assignment2

- Rhea Mae Edwards (**edwardrh**), Miles Curry (**currymi**) & Rutger Farry (**farryr**)
- CS325
- Dr. Xiaoli Fern
- 24 February 2016

Running the code

The script is built in Python 2, so running it should be simple on most computers:

```
python sequence_align.py {cost_file} {input_file}
```

Alternatively, you can put the script in the same directory as a cost file named `imp2cost.txt` and an input file named `imp2input.txt` and just run with no arguments:

```
python sequence_align.py
```

Generated output will be placed in: `./imp2output.txt`

Pseudocode

```
Given two strings s and t { // Need to turn s into t
                           // Use minimum editing operations
                  // D(m,n) = the edit distance between s1s2...si and t1t2...ti
    For i = 0 to m: D(i,0) = i
        For j = 0 to n: D(0,j) = j

            // Calculate edit distance
        // Remember alignment for visual output of computation (backtrace)
            For each i = 1...m
                For each j = 1...n
                    D(i,j) = min(
                        D(i-1,j) + cost(si, -),
                        D(i,j-1) + cost(-, tj),
                        D(i-1,j-1) + cost(si, tj)
                    )
                    Save backtrace to ptr(i,j)

            // Return sequences and cost
            Return D(m,n) and backtrace of ptr(m,n)
}
```

Runtime analysis

The overall runtime of this algorithm is O(mn), where m and n are the sizes of the two input sequences, respectively. Backtracing takes O(m+n) time, so the overall runtime for computing the cost and then backtracing is:

O(mn) + O(m+n) = O(mn)

Runtime graphs

These graphs were generated automatically using `matplotlib` by the `profiler.py` script in this directory. The graphs seen here were generated on a 2015 MacBook with a 1.2GHz Intel Core m5 processor.
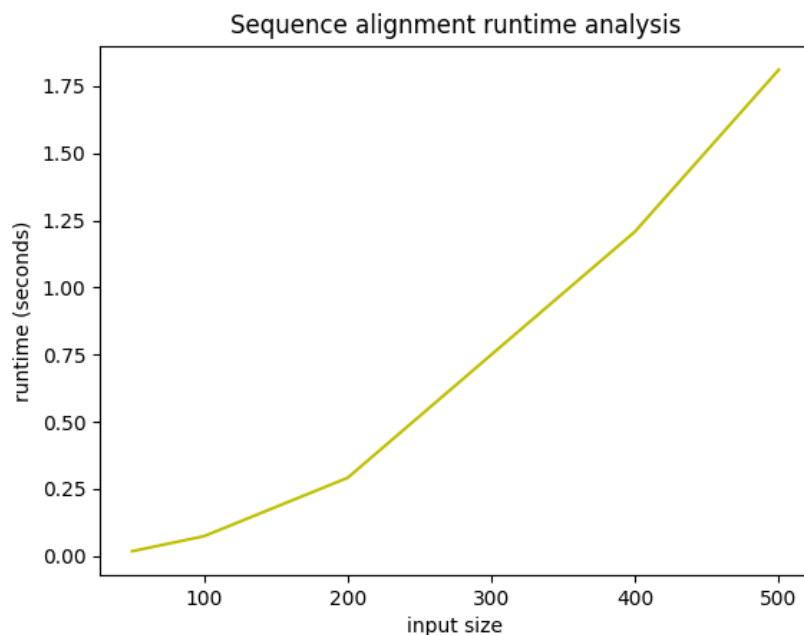

Figure 1: Runtime line graph

Graph interpretation and discussion

By looking at the logarithmic line chart, we can correctly infer the algorithm's O(mn) runtime (which appears to be O(n^2) since m = n in our tests).
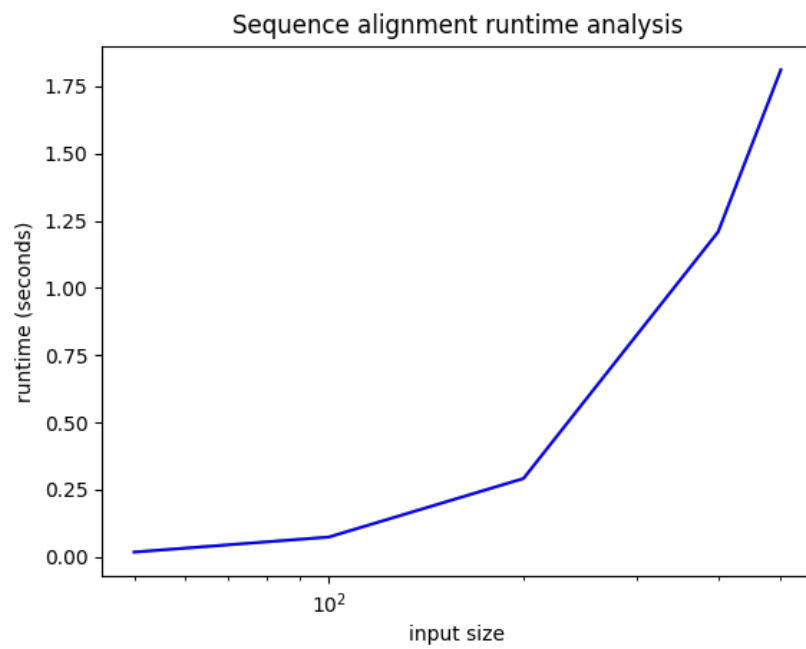
Figure 2: Runtime log line graph

Meta

If you're reading the PDF version of this file, we generated it from `README.md` using pandoc. You can update it from the README file using:

```
pandoc README.md --latex-engine=xelatex -o writeup.pdf
```