IPRO 303 FALL 2010

FINAL REPORT

Faculty Advisor: James Burstein

Sponsor Advisor: David Ko

Table of Contents

1. Executive Summary

1.1 Sponsor Background

NAVTEQ, the sponsor of this project, is one of the largest digital mapping companies in the world. They are in a constant phase of self-improvement. This semester, NAVTEQ asked IPRO 303, herein Chujio, to focus on social networking sites in order to update their maps using the "buzz" generated on these sites. The goal of this semester was to create a filtering software, that when attached, would update NAVTEQ's system using publicly available information on the web.

1.2 IPRO Background

The Inter-professional Project (IPRO) program was started at IIT in 1995 with aim of introducing students to the work styles and environments in a work place. Through promoting the use of work skills like teamwork, innovation and complex problem-solving, the IPRO program provides students with a practical experience that can have a huge impact on their professional and personal life. Through IPRO courses, students from different majors tackle different real-world problems that appeal to their interests. In certain scenarios, an engineering major may be important for the technical side of a problem, while other majors such as architecture may help with the design and visual representation of posters and presentations. In a real-world situation, the work place is usually comprised of individuals from different academic backgrounds.

1.3 Project Background

This is currently the third semester that NAVTEQ has sponsored a project with IIT. In the past this IPRO group was named Innovative Mapping. This semester's objectives took a different direction in that the team was asked to focus on improving NAVTEQ's abilities rather than changing the user's interface. Due to this change in direction, the team this semester has decided to change its name to Chujio. The term Chujio is a Swahili word for filter; this was chosen as the new team name because it applied to the goal of the project, but also foreshadows the global outreach as well as the diversity of the team members.

1.4 Current Work

After forming the three subgroups that make up the structure of Chujio, the teams began their planning work for the rest of the semester. This was achieved by learning about various deliverables and then forming tasks to achieve each one of them. The project was then split into two phases: the first one was collecting data from the social networking websites like Twitter, LinkedIn, Fwix, and the second phase consisted of actually writing the code that would filter out this data in a meaningful way. Both objectives were achieved as planned.

The purpose of this project was to create a way to continually add and update relevant content to digital mapping systems at NAVTEQ. Currently, in order to find information on a certain point of interests (POI), one would have to do web searches, make phone calls, or go to that place in person. If the place is not well known, new, or remodeled recently- there may not be relevant and up to date information about it on the web which would make it difficult for people to access information on the mapping software.

The task of Chujio is to make it easier to access this sort of information for users. Through public postings on social networking sites, details about these POIs can be used to provide ever changing and up to date information consistently on the map. The deliverable for this semester was to create a software that would constantly update live information, decide if it's relevant, and provide it to people who may use it, through NAVTEQ.

2. Purpose and Objectives

2.1 Company History

NAVTEQ, our sponsor is a Chicago-based provider of Geographic Information System (GIS) data - a dominant company in providing the base electronic navigable maps. NAVTEQ is a "world leader in premium-quality digital map data and content,"[1] and we are therefore very fortunate that NAVTEQ has decided to sponsor Chujio this term. NAVTEQ supplies information used for in-vehicle navigation systems, mobile navigation systems, navigation programs used in cell phones, and internet based navigation websites. The data that NAVTEQ provides is extensive, covering 77 countries and spanning 6 continents. Formed in 1985 in Silicon Valley California, it is one of the founders of digital mapping data and map routing.[2] With the help of a number of investors and owners, NAVTEQ was able to grow and establish offices in countries outside the United States by the early 1990's. It eventually moved its headquarters to the current location in the heart of Chicago. NAVTEQ currently has approximately 4,000 employees spread over 190 offices in 39 different countries, and was recently acquired by Nokia, the popular phone and electronics maker. [3] Currently, NAVTEQ's focus is on making their technology more community and pedestrian friendly.

2.2 Problem Statement

NAVTEQ presented our group with the challenge of using data from social networking sites to update NAVTEQ's map data to be more user friendly. From the publicly available data on the web it is possible to generate a mood graph with NAVTEQ's map data base that will allow a user to view points of interest (POI) to see what the mood is at the location.

