



Universidade Federal da Bahia  
Instituto de Matemática

Programa de Pós-Graduação em Ciência da Computação

## **IMPROVING SPATIAL QUERIES ACCURACY**

João Paulo Dias de Almeida

QUALIFICAÇÃO DE DOUTORADO

Salvador  
1 de julho de 2019



JOÃO PAULO DIAS DE ALMEIDA

## **IMPROVING SPATIAL QUERIES ACCURACY**

Esta Qualificação de Doutorado foi apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Doutor em Ciência da Computação.

Orientador: Frederico Araújo Durão

Salvador  
1 de julho de 2019



## RESUMO

to do

**Palavras-chave:** consulta espacial, linked data, LOD, personalização



## ABSTRACT

to do

**Keywords:** spatial query, linked data, LOD, personalization





# CONTENTS

<b>Chapter 1—Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Problem Statement . . . . .	1
1.3 Objectives of the Proposed Solution . . . . .	1
1.3.1 Specific Objectives . . . . .	1
1.3.2 Research Questions . . . . .	1
1.4 Methodology . . . . .	1
1.5 Statement of the Contributions . . . . .	1
1.6 Thesis Structure . . . . .	1
1.7 Summary . . . . .	1
<b>Chapter 2—Spatial Queries</b>	<b>3</b>
<b>Chapter 3—Linked Open Data</b>	<b>5</b>
<b>Chapter 4—Preliminary Experimental Evaluation</b>	<b>7</b>
4.1 Methodology . . . . .	7
4.1.1 Datasets . . . . .	7
4.1.2 Metrics . . . . .	7
4.1.3 Baseline . . . . .	7
4.2 Preliminary Results . . . . .	7
4.3 Discussion and Points of Improvement . . . . .	7
4.4 Summary . . . . .	7
<b>Chapter 5—Final Remarks</b>	<b>9</b>
5.1 Academic Life . . . . .	9
5.1.1 Completed Classes . . . . .	9
5.1.2 Publications and Participation in Scientific Conferences . . . . .	10
5.2 Schedule . . . . .	11
5.3 Summary . . . . .	13



## LIST OF FIGURES

5.1	Ph.D. activities since the course enrollment until thesis defense. . . . .	14
-----	--	----



## LIST OF TABLES



## LISTA DE SIGLAS

<b>PGCOMP</b>	Programa de Pós-Graduação em Ciência da Computação.....	10
<b>PGCA</b>	Programa de Pós-Graduação em Computação Aplicada.....	10





## **Chapter**

# **1**

## **INTRODUCTION**

### **1.1 MOTIVATION**

### **1.2 PROBLEM STATEMENT**

### **1.3 OBJECTIVES OF THE PROPOSED SOLUTION**

#### **1.3.1 Specific Objectives**

#### **1.3.2 Research Questions**

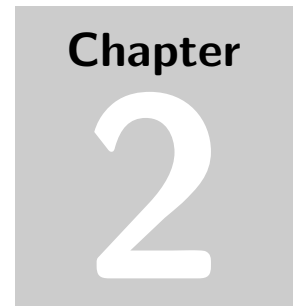
### **1.4 METHODOLOGY**

### **1.5 STATEMENT OF THE CONTRIBUTIONS**

### **1.6 THESIS STRUCTURE**

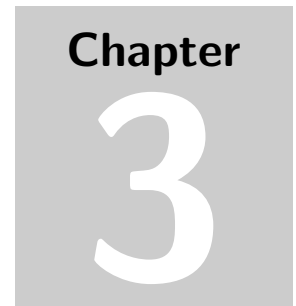
### **1.7 SUMMARY**





## **SPATIAL QUERIES**





## **LINKED OPEN DATA**



## **PRELIMINARY EXPERIMENTAL EVALUATION**

### **4.1 METHODOLOGY**

#### **4.1.1 Datasets**

#### **4.1.2 Metrics**

#### **4.1.3 Baseline**

### **4.2 PRELIMINARY RESULTS**

### **4.3 DISCUSSION AND POINTS OF IMPROVEMENT**

### **4.4 SUMMARY**





## Chapter

# 5

*This chapter presents the research stages until now. Classes completed and publications are described, as well as, the schedule.*

## FINAL REMARKS

### 5.1 ACADEMIC LIFE

During the Ph.D. several activities were developed to define the research topic in Information Retrieval (IR), like conferences presentations and submission of journal articles. In the early stages the focus was in the Ph.D. classes and curricular components to consolidate the theoretical background. Then, the teaching internship and the workshop presentations to the research group contribute to the teaching skills.

#### 5.1.1 Completed Classes

Each Ph.D. class gave a unique contribution to the research. In the sequence, is described each of these contributions.

