

Project: Text Generation (Part One)

Summary: In this laboratory, you will explore techniques for generating “random” texts.

Contents:

- Exercises
 - Exercise 0: Preparation
 - Exercise 1: Initial Explorations
 - Exercise 2: Your Own Words
 - Exercise 3: Capitalizing Sentences
 - Exercise 4: Other Sentence Structures
 - Exercise 5: Intransitive Verbs
- For Those With Extra Time

Exercises

Exercise 0: Preparation

a. Make your own copies of the following files. (Note that the easiest way to make your own copy is to right click on the file and select Save Link As ... or something similar. If the browser tries to add a .htm or html extension, delete that extension and do not add .ss.)

- `textgen.ss`
- `adjectives.txt`
- `articles.txt`
- `nouns.txt`
- `tverbs.txt`

b. Open your copy of `textgen.ss`. You will be working with this copy throughout this lab and the next.

Exercise 1: Initial Explorations

a. Generate ten “random” sentences using `(sentence)`.

b. What sentence structures do you see in those ten sentences?

c. Make a list of the parts of speech you see, what words appear for each part of speech you see, and the relative frequency of each word within its part of speech.

d. Suppose we were to generate ten “random” nouns using `(generate-part-of-speech 'noun)`. What words would you expect to see, and in what relative frequencies?

e. Look at `nouns` by opening it in DrScheme. Do the frequencies stated in that file match those you encountered in parts a and c?

Exercise 2: Your Own Words

a. Edit your copies of `adjectives.txt`, `nouns.txt`, and `tverbs.txt` to insert at least four more words in each. (You can edit each file in DrScheme, but you should not try to run it.) Note that you will also have to change the frequencies of the existing words to make the total be 1000. (You can check the frequencies with `(check-words-file filename)`.)

You should not spend more than five or so minutes updating these copies. (You'll have a chance to make more changes later.)

b. Generate ten new sentences. Did your sentences include your own words? If not, read through the code to see where `textgen.ss` looks for the words (hint, it's a variable called `root`) and then update it to use your words. (If you saved the files on your desktop, the new `root` should be `"/home/username/Desktop/"`.)

Exercise 3: Capitalizing Sentences

One issue with the `sentence` procedure is that it does not capitalize the sentences it writes.

a. Write a procedure, `(capitalize str)`, that capitalizes the first letter in `str`.

b. Update `sentence` to use this new procedure.

c. Generate ten or so sentences to ensure that your update has worked.

Exercise 4: Other Sentence Structures

a. Update `sentence` so that it randomly selects between two sentence structures. (You may choose the alternate structure.)

b. Generate ten or so sentences to ensure that your update has worked.

Exercise 5: Intransitive Verbs

a. Create a new words file, `iverbs.txt`, that contains a variety of intransitive verbs. Recall that each line of the file should have the form `("word" frequency)` and that the sum of the frequencies should be 1000.

b. Use `check-words-file` to make sure that your file has the correct format.

c. Update `sentence` to add a third kind of sentence, one that includes intransitive verbs.

For Those With Extra Time

If you find that you've finished all of the work, you might consider

- Playing with more variants (longer word lists, new sentence structures, etc.);
- Starting to read the second reading on text generation;
- Thinking about the lab report.

Copyright © 2007 Samuel A. Rebelsky. This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.