner coordinates (either upper left coordinate plus lower right coordinate or lower left coordinate plus upper right coordinate) or the center coordinates plus extent distances to each side (Kriegel and Seeger 1988). The grid file could be a four-dimensional grid, with the rectangle snapped to the closest cell in the grid file. On the other hand, the extraction of features goes through a computerized process to compute a set of values from the objects. For example, a hashing value is computed from the object using a hashing function. A centroid can also be computed from the object. The object can be represented as the first n principal components using principal-component extraction algorithms. These derived features can be used as indexed fields in a spatial database.

The next question for spatial indexing is how to handle the overlapping of spatial objects defined by the indexing spatial feature. Two schemes are available to deal with the partition: a clipping scheme (C-scheme) and a bounding scheme (OR-scheme) (Kriegel and Seeger 1988). For example, when an MBR is used as the spatial feature,

the coverage defined by one MBR may overlap with that of another MBR. One example is shown in Fig. 37.1. With the clipping scheme, the object is duplicated with both partitions when the partition line crosses the region. For example, Object R3 is duplicated in both partitions (Fig. 37.1a). With the OR-scheme, Object R3 is only included in one partition S1 (Fig. 37.1b). The advantages and disadvantages of the two schemes are described in Table 37.4.

The computerized data structures for spatial indexing are as follows:

- **Fixed grid index:** The simplest example is *uniform grid* scheme where the space is partitioned uniformly into regular grids by value ranges along each axis. The grid system can be predefined with specified intervals or units. Retrieval time for the closest spatial rectangle would be $O(1)$, and on average for any spatial

Fig. 37.1 Partition scheme for overlapping regions

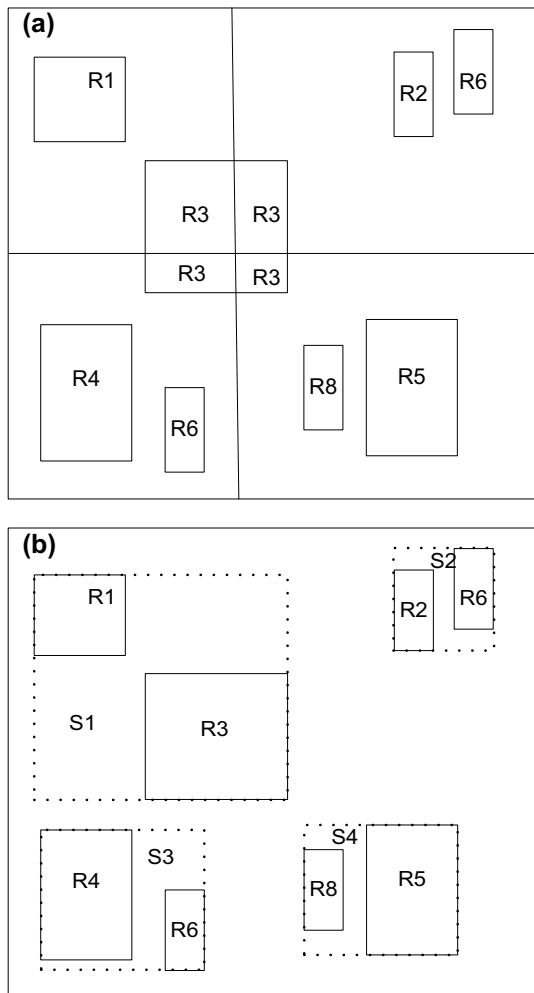


Table 37.4 Schemes for overlapping regions in a partition

Scheme	Pros	Cons
OR-scheme	Efficient storage utilization One file hosts both points and rectangles	Increased time to search, insert, or delete due to high overlap
C-scheme	Efficient inheritance of underlying point access methods One file hosts both points and rectangles	Duplications of MBRs Information redundancy

rectangle would be $O(nCells + n)$, where $nCells$ is the number of grid cells and n is the number of spatial objects, that is, the rectangles in the example. The memory requirement is $O(nCells + n)$.

