# Content Density

## Textifai

Semester: Fall 2017
Team Members: Ozias Gonet, Ash Johns, Tyler McMaster, Brandon Yannoni
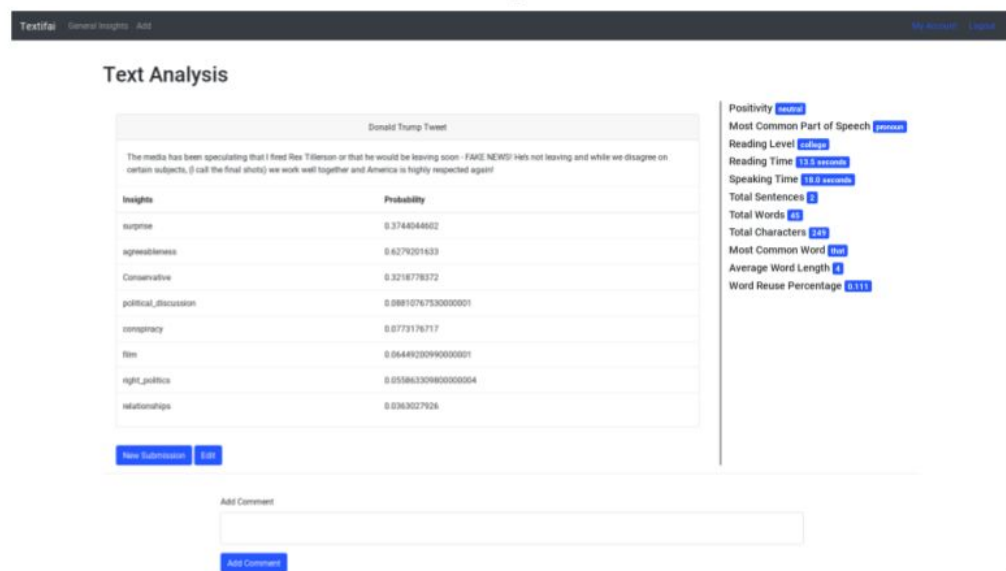Repository: CS326-important/space-deer
Slide presentation:
https://docs.google.com/presentation/d/1xlBn4p7xOuo-vzkFFDeeMHmlAtkaJpG4Z0XWieaMDjk/edit?usp=sharing
Single Slide:



## Overview

Textifai analyzes user-submitted text and generates statistics and insights based on the content of that text. The site uses NLP-based algorithms with the help of Indicoio and NLTK to perform analyses. The categories of insights displayed on the text analysis pages include subjects, emotion, personality, grammatical, and general. Users can submit their own text and then view the analysis, as well as view and comment on the analysis of any other user's submitted text. This application is innovative because it is an amalgamation of many state-of-the-art text analysis API features built for developers but brought to users in an intuitive way.

