LexiCloud

"Visualize the present."

Version 2.1
Last revised by Cara Magliozzi on 5/5/14

Disclaimer

Herein we detail both the current state of the product as well as future intentions for the product. Future intentions include implementation of the business model, linking of system components, and modification of the front-end. This document is subject to change.

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.

**This is the description of our intended product. Currently, on the front-end, our product generates word clouds based on user inputted strings of text. There is back-end implementation of natural language processing algorithms and a Twitter data retrieval script that are not linked with the front-end.*

Visual Output

This is a word cloud generated by our current product.
The input text is the "*To be or not to be...*" soliloquy in the "Nunnery Scene" of William Shakespeare's play Hamlet.

Non-Goals

- This application only interacts with the Twitter platform; it is not compatible with other social media platforms.
- Unlike Twitter, this will not display a trend feed, trending tags, or simple, unorganized search results.
- This application is not a social media platform. It will complement Twitter's existing functionality and will not be considered a Twitter competitor.

Business Model*

Free Version

- provide an appealing visualization of the most common words used in relation to the queried word or phrase

Premium Version

- registration and monthly subscription required
- perform more complex analysis on the data returned by premium user queries
- provide additional features:
  - save a search
  - track changes across saved searches
  - downloadable CSV files of all available metadata

*While we intend to implement this, we have not yet begun to do so.

User Scenarios: Standard

- Sara Phant is a busy college student. While she spends most of her time doing schoolwork or trying to stay active in clubs, she still likes to be kept in the social loop and stay in touch with what's going on. Miley Cyrus was recently the main subject of a supposedly scandalous 60 minutes interview and while Sara doesn't have the time to watch the entire thing, with a quick LexiCloud query of "Miley 60 minutes" she was able to deduce that the fact that she stormed off the set was NOTHING compared to the moment where she tried to kiss Leslie Stahl.
- Betty is planning her grandson, Zach's, fifth birthday party. Lately, Zach has been obsessed with collecting and trading Pokemon card's with his friends, so unquestionably the party is to be pokemon themed. When Betty went to order the cake for his party, she was faced with the problem of which Pokemon to have drawn in icing. Thankfully, Betty had LexiCloud to turn to. The cloud generated by entering Pokemon rendered "Pikachu" as the largest and therefore the word most associated with Pokemon. Zach's cake featured the chubby rodent in bright yellow icing.

User Scenarios: Premium

- Jack just launched a new social network startup, YourFriendSpace, but he knows that he must find a competitive advantage to compete with Facebook. To find his niche market, Jack compiles the data from his premium LexiCloud analysis to realize that single women with more than 7 cats make up his core user-base and can then reshape his site to better serve them.

- Brandon is a recent graduate who spends most of his spare time on his new hobby, daytrading. While he doesn't have all the sophisticated tools that the big hedge managers have access to, he was able to make proactive decisions on the direction of the market through the use of our system. Through proper use of the LexiCloud analytic data, Brandon was able to gauge the interest in certain stocks based on the the tweets from the many active traders on Twitter.

User Scenarios: Premium

- The National Broadcasting Company (NBC) broadcasts news, sports, and a variety of other programs. It is a mammoth task to collect and analyze data pertaining to each of its individual broadcasts, and even more specifically a specific story or piece of a program. One such program for which they are trying to better serve its viewers is the 6:30pm "Nightly News." This 30 minute segment clips right along with short news briefs of the most important news from the day. NBC is a premium user of LexiCloud. Employees of the social media department utilize the many data analytics features of LexiCloud to determine important statistics such as which states the program is most viewed in, who is tweeting about their show, which of their hashtags are most effective, and by looking at the word count which stories were most popular or controversial.

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 Tao → *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.

