Assignment 2: Original Proposal

Version 2.1
by Cara Magliozzi, Ricki Cohen, Sam Kolovson, Chris Scott, & Steven Tso
Last Updated: May 3rd, 2014

Slogan:
"Visualize the present."

Outside Field(s):
Natural Language Processing, Business, Data Analysis, Twitter

Problem Statement:
Twitter lacks robust external data analysis. Taking this into consideration, LexiCloud has decided to develop a product that will provide users with the analytic information they crave.

Sam Kolovson | 2/19/14

This problem is characterized by the following courses:
Computer Science:

- CS 446: Search Engines
- CS 585: Natural Language Processing

External Coursework:

- LING 401: Introduction to Syntax
- LING 397C: Linguistics and Literature
- LING 492B: Computational Linguistics
- MGMT 241: New Venture Creation

Cara Magliozzi | 5/3/14

Product Description:

Currently Twitter's Trending Topic feed only provides users with a sense of the most popular words and hashtags that are currently being mentioned, but does not provide the full scope of analytics possible. While Twitter provides a search function that matches a queried word with lists of strings including that word, it does not search or analyze related words.

In contrast, we at WordCloud intend to make use of all available data. Drawing on research in natural language processing, linguistics, and probability theory, we intend to offer a robust set of tools for our users. Through mining strings, we intend to extract the maximum amount of metadata possible while still being comprehensible to all users.

The free version of our product will provide users with an appealing visualization of the most common words used in relation to the queried word or topic. In the Age of Information, our social-media driven society craves information regarding what others are interested in. Our easily-accessible product will be stimulating for this casual audience.

For the more sophisticated users of our product, we will offer a premium service, which will provide extended analytical resources. They will not only see related "buzzwords," but will also be able to analyze the geographic locations of the tweeters, exact statistics about word usage, temporal analysis, and related images to the topic. All of these features will allow companies to find–and even predict–what's trending in order to better cater to their target audience or research a competitor.

Ricki Cohen | 2/19/14

Projected Timeline:

February 23rd - March 1st:

Research - stopword list, libraries to use, potential limitations, familiarizing with NLTK and Twitter API; Design document production

March 2nd - 8th:

Twitter API implementation; Prototypal site mock-ups

March 9th - 15th:
Visualization script development
March 16th - 22th:
Design revision; String-edit distance algorithm implementation
March 23rd - 29th:
Design revision; Connect front-end with visualization components
March 30th - April 5th:
Functionality testing & debugging
April 6th - 12th:
Connect all system components
April 13th - 19th:
Polishing and presentation preparation
April 20th - 30th:
Final presentations & demo

Cara Magliozzi & Ricki Cohen | 5/2/2014

Revenue Model:

We will be implementing a "freemium" model, which will generate revenue through two different types of users. Casual users will be able to generate word clouds from queries free of charge. This model will be funded by ad revenue. The premium package, which generates revenue through monthly subscription, will provide users with word results as well as the corresponding counts, the locations of the tweets, historical data and language change, and word types.

Cara Magliozzi & Ricki Cohen| 5/3/14

Associated Costs:

- Domain Name - $10
- Server - $300 ($50/month)
- Renting Office Space - $2400 ($400/month)
- Computers - $6,000
- Other Miscellaneous Office Supplies - $2,000
- Salaries - $144,000

Total: $154,710 for 6 months

Ricki Cohen, Cara Magliozzi, & Christopher Scott | 2/19/14

Meet the Team

Name:
Christopher Scott

Role:
Twitter Integration

> Chris has had prior experience with the Twitter API. Given this, he will be responsible for obtaining tweet data and providing it to the Linguistic Analysis team.

UMass CS Courses:

- CS 121: Introduction to Problem Solving with Computers
- CS 187: Programming with Data Structures
- CS 220: Programming Methodology
- CS 230: Computer Systems Principles
- CS 240: Reasoning Under Uncertainty

Other Courses:

- ISOM 397A: Intro to Entrepreneurship (including TA-ship)
- BDIC 397B: Leadership & Networking
- MGMT 241: New Venture Creation
- OIM 210: Intro to Business Information Systems

All four of the above courses are closely related however provide slightly altered views on the topics of Entrepreneurship and core business principles. Through participating in these courses, I have gained a thorough understanding of the world of entrepreneurship (tech and other), management and venture capital. I think this knowledge will be helping going forward in this course based on how you have described the grading system.

Bio:
I'm a junior year CS BA with a secondary concentration in Entrepreneurship. Most of my spare time is spent managing technology for TEDxUMassAmherst. I also like to spend time planning events for E-House (bi-weekly networking mixers hosted at my apartment), managing my personal investment portfolio of 5 years, playing soccer and basketball, reading (when I have the time) and plotting world domination with my three awesome roommates.