## User Interface

### Index

The index page is the first page the users sees when they go to the site. It consists of a navbar with the links to log in or sign up depending on their logged in state. The other links on the navbar are the Add link, and the General Insights link. The next part of the page is the search bar. It is a simple search that looks for matching text entries with the users query. After that there is the Recent Posts section, showing the most recently added posts to the site.
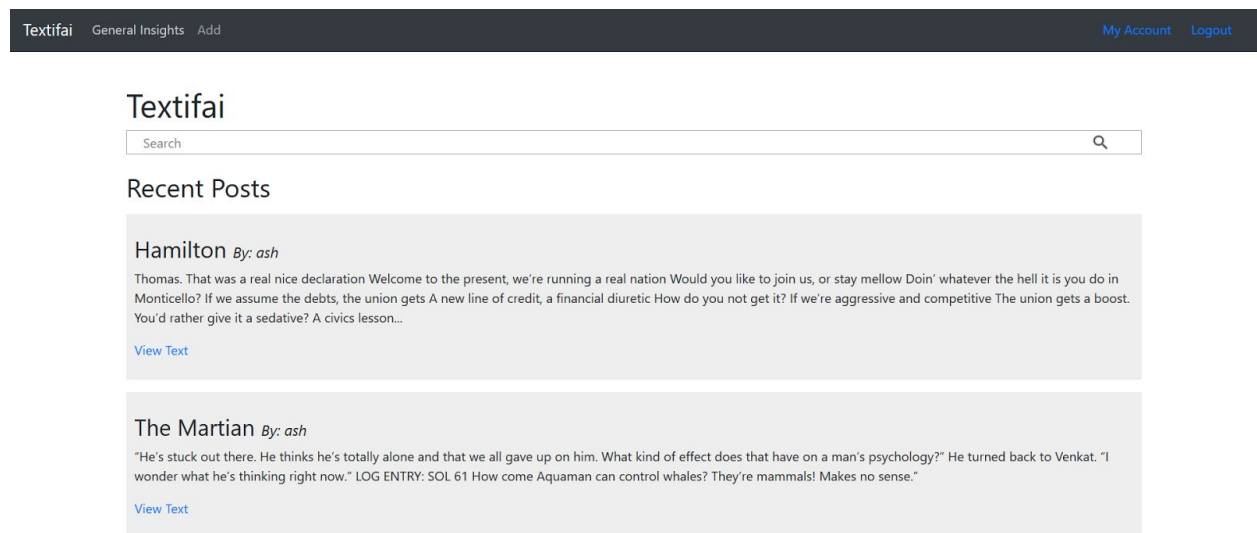


### Search Results

The search results view is displayed after a user submits a search query in the search form on the index page. It displays a list of all the submitted text that contains the text in the query. It's very similar to the list of texts on the index page, in that each result contains a title, the text content (or as much text content that fits), and a link to a unique featureoutput page that displays the analysis. If there aren't any text matches, it simply shows the search box and the text "No results found.". The user can also submit another search query from this page using an identical search form as the one from the index page.

## Textifai

Search                                                                                              🔍

**Search Results**

Some text to be analyzed *By: oz*

Entering in some text to be analyzed.

View Text

Testing search results *By: oz*

Some sample text.

View Text

## Login

The login page is where the user authenticates themselves. It can be accessed at any time from the toolbar at the top of every page if the user is not logged in. If a user is logged in, a link to log themselves out is displayed in its place.

## Textifai

Username: compsci326

Password: •••••••••

login

Lost password?

## General Insights

The General Insights view presents the user with table of global statistics which are compiled from each Text, User, and corresponding Insights. Some of these Global Insights are fixed and thus the are always available regardless of the content of the website. An example of this is the "Most Long-winded User" Global Insight. Others are created dynamically like, for example, "Average surprise." Global Insights which fall under this category may become present or not present on the page as the Users add data to the website. This allows for the comparison of Insight averages across dissimilar Texts and brief analysis of the Texts submitted to this website as a whole.