Technical Roadmap

Base

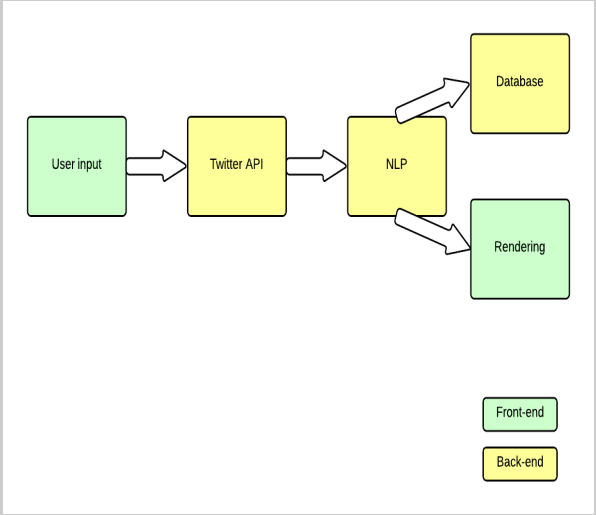
- Node.js/Express: backbone for this project
- Jade/ HTML / CSS: templating for front-end
- JavaScript: language used for this project
- Heroku: web hosting for cloud based application
- GitHub: source control and collaboration

Front-End

- AngularJS/D3: produces the cloud rendering
- C3/D3: alternative way for cloud rendering
- Twitter Bootstrap: front-end design
- Google Font API: front-end font for our site

