FootTraffic

A Traffic-Driven Location-Based Web Search Engine

Proposal

Department of Computer Science & Engineering
Texas A&M University

Authors:	*Emails:*
William Chen	wchen16@cse.tamu.edu
Eric Wood	ebw0178@cse.tamu.edu

October 21, 2011

1 Introduction

We are seeing an increasing amount of location sharing from social services Foursquare, Facebook, Google Latitude, etc., and we want to harness these traffic to recommend venues to users based on the traffic patterns. Most location-based search engines today rank results based on proximity, user rating, category, and popularity. We want to add another dimension to location-based searches, and that is ranking based on observed traffic similarities. This "*temporal dynamics embedded in the checkins* from location sharing services" [1], approach is introduced by a research paper authored by Zhiyuan Cheng, James Caverlee, Krishna Y. Kamath, and Kyumin Lee, and the fountain of FootTraffic search engine will be based on the ideas from the paper [2] aforementioned.

2 Purpose

Adding a traffic-driven dimension to location-based searches serves many purposes, and to illustrate, below are some user-case scenarios:

1. Reviews are relevant to products, because their quality doesn't fluctuate, and availability (inventory) is reflected real-time. For venues like restaurants, a raving review dated a year ago doesn't mean the restaurant isn't infested with rats and ultimately closed down today. If we observe a plummet of checkins, we can make our results more relevant and up to date.

2. Suppose you are new to an area and you wish to see what bars in the local area airs the NFL football games. You can review the traffic graph for the local bars and correlate the previous game days with the at-peak checkins

3. Tina and her friends are celebrating graduation, and they are looking for bars to hit late at night. What bars are at-peak and what bars will be at-peak when they arrive? [3]

4. John has moved to a new area and would like to know which basketball courts have a similar traffic pattern like the one at home, so he can meet new friends. [3]

A traffic-driven location-based search can easily handle these kind of queries that a rating- and review-based search cannot, whether it is by returning semantically correlated venues with similar traffic patterns, or by having traffic patterns to make a decision.

[1] "Toward Traffic-Driven Location-Based Web Search" by Zhiyuan Cheng, James Caverlee, Krishna Y. Kamath, and Kyumin Lee; CIKM 2011

[2] A copy of the paper can be found at: `http://faculty.cs.tamu.edu/caverlee/pubs/cheng11cikm.pdf`

[3] Example is taken from the paper

3 Design and Methods

3.1 Design Specifications

We plan on using the Ruby on Rails framework for several reasons. Not only will it allow us to rapidly build the web application, it provides an overall architecture for us to build the data processing on top of as well as an easy and somewhat transparent way to interact with our database. In addition, the MVC architecture makes modulization easy, and that is a huge advantage for us. We plan on using MariahDB, a fork of MySQL as our main database.

3.2 Development Resources

For our search engine, we will be reusing the FourSquare checkin data collected from the research paper which is 22 million checkins gathered from Twitter's public streaming, from October 2010 to January 2011. The data can be publicly downloaded at infolab.tamu.edu. Using this dataset, we can extract a wealth of information, including location, time, category of the venues, frequency, tags, reviews, features, photos, etc. Needless to say, this 3 GB of dataset is the integral piece that drives our search engine.

In addition, in our preliminary brainstorm, we have decided to use Google Maps API to plot the different venues in a geographical map, and Yelp API to show restaurant reviews. Other APIs will possibly used as we add more modules to different venues and categories.

Last but not least, our class materials on optimization, indexing, evaluation, compression, and many more will prove invaluable.

3.3 Design Envision

Users can submit queries just like they would to Google Maps, using specific locations or addresses, or broad categories. We'll combine the Google results with information gathered from our database of FourSquare checkins as well as any number of rating services. We will allow the user to sort their results based on popularity for current or future dates, allowing them to find the location that fits their needs best.

3.4 Design Challenges

The main challenge we face, like any search engine, is relevancy and speed (optimization). In building FootTraffic, we want to weigh our deciding factors (location, review, traffic, time, semantic relations, etc) at the optimal balance, and it may differ for the different types of queries. Also, keeping such a large dataset with millions of locations and checkin data over the course of several years is not trivial, and querying it in a timely and efficient manner makes

it even more difficult. These problems are not new, however with the correct approach, the amount of information is manageable.

4 Management Plan

4.1 Roles

To manage our time successfully, we have set a private Git repository on GitHub to manage our codebase. We plan to meet at least twice a week and continually have dialogues via IRC. We will split development and documentation responsibilities equally.

4.2 Schedule of tasks

	Tasks	Due Date
1	Submit proposal	10/21/2011
2	Proposal feedback	10/27/2011
-	Complete Milestone Walk	10/30/2011
-	Complete Milestone Jog	11/15/2011
3	Project checkpoint	11/15/2011
4	Two slides	11/22/2011
-	Complete Milestone Sprint	11/27/2011
5	Finish in-class presentation	11/29/2011
-	Complete Milestone Run	12/10/2011
6	Finish executive summary + demo	12/12/2011