Summer Research Plan

Bethany Anne Waldmann
H.R. Bright Building, 402D

REU-Computer Science 2009
Center for the Study of Digital Libraries
Department of Computer Science and Engineering

Mentor:
Dr. James Caverlee

Direct Supervisor:
Jeremy Kelley
Student Mentee

Research Title
Exploiting Geolocation Data on Twitter: Users as Sensors

Description
Although Twitter was built for social reasons, there is a lot of valuable information available for mining. Using locations given by users in their profile as well as in tweets (individual messages), patterns can be discerned which prove interesting if not useful.

Purpose
This research project interests me because of the vast number of applications it could have. For instance, tweets mentioning sickness could be used to detect outbreaks before doctors are aware - people are much more likely to casually mention that they do not feel well than to go to the doctor right away. Another possible use would be to get real-time updates in disasters - nobody knows what’s going on more than the people who are actually there. There are many more possibilities, and the project’s versatility greatly appeals to me.

Project Goals and Implications
The copious areas of research that can be combined with location data to produce an interesting and/or useful set of data will surely attract the attention of researchers and consumers in many different disciplines. Using a network created only for social purposes as a resource has only recently been harnessed to obtain previously unrecorded (or private) data, such the number of people traveling on flights from/to different places. Advancing this research has much potential in the area of information retrieval, using us (humans) as sensors in a variety of ways.

Personal Goals
By the end of the summer I expect to:
- Produce findings of my research project
- Be more familiar with Python
- Feel more comfortable on the command-line
- Experience being part of a research environment
- Meet new people (including possible career contacts)
- Come to a more firm decision regarding whether or not I will attend grad school

Approach
Steps
1. Crawl through mined Twitter data (on local machine) and extract username:place pairs, as well as username:tweetid:locationreferenced:timestamp sets
2. Analyze availability of geographical information
3. Determine area to focus on and document expectations
4. Implement chosen topic
5. Review effectiveness of implementation
6. Repeat steps 4-6 as time allows
Daily Activities

* Post activity and progress to Twitter (2-3 times daily)
* Communicate with mentors (as necessary)
* Check in modified code (as applicable)

Weekly Activities

* Write in research diary (1-3 times weekly)
* Prepare a synopsis of what has been accomplished in the past week
* Revise research slides (big picture, research challenges, related work, concrete steps, initial results and thoughts)
* Update research website with relevant links, files, etc.
* Attend lab meeting (Tuesdays at 3pm)

Method & Materials

* Machines (personal, distributed)
* Network tools (SSH, FTP, VPN)
* Programs (Python + libraries, LaTeX, Emacs)
* Version control (git, github)

Work Schedule

Precise hours will not necessarily be followed, but time in the office will resemble one of the following schedules each week for a total of 40 hours

* M-F 8 hours/day (roughly 9:00am-5:30pm)
* M-R 9 hours/day (roughly 9:00am-6:30pm)
  F 4 hours (roughly 9:00am-1:00pm)

Extra hours may also be worked, either in the office or at another location

Deliverables and Dates

* 06-11-09 Research plan due
* 06-15-09 Complete crawl through local Twitter data
* 06-17-09 Analysis of geographical information available
* 06-19-09 Chosen area to implement
* 07-09-09 Progress report due
* 07-23-09 Research abstract due
* 07-24-09 Review of effectiveness of implementation
* 07-27-09 Rough draft of research paper
* 07-29-09 Rough draft of poster
* 08-02-09 Research poster due
* 08-04-09 Poster session
* 08-06-09 Final research paper due