

Objectives

- Practice solving problems using the Map-Reduce paradigm.

The Lab

In this lab we will explore the Map-Reduce paradigm using a simplified Map-Reduce framework written in `python`.¹

1. The code we will use is in the `Lab8.zip` file. Download it and prepare a project.

There are two `python` files in the zip:

`MapReduce.py` is the implementation of the Map-Reduce framework. **Do not modify this file.**

`wordcount.py` uses Map-Reduce to count the number of words in a document (e.g. the Declaration of Independence).

Take a minute to study how the solution is implemented. What does `emit_intermediate` do?

2. Run the `wordcount.py`. (You may need to right-click and **Run** if using PyCharm.) Study the results and the source file `dec.json`. Do they make sense? You can modify `wordcount.py` to use the `test.json` file as input instead if you want to see results on a smaller dataset.
3. Then, for each of the following problems, modify `wordcount.py` (**only!**) to solve them.

Problem 1 count the number of words in the Declaration of Independence for each possible length (e.g. the number of 1-letter words, the number of 2-letter words, etc.).

Problem 2 count the number of *tiny*, *small*, *medium*, and *big* words in the Declaration of Independence, where *big* words have 10+ letters, *medium* words have 5-9 letters, *small* words have 2-4 letters, and *tiny* words have only 1 letter.

For all problems, don't worry about punctuation—we will consider `"world_"` and `"world"` to be different words (one with 6 letters, one with 5).

N.B. These problems are relatively easy to solve in `python` without involving Map-Reduce. However, we want to find a “Map-Reduce” solution, rather than a “python” solution. If you find yourself changing the framework, or doing all the work in either `mapper` or `reducer`, you may wish to reconsider your approach.

Deliverable

Write a brief report showing your successful completion of the lab. For each problem, include in your report:

- The output of your solution.
- A brief description of the intermediate form (What were the keys? the values?).

¹If you do not have a python-capable environment on your computer, you should install python (<https://www.python.org/downloads/>) and optionally an IDE (like PyCharm: <https://www.jetbrains.com/pycharm/download/>)