Prototyping a Personalized Contextual Retrieval Framework

Damien Palacio
Department of Geography
University of Zurich
Winterthurerstrasse 190
8057 Zurich
Switzerland
damien.palacio@geo .uzh.ch

Guillaume Cabanac
IRIT
University of Toulouse
118 route de Narbonne
F-31062 Toulouse cedex 9
France
guillaume.cabanac@univ-tlse3.fr

Gilles Hubert
IRIT
University of Toulouse
118 route de Narbonne
F-31062 Toulouse cedex 9
France
gilles.hubert@univ-tlse3.fr

Karen Pinel-Sauvagnat
IRIT
University of Toulouse
118 route de Narbonne
F-31062 Toulouse cedex 9
France
karen.sauvagnat@irit.fr

Christian Sallaberry
LIUPPA
University of Pau
Avenue de l'Université
64000 Pau
France
christian.sallaberry@univ-pau.fr

ABSTRACT

We introduce a framework for searching places according to user interests and spatial context. Our framework combines existing geo-tools or services (e.g., Google Places, Yahoo! BOSS Geo Services, PostGIS, Gisgraphy, Geonames) and ranks results according to features such as distance, popularity, and user preferences. We used this framework to participate in the TREC 2013 Contextual Suggestion Track.

Categories and Subject Descriptors

H.3 [**Information storage and retrieval**]: Information Search and Retrieval, Systems and Software

Keywords

Geographic Information Retrieval, Contextual Information Retrieval

1. INTRODUCTION

Location-based services rely on user location as a contextual clue to recommend places if he/she looks for points of interests (POIs) or nearby activities like restaurants or museums [4]. Nevertheless proximity should not be the only aspect to consider. Information such as user profile or previous queries can refine the context and be used by search engines to personalize results [5].

GIR'13, November 05 - 08 2013, Orlando, FL, USA

http://dx.doi.org/10.1145/2533888.2533935

We introduce a framework extending prior work [3]. The framework, based on Google Places, processes both user preferences and context. It ranked first among the 27 runs from 14 participants in the TREC 2012 Contextual Suggestion Track. This campaign deals with geopraphic user contexts, i.e. user location. Nevertheless, the framework had some limitations: result contents (i.e., webpages) are not checked to exclude inadequate pages (e.g., general information not related to a given location); user context is not considered to provide more accurate information about the places (orientation, distance in meter/time, others interesting POIs around, and so on). To overcome such limitations, the new framework triggers additional geo-tools and services (such as Yahoo! BOSS Geo Services, PostGIS, Gisgraphy and Geonames), as well as filtering and ranking processes based on all collected information (e.g., preferences, popularity, proximity). We tested the framework with the test collection provided by the TREC 2013 Contextual Suggestion Track, which is made up of users' profiles and contexts.

2. A PERSONALIZED CONTEXTUAL RETRIEVAL FRAMEWORK

The framework builds on a process flow composed of the following four main stages.

2.1 Preference Processing

Using the user's ratings on previous suggested places or examples, we first define categories of interest. These categories either correspond to those of Google Places or to our own defined categories (extended with their hyponyms using WordNet). Tying up with principles of content-based filtering [1, 2] we also build positive and negative term preferences for the user (Fig. 1, process 1).

2.2 Context Processing

Context processing (Fig. 1, process 2) relies on the user

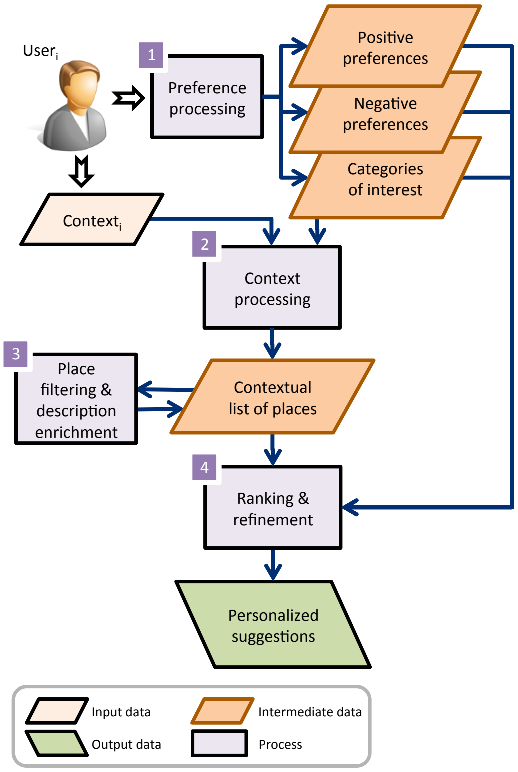

Figure 1: Architecture of our personalized contextual retrieval process

location (i.e., latitude and longitude) and the categories of interest for the user. As users' activities correspond to urban POIs (e.g., restaurants, museums), a dedicated external search engine such as Google Places can be used to retrieve potential suggestions. The external resource returns a list of relevant places described by: a place name, a website URL, and a relevance rank.