Our goal was to aggregate and integrate meaningful data from real time streaming content providers and social networks such as Foursquare and Twitter. To do this we needed to design a dynamic and easy-to-use algorithm that distinguishes the *noise* from the meaningful data, while at the same time gathering updated information regarding the acquired data. From this data, a 'mood-graph' was generated that could describe the details of certain POIs based on the acquired data in a way that the general population can understand and use to their benefit.

3. Organization and Approach

3.1 Team Structure

Before the start of the semester the group was introduced to the objectives of the project that were specified by the sponsor. As a group, the team then studied the objectives and preformed preliminary research about the target sources. At that point the group met with NAVTEQ at their corporate headquarters to discuss the planned outcome of the project. After our initial meeting it became evident that although this was a high energy team, the best way to accomplish our goals would be to split up into more specialized teams. The team leaders purposed at this point three subgroups that could share responsibilities from the main objectives. The three subgroups that were formed were the Design, Development, and Data sub-teams. In order to keep the teams focused the team leaders also appointed sub-team leaders that would be responsible for the outcome of their sub teams. Additionally, the sub-team leaders were responsible for ensuring communication between co-team leaders, and the project adviser.

The team members then volunteered to join each team based upon their strongest skills and abilities. This allowed the team to use everybody's abilities to their fullest to achieve our goals more quickly.

Data Team

The Data team was assigned to collect data that NAVTEQ provided to us and the raw data that is publicly available from the social networking websites. The team was given a list of POIs by NAVTEQ inn an excel file, all located within the Chicago loop. In order to use it in the algorithm, the team had to turn all the information into a relational database [Appendix A]. After selecting the columns useful to our project in the original excel file, the team was able to create a fully functional mySQL relational database. It was also the data team's decision to focus for (this part of the project) on two social media sources: Twitter and Foursquare. The former for the volume of data openly available to the team, despite the possibility of noise, and the latter for its location based friendly data format and also an increasing number of users in the last few months.

Development team

The Development team was responsible for planning and writing the code to implement the Chujio platform. Due to the prior experience of the members of the Development team, it was decided to implement Chujio as a Java Web Application. The team then developed an Object-Oriented modular software design to facilitate revision and future enhancements, and followed the Agile software development methodology to execute development in a quick, efficient and iterative manner. The Development team was also responsible for coordinating with the Data and Design teams as necessary to bring the project to fruition.

Design team

The Design team was assigned the parts needed for presentations for the IPRO office and NAVTEQ. Using image editing software, this sub-team brought to life everything that the team imagined. Through the use of image editing software and CAD, the group provided visualizations of the product, as well as displays that would communicate the team's work to any audience.

3.2 Team Meetings

Chujio met every Tuesday and Thursday afternoon. Meetings were utilized as a forum for members of the Chujio project to present reports on recent developments, to address obstacles faced within the team, and to discuss what these meant for the entire group. Solutions were then brainstormed and the process was repeated. Additionally, time in class was spent reviewing and revising the set schedule of events to make sure the issues outlined were relevant and realistic based on current progress. The next section contains a basic breakdown of the work process structure and the dates that Chujio planned to complete each step.

3.3 Activities

The team's aim was to deliver at least one working solution to NAVTEQ. A group of activities were involved in this process. These activities fit into two chronological phases:

Phase I (August-September) - Research. During this phase the students:

- Studied the requirements that were given by NAVTEQ
- Reached a consensus of specific milestones for this term
- Defined the requirements
- Proposed achievable and realistic solutions
- Met with NAVTEQ

Phase II (October-November) - Development and Testing. During this phase the students:

- Developed an algorithm for submission
- Tested the algorithm with test set of random raw data
- Applied improvements to the prototype(s) according to test results
- Delivered a working solution during a formal meeting with NAVTEQ

4. Analysis and Findings

4.1 Initial Research

The first thing the team decided to do was develop a list of social media website that could be worked with for the project. The team split up and each member researched different social media websites to be considered in the project.

