CHAPTER 1
INTRODUCTION

1.1 Overview

Reverse k nearest neighbours (RkNN) is a query that finds every data points that consider query point q as one of its k nearest neighbours. [1] shows that q has an area such that every point inside is the RkNN of q and every point outside is not the RkNN of q, this area is called influence zone.

There are some existing ways to answer RkNN query, the most common way is using point to point approach. Point to point approach will verify every objects whether they consider query point q as one of the k nearest point, therefore this approach is costly[2]. Another way to answer this is using region approach. Influence zone is one of the method that using region approach. As mentioned above, influence zone is an area such that every objects that lies inside that area will be considered as one of the RkNN of q. [3] proves that influence zone for RkNN where $k = 1$ is equal to first order voronoi diagram; thus pre-computed voronoi diagram can be used to compute the influence zone. However, as mentioned in [1], using pre-computed voronoi diagram to compute influence zone is not practical because voronoi diagram are expensive to compute and incur high space requirement.

This minor thesis provides a method to overcome the limitation of influence zone, which is inability to answer RkNN query where k value is dynamic. Suppose the value of k is changed, the influence zone must be recomputed to answer the new k value[1]. In this paper we introduce a concept that expand influence zone on $k = 1$ and we show that the expanded influence zone is able to answer RkNN even when the k value is dynamic without the needs to recompute it.

Consider a set of interest point $P = \{p_1, p_2, ..., p_n\}$ where p_i represents a point in euclidean space and a query point q. The areas $A = \{a_1, a_2, ..., a_n\}$ is a set of area such that every object that lies in a_i will consider q as one of its k nearest neighbour, where $k = i$. Every areas is a proper subset for its next area such that $a_i \subseteq a_{i+1}$; therefore every point that's considered as RiNN will also be considered as R$(i+1)$NN of q.

Expanded influence zone has various application in location based services, especially services that needs the value of k to be dynamic. Take following as example, suppose there is a taxi service that use Global Positioning System (GPS) and

online booking. Passenger will send a request to server containing its location and the amount of taxis they want to book. Server will then assign the passenger to taxis that consider the passenger as one of their RkNN where k is amount of taxis requested.

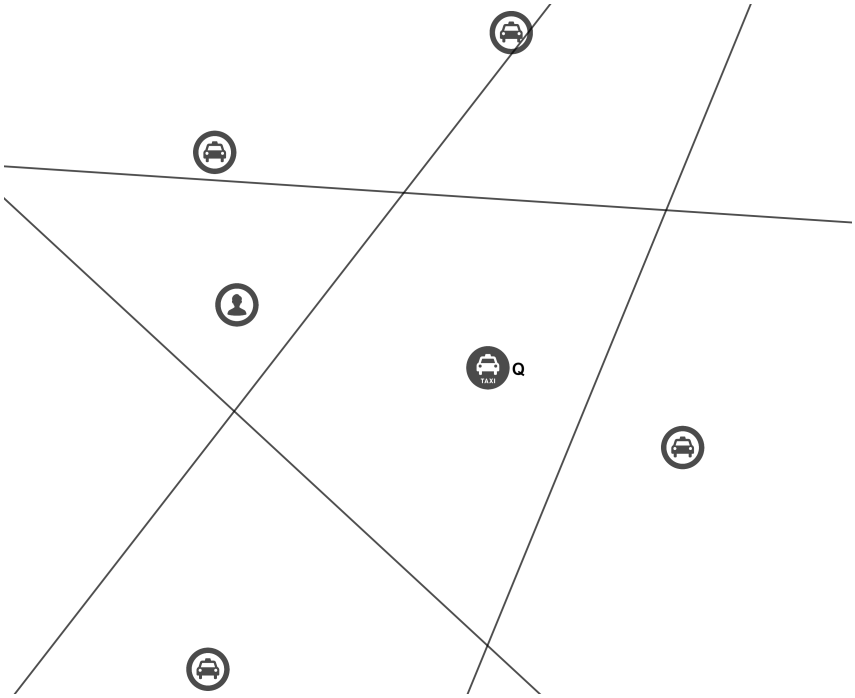

Figure 1.1: Influence Zone Expansion in location based service

Shown in figure 1.1, suppose there is an user that requested one taxi. Taxi Q (taxi with opaque background) will not get notified, as the user is not inside R1NN of Q. However, suppose the user change the number of requested taxis into two, zone recomputation is unnecessary. We can simply check whether the user lies inside R2NN of Q. If that was the case, then taxi Q will be notified.

This minor thesis is organised as follows. Chapter 2 provides some related works and previous research prior to this paper writing. Chapter 3 explains on system methodology and design used in this research. Section 4 describes about evaluation and result. Section 5 is the conclusion about the paper.

1.2 Problem Statement

Based on description above, the problems are formulated as follows.

1. How to compute influence zone expansion to answer dynamic k on RkNN query?
2. How does the influence zone expansion creation and query perform?

1.3 Objective

The objectives of this minor thesis are:

1. To implement influence zone expansion that can answer dynamic k on RkNN query.
2. Analyze the performance of influence zone expansion creation and query.

1.4 Scope

The scopes of this minor thesis are:

1. The point and object mentioned in this minor thesis are static point and object.
2. The point and object mentioned in this minor thesis are generated random with uniformly distributed value.

1.5 Hypothesis

The system will be able to compute influence zone expansion that can answer dynamic k on RkNN and influence zone expansion will outperform conventional influence zone on dynamic value of k. First, system will compute the influence zone expansion. Then, the computed influence zone is used to do query on various value of k.

1.6 Methodology

The minor thesis planning will be conducted as follows.

- Literature study
  Studying the literature that will be used as reference on this project.
- System analysis and design
  Analyse the method that will be used on computing influence zone expansion.
- Result analysis
  Testing the result and performance of the system.
- Report writing
  Test result will be documented and recorded according to reality.

1.7 Summary

This chapter explain what this minor thesis about. From the problem stated above, this minor thesis offer an answer. The main problem is how to answer RkNN with dynamic value of k efficiently. This minor thesis offer influence zone expansion to answer RkNN query on dynamic value of k.