- **Spatial hashing:** Because the distribution of spatial objects is often sparse, a uniform grid would result in many empty cells. A hash table can be used to store the index, and multi-level multi-key grid files can be used to index the multi-dimensional spatial data (Bentley and Friedman 1979).
- Spatial data partitioning trees
 - **Binary space partitioning (BSP) tree:** This is a general partition approach to partition space recursively into two convex sets using a hyperplane. It was developed as a general method in 3D video image processing (Schumacher et al. 1969). The k -dimensional binary search tree (k - d tree) is constructed by using one axis to split data at the median of the points along the axis (Bentley 1975). The Local Split Decision tree (LSD tree) is designed to handle both points and intervals (Henrich et al. 1989). The K-D-B tree is a derived tree structure that combines properties from the k - d tree and the B-tree (balanced tree) (Robinson 1981).
 - **Quad tree:** A quad tree builds a hierarchical representation of spatial data by dividing recursively into four quadrants (Finkel and Bentley 1974).
 - **Octree:** An octree is a hierarchical data structure that extends the quadtree to 3D, with all internal nodes having eight children (Meagher 1980).
 - **Balltree:** A balltree is “a complete binary tree in which a ball is associated with each node in such a way that an interior node’s ball is the smallest which contains the balls of its children” (Omohundro 1989).
 - **R-tree:** An R-tree uses a minimum bounding rectangle (MBR) to determine its children (Guttman 1984). It is a balanced tree. Its variant trees include the Hilbert R (Kamel and Faloutsos 1984), R + (Sellis et al. 1984), Priority R (Arge et al. 2008), R* (Beckmann et al. 1990), GiST (Hellerstein et al. 1995), and G-tree (Zhong et al. 2015).
 - **Metric tree:** The vantage-point tree (vp-tree) is a space-partitioning algorithm to construct a tree with a sphere-like bounding area to partition the metric space (Yianilos 1993). Each part is defined within a threshold to each vantage point. A multi-vantage-point tree (MVP tree) is a variant of vp-tree which uses more than one point to partition at each level (Bozkaya and Ozsoyoglu 1999). The cover tree algorithms construct a leveled tree where each parent covers the

extent of all children (Begelzimer et al. 2006). The Burkhard-and-Keller tree (BK-tree) is adapted to discrete space by arranging points that are close to each other (Burkhard and Keller 1973).

37.4 Search Algorithms

A spatial search in urban studies can be viewed from different perspectives and formulated differently for different subject domains. In this section, two perspectives are examined. First, from the perspective of geography, spatial search is treated as a technology and method, and typical spatial queries and corresponding search algorithms are reviewed. Second, from the perspective of urban economics and urban sociology, spatial search is treated as a form of decision-making, generalized spatial search is formulated with graph theory, and related search algorithms are reviewed.

37.4.1 Spatial Queries

The following are the common types of spatial search used in urban studies:

- **Nearest neighbor search:** This is termed the k -nearest neighbor (k -NN) search. Typical questions can be “Find the k stores that are closest to a given point or current location” or “Find the closest restaurant.”
- **Range search:** Range search is also common in urban studies. Example queries: “Find all the restaurants with 5 miles range” and “Find all the zones that can be reached between a half hour and a hour.”
- **Aggregate search:** Questions can be often asked in urban studies that involve spatial aggregation. Examples are: “Get the number of hospitals for travel distance zones of under 10, 10–50, 50–100, and above 100 miles” or “Find the total area of green space in an urban district.”

The k -NN search is well studied in computer science and geographic information systems (Knuth 1997). There are a suite of algorithms designed to solve the problem. There are two major categories of algorithms: exact search and approximate search. The simplest approach to find the k -nearest neighbors is sequential search that does not require any preprocessing of the spatial data (Bentley and Friedman 1979). The search time is $O(kn)$, where k is the dimension and n is the total number of features. The storage requirement is also $O(kn)$.