The initial criteria for considering a social media website in the project were as follows:

- Relatedness to the project
  - Different social media websites provided different points of interests in a city.
  - Popularity of the website
  - Some websites had an easily accessible API, making it easy to extract data from the site.
  - Content of noise in extractable data

Following research, the team discussed the different websites and voted on the ones that would be most appropriate for the project. After discussion, the websites fell into the following categories:

- Websites that could be worked with this semester
  - Scored well on the different criterion
- Websites that could be worked with in future projects
  - Required more work to handle the data
- Provided data that is irrelevant to the project and scored low on the criteria
  - Websites that could not be used for the project

The final list the team decided to work with this semester:

Social Site	Benefits	Downfalls	Used
Facebook	Accessible API	Not public	no
Twitter	Very popular	Lots of noise	yes
LinkedIn	Information is accurate and trustworthy	API only accessible to members	no
Buzz	Area search	Not Popular	yes
Flickr	Huge user base	Lots of noise	no
Foursquare	Noise is limited	Not a lot of accessible data	yes
Fwix	connects to social networks	Not easy to access information	no
Yelp	Built in 5 star rating system	Not a traditional Social Site	yes
Yahoo Pipes	Tool to help funnel data	Not a social Network	no

4.2 Development Resources

The development team used a multitude of available resources throughout the development of this project, including free and open source tools, frameworks, web services, and other tools (discussed below). The team also consciously designed the platform utilizing an Object-Oriented and modular approach to facilitate the continuation and expansion of the Chujio prototype by later semesters, if necessary.

Currently, the application uses several algorithms (both text and location-based) to determine if a point of interest is being mentioned in a social network status update and if any of several pre-determined attributes of a point of interest are being discussed. For example, the platform may determine that a user mentions a particular Starbucks location and that it was closed. In its current iteration, the Chujio platform will record an entry referred to as a 'flag', which denotes that the attribute of a given POI was mentioned. The platform cannot yet extract the content of what was said about the specific POI. This is to say that currently the platform can match a POI and determine that its hours of operation or location were discussed. However, it cannot, for example, distinguish the difference between saying that a location was closed for the day or closed permanently.

In addition, Chujio can represent the emotional content of social status messages graphically. As users make references to points of interest, any positively or negatively charged content is detected, and it is given a numerical score and attributed to the POI. This is then displayed as a colored section on a map around the point of interest in question.

The application is currently hosted on Google's App Engine cloud infrastructure at http://ipro303.appspot.com

This semester, Chujio has created a modular platform that addresses several complex goals, but given the time constraints of a single semester, many of the solutions are non-robust and the implementation of the algorithms often employ naïve approaches. This leaves much room for research to be done in order to implement these algorithms in a non-trivial way. Furthermore, the mood graph prototype is rather rudimentary at this point and warrants much work and consideration in its programmatic design, implementation, and functionality.

The pseudocode for general point of interest / attribute matching can be found in the Appendix under section 6.5.

A list of all the software frameworks, web services, and tools used and their functionality is discussed below:

- Play! – Java web application framework. Lends itself to rapid agile-style development.
- Google App Engine – Cloud web app Infrastructure. Includes the highly scalable Google datastore. Used to host the application / datastore
- Objectify – Persistance framework used to simplify interfacing with the datastore
- OpenAmplify – Speech processing web service. Extracts topics, intent, emotion, and speaker demographics from plain text.
- Google Maps javascript API – mapping web service, used for mood graph prototype
- Google Maps Geocoding API – geocoding web service, returns GPS coordinates from street addresses

- Git – Version control. Used to facilitate collaboration on project files and ensure backup and versioning
- Teambox – Free online project management / team collaboration software. Keeps all the team members in the loop about important developments regarding the progress of the project and also helps in keeping an organized list of all the tasks/goals to complete.
- Pivotal Tracker - Industry standard agile project management software. Used to plan software development, assign tasks, and track progress.

4.3 Data Gathering Resources

The Data sub-team had two sources of data to work with. The first source was provided by NAVTEQ, and contained lists of POIs and their different attributes that could be updated using the algorithm. The Data team took this information and reformatted it from its original Excel spreadsheet into a format that could be incorporated into the algorithm.

The second source was the data pulled from the social media websites. The Data team used the software tool Yahoo-Pipes to facilitate gathering of this data. Yahoo-pipe allowed the data sub team to organize the user input from different websites into one standard format, which could then be easily incorporated into the algorithm by the development team.