- **MATE64 - Scientific Seminars** - Professor Christina Von Flach organized a series of seminars where several professors or students in the late stage of their research presented interesting topics in computer science. The discussion after these presentations stimulate many ideas and inspired great solutions.
- **MATD74 - Algorithms and Graphs** - This class gave an overview of algorithms complexity and techniques. Professor Tiago de Oliveira Januário presented a series of challenging computational problems which students had to solve using techniques like dynamic programming, recurrences, or greedy algorithms. In addition, a paper entitled “Robot Routing: Genetic Algorithm Applied to Travelling Salesman Problem” was written to demonstrate how to solve an NP-hard problem.
- **MATE66 - Computer Science Research Fundamentals II** - The scientific methodology is one of the most important knowledge to develop solid research. This class was lectured by Luciano Rebouças de Oliveira who motivated the students to write scientific papers objectively and efficiently. Read and writing methods

were studied based on the book “Style - Toward Clarity and Grace” by Joseph M. Williams. As a result, the students produced a paper where the teacher evaluated the students’ writing skills. He simulated a journal submission process, thus all papers were evaluated using a blind review method. The paper wrote in this class was the start of the first paper published as a product of this research.

- **MATE32 - Topics on Computer Intelligence II** - Overview of machine learning algorithms focusing on automatic knowledge retrieval from datasets. Professors Ricardo Araújo Rios and Tatiane Nogueira Rios explained how to pre-process data properly and to analyze data using predictive probabilistic methods based on optimization. This class was focused on supervised learning algorithms and how to evaluate them.
- **MATE33 - Topics on Computer Intelligence III** - Professors Ricardo Araújo Rios and Tatiane Nogueira Rios taught about use clustering methods and unsupervised learning algorithms. The students were challenged with clustering problems and led to solving these problems using the main techniques available in the related literature. This class was essential to improve the skill of pattern identification on datasets.
- **MATE85 - Topics on Information Systems and Web I** - This class lectured by the advisor presented core topics on Linked Open Data and Web Semantic. The students developed projects using technologies related to this topic, like Jena and Protégé. This class did a substantial contribution to the research and software development.
- **MATA31 - Oriented Research** - This curricular component is used by the advisor to evaluate the research progress. Periodical feedbacks were given to the advisor who led the research to the correct path with suggestions and contributions.
- **MATA32 - Oriented Teaching Internship** - The advisor offered to the Ph.D. student the experience of teaching to undergraduate students. Additionally, the student gave support for lectures preparation and assisted the class on projects.
- **Classes Dispensed** - Some classes were valuable to the research but they were completed during the master degree on Programa de Pós-Graduação em Computação Aplicada (PGCA). For this reason, the following classes were dispensed by the Programa de Pós-Graduação em Ciência da Computação (PGCOMP):
  - MATE04 - Topics on Databases I
  - MATE10 - Topics on Computer Intelligence I

### 5.1.2 Publications and Participation in Scientific Conferences

It is important to the Ph.D. student to participate in scientific conferences and to publish the research results on relevant journals. Conferences are a channel to disclose and discuss

the research, exchanging information with other researchers from related areas. Following are described the published papers and the conferences attended.

- **WebMedia 2018 - Brazilian Symposium on Multimedia and the Web** - it is the main event of the theme in Brazil and an excellent opportunity for scientific and technical exchange among students, researchers and professionals in the areas of Multimedia, Hypermedia and Web. We published the paper (ALMEIDA; DURÃO, 2018) presenting the preliminary results of our method to enhance the SKPQ accuracy using Linked Open Data.
- **J.UCS 2018 - Journal of Universal Computer Science** - this journal is run by the J.UCS consortium consisting of research institutions from Austria, Germany, Guatemala, USA, and Pakistan. We published the paper (ALMEIDA; DURÃO; COSTA, 2018) detailing our approach to enhance the SKPQ and discussed all obtained results in the experimental evaluation.
- **AMCIS 2019 - Annual Americas Conference on Information Systems** - AMCIS is viewed as one of the leading conferences for presenting the broadest variety of research done by and for information technology academicians. Every year its papers and panel presentations are selected from over 700 submissions, and the AMCIS proceedings are in the permanent collections of libraries throughout the world. From a collaborative work we did a paper about exploiting Web features for relevance feedback. This paper has been accepted and it is in the publishing process.
- **WE.PGCOMP 2018** - this conference is a curricular component too. After the second year as a Ph.D. student, once per year, the student has to present his research progress to three professors of the doctoral program and to an audience. The research progress is evaluated by the professors who ask questions and make great suggestions for the research.
- **ESWA 2019 - Expert Systems With Applications** - is a refereed international journal whose focus is on exchanging information relating to expert and intelligent systems applied in industry, government, and universities worldwide. The personalization approach we described to improve the SKPQ will be reported in a paper which will be submitted to this journal.
- **Data Mining and Knowledge Discovery 2020** - the premier technical publication in the field, it is a resource collecting relevant common methods and techniques and a forum for unifying the diverse constituent research communities. We plan to submit a paper to this journal in 2020 describing new techniques and experimental evaluations.

