Project 2: Autocomplete – Analysis
Brian Jordan
Conclusions
How changing the size of the prefix changes runtime:
BruteAutocomplete
        The methods topMatch() and topMatches() for the class BruteAutocomplete only depend on the number N of elements of data fixing a complexity of O(N) for both methods. This occurs because both methods loop through all elements of N once. For this reason, size of the prefix should not affect the runtime of either of these methods based on the code. The data for topMatch() differs from this prediction because it showed a slight increase in runtime as the prefix increased. The data for the topMatches() follows the prediction closer by runtimes appearing random based on changing prefix lengths.
BinaySearchAutocomplete
        The methods topMatch() and topMatches() for the class BinarySearchAutocomplete should have runtimes that decrease as the length of the prefix increases because there will be a smaller set of words to compare the weights of if the prefix is longer. The degree to which this will affect the runtime will depend on the set of words. The data supports this prediction by showing an increase in runtime as the prefix size decreases.
TrieAutocomplete
        The methods topMatch() and topMatches() for the class TrieAutocomplete should have runtimes that increase as the length of the prefix because the methods need to search for each character of the prefix going down the tree. The data shows the opposite of this observation

How changing k affects runtime:
BruteAutocomplete
        As discussed above, runtime for the topMatches() method of the BruteAutocomplete class only depend on N so changing the value of k in the parameter should not have an effect on runtime. The data supports this prediction by runtimes being random but relatively the same as the value of k increases.
BinarySearchAutocomplete
        The method topMatches() of the BinarySearchAutocomplete class should not have its runtime affected by a change in the k parameter passed to it. This is due to the fact that k does not change the number of words the method must loop through to find the top weighted matches. It only effects the number of returned words. This prediction is supported by the data because as k changes, the runtimes only slightly change randomly but remain about the same.
TrieAutocomplete
        The method topMatches() of the TrieAutocomplete class should not have a change in runtime based on k because it only depends on the length of the substring.

How changing N affects runtime
        Changing N affects the runtime of the topMatch() and topMatches() methods of Brute Autocomplete linearly as discussed above. This is supported by the data, because the run times of the methods using fourletterwords.txt are about half that of the methods using fourletterwordshalf.txt. Changing N also increases the time of BinarySearchAutocomplete as well. It does not change the runtime of TrieAutocomplete

Total number of Nodes used for TrieAutocomplete

        Using the fourletterwords.txt file the TrieAutocomplete class creates 475255 nodes.

Using the fourletterwordshalf.txt file the TrieAutocomplete class creates 237628 nodes.

AutocompleterBenchmark Data

Using fourletterwords.txt

BruteAutocomplete

Found 456976 words
Time to initialize -   0.057408507
Time for topMatch("") - 0.001261739094
Time for topMatch("nenk") - 0.003168339933
Time for topMatch("n") - 0.001793930447
Time for topMatch("ne") - 0.001899560284
Time for topMatch("notarealword") - 0.003833707619
Time for topKMatches("", 1) - 0.00564757076
Time for topKMatches("", 2) - 0.00543564288
Time for topKMatches("", 4) - 0.00561036549
Time for topKMatches("", 8) - 0.00547202845
Time for topKMatches("", 16) - 0.0055313071
Time for topKMatches("", 32) - 0.00538704946
Time for topKMatches("", 64) - 0.00547412879
Time for topKMatches("", 128) - 0.00398283244
Time for topKMatches("", 256) - 0.00271105264
Time for topKMatches("nenk", 1) - 0.00249992614
Time for topKMatches("nenk", 2) - 0.00238801207
Time for topKMatches("nenk", 4) - 0.00225964146
Time for topKMatches("nenk", 8) - 0.00228902768
Time for topKMatches("nenk", 16) - 0.00233153364
Time for topKMatches("nenk", 32) - 0.00235203232
Time for topKMatches("nenk", 64) - 0.00222453979
Time for topKMatches("nenk", 128) - 0.00232319641
Time for topKMatches("nenk", 256) - 0.00240034936
Time for topKMatches("n", 1) - 0.00235988888
Time for topKMatches("n", 2) - 0.00229733704
Time for topKMatches("n", 4) - 0.00224640243
Time for topKMatches("n", 8) - 0.0024610281
Time for topKMatches("n", 16) - 0.00235273998
Time for topKMatches("n", 32) - 0.00230264973
Time for topKMatches("n", 64) - 0.00236611657
Time for topKMatches("n", 128) - 0.00237180385
Time for topKMatches("n", 256) - 0.00235967544
Time for topKMatches("ne", 1) - 0.00231770983
Time for topKMatches("ne", 2) - 0.00238334982
Time for topKMatches("ne", 4) - 0.00233346848
Time for topKMatches("ne", 8) - 0.00234718475
Time for topKMatches("ne", 16) - 0.00239747756
Time for topKMatches("ne", 32) - 0.00228719046
Time for topKMatches("ne", 64) - 0.00229096003
Time for topKMatches("ne", 128) - 0.00231416085
Time for topKMatches("ne", 256) - 0.00237731578
Time for topKMatches("notarealword", 1) - 0.00185561726
Time for topKMatches("notarealword", 2) - 0.00193602511
Time for topKMatches("notarealword", 4) - 0.00185102196
Time for topKMatches("notarealword", 8) - 0.00187922793
Time for topKMatches("notarealword", 16) - 0.00195102162
Time for topKMatches("notarealword", 32) - 0.00185888969
Time for topKMatches("notarealword", 64) - 0.00196432249
Time for topKMatches("notarealword", 128) - 0.00193509991
Time for topKMatches("notarealword", 256) - 0.0019619241