5. Conclusions and Recommendations

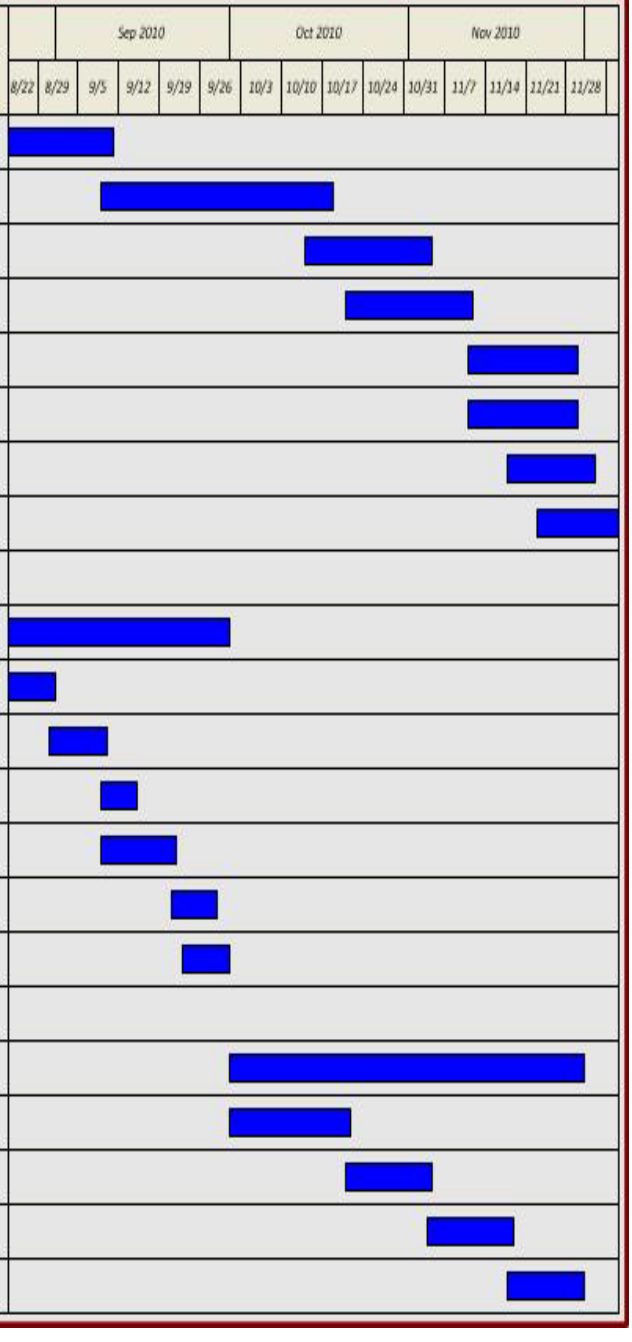
Through a multi-stage process, the team developed an algorithm that can be used by NAVTEQ to update their maps based on user input. In the first stage, members researched different social networking sites and were able to identify the sites that would be most helpful to use for perpetual updating of mapping systems. Once these sites were identified, the Data team was able to compile information from those sites into a working data source that would be used to test the algorithm. Concurrently, the Development team was able to pull from member resources to complete the algorithm that would take the information from that data source, and filter through it to separate the relevant information from the useless noise. Whether or not this solution is implemented or used is up to our sponsor, but the final working product matched the goals of the project. Chujio was successful in understanding the challenge given, identifying what could be worked on as a team, and putting the diverse backgrounds and knowledge of every individual member together into a solution that met the goals of both our sponsor and our team.

As the project will not end with this semester, the members of Chujio have recommendation for the future IPRO group. The group can improve the algorithm so that it can work with additional social media websites, such as Facebook and Flickr that were not used this semester. In addition, the algorithm can also be improved so that it is able to identify the content of the information from the social networking sites, rather than simply recognizing that a POI is mentioned and filtering that as relevant information. Both of these updates to the algorithm would provide a more extensive use of the algorithm, and better updatable information for NAVTEQ mapping systems. The content of the mood graphs can also be improved. Lastly, continuing groups can find a way to better authenticate the provided data prior to mapping updates so that NAVTEQ could benefit monetarily from the prototype.