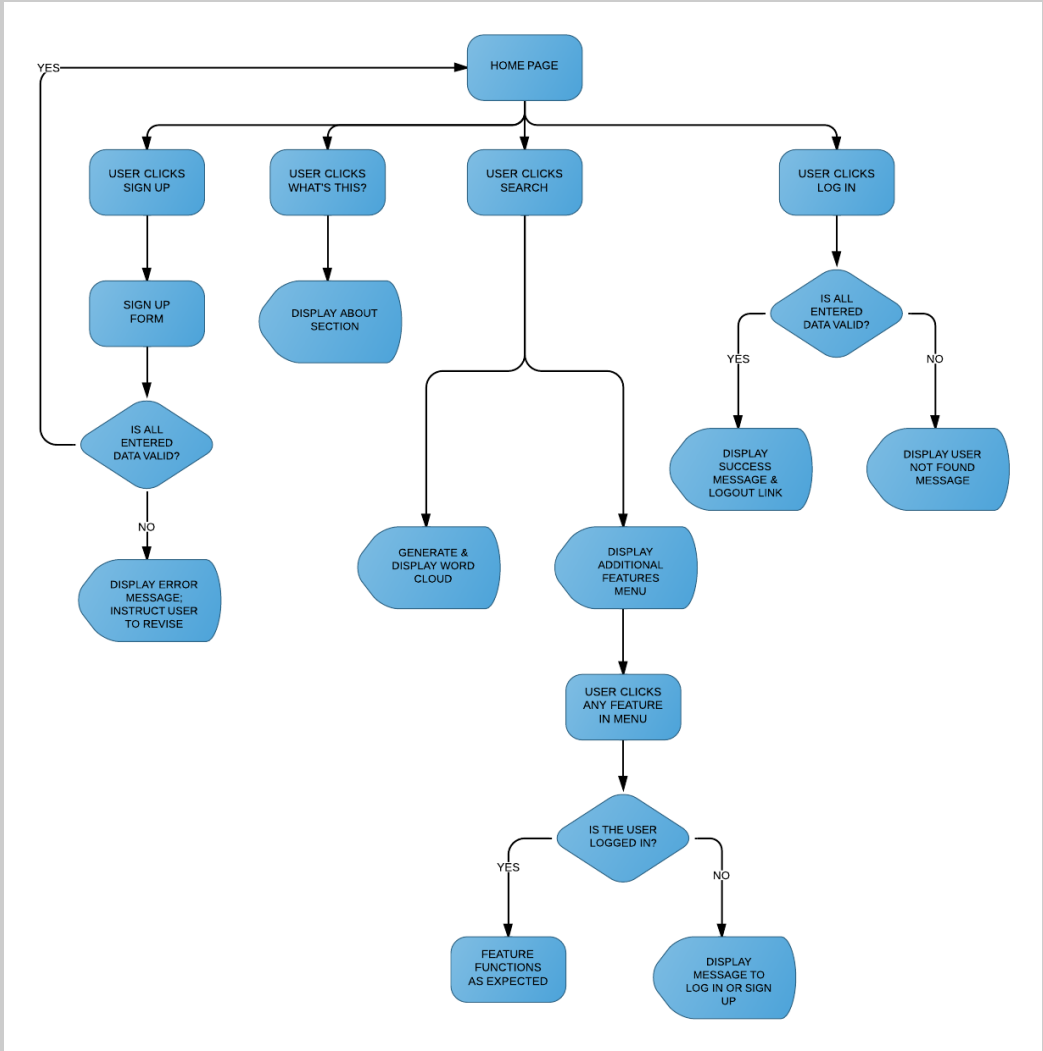
Back-End

- Twitter API: twitter script
- NLP: Damerau-Levenshtein edit distance, Soundex



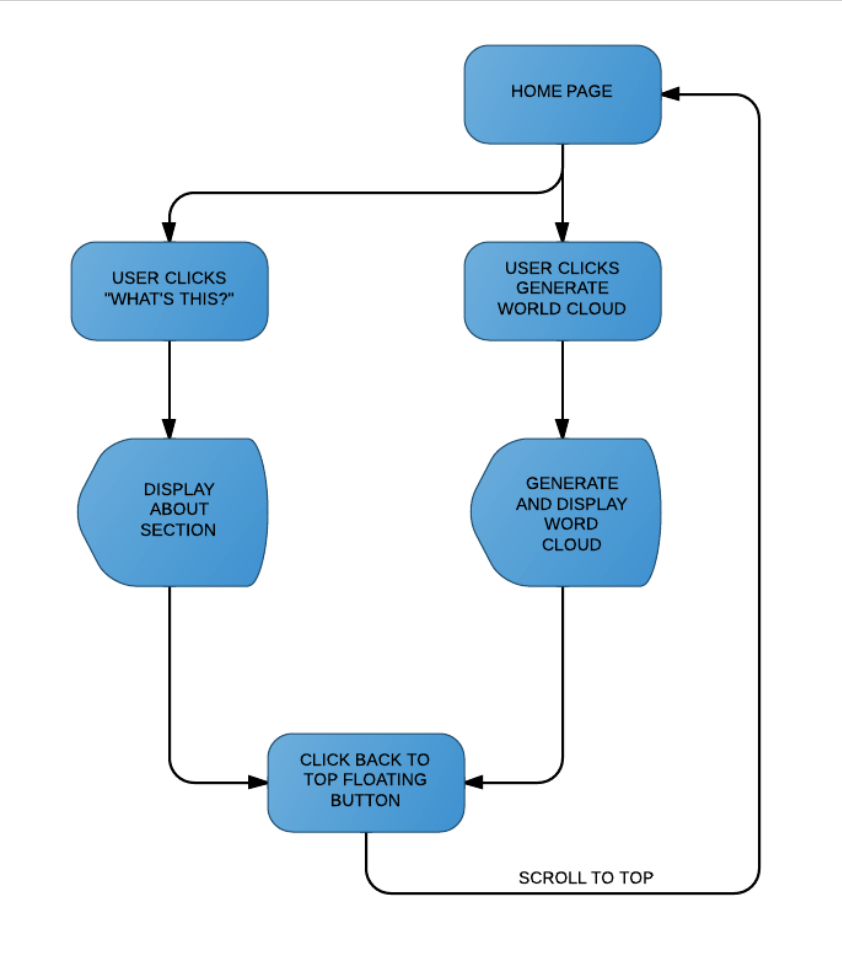
Flowchart

This flowchart details the intended flow of our product.

