

# Chapter 10

## An Interactive Personalized Spatial Keyword Querying Approach

**Xiangfu Meng**

*Liaoning Technical University, China*

**Lulu Zhao**

*Liaoning Technical University, China*

**Xiaoyan Zhang**

*Liaoning Technical University, China*

**Pan Li**

*Liaoning Technical University, China*

**Zeqi Zhao**

*Liaoning Technical University, China*

**Yue Mao**

*Liaoning Technical University, China*

### **ABSTRACT**

*Existing spatial keyword query methods usually evaluate text relevancy according to the frequency of occurrence of query keywords in the text information associated to spatial objects, without considering the degree of preference of users to different query keywords, and without considering semantic relevancy. To deal with the above problems, this chapter proposes an interactive personalized spatial keyword querying approach which is divided into two stages. In the offline processing stage, Gibbs algorithm is adopted to estimate the thematic probability distribution of text information associated to spatial objects, and then an LDA model is used for semantic expansion of spatial data set.*

## INTRODUCTION

Information query refers to the activity and process of searching, identifying and obtaining relevant facts, data and text for solving various problems. As an inseparable part of human social activities, people are more and more concerned about how to quickly and accurately find the information they need from the massive size of information sources. With the advent of GPS and other location-based service technologies, it becomes easier to obtain geographic spatial dimension information. As a result, more and more spatial objects with location information emerged on the Web, such as hotels, cafes and tourist attractions. These spatial objects are often referred to as Point of Interests (POIs). A spatial object contains the geographic location (usually in the form of latitude and longitude) and a textual description (such as object name, facility, comment, and so on). Spatial keyword query takes geographical location and keyword as parameters and returns spatial object that meets the query condition. Specifically, each POI/spatial object in the spatial database contains spatial location information and text information. Assuming the geographic location of a given user and a set of query keywords, the location-based service system returns the POIs related to the query both in spatial and textual from the spatial database. Now there are a large number of online resources, which can obtain large-scale geographic text objects, such as Google Places, yahoo, Foursquare, and other social networks, trip advisory groups and public comments, etc., which need technology to support the effective processing of spatial keyword query.

Traditional spatial queries and keyword search using a different indexing technology and query algorithms. To effectively deal with spatial keyword query, it is need to combine spatial index and text index as the spatial-text index, and to propose the corresponding query algorithms. According to the spatial index and text index combined with the different methods, existing query processing techniques can be roughly divided into three categories: loose coupling, spatial and text preference. Loose combination, which is for spatial data to establish a spatial index (generally the R-tree or Quad-tree), for the text data-based text index (usually Inverted index), respectively. There is no or only loose connection between the two types of indexes. During query processing, starting from spatial index and text index respectively, the objects that satisfy spatial constraint and text constraint are found, and then they are integrated. Spatial priority: in this kind of query processing technology, spatial text index is obtained by enhancing spatial index and adding text information. Spatial index mainly uses R-tree, while a few works use grid index. In this scenario, a spatial-text index is an enhancement of a text index (such as an Inverted file) that maps each keyword to a data structure that contains spatial information. In this paper, IR-tree index in the spatial priority scheme is used, which is combined by R-tree and Inverted list to achieve rapid filtering and expeditiously find the spatial objects required by users.

A spatial object  $o$  mainly contains two parts of information: spatial information and text information. Spatial information is usually represented by longitude and latitude while text information is the text description of the spatial object. The form of a spatial keyword query  $q$  is:  $q(loc, keywords, k, \alpha)$ , where  $q.loc$  stands for the query location,  $q.keywords$  is the set of query keywords,  $k$  is the number of results returned,  $\alpha \in [0, 1]$  is a weight coefficient. Currently, the commonly used scoring function for a spatial object  $o$  and a spatial keyword query  $q$  is as follows:

$$Score(o, q) = \alpha \cdot S_{spatial}(o.loc, q.loc) + (1 - \alpha) \cdot S_{text}(o.doc, q.keywords) \quad (1)$$

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/an-interactive-personalized-spatial-keyword-querying-approach/230690](http://www.igi-global.com/chapter/an-interactive-personalized-spatial-keyword-querying-approach/230690)

## Related Content

---

### XML Native Storage and Query Processing

Ning Zhang and Tamer M. Özsu (2010). *Advanced Applications and Structures in XML Processing: Label Streams, Semantics Utilization and Data Query Technologies* (pp. 1-17).

[www.irma-international.org/chapter/xml-native-storage-query-processing/41497](http://www.irma-international.org/chapter/xml-native-storage-query-processing/41497)

### Managing Research Data at the University of Porto: Requirements, Technologies, and Services

João Rocha da Silva, Cristina Ribeiro and João Correia Lopes (2013). *Innovations in XML Applications and Metadata Management: Advancing Technologies* (pp. 174-197).

[www.irma-international.org/chapter/managing-research-data-university-porto/73179](http://www.irma-international.org/chapter/managing-research-data-university-porto/73179)

### Mining Association Rules

Mihai Gabroveanu (2009). *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches* (pp. 647-673).

[www.irma-international.org/chapter/mining-association-rules/35878](http://www.irma-international.org/chapter/mining-association-rules/35878)

### Extension to UML Using Stereotypes

Daniel Riesco, Paola Martellotto and German Montejano (2003). *UML and the Unified Process* (pp. 273-293).

[www.irma-international.org/chapter/extension-uml-using-stereotypes/30547](http://www.irma-international.org/chapter/extension-uml-using-stereotypes/30547)

### The Debugging of Rule Bases

Valentin Zacharias (2009). *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches* (pp. 302-325).

[www.irma-international.org/chapter/debugging-rule-bases/35864](http://www.irma-international.org/chapter/debugging-rule-bases/35864)