Google Research Awards Proposal

1. Overview

Title:	**Exploiting Open Data for Improving Spatial Keyword Query Applications**
Primary investigator:	**Frederico Araújo Durão**
Affiliation:	Mathematic Institute - Federal University of Bahia (UFBA)
	Av. Adhemar de Barros, Ondina, Salvador, BA, Brazil, fdurao@ufba.br
	+55 71 99613-6523

2. Proposal

Abstract

This proposal aims at proposing strategies to improve spatial-keywords queries results. First, we propose to exploit Linked Open Data (LOD) datasets (e.g., DBpedia and LinkedGeoData) to improve the points of interest (POI) description. Then, we aim at personalizing the query results to present the best POIs for the underlying user. By exploiting reviews on POIs, the system identifies the object which best satisfies the user and re-order the rank concerning for the user preference.

Problem Statement and Research Goals

The problem this proposal aims at solving is the incapability of single spatial keyword querying to satisfy user needs by neglecting important and available information. Usually, these queries manipulate objects with short textual description, hindering the query capability to identify objects to satisfy the user need. Moreover, traditional keyword queries just consider the user information need depicted by query keywords, overlooking that users have personal preferences even when they type the same query keywords.

Tackling this problem, we propose a location-based solution that exploits the benefits of a LOD dataset to enhance the object's textual description. LOD enables applications to navigate along with links into related data sources. By navigating these links, the approach obtains an enhanced textual description for the objects. In addition, the query results are personalized to consider users unique preferences. A classifier exploits user reviews to define the user preference for each object in the query result.

Work Description and Expected Outcomes

In order to improve spatial keyword queries, the solution combines query personalization with the object's description enhancement. This way, the best item for the user is presented first, improving the user experience with the system. We already built a search engine prototype able to apply the solution to the top-k Spatial Keyword Preference Query [2]. This query searches for points of interest based on other objects (features) in their spatial neighborhood.

Figure 1 describes an overview of our approach where the query result is improved automatically. Given a user query, the search engine first searches for features inside each point of interest's neighborhood. Then, it enhances the feature's textual description using LinkedGeodata and DBpedia (LOD datasets). A score function defines the relevance of each feature description based on the query keywords. Subsequently, the objects are ordered according to their scores and sent to the personalization algorithm. Finally, the personalization algorithm re-order the objects based on the user reviews and send the results to the user.
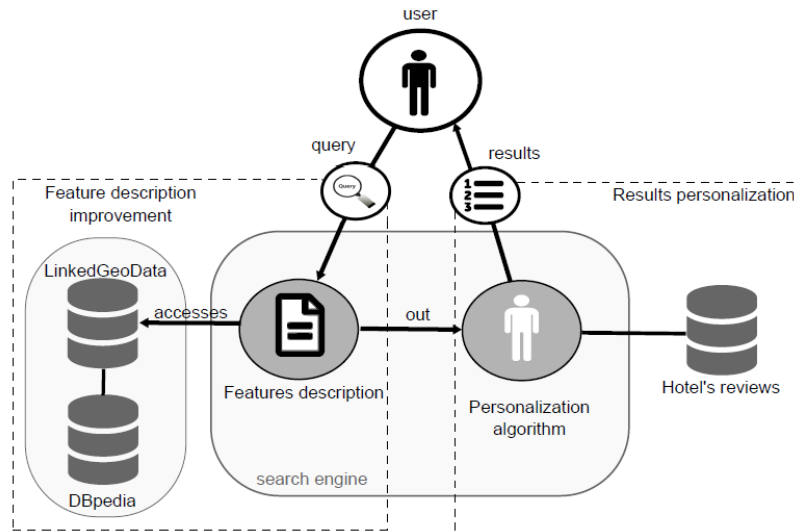

Figure 1. Overview of the approach to improve query results.

