11-601 Coding & Algorithms Bootcamp

Node.js Document Summary Statistics Application

**Task:**

In this task, you will utilize Node.js to create a simple command line application that provides summary statistics for a given dataset. This will not require using Node.js as a web server, but rather will use it as a command line application only. The application you are building will take the relative path to a text file (.txt) as a command line argument. Your application should then process the text file to generate the following descriptive statistics about the document:

1. The number of words in the document
2. The number of lines
3. A list of all the words with the count of their occurrence, in descending order
4. A list of the ten most common trigrams
5. A listing of the edit distances between the most common trigram and the next ten most common trigrams, in order of distance value. This should include the trigrams being compared and the edit distance value.

The output of your program should be a properly formatted HTML file that can be viewed in any modern browser. You do not need to serve the HTML via HTTP server, just generate the HTML report and store it in the same directory as the application is run from. The name of the HTML report should include the name of the data file provided as input.

This assignment requires that you complete the entire application in Node.js and JavaScript. You may not use low-level unix/windows command line helper utilities (such as wc).

*Working or studying together requires disclosure in your assignment submission, including names of all participants.* **Copying or sharing code is always forbidden.**

**Data:**

For development, use the text file downloadable at:

http://www.gutenberg.org/cache/epub/45/pg45.txt

We will verify your code using this document

**Minimum deliverables:**

Source code of your Node.js application

HTML report output page

Submit your assignments as .zip files through BlackBoard

**Due Date:**

December 8, 2015

**Readings and Resources:**

*Node.js:*

https://nodejs.org/

https://nodejs.org/en/docs/

https://nodejs.org/api/

*Node.js word counting:*

http://stackoverflow.com/questions/18287323/split-and-count-words-in-string-node-js

*Node.js file operations:*

http://www.tutorialspoint.com/nodejs/nodejs\_file\_system.htm