Spatial indexing can be used in preprocessing the data, creating a data structure that can be easily retrieved. BSP-trees, metric trees, and R-trees are three types of commonly used tree data structures in indexing spatial data. The kd -tree, one of the BSP-trees, uses axial rays to partition (ending up as rectangles), while the vp-tree, one of the metric trees, uses equidistance circles to partition data. The R-tree structure uses rectangles but has a focus on keeping the geographic object in a hierarchical

structure. Most of these data structures lead to improvements by reducing the time to search to approximately $O(\log n)$ on average.

Different geographic information systems may support different spatial indexing algorithms. The R-tree and its variants are the most popularly implemented spatial indexing algorithms in geographic information systems, including PostGIS, MySQL, and Oracle. A grid-based spatial indexing scheme is popularly implemented in many geospatial databases, including Esri geodatabase, Oracle, and Microsoft SQL, due to its data-driven spatial indexing scheme.

Spatial search (k -NN, range search, or aggregate search) has been applied in many urban studies. Alternative site selections, such as the “spatial search” of Massam (1980), analyze spatial interactions and require range searches to assess the effect of selecting one alternative over another. For example, a firm searching for a location may consider the labor force that is available within a certain distance of each alternative location. In choosing a location for a retail store location, the analyst may need to conduct spatial queries on household purchasing power within a certain distance of each of the location alternatives. The results of such spatial queries would help in evaluating alternatives and making better plans.

37.4.2 *Spatial Search with Graph Theory*

Spatial search can be seen as a decision problem in urban studies, especially those studies with roots in economics. Economic Search Theory is well studied and has been used in studies of urban migration, urban markets, and urban agglomeration effects (Meier 2009,2010). Adding the spatial context, a generalized spatial search model can be formulated (Meier, 1995,2010). The spatial search problem is effectively defined within a connected graph. The vertices of the connected graph are alternatives at discrete locations in two-dimensional space. The edge connecting two vertices represents the cost, which may be a function of distance. The goal is to maximize the expected utility when the decision is to move from one vertex to another. Each alternative may be visited once.

The model of spatial search results from the tight bounding and integration of spatial context with a domain-specific model. In economics, this spatial model is tightly integrated with a model of economic search. This approach of integrating the spatial context with models in urban studies effectively converts the spatial search problem into an optimization problem on a graph.

The traveling salesperson problem is NP-hard. However, most problems in urban studies have a limited size, making them soluble. There are also heuristics to help in solving the optimization problem efficiently.

With the conversion of the spatial search problem to an optimization problem in a graph, the commonly used graph search algorithms become applicable to the spatial search model. These algorithms include breadth-first search, depth-first search, greedy best-first search, heuristic A*, and Dijkstra’s shortest path algorithm.

The spatial search model has found applications in market area analytics, firm location, urban effect analysis, and urban modeling (Meier 1995). The simple distance or fuel cost-based spatial search model may be used in urban transportation planning and commercial truck routing (Zarezadeh et al. 2018; Moreno-Monroy and Posada 2018; Monte et al. 2018).

37.5 Distributed Search and Interoperability in the Web Environment

The abundance of geospatial information has grown beyond anyone's ability to manage it be properly. The introduction of live sensors and fast updating of information also suggests that the monolithic geographic information system cannot satisfy the requirements of spatial search in urban studies. Yet the data resources available for urban studies continue to grow.

There are several approaches to enable spatial search and geoprocessing to leverage the growing volume of information for urban studies. First, the information can be harvested and ingested into a local spatial catalog system through the harvesting of spatial metadata and data from different sources. The local spatial catalog system has to manage all the information. Each harvester may be updated or re-started (if incremental harvest is not supported by remote services). After each harvest, spatial indexing needs to be updated or re-built. The advantage for such a system is that the existing spatial indexing techniques are already supported. The major drawbacks are that the data can grow out of control and are not always current.

