

Jacek Radajewski

SCIE2100 Prac 4

Questions 1 & 2

Q1

Time taken to run the triplet alignment was approximately 0.62 seconds. Pairwise alignment on the other hand completed in 0.005 seconds.

Q2

In this question we have doubled the sequence length and it has taken 4.74 seconds to calculate the global alignment of the triplet. The time taken to execute the algorithm has increased by $4.74 / 0.62$ which is around 7.65. In triplet alignment we are processing a three dimensional matrix so by increasing the the length of our sequence by a factor of 2 we increase the total number of the cells in the matrix by a factor of $2^3 = 8$. The actual increase in execution time was 7.65.

I have written code which calculates the execution time for computing triplet global alignment for sequences of length from n to $10n$. These were graphed next to expected execution times and shown below. The expected calculations for $2 \leq n \leq 10$ is based on $n=1$.

Code

```
...
```

Created on 15/04/2014

@author: s4361277

```
'''
from sequence import *
import time

seqA1 = Sequence("AGCGCGATTATATAAGACGGACGGCTAAAG")
seqB1 = Sequence("AGCGGATATTTATATCGCACGACGACTACG")
seqC1 = Sequence("GGATCGATTATATAGCCTGGACGAGACATG")

seqA2 = Sequence("AGCGCGATTATATAAGACGGACGAGACGACAGCGCGATAGACGAGACGGACGAGACGACT")
seqB2 = Sequence("GGATCGATTATATAGCCTGGACGAGACATGGGATCGAAGACGACCTGGACGAGACATGAC")
seqC2 = Sequence("AGCGGATATTTATATCGCACGACGACTACGAGCGGAAGACGAGTTTCGCACGACACTACG")

matrix = readSubstMatrix("dna.matrix", DNA_Alphabet)

print "pairwise align of short sequences (Q1)"
start_time = time.time()
alignGlobal(seqA1, seqB1, matrix, -1)
end_time = time.time()
print("Elapsed time was %g seconds" % (end_time - start_time))

print "triplet align of short sequences (Q1) ..."
start_time = time.time()
tripletAlignGlobal(seqA1, seqB1, seqC1, matrix, -1)
end_time = time.time()
print("Elapsed time was %g seconds" % (end_time - start_time))

print "triplet align of long sequences (Q2) ..."
start_time = time.time()
tripletAlignGlobal(seqA2, seqB2, seqC2, matrix, -1)
end_time = time.time()
print("Elapsed time was %g seconds" % (end_time - start_time))

# the following code provides a more detailed analysis of algorithm time
# the output was use to genmerate a comparison graph of expected vs real
# execution times

print "triplet align: time vs sequence length (Q2) ..."
sequence_string_fragment_1 = "AGCGCGATTATATAAGACGGACGGCTAAAG"
sequence_string_fragment_2 = "AGCGGATATTTATATCGCACGACGACTACG"
sequence_string_fragment_3 = "GGATCGATTATATAGCCTGGACGAGACATG"

sequence_string_1 = ""
sequence_string_2 = ""
sequence_string_3 = ""

for length_multiplier in range(1, 11):
    sequence_string_1 += sequence_string_fragment_1
    sequence_string_2 += sequence_string_fragment_2
    sequence_string_3 += sequence_string_fragment_3
    start_time = time.time()
    tripletAlignGlobal(Sequence(sequence_string_1), Sequence(sequence_string_2),
                      Sequence(sequence_string_3), matrix, -1)
    end_time = time.time()
    # output produced can be redirected into a CSV file
    print length_multiplier, ",", end_time - start_time
```

Execution

Screen shot below shows the execution of the code for questions 1 and 2.

```
Terminal
File Edit View Search Terminal Help
jacekrad@z400 ~/var/github/prac04/src $ time python question1_2.py
pairwise align of short sequences (Q1)
Elapsed time was 0.00509787 seconds
triplet align of short sequences (Q1) ...
Elapsed time was 0.62357 seconds
triplet align of long sequences (Q2) ...
Elapsed time was 4.74108 seconds
triplet align: time vs sequence length (Q2) ...
1 , 0.580306053162
2 , 4.61684989929
3 , 15.7325119972
4 , 36.8982918262
5 , 72.221380949
6 , 125.038820982
7 , 200.297755003
8 , 294.624313831
9 , 427.168880939
10 , 577.339603901

real    29m20.062s
user    29m18.732s
sys      0m0.336s
```