The LOD datasets will be accessed by their respective SPARQL endpoint to process the query. SPARQL is a query language that can be used to express queries across diverse LOD sources. A SPARQL endpoint is used to enable users to query a knowledge base via SPARQL. DBpedia and LinkedGeoData endpoints can be accessed at http://dbpedia.org/snorql/ and http://linkedgeodata.org/sparql, respectively. Both DBpedia and LinkedGeoData have public access.

In order to personalize the query, we will build user profiles and choose the suitable classifier to work with these profiles. The user profile will be composed of the user's past reviews and a label indicating whether the review is positive or negative. The Google Places API can be used to extract the user reviews and ratings required to build the user profile. A classifier, trained with the user profile, classifies each POI review in the query result as good or bad to the query user. In fact, the classifier compares the query user review with the ones in the database. Whether the POI review is similar to the query user review, it receives a value of 1 (good); otherwise, it receives 0 (bad).

At the end of this project, we expect to build a search engine capable of improving spatial-keyword query results automatically. Leastwise, the described approaches may improve the Spatial Keyword Preference query results, but we hope to build a generic search model able to improve any spatial-keyword query. Moreover, the search engine will be used to build a prototype Web service that, given a set of keywords, and a city of interest, generates a list of POIs that satisfy the user need in that city.

Relation to Prior Work

Several studies employ LOD datasets to improve textual descriptions of spatial objects. [5] describe an augmented reality browser that uses LOD to enhance the description of objects. Similarly, [6] employ LOD to improve POIs description, but they use voluntary users to generate, update and revise the POI's description. On the other hand, [2] discuss a detailed approach on how to improve a POI description automatically, using LOD.

Personalizing services offer relevant information tailored to user preferences and behavior. Some approaches deal with personalization of spatial query results employing methods based on the user's current location and schedule [7], or user weights on POIs [4], or user formulary describing her preferences [3].

References

[1] ALMEIDA, J. P. D de; DURÃO, F. A.; COSTA, A. F. Enhancing Spatial Keyword Preference Query with Linked Open Data, v. 24, n. 11, p. 1561-1581, 2018.

[2] ALMEIDA, J. P. D de; ROCHA-JUNIOR, J. B. Top-k Spatial Keyword Preference Query. Journal of Information and Data Management - JIDM, v. 6, n. 3, p. 162-177, 2016.

[3] BOUHANA, A. et al. An Ontology-based CBR Approach for Personalized Itinerary Search Systems for Sustainable Urban Freight Transport. Expert Systems with Applications. Elsevier, v. 42, n. 7, p. 3724-3741, 2015.

[4] DAI, J. et al. On Personalized and Sequenced Route Planning. World Wide Web, Springer, v. 19, n. 4, p. 679-705, 2016.

[5] HEDGE, V. et al. Utilising Linked Data for Personalized Recommendation of POI's. International AR Standards Meeting, Barcelona, Spain. [S.l.: s.n.], 2011.

[6] KARAM, R.; MELCHIORI, M. Improving geo-spatial linked data with the wisdom of the crowds. ACM. Proceedings of the joint EDBT/ICDT 2013 workshops. [S.l.], p. 68-74, 2013.

[7] KWON, O.; SHIN, M. K. Laco: A Location-aware Cooperative Query System for Securely Personalized Services. Expert Systems with Applications, Elsevier, v. 34, n. 4, p. 2966-2975, 2008.

3. Proposal Context

João Paulo is currently a PhD candidate under a Brazilian Postgraduate scholarship (FAPESB). This grant proposal is related to João Paulo's PhD thesis topic on the Semantic Web, contributing with a personalization method to improve spatial-keyword queries, providing a Web Service prototype to search for points of interest, and datasets for other researchers evaluate and compare systems which employ spatial-keyword queries.

4. Budget

We request the value to support one PhD student for one year (US$ 1,200 monthly), as well as a modest conference travel support for the student (US$ 3,000).

5. Data Policy

All developed code and data sets will be made available at GitHub or another appropriate public repository. All papers derived from this work will be published preferentially in open-access journals.