There is a second column in this view which shows the current Users contributions to the General Insights as a whole. This is an assortment of Insights compared with the General Insights displayed on the left. If there is no current User, a login suggestion is displayed instead.
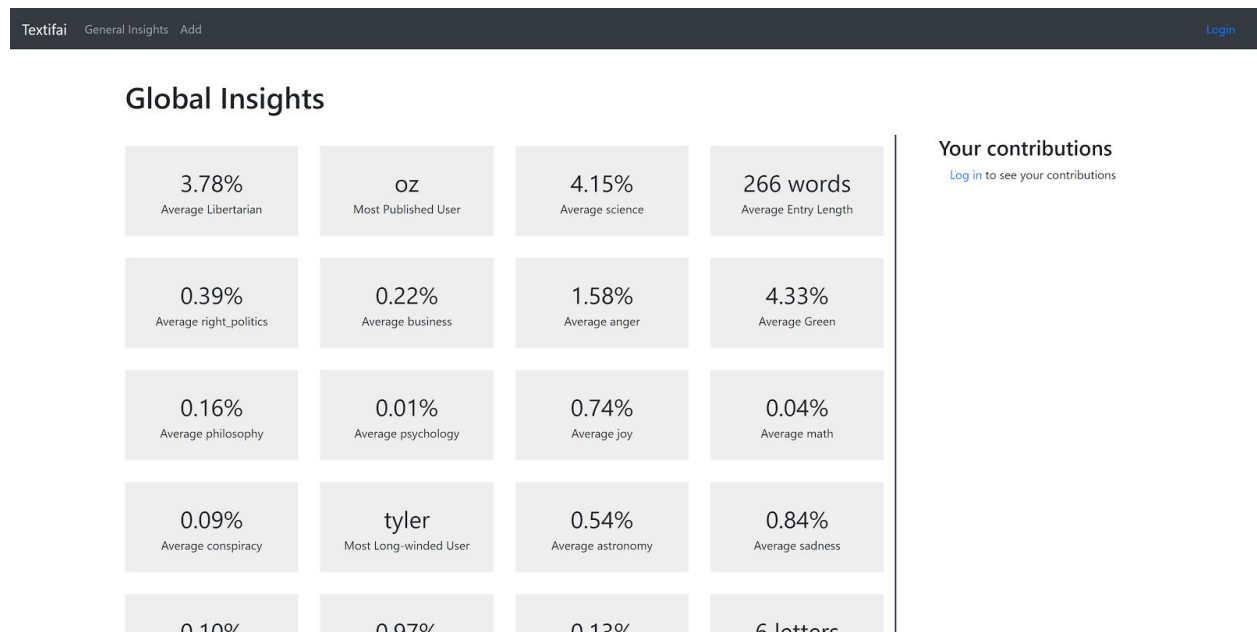


## Account

The account page allows a user to take a glimpse at their most recent Text submissions and comments, as well as analytics about their writing specifically. When first loaded it only displays 3 of each, but the user can also click to see their entire submission history. The account page is also where a user is first redirected when they log in, making it serve as a "home" of sorts.

# Your Profile

Welcome to your user page, *compsci326!*

## Recent Content                                                    See All
*What have you written recently?*

### Donald Trump Tweet *Dec. 10, 2017, 2:52 p.m.*

The media has been speculating that I fired Rex Tillerson or that he would be leaving soon - FAKE NEWS! He's not leaving and while we disagree on certain subjects, (I call the final shots) we work well together and ...

### d *Dec. 10, 2017, 2:52 p.m.*

i never repeat myself

### WOWOWO *Dec. 10, 2017, 2:52 p.m.*

you can't feel me like you can't feel me like you can't feel me like you can't feel me like you can't feel me like you can't feel me like you can't feel me like you can't feel me like ...

| Recent Comments | Writing Analytics |
| --- | --- |

Monday, December 11, 2017

## Text Input

The Text Input page is displayed after clicking the "Add" button on the navbar. Either select text form the "Sample text" dropdown menu, or enter a title and text content manually. Once the "Analyze" button is clicked, all the form data is sent to the backend and immediately analyzed. The user is then redirected to the feature output page to view the analysis of the text they just submitted.

## Text Analysis Input

Not sure what to write? Try Textifai with one of the samples from the dropdown menu below.

Sample text ▾

Enter Title

Enter text

Analyze

## Feature Output

The Feature Output page is displayed immediately after the user clicks "Analyze" on the Text Input page. This page can also be accessed by clicking submissions on the user account page, recent submissions on the index page, and query results on the search page. It contains a primary card with the submission title and text content. Immediately below the content is a list of insights, which are name, probability pairs. These insights are generated from various algorithms in the backend which use Indico.io and NLTK. The right column contains insights that are much more grammatical in nature, such as reading level and most common part of speech. Below the insights is a comment section, where users can discuss the text submission.. Note that users must be authenticated to comment, and must also be the original author to edit the submission.

