CS 410 Project Report - BuddyMeet

Wenbo Yu (wenboyu2) | Xiaojie Zhang (zhang369)

Abstract

Buddymeet is a project of building a platform for students in U of I to explore and initiate events/activities. The activities are either user posted, or collected from public resources, for example, UIUC university calendar. Combined with our specialized search engine, and recommender our web service can help user find their interested events easily and accurately.

Related Work

There do exist projects which have similar event search functionality, such as EventFinda. Also, Facebook supports event pages and users can search the events in the search bar. However, all the current event finder webapps are searching for general events, like concerts and conference. Our project more focus on the school related events, for examples, group meetings, review sections, and free food events!!! Therefore, the website will be extremely attractive to college students, since it is more related to their daily lives.

Novelty and Challenge

Buddymeet is a fusion of public resource and private resource. On our website, user can make the most of the activities nearby. It is not only a website for people to share private events but also discover public events they have access to but may never get to know. Besides, we have a recommender system, based on the past events user has created or joined, and the interests he/she selected in our user profile setting, we can periodically deliver the most recent events he/she may be interested in. User experience is optimised by learning from users, and thus users can find events with minimal efforts. Also, we built a simple social network, to help people with similar interests or have attended many events together to get in touch.

The challenge is that our data set might not be big enough, since we are mainly focus on the on-campus events. We plan to crawl school events from the web, and user created event data to cover the shortage of data source.

Methods

Data collection

<u>Data format:</u>

```
  detail character varying(255),
  category character varying(255),
  title character varying(255),
  time character varying(255),
  date character varying(255),
  location character varying(255),
  url text,
  user_created boolean,
  source character varying(255)
```

Created by user

User are allowed to create their own event in the system. The events created by user are searchable and weight more than the crawled events. user_created = True, user_created = True.

Crawled by web crawler

We use Scrapy to crawl the data from public websites. Scrapy is a python based screen scraping web crawling framework.
We implemented the spiders to crawl the data from the following public websites.

UIUC gerneral event calender: http://illinois.edu/calendar/list/7
Champaign County evetn calender: http://www.visitchampaigncounty.org/calendar/
Wadoo, UIUC free food: http://uiucfreefood.com/

The data is crawled with the following format:
title, date, time, location, url and detail.
We stored the data crawled from these websites to local machine as a raw files named by timestamp. Eg. Below is the first event crawled from wadoo in file 2014-5-4_21:24.txt.

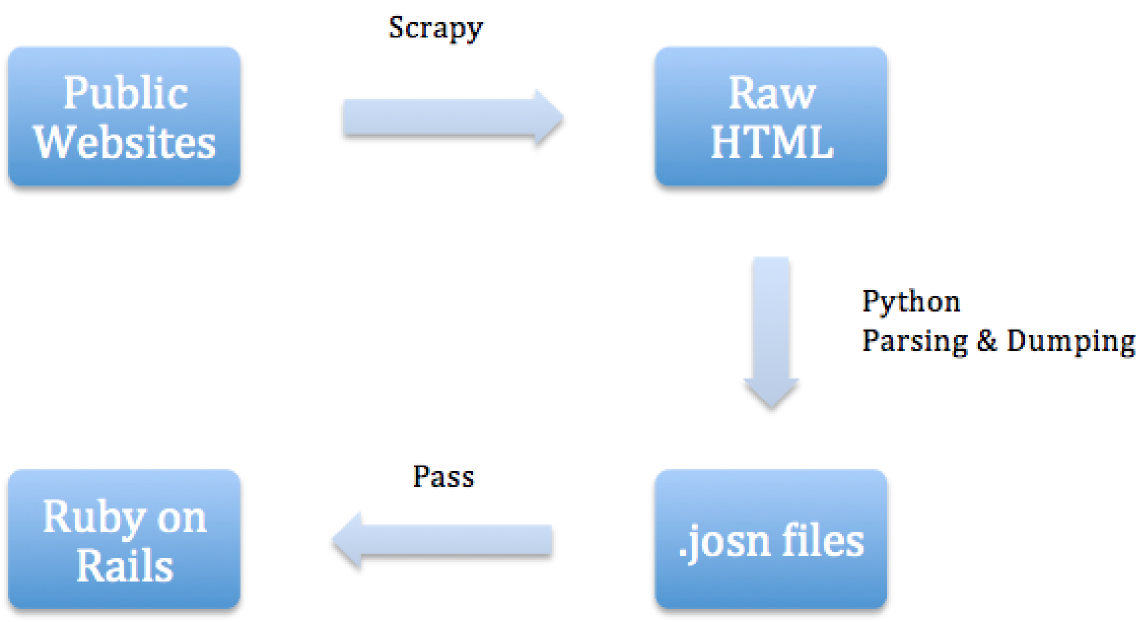
```
~event1~
IEI Reception
May 7, 2014
7:00 pm - 10:00 pm
Illini Union Ballroom
htttp://illinois.edu/calendar/detail/3126?eventId=31367865&amp;calMin=201402&amp;cal=20140224&amp;skinId=1
Come celebrate the end of the semester with the IEI!
…
```

