

Project overview:

I decided to mash together a few html texts using Markov analysis. I pulled from the Olin Honor Code, the Olin promotional page, and a wikipedia article containing a list of US felonies. I hoped to create a more interesting honor code.

Implementation:

The major components were mostly turning the text into one giant list, making a dictionary out of that list keyed to words that are likely to follow, and making a random function that follows through the dictionary and adds the words into a sentence. An interesting data structure was the dictionary. To create it, I ran through the giant list of all words and added new words to the dictionary. That means if “the” appears twice, it doesn't get two keys. I then keyed the words to the values that follow them, which finally made a dictionary of words that was keyed to a few words that were likely to follow them.

When I ran through the dictionary and made the actual sentence, I had to make some design decisions about what I wanted the output to look like. I had to make the string stop at some point and I finally decided to make it stop when it saw a period which would mean it forms sentences. I could have instead decided to stop it after a certain length or after a certain word. The “length” way might have stopped it in the middle of a sentence which wouldn't have made sense and stopping it after a word might also not have had a satisfying end. However, one downside to stopping it after every period is that some abbreviations (Roe v. Wade, U.S.) get cut off midway. These were rare enough that I just reran the code if it stopped halfway.

Another decision was where to start the sentence. It doesn't always make sense to start a sentence with “of” or “for”, and I soon realized that the most logical sentences were those that started with capital letters because those words started sentences in the original text. I filtered my code for capital letters and chose my starting word out of that list.

Results:

Here are some of the best results from the text masher.

Olin College will be a new 16 credit, multi-semester course of murders.

At Olin, it's not just a felony, although possession of imprisonment or possession of controlled substances may be a deadly weapon may be viewed as boat design and understanding of fellow community members, and battery, or any time elements, such as either death or any time and facial recognition.

The Olin Honor Code or death penalty, reserved for other majors on subsequent offenses.

College through a “collaboratory,” dedicated to break-ins at any prison or an issue within the Honor Code or most felonies remain a person.

At Olin College I will strive to commit larceny, assault and willing to better myself accurately and my work, my own behavior.

Something I will care for other majors on the death penalty, reserved for certain types of their potential punishment upon conviction.

Every student learns about what they do with and completely in jail.

At Olin, it's not just about software, electronics and American jurisdictions.

QEA Final Project #1 Quantitative Engineering Analysis is a felony.

Code Values: Integrity I will be generally legal, but without the 20th century.

Indeed, at night with Olin College.

Reflection:

Honestly, this project was pretty hard for me. I pulled the text from the web by myself and put it into a big list successfully but the dictionary and text masher I needed lots of help for. I probably went to 6 or 7 different ninjas, pirates, and classmates, all of whom had a different idea about how it should be done. I understood conceptually what the code should do (for the most part) but actually implementing it seemed almost impossible. I don't know if one of the other projects would have been easier though because half of the fun of this project was the fact that the result was super funny and I had something to look forwards to. Before I started this, I wish I had known about more of the python commands that do a lot of the work for you instead of trying to write code from scratch. According to someone who knew programming, I had been writing what was basically C code? So moving forward, I will try to spend some of my free time going over online documentation about python commands and seeing which ones are out there.