6. Appendix

6.1 Gantt Chart

ID	Task Name	Start	Finish
1	Project Plan	8/24/2010	9/10/2010
2	Midterm Review Presentation	9/9/2010	10/18/2010
3	Ethics Reflective Report	10/14/2010	11/4/2010
4	Final Project Report (Draft)	10/21/2010	11/11/2010
5	IPRO Abstracts/Brochure	11/11/2010	11/29/2010
6	Poster	11/11/2010	11/29/2010
7	Final Presentation	11/18/2010	12/2/2010
8	Final Project Report (Final)	11/23/2010	12/6/2010
9			
10	Phase 1	8/24/2010	9/30/2010
11	Study and Analyze Requirements	8/24/2010	8/31/2010
12	Conduct Market research	8/31/2010	9/9/2010
13	Discuss research/surveys results	9/9/2010	9/14/2010
14	Reach Consensus	9/9/2010	9/21/2010
15	Define Requirements	9/21/2010	9/28/2010
16	Propose Achievable solution	9/23/2010	9/30/2010
17			
18	Phase 2	10/1/2010	11/30/2010
19	Develop an algorithm	10/1/2010	10/21/2010
20	Test Algorithm with raw data	10/21/2010	11/4/2010
21	Apply Improvements to prototype	11/4/2010	11/18/2010
22	Deliver working solution	11/18/2010	11/30/2010



6.2 Budget

The following is a list of materials and costs that were used to complete the semester's project

IPRO303 BUDGET	
FALL 2010	
Printing	$150
IPRO Day Supplies	$250
Team Building*	$80
Total	**$480**

*"Team Building" is any light competition to promote fellowship. Event included food.

6.3 Team Roster

At the beginning of the semester the team created a roster that could easily explain what assets each member of the team brought to Chujio. After creating this roster, the team was able to create a solution to the problem utilizing all the assets that each individual brought to the table. The team roster also gave the team the ability to look at the available skills so that tasks could be created that honed in on the specific strengths of each individual.

Aditi Kumar
Year: Third Year
Major: Computer Science
Minor: Business
Description: Aditi's technical knowledge and skills in the area of computer programming are not only helpful in terms of software development but also provide a significant base when it comes to systematic planning, logical thinking, problem solving, and creative thinking. She aims to use those skills to add value and achieve the team's project goals for this semester.

Christopher Curtis
Year: Fifth Year
Major: Computer Science
Description: Christopher's particular interests include engineering software and mobile applications which enables him to give practical insights and creative solutions to the various issues faced by Chujio. Currently working as a mobile software developer and with experience in web design Christopher plans to help design a fresh, innovative social product for the NAVTEQ group.

Dmitriy Vysotskiy
Year: Fourth Year
Major: Electrical and Computer Engineering
Description: Dmitriy has years of experience working in team environments and has previously collaborated on a social project. His skills include digital design, product management and team administration. Dmitriy also has an interest in the growing mobile market and social interaction. He hopes to harness his skills and interests to guide his team in providing NAVTEQ with an innovative new product.

Ernesto Ramirez
Year: Fifth Year
Major: Computer Engineering and Computer Science
Description: Ernesto wishes to use his knowledge of programming to help develop software that will be user friendly. He is a part of an Illinois Institute of Technology and Dominican University joint program. This will broaden the demographics of the project and bring a different concept to the group. He also has experience with some of the methods that will play a major role to the project. He is also a returning member of the previous IPRO and will bring that knowledge with to make this IPRO a success.

Hashem Abu-Amara
Year: Fourth Year
Major: Electrical Engineering
Minor: Psychology
Description: Hashem's technical background adds to the team's ability of resolving problems and engineering solutions and also enabling him to grasp new concepts. Hashem's minor in psychology adds to the team a new way of thinking with a different perspective and through experience from areas like motivation and success and industrial psychology would help to bring the team together on working more efficiently and effectively. Through his experience in presentation, communication and people skills from previous projects he hopes to help the team achieve the goals and objectives of this semester.

Jason Chun
Year: Fourth Year
Major: Computer Engineering
Description: Jason's computer background and analytical skills allow him to see straight to the root of a problem. He will use these skills to help keep the team on track and quickly work through what may seem like difficult issues. His off the wall thinking will help to bring new and unique ideas to the table and his natural tendency for simplicity will help to stream line those ideas.

John Jewell
Year: Fourth Year
Major: Architecture
Description: John's experience with hand drawing allows him to communicate with the group in a visual manner. As well, he is strong in art work and layout design and will be helping the team in the presentation boards and possibly schematic drawings.