2.3 Place Filtering and Enrichment

This process (Fig. 1, process 3) consists of two phases. First, it filters inadequate pages (URLs) with geoparsing tools (Yahoo! BOSS Geo Services). A page is discarded when no relevant placename is detected within its text. Considering all user contexts, overly frequent URLs are also blacklisted (e.g., global websites like "mcdonalds.com" that does not inform about a particular restaurant but list all existing restaurants). Second, the descriptions of remaining places are enriched. We use distance computing tools (Google MatrixDistance) to get foot distance and car distance between the user location and that of the retrieved places. Other geotools (PostGIS) are used to calculate Euclidian distance and orientation details between the user location and the retrieved places. Finally, a set of nearby POIs (generated by Gisgraphy with Geonames) is associated with each place.

2.4 Suggestion Ranking and Refinement

This process (Fig. 1, process 4) generates descriptions summarizing all the data found. It also ranks places according to the initial context and user profile. Two main criteria are used for calculating a place score: the *category confidence*, the sum of the confidence scores of the positive categories shared by the user context and the place retrieved (Fig. 1, process 3); the *positive (resp. negative) cosine*, the cosine score between the vectors representing the positive (resp. negative) user preferences and the place description (Fig. 1, process 3). Both are weighted by 1/3. Four other criteria contribute to the score ranking with a weight of 1/12: *rating*, the mean score given by Google Places users (Fig. 1, process 2); *rank*, the rank of the place in Google Places (Fig. 1, process 2); *shared categories*, the number of categories in common between the user context and the place retrieved (Fig. 1, process 3); *Euclidian distance*, the distance between the user location and that of the place (Fig. 1, process 3).

The suggestion issued reports the results of the third stage. For example: *Rank: 1 • Title: "Bennigan's Grill and Tavern" • Description: "Type of place: bar, establishment, food, restaurant. This place is about 0.2 Km Southwest from here (1 min by car with no traffic). Address: 700 East Adams Street, Springfield. There are 11 POIs around: 2 Hotels, 3 Libraries, 4 Parks, 1 PostOffice, 1 Religious" • URL: . . .* A query for such a result could be: "Springfield", "restaurant,. . . , museum" (categories extracted from user's profile).

3. CONCLUSION

The personalized contextual framework introduced is based on a modular architecture. It combines existing geo-tools and services, while embedding filtering and ranking processes based on collected information (e.g., preferences, popularity, proximity). We experimented it with the collection provided by the TREC 2013 Contextual Suggestion Track.

As preliminary results, we already know that our framework has a P@5 of 0.3112. This is comparable to the results we got at the previous campaign. Among the 223 queries, our system has the highest P@5 for 10.3% and is at least equal to the median for 82% of the queries. Full evaluation results, in particular comparison of participants, are to be released in November 2013 (`http://trec.nist.gov`). Future work will tackle ranking algorithms giving a modulated weight to the different arguments involved.

4. REFERENCES

[1] N. J. Belkin and W. B. Croft. Information filtering and information retrieval: two sides of the same coin? *Commun. ACM*, 35(12):29–38, 1992.

[2] K. Hoashi, K. Matsumoto, N. Inoue, and K. Hashimoto. Document filtering method using non-relevant information profile. In *Proceedings of the 23rd annual international ACM SIGIR*, pages 176–183, New York, NY, USA, 2000. ACM.

[3] G. Hubert and G. Cabanac. IRIT at TREC 2012 Contextual Suggestion Track. In *Proceedings of the 21th Text REtrieval Conference*, pages 1–8, Gaithersburg, MD, USA, 2012.

[4] B. Jiang and X. Yao. Location-based services and GIS in perspective. *Computers, Environment and Urban Systems*, 30(6):712 – 725, 2006.

[5] M. Speretta and S. Gauch. Personalized search based on user search histories. In *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, pages 622–628, 2005.