

Profiling

Caden Corontzos

I used callgrind to profile my code. Callgrind finds out how many instructions are associated with each line of code, which gives an indication of how long each line will take. I looked at some of the trouble spots that Eitan and I found to see if we could see a noticeable difference in runtime before and after each change.

I used the two groups of DNA I had found the other day as benchmarks.

Here is the data before any optimizations.

Table 1: Corpus 1

| File.Name | Original.File.Size | Compressed.Size | Compression.Ratio | Compression.Time | Decompression.Time |
|--------------------|--------------------|-----------------|-------------------|------------------|--------------------|
| DNACorpus1/chmpxx | 121024 | 43516 | 2.781 | 298 | 58 |
| DNACorpus1/chntxx | 155844 | 58336 | 2.671 | 410 | 89 |
| DNACorpus1/hehcmv | 229354 | 85526 | 2.682 | 607 | 112 |
| DNACorpus1/humdyst | 38770 | 15300 | 2.534 | 95 | 25 |
| DNACorpus1/humghcs | 66495 | 25552 | 2.602 | 173 | 35 |
| DNACorpus1/humhbb | 73308 | 28134 | 2.606 | 184 | 44 |
| DNACorpus1/humhdab | 58864 | 22699 | 2.593 | 149 | 41 |
| DNACorpus1/humprtb | 56737 | 21902 | 2.590 | 157 | 34 |
| DNACorpus1/mpomtgc | 186609 | 70254 | 2.656 | 535 | 110 |
| DNACorpus1/mtpacga | 100314 | 36862 | 2.721 | 238 | 59 |
| DNACorpus1/vaccg | 191737 | 70067 | 2.736 | 455 | 105 |

Table 2: Corpus 2

| File.Name | Original.File.Size | Compressed.Size | Compression.Ratio | Compression.Time | Decompression.Time |
|-----------------|--------------------|-----------------|-------------------|------------------|--------------------|
| DNACorpus2/AeCa | 1591049 | 556535 | 2.859 | 4376 | 799 |
| DNACorpus2/AgPh | 43970 | 17442 | 2.521 | 112 | 27 |
| DNACorpus2/AnCa | 142189675 | 43665091 | 3.256 | 398309 | 60476 |
| DNACorpus2/BuEb | 18940 | 7893 | 2.400 | 57 | 13 |
| DNACorpus2/DaRe | 62565020 | 19586457 | 3.194 | 162491 | 28269 |
| DNACorpus2/DrMe | 32181429 | 10619042 | 3.031 | 85115 | 16587 |
| DNACorpus2/EnIn | 26403087 | 8609993 | 3.067 | 72791 | 10172 |
| DNACorpus2/EsCo | 4641652 | 1593404 | 2.913 | 11278 | 2069 |
| DNACorpus2/GaGa | 148532294 | 46851765 | 3.170 | 390118 | 56564 |
| DNACorpus2/HaHi | 3890005 | 1306708 | 2.977 | 9912 | 1648 |
| DNACorpus2/HePy | 1667825 | 566972 | 2.942 | 4217 | 789 |
| DNACorpus2/HoSa | 189752667 | 57200209 | 3.317 | 518318 | 65483 |
| DNACorpus2/OrSa | 43262523 | 14148071 | 3.058 | 114892 | 17376 |
| DNACorpus2/PlFa | 8986712 | 2895744 | 3.103 | 24248 | 4125 |
| DNACorpus2/ScPo | 10652155 | 3590856 | 2.966 | 28334 | 4485 |
| DNACorpus2/WaMe | 9144432 | 3112000 | 2.938 | 25484 | 3904 |
| DNACorpus2/YeMi | 73689 | 27235 | 2.706 | 183 | 45 |

I then tested my implementation with callgrind. I encoded HaHi from DNA Corpus 2 to see what lines are taking long.