BinarySearchAutocomplete

Found 456976 words
Time to initialize - 0.028837956
Time for topMatch("") - 5.63920613E-4
Time for topMatch("nenk") - 1.30705E-6
Time for topMatch("n") - 1.4913115E-5
Time for topMatch("ne") - 2.350927E-6
Time for topMatch("notarealword") - 1.251824E-6
Time for topKMatches("", 1) - 0.00477854418
Time for topKMatches("", 2) - 0.00584068124
Time for topKMatches("", 4) - 0.00905636692
Time for topKMatches("", 8) - 0.01166161353
Time for topKMatches("", 16) - 0.01498259786
Time for topKMatches("", 32) - 0.01757510458
Time for topKMatches("", 64) - 0.0202710071
Time for topKMatches("", 128) - 0.02402096847
Time for topKMatches("", 256) - 0.02717570118
Time for topKMatches("nenk", 1) - 2.60347E-6
Time for topKMatches("nenk", 2) - 8.2255E-7
Time for topKMatches("nenk", 4) - 7.8024E-7
Time for topKMatches("nenk", 8) - 7.1251E-7
Time for topKMatches("nenk", 16) - 7.1989E-7
Time for topKMatches("nenk", 32) - 7.3514E-7
Time for topKMatches("nenk", 64) - 7.9492E-7
Time for topKMatches("nenk", 128) - 9.2034E-7
Time for topKMatches("nenk", 256) - 1.135E-6
Time for topKMatches("n", 1) - 1.8335673E-4
Time for topKMatches("n", 2) - 2.2197133E-4
Time for topKMatches("n", 4) - 3.2334744E-4
Time for topKMatches("n", 8) - 4.8972403E-4
Time for topKMatches("n", 16) - 6.2486842E-4
Time for topKMatches("n", 32) - 7.1367833E-4
Time for topKMatches("n", 64) - 8.4821324E-4
Time for topKMatches("n", 128) - 9.8449923E-4
Time for topKMatches("n", 256) - 0.00105835172
Time for topKMatches("ne", 1) - 1.167643E-5
Time for topKMatches("ne", 2) - 1.259943E-5
Time for topKMatches("ne", 4) - 1.690659E-5
Time for topKMatches("ne", 8) - 2.157757E-5
Time for topKMatches("ne", 16) - 2.22287E-5
Time for topKMatches("ne", 32) - 2.637077E-5
Time for topKMatches("ne", 64) - 3.058931E-5
Time for topKMatches("ne", 128) - 3.599521E-5
Time for topKMatches("ne", 256) - 4.562756E-5
Time for topKMatches("notarealword", 1) - 5.7565E-7
Time for topKMatches("notarealword", 2) - 3.2228E-7
Time for topKMatches("notarealword", 4) - 3.8427E-7
Time for topKMatches("notarealword", 8) - 4.3968E-7
Time for topKMatches("notarealword", 16) - 4.3985E-7
Time for topKMatches("notarealword", 32) - 4.5585E-7
Time for topKMatches("notarealword", 64) - 3.3361E-7
Time for topKMatches("notarealword", 128) - 2.66E-7
Time for topKMatches("notarealword", 256) - 3.2397E-7

TrieAutocomplete

Found 456976 words
Time to initialize - 0.104590736
Created 475255 nodes
Time for topMatch("") - 4.653431E-6
Time for topMatch("nenk") - 2.48762E-7
Time for topMatch("n") - 1.186121E-6

