LexiCloud

"Visualize the present."

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

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.

How do we distinguish ourselves from Twitter?

- Twitter trends only provide most popular words and hashtags
- Searches only return string matches
- No additional analysis
- LexiCloud mines additional metadata
- We are not worried about what is trending
- We show related “buzzwords”

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.

Goals → Achieved

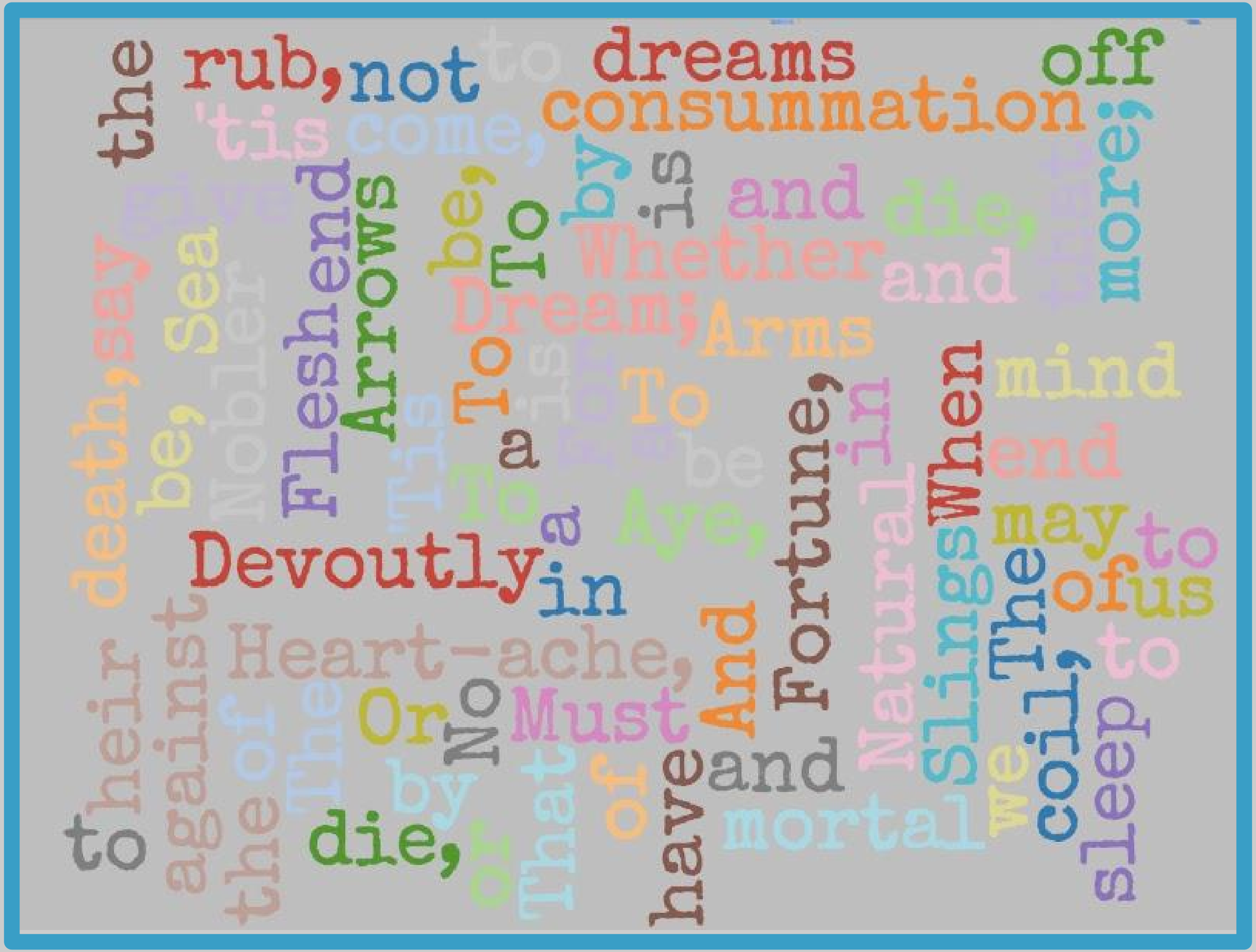
- implement a front-end
- dynamically render a word cloud generated from a user inputted string
- query Twitter to obtain tweet data
- implement Damerau-Levenshtein & Jaro-Winkler string-edit distance, Metaphone, and Soundex algorithms
- pushed app to Heroku

Struggles & Limitations

- the Twitter script, including API limitations on query frequency & number of allowed returns
- learning 3rd party software
- connecting system components
- modifying D3 directive from statically to dynamically typed
- JavaScript errors
- determining what is accomplishable given the timeframe
- coordinating team meetings

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.



Demo

Utilized Software

- Node/Express
- AngularJS
- JavaScript
- HTML / CSS
- Bootstrap
- Heroku

- D3
- Twitter API
- GitHub
- Google Font API

Goals → Future

- refine the front-end
- link all system components:

  user input → twitter.js → nlp.js → cloud generator → front-end output
- implement additional natural language processing (e.g. Twitter part of speech tagger)
- add sections explaining the different linguistic analyses used
- implement user registration and login
- incorporate the business model
- world peace

Q&A

Thank you for listening!