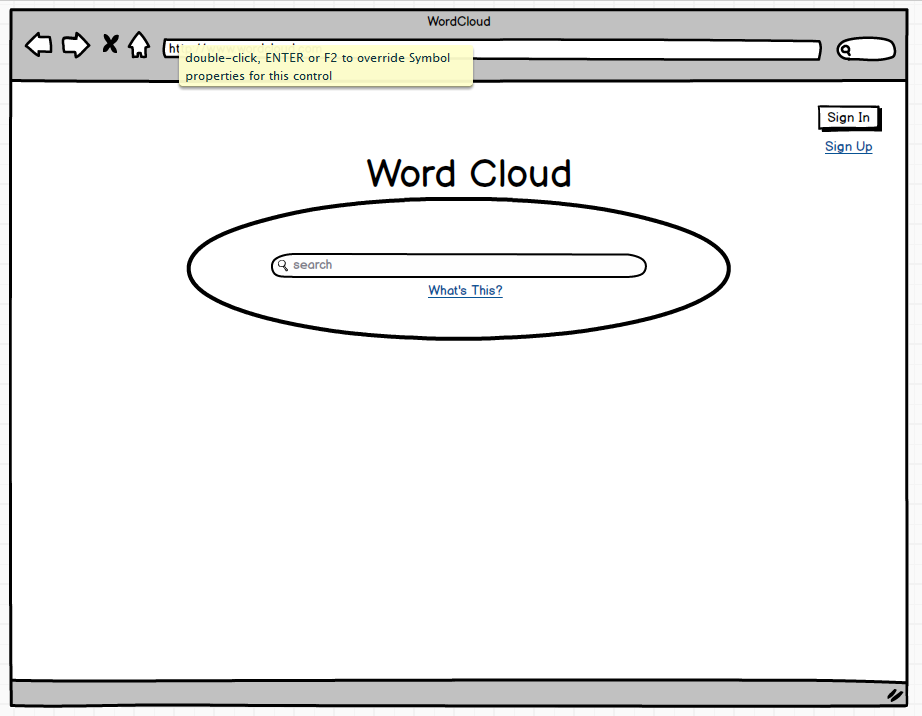


Flowchart

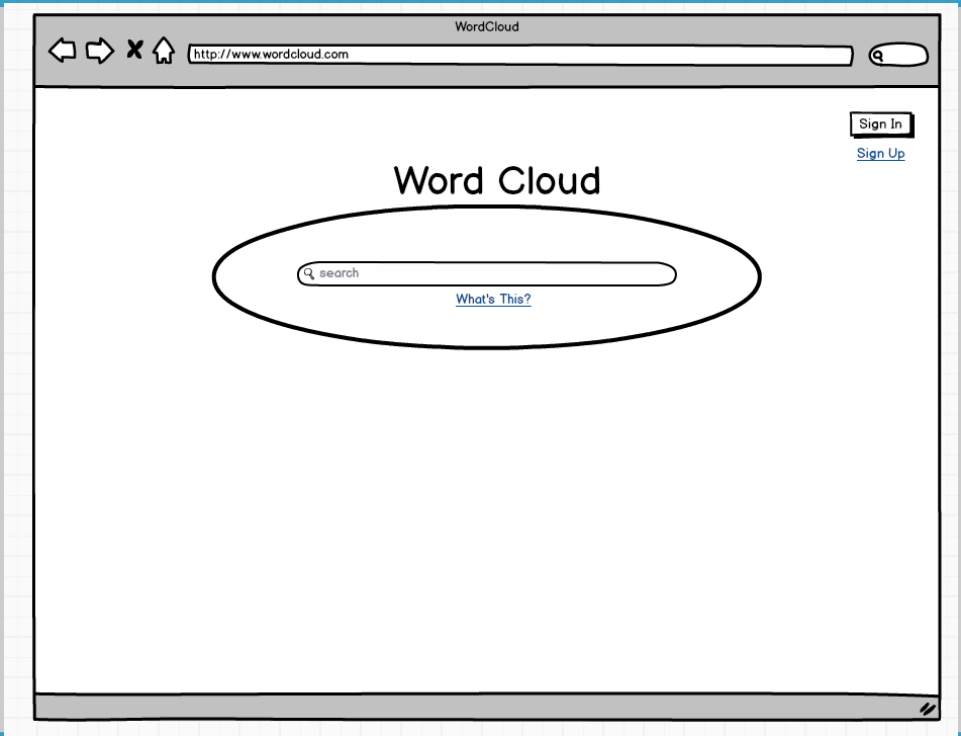
This flowchart details the current flow of our product.



