Journal of Computer and System Sciences 81 (2015) 1508–1531

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

A taxonomy for region queries in spatial databases

David Taniar [a,*], Wenny Rahayu [b]

[a] Clayton School of Information Technology, Monash University, Australia
[b] Department of Computer Science and Computer Engineering, La Trobe University, Australia

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form 14 December 2014
Accepted 14 December 2014
Available online 24 December 2014

Keywords:
Spatial databases
Spatial queries
Range queries
Region queries
Query processing

ABSTRACT

In spatial databases, there are two basic types of queries, namely *nearest neighbour queries* (*k*NN) and *range queries*. Spatial range queries are not only finding objects of interest within a certain range or radius, but feature a wide spectrum: from finding objects of interest to forming the range (or region). Therefore, in this paper, we coin a term "Region Queries" to indicate a broad category of spatial range queries. It is imperative to understand the full capabilities of region queries, before starting to work on processing and optimising such queries. The aim of this paper is to show a complete picture of region queries. In this study, we present taxonomy of region queries, comprising of three categories: (*i*) finding objects of interest, (*ii*) forming regions, and (*iii*) determining centroids. These three query types form a comprehensive view of what region queries are about.

1. Introduction

Range queries in spatial databases have been widely known [25,26,28]. A typical example of a spatial range query is "to find all Chinese restaurants in a radius of 5 km". This kind of queries has the basis of *region*. In this paper, we would particularly like to study region queries in spatial databases.

Region is an important topic, not only in spatial databases, but also in many other areas. The concept of region is naturally found almost everywhere. For example, country, house, suburb, etc., all concern about region. In human or animal, there is a *territory*, which is basically a region.

In the context of technology, such as mobile technology, there is a new notion of *geofencing*, which is also a form of region. Geofencing is a kind of virtual fencing or virtual perimeter of a particular object of interest. This fence can be a predefined set of boundaries or dynamically built.

The problem of regions or areas have been studied in *computational geometry* or geometry, in general. For example, polygon is a spatial region which is one of the basic concepts in geometry, or coin placement problem which has been extensively addressed in computational geometry.

The concept of region has now become even more important in spatial (or geo-spatial) information systems, due to the widely available online maps, and ubiquitous mobile devices [25,26]. These technologies have created a new wave of applications, not existing before, including tourism, urban planning, mobile workforce, etc.

The focus of this paper is *spatial queries*, especially those involving region. In a spatial query like the Chinese restaurant example mentioned earlier, the region is specified by the radius (or distance). In general, the specified region can be any-

* Corresponding author.
E-mail addresses: David.Taniar@monash.edu (D. Taniar), W.Rahayu@latrobe.edu.au (W. Rahayu).

http://dx.doi.org/10.1016/j.jcss.2014.12.025
0022-0000/

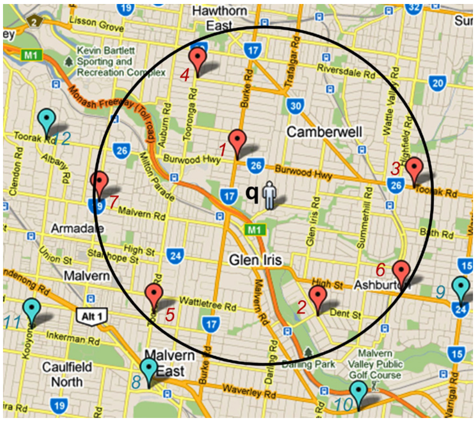

Fig. 1. Region query (Euclidean).

thing; not necessarily bounded by a radius. Therefore, in this paper, we use the term "region queries", rather than range queries, since region queries are more general, and have wider usage.

Region queries basically consist of three main elements: (*i*) objects of interest, (*ii*) query point, and (*iii*) region. Objects of interest are often the results of the query located in the specified region, whereas the query point is the location where the query is invoked and the results are relative to the query location — this is often called a *location-dependent* or *location-based* system.

We have previously published a similar work in this journal, but with a focus on *nearest neighbour* (*k*NN) [22], in which we presented a comprehensive view of nearest neighbour queries in spatial databases. In this paper, we study region queries. Apart from *k*NN queries, region queries are one of the most common queries in spatial databases, which are widely used in location-based services.

As the study of spatial region queries is very big, in this paper, we would like to confine our study to static objects (i.e. objects of interest that do not move) and static queries (i.e. the query is once off). The study of moving objects (as opposed to static objects) is a big topic by itself, and we have defined a taxonomy for this [7], where we defined various potential queries involving moving objects. When the query point is moving, also known as *continuous queries* (as opposed to static queries) [33], it is about the changing of query results because the query point moves. When the query moves, objects of interest that have been relevant to the previous query location might become obsolete, and the opposite might also be true. The study of continuous or moving queries are also excluded in this paper. We also limit our study to an outdoor environment (e.g. geospatial), rather than an indoor environment. In an indoor environment, there are many new elements, such as multi floor, which raise new complexities [8,6,9].

The proposed region query taxonomy presented in this paper is divided into three main categories; all of which have the basis of objects of interest, query point, and region. Based on these three main elements, we categorise region queries into: (*i*) finding objects of interest, (*ii*) forming or finding regions, and (*iii*) determining centroids.

2. Category I: finding objects of interest

The first category of region queries is finding objects of interest — to be more precise: "given a region, find all objects of interest in the region". We define six different types of region queries in this category: (*i*) traditional region queries, (*ii*) approximate region queries, (*iii*) constrained region queries, (*iv*) clustered objects region queries, (*v*) outer/inner fence object queries, and (*vi*) inverse range queries.

2.1. Traditional region queries

The Chinese restaurant example mentioned earlier is a typical traditional spatial region query — that is given the query location and a radius, retrieve all objects within this area. Fig. 1 shows an example of a traditional region query. The query point is denoted by q, which is the centre of the region. In this case, the query retrieves 7 objects of interest with flags $1, 2, \ldots, 7$. Other objects outside this region are not retrieved by the query. In this example, the radius of the search is specified in a Euclidean distance — a straight and direct distance measure from the query point q.

In a geospatial database setting, there is an underlying *road network* (or *spatial road network*) where spatial objects are located. Therefore, the distance between two objects is a network distance which is the shortest road (or shortest path) between these objects. Fig. 2 shows the example of the region query, but uses a road network distance, instead of Euclidean distance. Noticeable, that in this case, objects 4 and 6 (which are inside the region) are out of reached, because their road network distance is beyond the specified distance.