Programming Exercise

for Software Developer Candidates

Bigram Parsing

Create an application that can take as input any text file and output a histogram of the bigrams in the text.

A bigram is any two adjacent words in the text. A histogram is a kind of graphical representation of numerical data.

A well-formed submission will be executable and accompanied by tests. You may do this in any language and use any framework or data structures to read files, build the output, and run tests. **The submission should be your own work.** Great candidates go above and beyond the problem statement and think about edge cases and failure.

Example

Given the text

"The quick brown fox and the quick blue hare."

The bigrams and their counts would be

- "the quick": 2
- "quick brown": 1
- "brown fox": 1
- "fox and": 1
- "and the": 1
- "quick blue": 1
- "blue hare": 1

Submission

You will be provided a folder where you can upload your submission or you may provide a URL where your submission can be accessed.

Tips

Pick your own text to use as an input. There are several edge-cases in this problem that our sample text does not expose. It's up to you to find out what those are and handle them (just like you would be expected to do in any professional setting).

Present your best work. We'll use this as a representation of the code you'd write and the deliverables you'd present if we were colleagues.

Handle this problem the way you'd handle any other problem you've been given professionally. This is not just a "can you code" exercise. It's a "how would this person handle this problem if they were hired today?"

Give us a buildable solution. The first thing we're going to do is try to run your code. If we can't run it, it's really hard to evaluate how good your solution was.

Don't reinvent the wheel. (Unless of course you'd reinvent the wheel if you were hired).

Questions

If you have questions, please address them to toni.ham@vaeit.com; we'll be available to turn around answers during work day hours.