



DUTH AT TREC 2013 CONTEXTUAL SUGGESTION TRACK

{ G. DROSATOS, G. STAMATELATOS, A. ARAMPATZIS AND P. S. EFRAIMIDIS }

DEMOCRITUS UNIVERSITY OF THRACE, GREECE
ATHENA RESEARCH & INNOVATION CENTER, GREECE



ABSTRACT

We present an approach for context processing that comprises a carefully designed and fine-tuned POI (Point Of Interest) data collection technique, a crowdsourcing approach to enrich our data collection and two radically different approaches for suggestion processing (a k-NN classification-based and a Rocchio-like).

In the *context processing*, we collect POIs from three popular place search engines, Google Places, Foursquare and Yelp. The collected POIs are enriched by adding snippets from the Google and Bing search engines using crowdsourcing techniques.

In the *suggestion processing*, we propose two methods:

1. The first submits each candidate place as a query to an index of rated examples and scores it based on the top- k user's ratings.
2. The second method is based on Rocchio's algorithm and uses the rated examples per profile to generate a personal query which is then submitted to an index of places.

INTRODUCTION

TREC 2013 is the second year that the Contextual Suggestion Track is running. The track's goal is to investigate search techniques considering as context only the user's location, as well as considering user interests via personal preferences and past history. In other words, the track focuses on one situation: a user with a mobile device with limited interaction but some sort of user profile; who is in a strange town; and who is looking for something to do. There is no explicit query: the implicit query is "Here I am, what should I do?".

REFERENCES

- [1] Thomas M. Cover and Peter E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21-27, 1967.
- [2] J. J. Rocchio. Relevance feedback in information retrieval. In Gerard Salton, editor, *The SMART Retrieval System - Experiments in Automatic Document Processing*, pages 313-323. Prentice Hall, Englewood Cliffs, NJ, 1971.

CONTACT INFORMATION

Web: <http://utopia.duth.gr/~gdrosato>
Email: gdrosato@ee.duth.gr
Phone: +30 2541079756
vCard:



CONTEXT PROCESSING

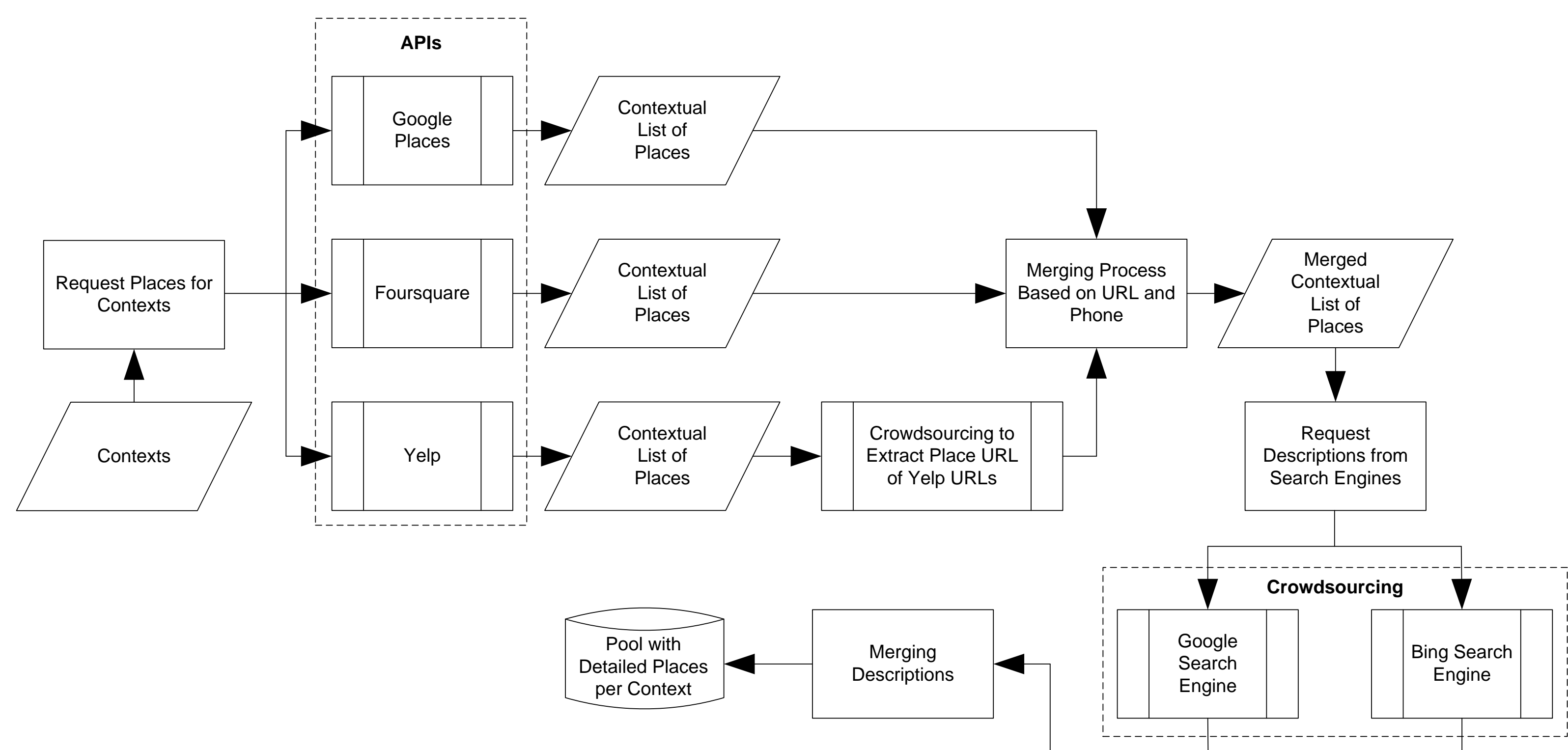


Figure 1: A system flowchart of the context processing.

