experience and performance) ranking result in the structure. The classical structure of searching includes three submodules, which are *crawler, indexer,* and *query*. *Crawler* downloads webpages and *indexer* build indices for those words in webpages, then *query* responds to user input by returning the most related pages.

With this basic model in hand, here is a possible usage scenario of open data in searching. In the first step of crawling, the crawling from both online and offline open data (such as geographic databases, Yellow Pages brochure, social network reviews, etc.) should be performed. Since the local area often has a limited range of business unit candidates within a certain radius, it is possible to collect information from multiple aspects and resources, when the single resource cannot generate a large enough document set about business units. The second step is indexing. This necessitates execution of the challenging task of merging multiple descriptions of the same entity, acquired from diverse information resources. The third step is the modified design that incorporates local search. Since GPS on mobile devices enables a real-time location record, a user's current location can trigger the preference estimation model given in Section 3. The model will use the information of surrounding business units acquired from geographic open data. Semantic open data can also work in the matching of query and business category. The model will then produce a list of nearby business units with their preference values for a user. The last step is the response to user's query. Traditional ranking results relate the semantic similarity between the input string and the candidate document with the index. Here we have another preference list based on the additional estimation model in Section 3. A suitable mix of the two methods should improve the searching performance.

Advantages. The use of open data in preference estimation could solve the problem of insufficient web-available information about local business units. Besides, the added step of geographic analysis can also serve for a local recommendation system before the user's query. Meanwhile, the structure leaves room for incorporating other types of open data.

5. CONCLUSION

In this paper, we analyze the patterns of relationships between geographic features derived from open geographic data and user preference, and describe a preference model that incorporates several detailed geographic features. We discuss the potential improvement about the structure of local search for better preference estimation. The initial analysis tend to support that open data, and especially geographic open data, can be a powerful factor to estimate user preference, and local search incorporating a parser of geographic features might overcome a lack of descriptive words about business units.

Future possible directions of work include: (1) Collecting real query logs that track movement and evaluate the preference model and the revised local search. (2) Finding and determining more helpful geographic features. (3) Mining the patterns encoding the relationship geographic features in Section 2 and the preferences. At the same time, working on different scales of local data in terms of the size of a city and the radius of address query around a business unit. (4) Finalizing several sub-parts in local search algorithm, including the merge of information about the same entity, the cooperation of preference estimation model and traditional ranking method.

6. REFERENCES

[1] Ahlers, D. Business entity retrieval and data provision for yellow pages by local search. In *IRPS Workshop* (2013).

[2] Berberich, K., König, A. C., Lymberopoulos, D., and Zhao, P. Improving local search ranking through external logs. In *Proc. of the SIGIR* (2011), ACM, pp. 785–794.

[3] Church, K., and Smyth, B. Who, what, where & when: a new approach to mobile search. In *Proceedings of the 13th international conference on Intelligent user interfaces* (2008), ACM, pp. 309–312.

[4] Dragut, E. C., Dasgupta, B., Beirne, B. P., Neyestani, A., Atassi, B., Yu, C., and Meng, W. Merging query results from local search engines for georeferenced objects. *ACM Transactions on the Web (TWEB) 8*, 4 (2014), 20.

[5] Gan, Q., Attenberg, J., Markowetz, A., and Suel, T. Analysis of geographic queries in a search engine log. In *LocWeb Workshop* (2008), ACM, pp. 49–56.

[6] Geopy-1.11.0. Python geocoding toolbox. `https://pypi.python.org/pypi/geopy/`. Accessed Dec.15, 2015.

[7] Lv, Y., Lymberopoulos, D., and Wu, Q. An exploration of ranking heuristics in mobile local search. In *Proc. of the SIGIR* (2012), ACM, pp. 295–304.

[8] Lymberopoulos, D., Zhao, P., Konig, C., Berberich, K., and Liu, J. Location-aware click prediction in mobile local search. In *Proc. of the CIKM* (2011), ACM, pp. 413–422.

[9] Teevan, J., Karlson, A., Amini, S., Brush, A., and Krumm, J. Understanding the importance of location, time, and people in mobile local search behavior. In *Proc. of the MobileHCI* (2011), ACM, pp. 77–80.

[10] Yelp. Yelp data challenge dataset. `https://www.yelp.com/dataset_challenge/dataset/`. Accessed Dec.2, 2015.

[11] Zhou, Y., Xie, X., Wang, C., Gong, Y., and Ma, W.-Y. Hybrid index structures for location-based web search. In *Proc. of the CIKM* (2005), ACM, pp. 155–162.