Second, the information is harvested, integrated, and indexed in a distributed manner. In this case, the local catalog system is replaced with a distributed catalog that clusters multiple cloud-computing instances. Each cloud-computing instance may handle a strip of information. A distributed spatial indexing scheme needs to be adopted to support the spatial search in such a distributed system (Priya and Kalpana 2018). The advantage for such a system lies in its capability to handle large datasets in a scalable cloud-computing environment. The major limitations are: (1) the freshness of the metadata and data cannot be warranted, (2) the remote services may not allow the duplication of their metadata and data for various reasons, and (3) the maintenance of a large distributed spatial catalog system can still be a challenge, and the distributed spatial search capability is still in development.

Third, a federated spatial catalog system can be adopted to support the on-the-fly integration of distributed search (Shao et al. 2013; Bai et al. 2007). The development of a federated spatial catalog depends on the adoption of open geospatial standards. The standard interface and response from catalogs make it possible to do translation on the fly. The idea of federated catalog is to set up a series of plug-in translators that handle the translation of request to and response from the remote catalog services. When a user sends in a spatial query, the query request is first translated into a format that matches the remote server and the translated request is sent out. The

response from the remote service is then translated and integrated in the mediator to be sent back to the user. The advantages of such a federated catalog are: (1) it does not need extensive resources in manage the metadata and data since most of the resources are still maintained by the original provider; (2) the contents are in complete synchronization with remote services; and (3) spatial search is completed in a distributed environment. The drawbacks are: (1) the spatial search function and responses are tied to what the remote services offer, and (2) duplicates may not be removed properly if two remote services offer the same content.

37.6 Trends

The spatial search problem is a hard problem to solve. The performance of current solutions is acceptable only because either one of the following assumptions stands: (1) the size of data is limited, (2) optimal heuristics exist for the dataset, or (3) the best option executes in an acceptable time. This section reviews two frontiers in solving the spatial search problem: a quantum spatial search algorithm and semantic spatial search.

Quantum algorithms have emerged in solving the spatial search problem with improvements. Quantum computing is seen as the future of computing, to improve non-deterministic algorithms that consider multiple superpositions of states (Venegas-Andraca 2008; Chakraborty et al. 2016; Ambainis 2008). The spatial search problem is seen as one of the hard problems to be solved with classic computers (Meier 1995,2010), or as a decision problem to find the target vertex in a connected graph (Meier 1995). In a fully connected lattice graph of n vertices, the worst time to find the marked target is $O(n \log n)$ using a random walk in a classic computer. New algorithms in quantum computing have shown that the search can be improved many fold with quantum random walks (Portugal 2018). A discrete-time quantum walk (DTQW) algorithm improved the time to $O(\sqrt{n \log n})$ (Ambainis et al. 2005). A controlled quantum walk (CQW) algorithm on a lattice using an ancilla qubit improved the time complexity to $O((n \log n)^{1/2})$ (Tulsi 2008). An improved version of DTQW also achieved the same time complexity (Ambainis et al. 2015). Portugal described an approach to the design of quantum algorithms for the spatial search problem that explains how Grover's algorithm (Grover 1996), the quantum algorithm for searching a database, "can be seen as a spatial search problem on the complete graph with loops using the coined model and on the complete graph without loops using the staggered model" (Portugal 2018).

The application of semantic technology improves the accuracy of spatial search with more explicit spatial semantics. Most current spatial search solutions treat spatial objects as a spaceless point. Spatial extents and spatial relationships are not taken into full consideration with current solutions. The augmentation of linked geodata (Stadler et al. 2012) with spatiotemporal semantics enables a semantic spatial search (Neumaier and Polleres 2019). A Transportation ontology domain can be added to a semantic-based public transportation geoportal to support semantic spatial search

on concepts, relationships, and individuals (Gunay et al. 2014). Ontology provides additional semantic constraints in semantic spatial search (Jones et al. 2004,2001). A spatial entity can be described by its sub-components, and the search for a spatial entity can be modeled as a multi-component spatial search problem (MCSSP) (Menon and Smith 1989, Menon 1990). This effectively formulates the spatial search problem as a constraint satisfaction problem (CSP) in computer science. The suite of heuristic CSP algorithms can be applied to help in finding the best match, including backtracking, graph-based backjumping, arc consistency, and forward checking (Frost 1997).