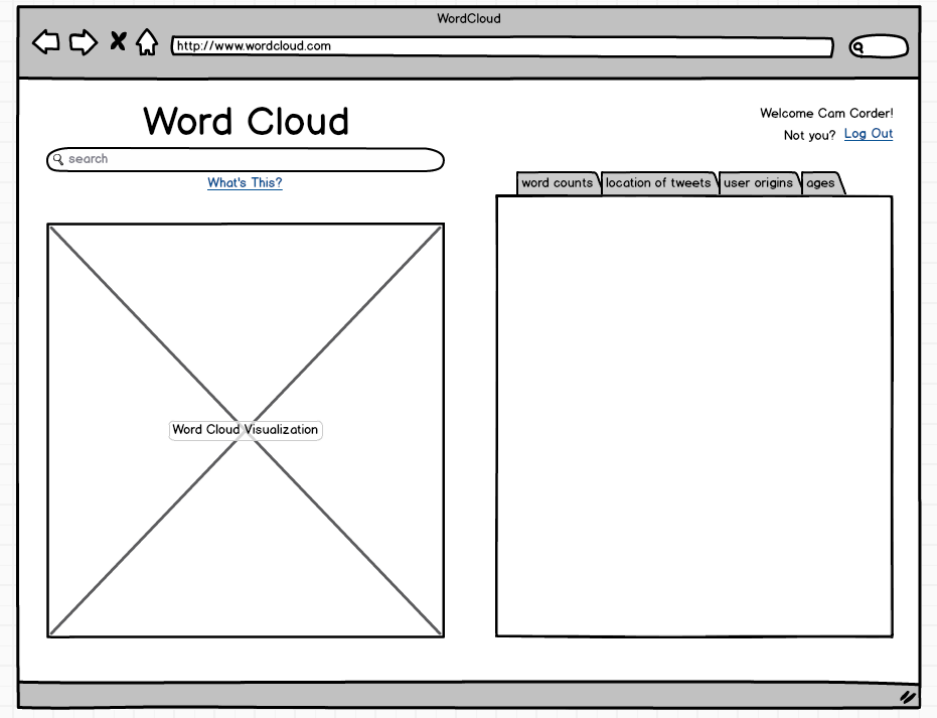
Screen-By-Screen: Home Page



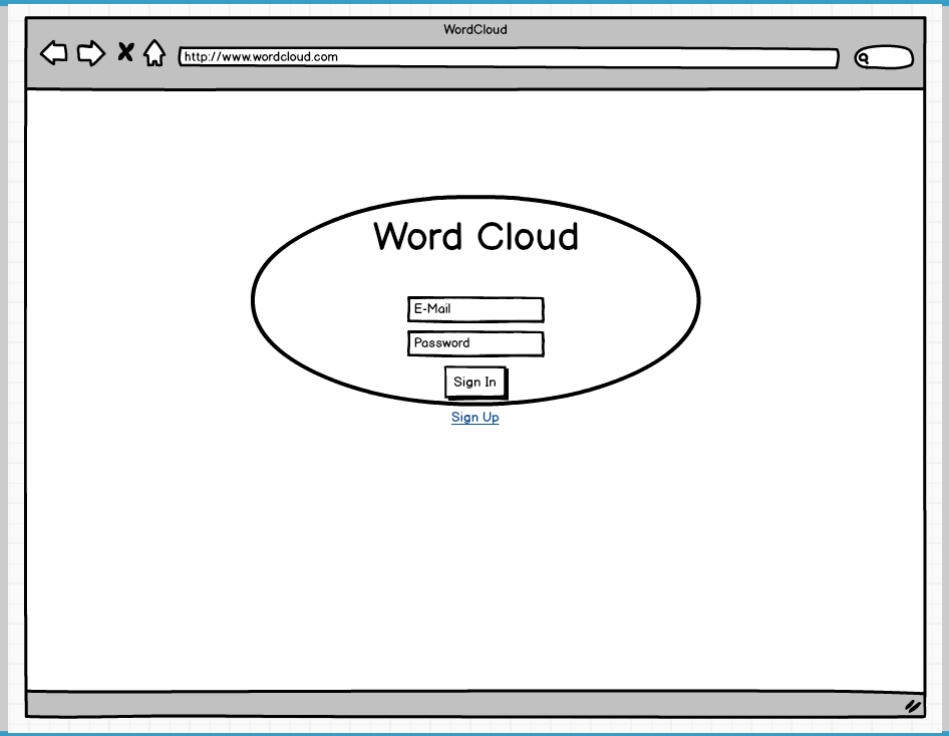
Screen-By-Screen: Home Page & About



Screen-By-Screen: Search Results



Screen-By-Screen: Sign In



Screen-By-Screen: Sign Up