Kalman Varga
Year: Fourth Year
Major: Architecture
Description: Kalman's background in architecture & design brings a different eye to this team. He will be contributing his design efforts in the overall appeal of his teams' work. He is very personable, communicates well with others and wants to use these strengths to give presentations, organize team members, and contribute to public speaking of the team.

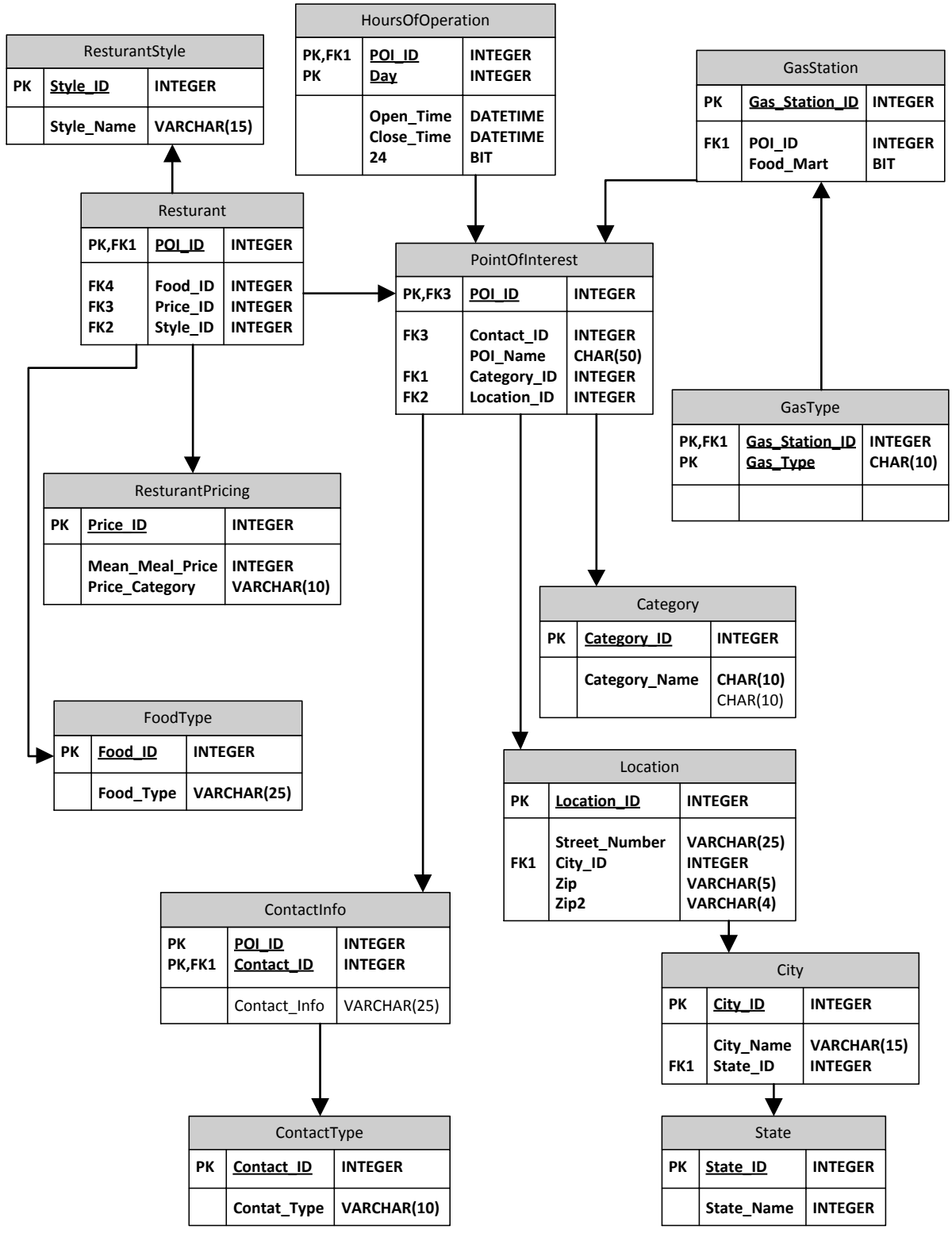
Mark Michael
Year: Fourth Year
Major: Computer Information Systems
Minor: Business Management and Information Security
Description: Mark has a strong background with technology, and learning social media development. Also this is his third time working with NAVTEQ and he helped developed the objectives of this semester. After the success of the previous semesters he is using the skills and knowledge he gained from those projects to provide the perfect solution to the project at hand.