Textifai   General Insights   Add                                                                    My Account   Logout

## Text Analysis

| Donald Trump Tweet |
|---|

The media has been speculating that I fired Rex Tillerson or that he would be leaving soon - FAKE NEWS! He's not leaving and while we disagree on certain subjects, (I call the final shots) we work well together and America is highly respected again!

| Insights | Probability |
|---|---|
| surprise | 0.3744044602 |
| agreeableness | 0.6279201633 |
| Conservative | 0.3218778372 |
| political_discussion | 0.08810767530000001 |
| conspiracy | 0.0773176717 |
| film | 0.06449200990000001 |
| right_politics | 0.055863309800000004 |
| relationships | 0.0363027926 |

[New Submission] [Edit]

Positivity neutral
Most Common Part of Speech pronoun
Reading Level college
Reading Time 13.5 seconds
Speaking Time 18.0 seconds
Total Sentences 2
Total Words 45
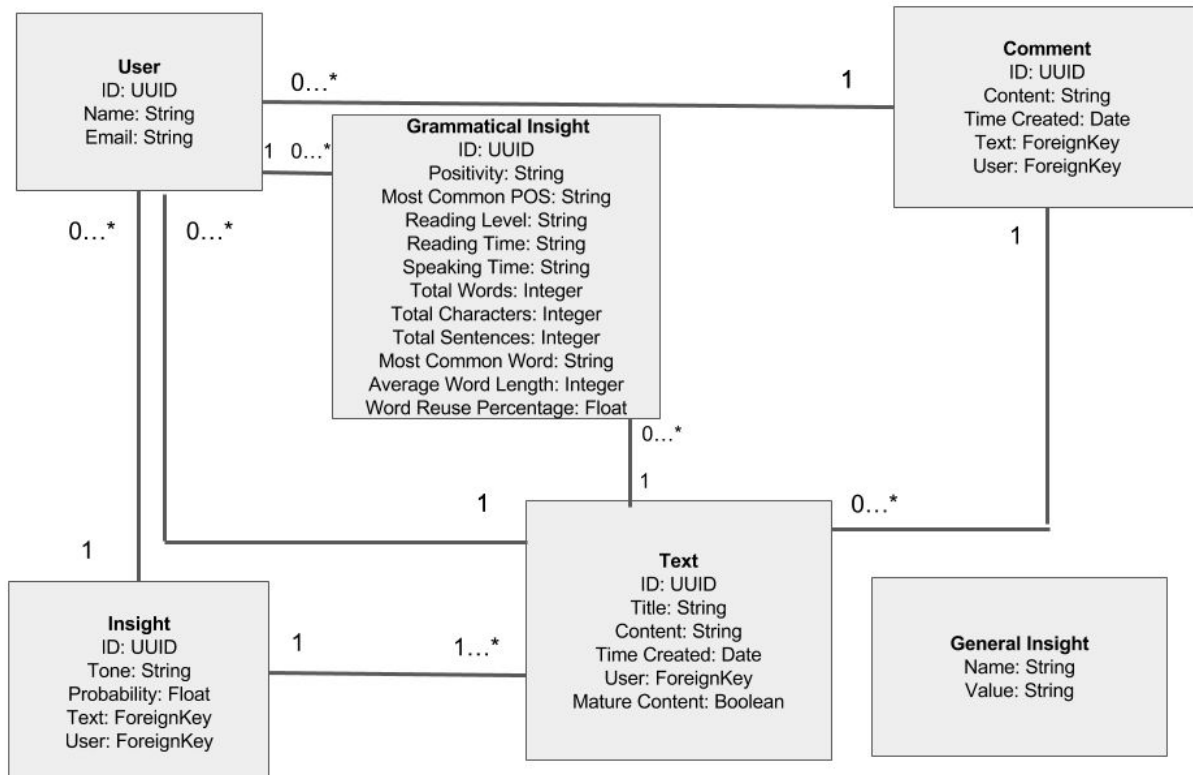Total Characters 249
Most Common Word that
Average Word Length 4
Word Reuse Percentage 0.111