Then, we write a python script to load the files and parsing (by “\n”) the format to form a json object with events array. Below is the example of generated 2014-5-4_21:32.json:

```
{"events": [
{
   "category": "food",
```

    "url":
"http://illinois.edu/calendar/detail/3126?eventId=31367865&amp;calMin=201402&amp;cal=20140224&amp;skinId=1",
    "location": "Illini Union Ballroom",
    "title": "IEI Reception",
    "date": "May 7, 2014",
    "time": "7:00 pm - 10:00 pm ",
    "detail": "Come celebrate the end of the semester with the IEI!",
    "user_created": false,
    "source": "UIUC Free Food"
},
….. // other events
}

Then, we pass the .json file to the Ruby on Rails, the load the data to the search engine. The data stream was shown as the graph below:



Search Engine

The search engine is powered by solr. Since the web service is built with Ruby on Rails, connection to solr is accomplished with sunspot. First, fill the collected data into the MySQL database. Then we use sunspot to index the data to make it ready for search. Inside rails event module, some attributes are selected for full-text search, and in the corresponding event controller, we customize the ranking by adding a few rules for boosting scores for certain data:

1. If query word is found in the “title” of the event, boost score by 1.0
2. If query word matches the “category” of the event, boost score by 2.0
3. If an event is user created (not crawled from public resource), boost score by 0.5

By adding this rules we make sure the ranking is optimized, and make the events user most interested in show up in front. We believe if the query word appears exactly in the category, the event should be a very good match; if the query word appears in the title, since the title is in some degree a short summarization of the event details, the event should have advantage in ranking. Also, to encourage users post and join more user created events through our web service, a rank boost is also applied to all user created events. We believe given similar description of an event, a user may be more interested in a user created event rather than public event collected by us.