37.7 Conclusion

Spatial search has been one of the most intensively researched topics in urban studies, and can be traced back to a pre-computer era. The classic spatial search in dealing with connectivity between spatial objects or entities has been thoroughly researched and supported by most geographic information systems. The spatial search problem can be integrated with models in urban studies to put the research in spatial context. Extending studies with spatial dimensions increases the complexity of problem solving. In a fully connected graph depicting the relationships among entities in a spatial context, the problem is NP-complete and is therefore difficult to solve. However, in actual applications in urban studies, the data size is often manageable and heuristics can be applied to solve the spatial search problem within a reasonable time interval.

New developments in alternative computing environments shed light on solving the spatial problem more efficiently. One of the most researched alternatives is to leverage random walk with quantum computing. Several algorithms have been proposed to solve the spatial search problem efficiently with quantum walks. Another frontier is the use of semantic Web technology in dealing with big data and heterogeneous data in the spatial context.

References

- Ambainis A (2008) Quantum random walks—new method for designing quantum algorithms. In: Geffert V, Karhumäki J, Bertoni A, Preneel B, Návrat P, Bieliková M (eds) SOFSEM 2008: theory and practice of computer science. Springer, Berlin, pp 1–4. https://doi.org/10.1007/978-3-540-77566-9_1
- Ambainis A, Bačkurs A, Nahimovs N, Ozols R, Rivosh A (2013) Search by quantum walks on two-dimensional grid without amplitude amplification. In: Iwama K, Kawano Y, Murao M (eds) Theory of quantum computation, communication, and cryptography. Springer, Berlin, pp 87–97. https://doi.org/10.1007/978-3-642-35656-8_7

