

DOCTORAL THESIS

Answering why-not questions on spatial keyword top-k queries

Chen, Lei

Date of Award:
2016

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

With the continued proliferation of location-based services, a growing number of web-accessible data objects are geo-tagged and have text descriptions. Spatial keyword top- k queries retrieve k such objects with the best score according to a ranking function that takes into account a query location and query keywords. However, it is in some cases difficult for users to specify appropriate query parameters. After a user issues an initial query and gets back the result, the user may find that some expected objects are missing and may wonder why. Answering the resulting why-not questions can aid users in retrieving better results and thus improve the overall utility of the query functionality. While spatial keyword querying has been studied intensively, no proposals exist for how to offer users explanations of why such expected objects are missing from results. In this dissertation, we take the first step to study the *why-not* questions on spatial keyword top- k queries. We provide techniques that allow different revisions of spatial keyword queries such that their results include one or more desired, but missing objects. Detailed problem analysis and extensive experimental studies consistently demonstrate the effectiveness and robustness of our proposed techniques in a broad range of settings.

Keywords: Spatial Keyword Top- k Query, Why-Not Question, Query Refinement

Table of Contents

Declaration	i
Abstract	ii
List of Publications	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of Figures	x
Chapter 1 Introduction	1
1.1 Preference-Adjusted Why-Not Spatial Keyword Top- k Query	2
1.2 Keyword-Adapted Why-Not Spatial Keyword Top- k Query	3
1.3 Direction-Aware Why-Not Spatial Keyword Top- k Query	5
1.4 Summary	6
Chapter 2 Literature Review	7
2.1 Spatial Keyword Query Processing	7
2.2 Why-Not Query Processing	9
Chapter 3 Preference-Adjusted Why-Not Spatial Keyword Top-k Query	11
3.1 Preliminaries and Problem Definitions	11

3.1.1	Spatial Keyword Top- k Query	11
3.1.2	Preference-Adjusted Why-Not Spatial Keyword Query	13
3.1.3	Baseline Algorithm	14
3.2	Properties and Basic Algorithm	16
3.2.1	Projection of Objects	16
3.2.2	Query Modification and Ranking Updates	18
3.2.3	Basic Algorithm for Why-Not Spatial Keyword Query	20
3.3	Bound and Prune Algorithm	21
3.3.1	A Hybrid Index: BIR-Tree	22
3.3.2	Optimized Query Processing by Estimated Bounds	24
3.3.3	Derivation of $\hat{DP}(N, e)$ and $\check{DP}(N, e)$	25
3.4	Handling Multiple Missing Objects	33
3.4.1	Identifying Working Parts	34
3.4.2	Why-Not Spatial Keyword Query Processing for Multiple Missing Objects	36
3.5	Empirical Study	37
3.5.1	Experimental Setup	37
3.5.2	Empirical Results	39
3.6	Summary	43
Chapter 4 Keyword-Adapted Why-Not Spatial Keyword Top-k Query		44
4.1	Preliminaries and Problem Definition	44
4.1.1	Preliminaries	44
4.1.2	Keyword-Adapted Why-Not Spatial Keyword Query	45
4.2	Basic Why-not Answering Algorithm	48
4.2.1	Problem Analysis	48
4.2.2	Basic Algorithm	48
4.2.3	Optimizations	50
4.3	Bound-and-Prune Algorithm	53
4.3.1	Preliminaries: KcR-Tree	55

4.3.2	Properties of KcR-tree	56
4.3.3	Optimized Bound-and-Prune Algorithm	60
4.3.4	Strategic Use of Algorithm 5	61
4.4	Multiple Missing Objects and Approximate Algorithm	64
4.4.1	Multiple Missing Objects	64
4.4.2	Approximate Algorithm	65
4.5	Empirical Study	65
4.5.1	Experimental Setup	66
4.5.2	Experimental Result	67
4.6	Summary	74
Chapter 5 Direction-Aware Why-Not Spatial Keyword Top-k Query		76
5.1	Preliminaries and Problem Formulation	76
5.1.1	Preliminaries	77
5.1.2	Direction-Aware Why-Not Spatial Keyword Query	79
5.1.3	Baseline Algorithm	80
5.2	Answering Why-Not Questions: Special Case	81
5.2.1	Case Analysis	82
5.2.2	Ranking Updates and Direction Modification	82
5.2.3	Answering Why-Not	85
5.3	Answering Why-Not Questions: General Case	88
5.3.1	Case Analysis	88
5.3.2	Answering Why-Not	92
5.4	Handling Multiple Missing Objects	95
5.5	Empirical Study	97
5.5.1	Experimental Setup	98
5.5.2	Experimental Results	99
5.6	Summary	105

Chapter 6	Conclusions and Future Work	107
6.1	Contributions	107
6.2	Possible Directions for Future Work	108
Bibliography		110
Curriculum Vitae		114