Name:
Cara Magliozzi

Role:
Linguistic Analysis, UI Design, & Document Revision

Cara will work with Ricki to perform linguistic analysis on the obtained tweet data. Additionally, she will work with Sam on front-end design choices and with Ricki on maintaining and revising all required documents throughout the course of the project.

UMass CS Courses:

- CS 121: Introduction to Problem Solving with Computers
- CS 187: Programming with Data Structures
- CS 230: Computer Systems Principles
- CS 240: Reasoning Under Uncertainty

Other Courses:

- LING 401: Introduction to Syntax
- PSYCH 330: Behavioral Neuroscience

These courses were particularly impactful because they provided a formal understanding of topics I have always found fascinating as well as a foundation upon which I can build. I am highly interested in algorithmic nature of human language, neural networking, and the interplay between humans and artificial intelligence.

Bio:
I am a 23 year-old non-traditional junior pursuing a double-major in Computer Science (BS) and Linguistics & Psychology (Joint Major, BA). I work for the Fine Arts Center as a technology intern where my responsibilities include supporting both the staff and computing equipment, assisting in the production of FAC websites, and managing client accounts. As a result of this, I have experience with HTML, CSS, and mobile website production. I also work as a grader for CS 121: Introduction to Problem Solving with Computers, a peer advisor for undergraduate Computer Science students, and am a co-chair of the Computer Science Women's Group.

Name:
Samantha Kolovson

Role:
UI Design & Implementation

Sam will be responsible for front-end implementation. She will collaborate with Cara on design decisions.

UMass CS Courses:

- CS 187: Programming with Data Structures
- CS 220: Programming Methodology
- CS 230: Computer Systems Principles
- CS 240: Reasoning Under Uncertainty
- CS 250: Introduction to Computation
- CS 311: Introduction to Algorithms (in progress)

Other Courses:

- MATH 235: Linear Algebra
- MATH 233: Calculus III

These courses provided me we a deeper understand of Mathematics, which has consistently been one of my strong suits. Calculus III delved into the more advanced topics in Calculus, particularly conceptualizing in three dimensions. Linear Algebra was more impactful to my Computer Science education in that I learned how to manipulate information in matrices and vector spaces.

Bio:

I am a sophomore CS BS major. While I spend a majority of my time outside class fulfilling my commitment to the varsity rowing team, I also work at OIT as a Junior Web Developer. There I collaborate with another student employee on an interface for managing university Drupal sites. This is just starting to be used by university relations.

Name:

Steven Tso

Role:

Data Visualization & Code Structure

Steven will determine what 3rd party libraries will be used, how the visualization will occur, and how the project will be structured on the back-end.

UMass CS Courses:

- CS 121: Introduction to Problem Solving with Computers
- CS 187: Programming with Data Structures
- CS 250: Introduction to Computation

- CS 311: Introduction to Algorithms

Other Courses:
- ECE 354: Computer Systems Lab II
- SCH-MGMT 491E: Entrepreneurial Field Experience

Computer Systems Lab II was helpful since it taught me nothing ever goes as planned. We were working with FPGAs that had very bad compiler support. This class will help me in this class since I understand how to cope when technical things do not work as planned. Also the entrepreneurial field experience was impactful since that class taught me that sometimes your client does not know what they want so sometimes you have to push them to go a certain way.

Bio:
I am a senior computer engineer with a minor in business. I have a solid background in embedded systems from pcb design to driver integration. I am taking this class since I believe the web is the future and I am very interested in knowing more about web programming. I am also part of TEDxUMassAmherst developing NFC bands for attendees. I have basic experience with web programming via Heroku and bootstrap.

Name: Richelle “Ricki” Cohen

Role:
Linguistic Analysis & Document Revision

Ricki will work with Cara to perform linguistic analysis on the obtained tweet data. Additionally, she will work with Cara on maintaining and revising all required documents throughout the course of the project.

UMass CS Courses:
- CS 121: Introduction to Problem Solving with Computers
- CS 187: Programming with Data Structures
- CS 220: Programming Methodology
- CS 240: Reasoning Under Uncertainty
- CS 250: Introduction to Computation
- CS 585: Introduction to Natural Language Processing
- CS 311: Introduction to Algorithms (in progress)

Other Courses:
- LING 397C: Linguistics and Literature
- ENG 329H: Tutoring Writing

These two outside courses taught me to think critically and analytically about written words. They gave me the tools to understand writing in a scientific light, as well as sharing this insight with others.

Bio:
I am a junior double-majoring in linguistics and computer science who is also interested in the impact literature has on language (and vice-versa). When not in classes, I work as a tutor at the Writing Center, where I collaborate with other students on content, clarity, and style in their writing to strengthen their skills as writers. I am also currently working with Professor Eitan Mendelowitz at Smith College to help complete his Global Proverbs Project, where I am refining a sentence generator to create 'novel' proverbs.