Melanie Koto
Year: Third Year
Major: Biology
Minor: Math and Science Education
Description: As a future teacher, Melanie has a great interest in serving people. That desire leads her to this IPRO group with the expectation that she will help propose new ideas and programs that anyone, even the "average Joe", may be able to utilize in travel. The development of using data from individuals' everyday information is also appealing. As a user of social networking sources, Melanie feels that she can provide helpful input in researching and gathering data for the project.

Patrick Tagny
Year: Fourth Year
Major: Computer Information Systems and Applied Mathematics
Description: Patrick is particularly interested in Software Engineering and ways to turn those products into revenues for companies. Regarding Chujio, he is mostly excited about the combination of available social tools with consumer digital maps to get, in the end, a useful tool for consumer as well as a profitable product for the sponsor, NAVTEQ.

Rishi Kushaare
Year: Fourth Year
Major: Business Administration and Applied Science in Engineering.
Description: Rishi's talents and strengths include business development, communication / keynote presentation, marketing, projection analysis. Also he has experience in business development, analysis, product research and development at HotSet Corp and Newell Rubbermaid (Fortune 500) company.

6.4 Database Schema



6.5 Main Algorithm

MAIN()

1. statuses[ ] .COLLECTNEWSTATUSESFROMINTERNET()
2. CHUJIO(statuses[])

CUHIO (statuses[])

1. for each s in statuses[]
2. poi .MATCHPOI(s)
3. attributes[] .EXTRACTATTRIBUTECHANGES(s)
4. poi.UPDATE(attributes[])
5. mood . EXTRACTMOOD(s)
6. MOODGRAPH.UPDATEDATA(poi, mood)

MATCHPOI(s)

1. P . [] (holds possible point of interest matches)
2. flters . load flters
3. for each f in flters
4. P.APPEND(f.APPLY(s))
5. poi . CHOOSEBEST(P[])
6. return poi

EXTRACTATTRIBUTECHANGES(s)

1. A . [] (holds attribute changes)
2. flters . load flters
3. for each f in flters
4. A.APPEND(f.APPLY(s))
5. return A[]

UPDATEPOI(poi, attributes[])

1. for each a in attributes[]
2. poi.UPDATE(a)

6.6 Works Cited

1. Facebook. Web. 2 Dec. 2010. < http://www.facebook.com/>
2. Flickr. Web. 2 Dec. 2010. < http://www.flickr.com/>
3. Four Square. Web. 2 Dec. 2010. < http://foursquare.com/>
4. Fwix. Web. 2 Dec. 2010. < http://fwix.com/>
5. Google Buzz. Web. 2 Dec. 2010. < http://www.google.com/buzz>
6. "About IPRO" *Inter-professional Project Program Site*. 02 Dec. 2010 <http://www.ipro.iit.edu/about>
7. LinkdIn. Web. 2 Dec. 2010. < http://www.linkedin.com/>
8. *NAVTEQ Corporate - About Us*. Web. 6 Sept. 2010. <http://corporate.navteq.com/index.html>.
9. "NAVTEQ Corporate - Global Leadership." *NAVTEQ Corporate - About Us*. Web. 10 Sept. 2009. <http://corporate.navteq.com/global.html>.
10. "NAVTEQ Corporate - History." *NAVTEQ Corporate - About Us*. Web. 6 Sept. 2010. <http://corporate.navteq.com/history.html>.
11. "IPRO 303 Innovative Mapping Project Plan Spring 2010 ." *IPRO iGroups*. IPRO, 02 Feb 2010. Web. 6 Sep 2010. http://igroups.iit.edu/download.php?id=75557.
12. Twitter. Web. 2 Dec. 2010. < http://twitter.com/>
13. Yahoo Pipes. Web. 2 Dec. 2010. < http://pipes.yahoo.com/pipes/>
14. Yelp! Web. 2 Dec. 2010. < http://www.yelp.com/chicago>