- Ambainis A, Kempe J, Rivosh A (2005) Coins make quantum walks faster. In: Proceedings of the Sixteenth Annual ACM-SIAM symposium on discrete algorithms. Society for industrial and applied mathematics, pp 1099–1108
- Arge L, Berg MD, Haverkort H, Yi K (2008) The priority R-tree: a practically efficient and worst-case optimal R-tree. *ACM Trans Algorithms* 4:1–30
- Bai Y, Di L, Chen A, Liu Y, Yaxing W (2007) Towards a geospatial catalogue federation service. *Photogrammetric Eng Remote Sens* 73:699–708
- Ballatore A, Kuhn W, Hegarty M, Parsons E (2016) Special issue introduction: spatial approaches to information search. *Spatial Cognit Comput* 16:245–254. <https://doi.org/10.1080/13875868.2016.1243693>
- Beckmann N, Kriegel HP, Schneider R, Seeger B (1990) The R*-tree: an efficient and robust access method for points and rectangles. In: Proceedings of the 1990 ACM SIGMOD international conference on management of data. ACM Press, Atlantic City, pp 322–331. <https://doi.org/10.1145/93597.98741>
- Bentley JL (1975) Multidimensional binary search trees used for associative searching. *Commun ACM* 18:509–517. <https://doi.org/10.1145/361002.361007>
- Bentley JL, Friedman JH (1979) Data structures for range searching. *ACM Comput Surv* 11:97–409. <https://doi.org/10.1145/356789.356797>
- Beygelzimer A, Kakade S, Langford J (2006) Cover trees for nearest neighbor. In: Proceedings of the 23rd international conference on machine learning. ACM, Pittsburgh, pp 97–104
- Bowen WM, Dunn RA, Kasdan DO (2010) What is “urban studies”? Context, internal structure, and content. *J Urban Affairs* 32:199–227. <https://doi.org/10.1111/j.1467-9906.2009.00474.x>
- Bozkaya T, Ozsoyoglu M (1999) Indexing large metric spaces for similarity search queries. *ACM Trans Database Syst* 24:361–404. <https://doi.org/10.1145/328939.328959>
- Broome FR, Meixler DB (1990) The TIGER data base structure. *Cartography Geogr Inform Syst* 17:39–47. <https://doi.org/10.1559/152304090784005859>
- Burkhard WA, Keller RM (1973) Some approaches to best-match file searching. *Commun ACM* 16:230–236
- Chakraborty S, Novo L, Ambainis A, Omar Y (2016) Spatial search by quantum walk is optimal for almost all graphs. *Phys Rev Lett* 116:100501. <https://doi.org/10.1103/PhysRevLett.116.100501>
- Davis BA, George JR, Marx RW (1992) TIGER/SDTS: standardizing an innovation. *Cartography Geogr Inform Syst* 19:321–327. <https://doi.org/10.1559/152304092783762164>
- Dimou M, Schaffar A (2009) Urban hierarchies and city growth in the Balkans. *Urban Studies* 46:2891–2906
- Dueker KJ (1974) Urban geocoding. *Ann Assoc Am Geogr* 64(2):318–325
- Farnsworth G, Curry JP (1992) The DIME geocoding system. U.S, Bureau of the Census
- Finkel RA, Bentley JL (1974) Quad trees: a data structure for retrieval on composite keys. *Acta Informatica* 4:1–9. <https://doi.org/10.1007/BF00288933>
- Frost DH (1997) Algorithms and heuristics for constraint satisfaction problems. Ph.D. dissertation, University of California, Irvine. <https://www.ics.uci.edu/~dechter/publications/R69.pdf>
- Galdi D (2005) Spatial data storage and topology in the redesigned MAF/TIGER system. U.S, Census Bureau Geography Division
- Goldberg DW (2017) Geocoding. In: Richardson D, Castree N, Goodchild MF, Kobayashi A, Liu W, and Marston RA (eds) International encyclopedia of geography: people, the Earth, environment and technology. Wiley, Oxford, pp 1–12. <https://doi.org/10.1002/9781118786352.wbieg1051>
- Goldberg DW, Wilson JP, Knoblock CA (2007) From text to geographic coordinates: the current state of geocoding. *URISA J* 19:33–47
- Grover LK (1996) A fast quantum mechanical algorithm for database search. In: Proceedings of the twenty-eighth annual ACM symposium on theory of computing, pp 212–219
- Gunay A, Akcay O, Altan MO (2014) Building a semantic based public transportation geoportal compliant with the INSPIRE transport network data theme. *Earth Sci Inf* 7:25–37. <https://doi.org/10.1007/s12145-013-0129-z>

