Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)

Urban Maps of Social Activity

Stewart Whiting, Omar Alonso,
Vasileios Kandylas, Serge-Eric Tremblay
Microsoft
stewh@outlook.com,{omalonso, vakandyl, sergetr}@microsoft.com

Abstract

Recreational queries from users looking for ideas of what to do or see and where to go are very common in desktop, mobile and, increasingly, contextual search scenarios. Such queries typically contain {what, where, when} components as the user seeks future activities, whether in real-time or future trip planning. Often, the user will have additional constraints and requirements for what they are seeking, such as suitability for kids, budget, for a romantic date, etc. Currently, simple recreational queries (e.g. "restaurants in Mountain View" are served by static local results, or the Points of Interest thumbnail carousel (e.g. "things to see in San Francisco"). More complex recreational queries, such as "romantic places to eat in San Francisco friday night", or "educational places to visit with kids nearby" require the user to click on the 10 blue links to read through articles from sites such as TripAdvisor and WikiTravel to satisfy their needs. Employing location based social network data, we construct urban maps of social activity for answering recreational queries using a model based on social, geographical, and temporal information. We demonstrate the feasibility of our approach using a data set of 1B check-ins.

Introduction

"What should I do?" (i.e. activity, or, what/where/when) queries from people looking for recommendations on places to visit, eat, drink, shop etc., are prevalent as people seek real-time and future ideas – both from desktop and increasingly, mobile search.

Currently, some activity queries are served by answers containing either static recommendations of the most common points of interest (e.g., "what to do in San Francisco" presents a carousel of POI thumbnails), or a static list of relevant Yelp places ranked by review (e.g., "pizza in San Jose"). Both these existing approaches disregard several implicit (e.g., the user's local query time and location) and explicit signals (e.g., temporal qualifier such as "this weekend" expressed in the query) which define the user activity motivations and therefore expectations – and consequently, what they consider as relevant activity recommendations (e.g., is the place open at the time the user intends?). Hence, out-of-season or currently closed places (e.g., beaches) will always

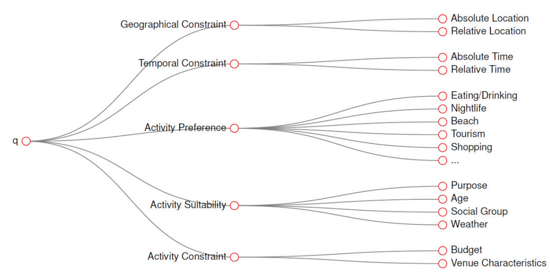

Figure 1: Top level taxonomy of recreational queries.

rank highly, regardless of their contextual relevance. Therefore, users often need to consult multiple web search results to distill suitable ideas. Relevance could be characterized as is the place open at the time the user wants to go? Is the place bustling with people at that time? Is the place a suitable choice given the inclement weather conditions?

Using 1B check-ins from Foursquare and Facebook with a solid understanding of periodic activity patterns, we built a temporal model of things to do in cities across the world. The work presented uses time, social signals, and location so users can get the best recommendations. Our proposed relevance model incorporates the following factors: 1) Spatial (i.e. city-based), 2) Temporal (e.g. (i) periodic: time of day, day of week, weekday/weekend, season, etc., and (ii) trending: short-term popularity such as events), and 3) Taxonomy of aspects (e.g., geography, time, activity preference, activity suitability, and activity constraint).

Related similar work includes a characterization of urban environments using Foursquare and call detail records (Noulas, Mascolo, and Frías-Martínez 2013) and a cluster-based taxonomy for three European cities (Ruiz et al. 2015).

