



LexiCloud

“Visualize the present.”

Version 2.0

Last revised by Cara Magliozzi on 5/3/14

Meet the Team

Cara Magliozzi → *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.

Ricki Cohen → *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.

Sam Kolovson → *UI Design & Implementation*

Sam will collaborate with Cara on design decisions. She is responsible for the user experience and implementing the front-end.

Chris Scott → *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.

Steven Tso → *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.

Inspiration

Computer Science:

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

External Coursework:

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

Social media sites like [Twitter](#) have had a tremendous effect on the English language and how we communicate. Taking this into consideration, we wanted to develop a tool that performed linguistic analysis on the tweet data. Additionally, we wanted to be able to generate appealing data visualizations based on this data.

Product Overview: 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.

Product Overview: 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 [LexiCloud](#) 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.

How do we distinguish ourselves from **Twitter**?



- **Twitter** lets you search words and topic, but just returns tweets without processing them.
- **Twitter** can tell you what is **trending** – that is not our concern; regardless of the popularity of a topic, we analyze the associated “buzzwords”.

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

Cost & Revenue

How much will it cost to complete this project?

- Domain Name - \$10
- Server - \$300 (\$50/month)
- Office Space - \$2400 (\$400/month)
- Computers - \$6000
- Miscellaneous Office Supplies - \$2000
- Salaries - \$144,000

Total: \$154,710 for 6 months

How will this product generate revenue?

- Free: basic word cloud, ads
- Premium: heavy data analysis; downloadable spreadsheet; no ads
 - all words and corresponding count
 - locations of tweets
 - change over time
 - word type (nouns, adjectives, verbs, etc.)