Add Comment

[                                                    ]

[Add Comment]

## Data Model



The User model was originally going to be custom and extend the default Django User model, but we later found out that we didn't have any use for it and that the default Django User would suffice. It's included in this diagram because of the many foreign key relations it has with many other models. The Text model represents analyzed text that's displayed on the featureoutput page. The Text attributes are fairly self explanatory, and the analysis itself is represented in the form of insights as separate models with foreign keys. The Insight model is the first representation of text analysis; it has foreign keys to User and Text, and contains fields for tone and probability. The GrammaticalInsight model also contains attributes for text analysis output, such as positivity and reading level. Insights and GrammaticalInsights are displayed on the featureoutput page along with Text. The Comment model represents comments that are found on the featureoutput page. The GeneralInsight model represents the data that's found on the General Insights page. GeneralInsights are global statistics, where each statistic is a unique name and value pair based off of all the submitted text and user activity, such as the percentage of text that contains right wing politics, fear, philosophy, etc.

## URL Routes & Mappings

| URL route | Description | Authentication |
| --- | --- | --- |

| | | |
|---|---|---|
| index | Used for the landing page. Contains a feed of recently submitted text. | No authentication required for viewing this page. |
| textinput | This page is accessed by clicking "Add" on the navbar. It contains a title field and a content field. The text entered here is sent to the backend for analysis. | The user must be authenticated to view this page. If the user isn't authenticated, they're redirected to the login page. |
| featureoutput/<Text.m_id> | This page is used for the text analysis, and the url will have a unique text id. Users can access this page from clicking  text on the index page, the account page, search results, or immediately after submitting text. | No authentication required for viewing this page. However, if the user wishes to edit the submission, the user must be authenticated and the original author. In addition, the user must be authenticated to add a comment to this page. |
| account | A profile page for the authenticated user. It Contains information about submissions. | The user must be authenticated to view this page, and will be redirected here immediately after logging in. |
| general-insights | Navigate here by clicking the General Insights button on the navbar. It contains global statistics. | No authentication required, but the user must be authenticated to see how their submissions contribute to the overall statistics. |
| texts | Navigate here by clicking "See All" on the account page. This displays a list of all the user's submitted text. | The user must be authenticated to view this page. |
| searchresults | This page is displayed after submitting a search query on the index page. It lists all text submissions that contain the text in the user's query. | No authentication is required here. |

## Authentication/Authorization

Users are authenticated through a typical login/logout page. If they aren't authenticated, they can't access pages that describe the logged in user (their profile page, their user-specific insights, etc.), but they can still see the front page and texts. Users must be authenticated to add text, and are redirected to the login page if they click the Add button on the navbar. On the General Insights page, users must be authenticated to view how their submissions contribute to the overall statistics. On the text analysis page for a given submission, users must be authenticated to add comments, and must also be the original author to edit the submission. Users must also be authenticated to view the My Account page, and are redirected to that page immediately after logging in.

## Team Choice

Our team choice was simply to deploy the website using a service like Heroku. We wanted to see our website live, as the more users it has, the more useful it is as a writing aggregation service and the more interesting it is to new users. Additional steps had to be followed to deploy the app successfully. These steps included adding a procfile, dependency files for both nltk and pip packages, as well as some minor changes in wsgi.py and settings.py. Please note that we're using Heroku's free tier, so the web dyno will sleep if there isn't any traffic in a 30 minute period. This means that there may be a short delay before the website loads, since the web dyno takes a little bit of time to become active again. It can be accessed via this url: http://textifai.herokuapp.com

## Conclusion

Each of our team members learned something while creating Textifai, but each of us had a different thing to learn. Whether that was the usage of Django, the implementation of NLP libraries, or the understanding of HTML and CSS, each member was inexperienced in some field but became more proficient at it as the application progressed. Our main issue may well have been that the majority of the website's implementation was piled up at the end of the semester; if we had known that ahead of time we would have spaced out the work more evenly.