Urban Maps Demo

In contrast to traditional approaches for recommending a POI based on past activity, our goal is to algorithmically match venues to searches based on crowd sensing around the world in the form of check-ins and tips. Simultaneously,

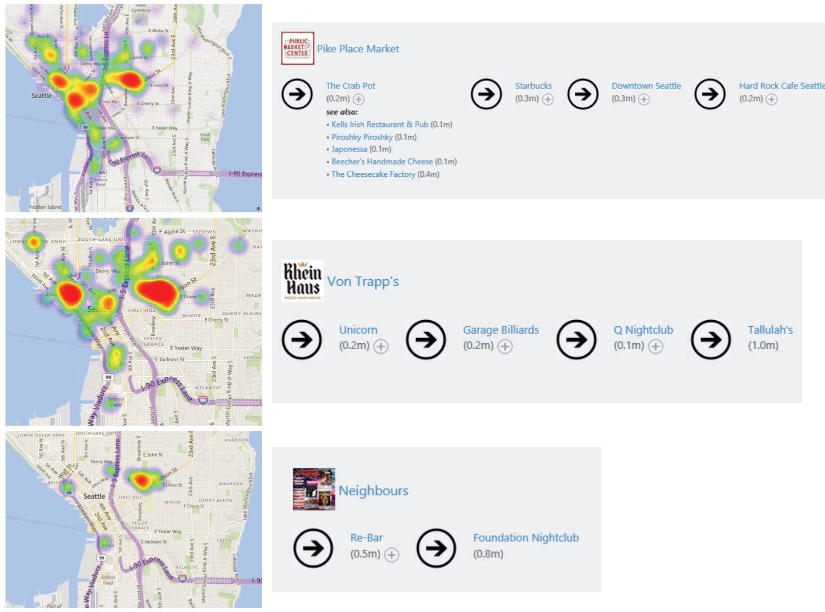

Figure 2: Urban maps for Seattle for the query "things to do in seattle": 1) Things to do over the weekend showing the famous Pike Place Market as first option and suggested trails of nearby attractions along with their distance, 2) Results for restaurants at night and recommendations for what do to after dinner, 3) Night clubs.

people are looking for things to do and understanding the what/where/when patterns allows us to rank appropriately. Finally, temporal information is used, in conjunction with location, to generate the maps.

We start the data mining process as follows. We sample the query logs of the Bing search engine where queries contain a pattern seed (e.g., "things to", "places to", "what to", etc.). We then annotate queries with location information using an internal tagger with the goal of selecting a recreational sample that contains the seed patterns with the presence of location entities, very similar to a geographical query. Examples of such queries are "things to see in San Francisco", "places to visit in Seattle", "what to do in Paris". Additionally, we follow the same process using browser log behavioral data (e.g., queries, clicks, and visited links).

With all these data, we produce a data-driven taxonomy of query constraints that allow us to provide contextual recommendations based on time and location. The taxonomy contains the following top-level aspects: geography (e.g., "near", "around"), temporal (e.g., "now", "this weekend"), activity preference (e.g., "eating", "drinking"), activity suitability (e.g., "romantic", "kids"'), and activity constraint (e.g., "cheap", "free"). Figure 1 shows more details.

We model relevance using check-in data by location, time, and tips. Check-in data tell us *what* places are popular in what areas and check-in times tell us *when* places are popular. Finally, tips or reviews that are written in English are useful for extracting more information about a place (e.g., "good for kids", "are romantic", etc.) in the proposed taxonomy. The underlying framework is described in (Whiting and Alonso 2016).

We also introduce suggested trails of places to visit. The trails are computed based on paths between POIs incorporating (1) place popularity, (2) POI category to POI category transition popularity, and (3) distance between places. It is built upon re-ranking the current results to find logical trails between them. Thus, it takes into consideration the popularity of venues in the temporal context and aspects. Figure 2 shows search results for things to do in Seattle with different aspects and constraints.

Conclusion

We showed how to use check-in data to build maps of social activity to recommend places for recreational queries, a vertical experience that web search engines satisfy by offering a manually curated list of POIs. The derived data set described in this demo is currently deployed as part of Bing carousel where users can try such queries but with a limited set of aspects.

References

Noulas, A.; Mascolo, C.; and Frías-Martínez, E. 2013. Exploiting Foursquare and Cellular Data to Infer User Activity in Urban Environments. In *Proc. of IEEE ICMDM*, 167–176.

Ruiz, C. K. V.; Quercia, D.; Bonchi, F.; and Fraternali, P. 2015. Taxonomy-Based Discovery and Annotation of Functional Areas in the City. In *Proc. of ICWSM*, 445–453.

Whiting, S., and Alonso, O. 2016. SGT Framework: Social, Geographical and Temporal Relevance for Recreational Queries in Web Search. In *Proc. of SIGIR*, 1001–1004.