SUGGESTION PROCESSING

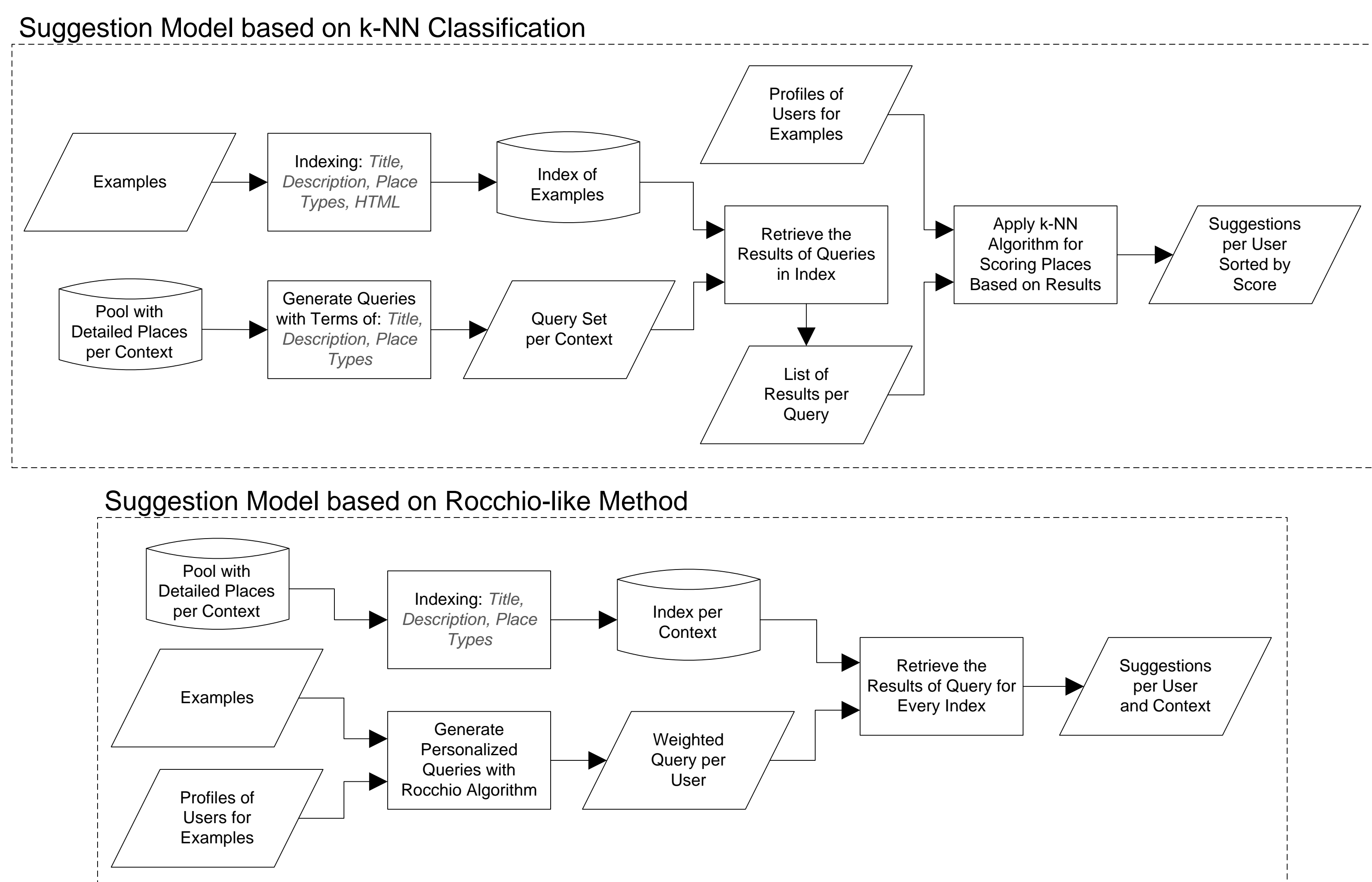


Figure 2: An overview of the proposed suggestion models.

- *Suggestion model based on k-NN classification method:* In this model, we propose the usage of places as queries that are requested in the index of examples and we score the results by using tf-idf weights on k-NN algorithm [1]. The scoring of results is based on the profiles of users. (Run DuTH_A)
- *Suggestion model based on modified Rocchio relevance feedback method:* In this model, we use the examples and profiles to generate a weighted personal query per user that is requested then in the index of places per context. For the weights of terms in personal query, we use the Rocchio algorithm [2]. (Run DuTH_B)

RESULTS

	P@5	MRR	TBG
<i>Runs:</i>			
DuTH_A	0.3283	0.4836	1.3109
DuTH_B	0.4090	0.5955	1.8508
<i>Difference:</i>			
DuTH_B vs _A	+24,58%	+23,14%	+41,19%

Table 1: Mean of results over all the profiles and contexts for P@5, MRR and TBG measures.

Runs	Median-or-better			Best		
	P@5	MRR	TBG	P@5	MRR	TBG
DuTH_A	189	175	151	25	86	22
DuTH_B	209	206	185	47	114	40
Total: 223 judged context-profile pairs						

Table 2: Number of context-profile pairs with Median-or-better and Best scores per measure.

CONCLUSION

In the TREC evaluation results, both approaches seem very promising. DuTH_B (i.e. the Rocchio-like approach) performed better than DuTH_A. Compared to other groups, DuTH_B scored almost firmly above the median (in P@5 and MRR) and achieved the best results in almost half of the judged context-profile pairs (at MRR).

ACKNOWLEDGEMENTS

The present work was partially funded by the project ATLAS (Advanced Tourism Planning), GSRT/CO-OPERATION/11SYN-10-1730.