- Guttman A (1984) R-trees: a dynamic index structure for spatial searching. In: Proceedings of the 1984 ACM SIGMOD international conference on management of data—SIGMOD '84. ACM Press, Boston, p 47. <https://doi.org/10.1145/602259.602266>
- Harris R, Smith ME (2011) The history in urban studies: a comment. *J Urban Affairs* 33:99–105. <https://doi.org/10.1111/j.1467-9906.2010.00547.x>
- Hellerstein JM, Naughton JF, Pfeffer A (1995) Generalized search trees for database systems. In: Proceedings of the 21th International conference on very large data bases. Morgan Kaufmann, pp 562–573
- Henrich A, Six HW, Hagen F, Widmayer P (1989) The LSD tree: spatial access to multidimensional point and non-saint objects. In: 5th very large databases conference. pp 45–54
- Janowicz K, McKenzie G, Hu Y, Zhu R, Gao S (2019) Using semantic signatures for social seing in urban environments. In: Antoniou C, Dimitriou L, Pereira F (eds) *Mobility patterns, big data and transport analytics*. Elsevier, New York, pp 31–54
- Jones CB, Abdelmoty AI, Finch D, Fu G, Vaid S (2004) The SPIRIT spatial search engine: architecture, ontologies and spatial indexing. In: Egenhofer MJ, Freksa C, and Miller HJ (eds) *Geographic information science*. Springer, Berlin. pp 125–139. https://doi.org/10.1007/978-3-540-30231-5_9
- Jones CB, Alani H, Tudhope D (2001) Geographical information retrieval with ontologies of lace. In: Proceedings of the international conference on spatial information theory: foundations of geographic information science. Springer. pp 322–335
- Kamel I, Faloutsos C (1994) Hilbert R-tree: an improved R-tree using fractals. In: Proceedings of the 20th international conference on very large data bases. Morgan Kaufmann. pp 500–509
- Kriegel H, Seeger B (1988) Techniques for design and implementation of efficient spatial access methods. In: Proceedings of the 14th VLDB Conference. pp 360–370
- Knuth DE (1997) *The art of computer programming*. Addison-Wesley, Reading
- Lu H, Ooi BC (1993) Spatial indexing: past and future. *IEEE Data Eng Bull* 16:16–21
- Maier G (1995) Spatial search: structure, complexity, and implications. *Physica-Verl, Heidelberg*
- Maier G (2009) Spatial prices, search behavior and location. *Ann Reg Sci* 43:27–48. <https://doi.org/10.1007/s00168-007-0203-x>
- Maier G (2010) The spatial search problem: structure and complexity. *Geogr Anal* 25:242–251. <https://doi.org/10.1111/j.1538-4632.1993.tb00295.x>
- Massam BH (1980) *Spatial search: application to planning problems in the public sector*. Pergamon Press, Oxford
- Meagher D (1980) Octree encoding: a new technique for the representation, manipulation and display of arbitrary 3-D objects by computer. Rensselaer polytechnic instiute
- Menon S (1989) *Spatial search for multicomponent objects in a geographic information system using symbolic models and hierarchical data structures*. PhD dissertation, University of California, Santa Barbara
- Menon S, Smith TR (1989) Search for multi-component objects in large spatial databases. In: Litwin W, Schek HJ (eds) *Foundations of Data Organization and Algorithms*. Springer, Berlin. pp 246–251. https://doi.org/10.1007/3-540-51295-0_134
- Mesev V (2005) Identification and characterisation of urban building patterns using IKONOS imagery and point-based postal data. *Comput Environ Urban Syst* 29:541–557. <https://doi.org/10.1016/j.compenvurbsys.2005.01.003>
- Miller HJ (1992) Consumer spatial search behavior: a general modeling framework and operational models. Paper presented to the Association of American Geographers, San Diego, April
- Monte F, Redding SJ, Rossi-Hansberg E (2018) Commuting, migration, and local employment elasticities. *Am Econ Rev* 108:3855–3890. <https://doi.org/10.1257/aer.20151507>
- Moreno-Monroy AI, Posada HM (2018) The effect of commuting costs and transport subsidies on informality rates. *J Dev Econ* 130:99–112. <https://doi.org/10.1016/j.jdevco.2017.09.004>
- Neumaier S, Polleres A (2019) Enabling spatio-temporal search in open data. *J Web Semant* 55:21–36. <https://doi.org/10.1016/j.websem.2018.12.007>
- Omohundro SM (1989) *Five balltree construction algorithms*. International Computer Science Institute, Berkeley