Recommender

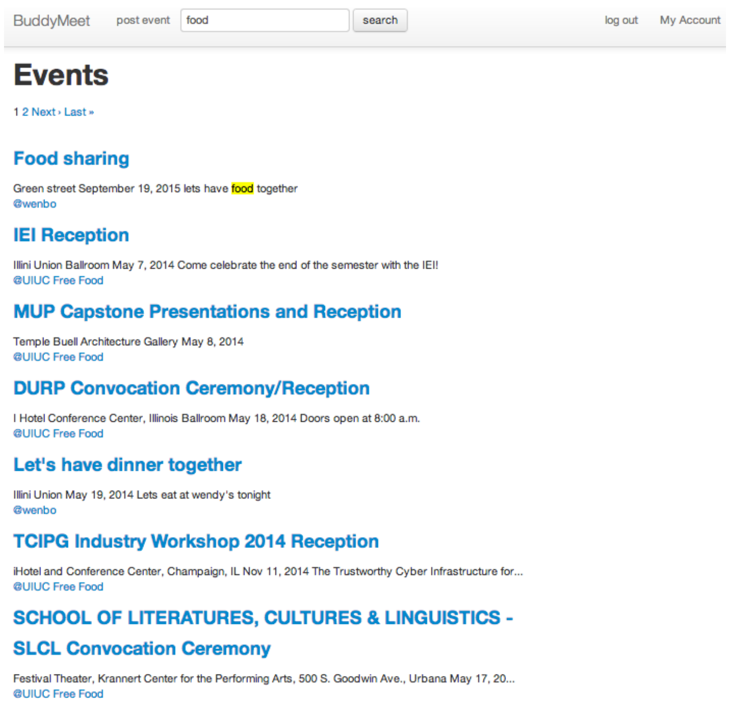
After registration, user is asked to select from a list of categories which they are most interested in. This can also be updated later in the user profile. If a user is interested in an event, he/she can click the button “Join” under the event. We regularly send user emails containing activities based on other user with similar interest distribution, or has joined many same events in the past are interested in. Each user has a vector indicating relevance a category to him/her, computed by the interest he/she marked, and the events he/she joined. Each event also has a vector describing relevance to different categories. By joining different events, both user’s interest vector and event’s interest vector is updated. The recommender takes events with smallest euclidean distances, and give to user as recommendation.

Social network

A simple social network is built, where users can follow each other, get to know what activities his/her friends are joining, invite people to events.

Evaluation/Sample Results

In the database we have generated multiple user created activities for evaluating the search engine, along with public events crawled from public sources. The search engine are designed to provide user not only the events closest to their query word, but also events they are most interested in and very likely to attend. In the dataset,



Events

1 2 Next › Last »

Food sharing

Green street September 19, 2015 lets have food together
@wenbo

IEI Reception

Illini Union Ballroom May 7, 2014 Come celebrate the end of the semester with the IEI!
@UIUC Free Food

MUP Capstone Presentations and Reception

Temple Buell Architecture Gallery May 8, 2014
@UIUC Free Food

DURP Convocation Ceremony/Reception

I Hotel Conference Center, Illinois Ballroom May 18, 2014 Doors open at 8:00 a.m.
@UIUC Free Food

Let's have dinner together

Illini Union May 19, 2014 Lets eat at wendy's tonight
@wenbo

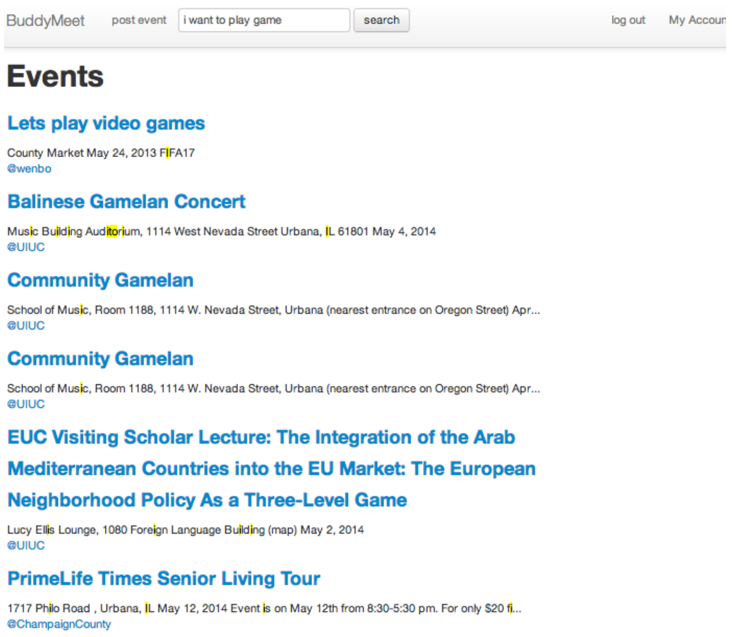
TCIPG Industry Workshop 2014 Reception

iHotel and Conference Center, Champaign, IL Nov 11, 2014 The Trustworthy Cyber Infrastructure for...
@UIUC Free Food