## 5.2 SCHEDULE

Until now we have proposed two techniques to improve spatial keyword queries accuracy. There were several experimental evaluations demonstrating query improvement. How-

ever, there is room to further improve our work. Under those circumstances, Figure 5.1 describes the activities of this research. Activities A01 to A10 was developed before the qualification exam, while A10 to A17 will be developed before the thesis defense.

- **Activity A01 - Academic Classes** - describes the time consumed to complete all classes demanded by the doctoral program. During this time the theoretical background was consolidated. All classes have been described in Section 5.1.1
- **Activity A02 - Workshops** - meetings involving the research members of the RecSys group. Each meeting consists of members presentation about their research topics followed by practices exercises. There was a workshop about SPARQL queries and another about Jena and Web Semantic.
- **Activity A03 - Defining research topic** - studies related to the research topic definition. Many papers were read and drafts were written during this process.
- **Activity A04 - Preliminary experiments** - the first rounds of experiments on our approach based on LOD to improve spatial keyword queries.
- **Activity A05 - J.UCS 2018** - the first publication inside the doctoral program. This activity consumed much time because it was the first related to this topic. In this paper, we talked about the benefits of using LOD to enhance textual descriptions of objects and how it improves spatial keyword queries. Several evaluation experiments were described and analyzed.
- **Activity A06 - Query personalization** - after the first publication, the work on the second technique was started. In this activity we started the work on the personalization of spatial keyword queries. The experimental evaluation indicates a relevant improvement on these queries accuracy.
- **Activity A07 - WebMedia article 2018** - the first paper submitted to a conference during the doctoral research. Motivated by the late review response on the J.UCS article, in this paper, we described some of the experiments on the LOD enhancement. The conference was a pleasurable experience where was possible to discuss the research topic with experienced researchers.
- **Activity A08 - AMCIS 2019** - in this activity, the article was revised and the journal's reviewers demands were attended.
- **Activity A09 - ESWA 2019** - the second paper describing the personalization approach is submitted and we are waiting for the journal response.
- **Activity A10 - Qualification** - write qualification text and prepare the qualification presentation.
- **Activity A11 - New optimization research** - explore new methods to improve spatial keyword queries. New algorithms or hybrid solutions can be proposed in this stage.

- **Activity A12 - Implementation of the new optimization** - this activity will implement the strategies or algorithms defined in the previous activity.
- **Activity A13 - Experiment results** - after the development stage it is necessary to evaluate the algorithms or method developed.
- **Activity A14 - Experiment analysis** - the experimental results are analyzed and compared with previous results, as well as, with the results reported in the literature.
- **Activity A15 - Preliminary results paper** - another paper will be published introducing the new approach and the results obtained so far. It is possible to submit on paper to a journal and another to a conference in this stage.
- **Activity A16 - Thesis** - activity defined to write the thesis.
- **Activity A17 - Thesis defense** - designated to thesis presentation and research end.

The schedule of activities executed since the course enrollment together with the ones that will be executed after the qualification exam are presented in Figure 5.1.

### 5.3 SUMMARY

The aforementioned chapter described the student life academy until now. It was depicted the classes completed and how they contributed to the research. In addition, it is possible to see the timeline between the publications and the stages before them. On the other hand, the schedule illustrates the time left to the thesis defense and future works followed by each scheduled date.

Date		Activities																
Year	Month	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16	A17
2016	Nov																	
	Dec																	
2017	Jan																	
	Feb																	
	Mar																	
	Apr																	
	May																	
	June																	
	July																	
	Aug																	
	Sept																	
	Oct																	
	Nov																	
	Dec																	
2018	Jan																	
	Feb																	
	Mar																	
	Apr																	
	May																	
	June																	
	July																	
	Aug																	
	Sept																	
	Oct																	
	Nov																	
	Dec																	
2019	Jan																	
	Feb																	
	Mar																	
	Apr																	
	May																	
	June																	
	July																	
	Aug																	
	Sept																	
	Oct																	
	Nov																	
	Dec																	
2020	Jan																	
	Feb																	
	Mar																	
	Apr																	
	May																	
	June																	
	July																	
	Aug																	
	Sept																	
	Oct																	

**Figure 5.1** Ph.D. activities since the course enrollment until thesis defense.

## BIBLIOGRAPHY

ALMEIDA, J. P. D. de; DURÃO, F. A. Improving the spatial keyword preference query with linked open data. In: SBC. *Anais Estendidos do XXIV Simpósio Brasileiro de Sistemas Multimídia e Web*. [S.l.], 2018. p. 19–24.

ALMEIDA, J. P. D. de; DURÃO, F. A.; COSTA, A. F. da. Enhancing spatial keyword preference query with linked open data. v. 24, n. 11, p. 1561–1581, nov 2018.