- Owusu C, Lan Y, Zheng M, Tang W, Delmelle E (2017) Geocoding fundamentals and associated challenges. In: Karimi HA, Karimi B (eds) *Geospatial data science techniques and applications*. CRC Press, Boca Raton. <https://doi.org/10.1201/b22052>
- Panasyuk A, Yu ESL, Mehrotra KG (2019) Improving geocoding for city-level locations. In: *Proceedings, 2019 IEEE 13th international conference on semantic computing (ICSC)*. pp. 416–421. <https://doi.org/10.1109/ICOSC.2019.8665524>
- Paull D (2003) A geocoded national address file for Australia: the G-NAF what, why, who and when. PSMA Australia Limited, Griffith
- Portugal R (2018) Spatial search algorithms. In: *Quantum Walks and Search Algorithms*. Springer, Cham. pp. 175–200. https://doi.org/10.1007/978-3-319-97813-0_9
- Priya M, Kalpana R (2018) Distributed processing of location based spatial query through vantage point transformation. *Future Comput Informatics J* 3:296–303. <https://doi.org/10.1016/j.fcij.2018.09.002>
- Robinson JT (1981) The K-D-B-tree: a search structure for large multidimensional dynamic indexes. In: *Proceedings of the 1981 ACM SIGMOD international conference on management of data*. pp. 10–18
- Schumacher R, Brand B, Gilliland M, Sharp W (1969) Study for applying computer-generated images to visual simulation. U.S. Air Force Human Resources Laboratory
- Sellis T, Roussopoulos N, Faloutsos C (1987) The R+-tree: a dynamic index for multi-dimensional objects. In: *Proceedings of the 13th international conference on very large data bases (VLDB)*
- Shao Y, Di L, Bai Y, Wang H, Yang C (2013) Federated catalogue for discovering Earth observation data. *Photogrammetrie—Fernerkundung—Geoinformation* 1:43–52. <https://doi.org/10.1127/1432-8364/2013/0157>
- Stadler C, Lehmann J, Höffner K, Auer S (2012) LinkedGeoData: a core for a web of spatial open data. *Semantic Web* 3(4):333–354. <https://doi.org/10.3233/SW-2011-0052>
- Tobler W (1972) Geocoding theory. In: *Proceedings of the national geocoding conference*
- Trainor T (2003) U.S. Census Bureau geographic support: a response to changing technology and improved data. *Cartography Geogr Inform Sci* 30:217–223. <https://doi.org/10.1559/152304003100011054>
- Trainor TF (2005) The MAF/TIGER enhancement program: the mechanics and maintenance of a large-scale national spatial database. *Govern Publ Rev* 13:181–201
- Tulsi A (2008) Faster quantum-walk algorithm for the two-dimensional spatial search. *Phys Rev A* 78:012310. <https://doi.org/10.1103/PhysRevA.78.012310>
- Venegas-Andraca SE (2008) Quantum walks for computer scientists. *Synth Lect Quantum Comput* 1:1–119. <https://doi.org/10.2200/S00144ED1V01Y200808QMC001>
- Yianilos PN (1993) Data structures and algorithms for nearest neighbor search in general metric spaces. In: *Proceedings of the fourth annual ACM-SIAM symposium on discrete algorithms*. Society for industrial and applied mathematics, Austin. pp. 311–321
- Zandbergen PA (2008) A comparison of address point, parcel and street geocoding techniques. *Comput Environ Urban Syst* 32:214–232. <https://doi.org/10.1016/j.compenvurbsys.2007.11.006>
- Zarezadeh Z, Benckendorff P, Gretzel U (2018) Lack of progress in tourist information search research: a critique of citation behaviour and knowledge development. *Current Issues Tourism* 22(19):1–15. <https://doi.org/10.1080/13683500.2018.1459511>
- Zhong R, Li G, Tan KL, Zhou L, Gong Z (2015) G-Tree: an efficient and scalable index for spatial search on road networks. *IEEE Trans Knowl Data Eng* 27:2175–2189. <https://doi.org/10.1109/TKDE.2015.2399306>



Liping Di is the professor and founding director of the Center for Spatial Information Science and Systems (CSISS), George Mason University. He served as the Chair for the US National Committee of the Geographic Information Standards from 2010–2016 and is currently the convener of the ISO TC 211 Working Group 9. Dr. Di has been engaged in geoinformatics research for over thirty years. He has achieved over 450 publications and has received more than \$56 million in research grants.



Eugene Genong Yu is a Research Associate Professor in geographic information sciences and remote sensing at the Center for Spatial Information Science and Systems at George Mason University, USA. His research interests include geographic information systems, remote sensing, intelligent image understanding, sensor Web, semantic Web, computational vision, and robotics. He is a member of the American Geophysical Union, the IEEE Geoscience and Remote Sensing Society, and the IEEE Computer Society.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