SCHOOL OF LITERATURES, CULTURES & LINGUISTICS - SLCL Convocation Ceremony

Festival Theater, Krannert Center for the Performing Arts, 500 S. Goodwin Ave., Urbana May 17, 20...
@UIUC Free Food

we have public events with word “food” in searchable parts but not in food category, special public events we identified as “free food events”, assigned “food” as category, also food events created by users. When query with “food”, search results is as appears on the right. The result ranked first is user created result, which has keyword “food” in title, details, and category. These facts reward it with 3.5 boost score, and it is expected to appear at top. Among the top ranking events are some public events collected by us with category “food”, and some other user created events in “food” category. This result satisfies our expectation. To test the capability of substring search and stop word removal, we used query “i want to play game”. Result of the query is show as on the left. The ranked first event is user created, and has word “games” in the title, which contains “game” as substring. Notice none of the query results contain all of the query words “i want to play game”. This is because the words “i want to play” are listed as stop words, and we enabled stop word removal in solr. We used stop word list *stop-words_english_1_en.txt* from https://code.google.com/p/stop-words/. And since substring searching is enabled words such as “games”, “Gamelan” are considered as matches to query word “game”. The recommender system is hard to test because we have very limited user group and the events data is in limited categories. Meaningful recommendation is possible only after the event dataset grows larger, and enough users’ event joining record is collected, as they are the essential elements for computation.



Events

Lets play video games

County Market May 24, 2013 FIFA17
@wenbo

Balinese Gamelan Concert

Music Building Auditorium, 1114 West Nevada Street Urbana, IL 61801 May 4, 2014
@UIUC

Community Gamelan

School of Music, Room 1188, 1114 W. Nevada Street, Urbana (nearest entrance on Oregon Street) Apr...
@UIUC

Community Gamelan

School of Music, Room 1188, 1114 W. Nevada Street, Urbana (nearest entrance on Oregon Street) Apr...
@UIUC

EUC Visiting Scholar Lecture: The Integration of the Arab Mediterranean Countries into the EU Market: The European Neighborhood Policy As a Three-Level Game

Lucy Ellis Lounge, 1080 Foreign Language Building (map) May 2, 2014
@UIUC

PrimeLife Times Senior Living Tour

1717 Philo Road , Urbana, IL May 12, 2014 Event is on May 12th from 8:30-5:30 pm. For only $20 fi...
@ChampaignCounty

Conclusions & Future Work

This project is a very good opportunity to apply the knowledge we learnt from the lectures and mps. We got the data from public events websites, parsed the data and then dump event information into a .json file. After that, we loaded the .json file to the search engine.

For the future work, we are considering to make the crawler a automation framework that can crawl larger amount of websites every few hours to make the data up-to-date. The idea is to create a python script that can be kick off every few hours on the server. The script will first start the Scarpy framework and get the raw HTML, after crawling data to the local machine, start parsing and dumping, then sent .json files to the search engine. Everything can be implemented in an automation framework and can run by itself. Also, we are thinking about to add some fault tolerant features. For example, if sometimes the target websites for Scrapy spider is invalid, we can not generate valid data for the python script to create json object. In this case, the new server script can detect this kind of error and generate error logs, then send email to the administrators.

A problem we encountered as scraping data from websites is that the formats of date and time are different. In the future we may find a solution to translate different formats into standard SQL datetime type. This is necessary because time is essential for an event, and it is only effective when expressed in standard way.

Appendix

Individual Contributions

Wenbo Yu :

1. Front-end and back-end construction with ruby on rails
2. Configuration of search engine and recommender system

Xiaojie Zhang:

1. Scrapy Crawler framework setup
2. Scrapy spider implementation
3. Python script for parsing and dumping raw HTML data.

References:

http://scrapy.org/
https://code.google.com/p/stop-words/
http://illinois.edu/calendar/list/7
http://www.visitchampaigncounty.org/calendar/
http://uiucfreefood.com/
http://www.eventfinder.co.nz/