Time for topMatch("ne") - 2.185949E-6
Time for topMatch("notarealword") - 2.60784E-7
Time for topKMatches("", 1) - 4.574626E-5
Time for topKMatches("", 2) - 2.001494E-5
Time for topKMatches("", 4) - 5.458613E-5
Time for topKMatches("", 8) - 3.34548E-5
Time for topKMatches("", 16) - 5.621542E-5
Time for topKMatches("", 32) - 1.1847182E-4
Time for topKMatches("", 64) - 2.3761505E-4
Time for topKMatches("", 128) - 4.5737045E-4
Time for topKMatches("", 256) - 8.6404725E-4
Time for topKMatches("nenk", 1) - 3.64976E-6
Time for topKMatches("nenk", 2) - 1.09283E-6
Time for topKMatches("nenk", 4) - 1.61898E-6
Time for topKMatches("nenk", 8) - 1.06413E-6
Time for topKMatches("nenk", 16) - 1.60104E-6
Time for topKMatches("nenk", 32) - 1.65198E-6
Time for topKMatches("nenk", 64) - 1.58638E-6
Time for topKMatches("nenk", 128) - 1.85812E-6
Time for topKMatches("nenk", 256) - 2.9123E-7
Time for topKMatches("n", 1) - 3.03349E-6
Time for topKMatches("n", 2) - 4.97538E-6
Time for topKMatches("n", 4) - 8.50964E-6
Time for topKMatches("n", 8) - 1.63187E-5
Time for topKMatches("n", 16) - 3.486143E-5
Time for topKMatches("n", 32) - 6.675775E-5
Time for topKMatches("n", 64) - 1.2947413E-4
Time for topKMatches("n", 128) - 1.8102084E-4
Time for topKMatches("n", 256) - 3.0126265E-4
Time for topKMatches("ne", 1) - 9.6577E-7
Time for topKMatches("ne", 2) - 1.05949E-6
Time for topKMatches("ne", 4) - 1.82537E-6
Time for topKMatches("ne", 8) - 3.85086E-6
Time for topKMatches("ne", 16) - 7.2076E-6
Time for topKMatches("ne", 32) - 1.229911E-5
Time for topKMatches("ne", 64) - 1.818319E-5
Time for topKMatches("ne", 128) - 3.040511E-5
Time for topKMatches("ne", 256) - 5.75787E-5
Time for topKMatches("notarealword", 1) - 9.217E-7
Time for topKMatches("notarealword", 2) - 1.552E-7
Time for topKMatches("notarealword", 4) - 2.5626E-7
Time for topKMatches("notarealword", 8) - 1.5413E-7
Time for topKMatches("notarealword", 16) - 2.2476E-7
Time for topKMatches("notarealword", 32) - 1.5331E-7
Time for topKMatches("notarealword", 64) - 2.6304E-7
Time for topKMatches("notarealword", 128) - 1.6641E-7
Time for topKMatches("notarealword", 256) - 1.5554E-7

Using fourletterwordshalf.txt
BruteAutocomplete
Found 228488 words
Time to initialize - 0.033927246
Time for topMatch("") - 5.94264946E-4
Time for topMatch("aenk") - 0.002274728814
Time for topMatch("a") - 5.59578229E-4
Time for topMatch("ae") - 5.71715822E-4
Time for topMatch("notarealword") - 0.00175254584
Time for topKMatches("", 1) - 0.00246357609
Time for topKMatches("", 2) - 0.0022055974
Time for topKMatches("", 4) - 0.00246322745
Time for topKMatches("", 8) - 0.00254120894

Time for topKMatches("", 16) - 0.00229317396
Time for topKMatches("", 32) - 0.00232015974
Time for topKMatches("", 64) - 0.0023869788
Time for topKMatches("", 128) - 0.00236809317
Time for topKMatches("", 256) - 0.00250593311
Time for topKMatches("aenk", 1) - 0.00240576099
Time for topKMatches("aenk", 2) - 0.00229201446
Time for topKMatches("aenk", 4) - 0.00219376338
Time for topKMatches("aenk", 8) - 0.00229343913
Time for topKMatches("aenk", 16) - 0.00221701316
Time for topKMatches("aenk", 32) - 0.00230144003
Time for topKMatches("aenk", 64) - 0.00236209915
Time for topKMatches("aenk", 128) - 0.00231480233
Time for topKMatches("aenk", 256) - 0.00236280492
Time for topKMatches("a", 1) - 0.00226572537
Time for topKMatches("a", 2) - 0.00225259922
Time for topKMatches("a", 4) - 0.00229653985
Time for topKMatches("a", 8) - 0.00231456516
Time for topKMatches("a", 16) - 0.00230711616
Time for topKMatches("a", 32) - 0.00244002962
Time for topKMatches("a", 64) - 0.00231349241
Time for topKMatches("a", 128) - 0.00242362466
Time for topKMatches("a", 256) - 0.00244691179
Time for topKMatches("ae", 1) - 0.00222351182
Time for topKMatches("ae", 2) - 0.00230751121
Time for topKMatches("ae", 4) - 0.00205470791
Time for topKMatches("ae", 8) - 0.00106135771
Time for topKMatches("ae", 16) - 0.00111478218
Time for topKMatches("ae", 32) - 0.00106958652
Time for topKMatches("ae", 64) - 0.00109777756
Time for topKMatches("ae", 128) - 0.0010970924
Time for topKMatches("ae", 256) - 0.0011892655
Time for topKMatches("notarealword", 1) - 8.3964376E-4
Time for topKMatches("notarealword", 2) - 8.5785997E-4
Time for topKMatches("notarealword", 4) - 9.1992775E-4
Time for topKMatches("notarealword", 8) - 8.5116277E-4
Time for topKMatches("notarealword", 16) - 8.7941175E-4
Time for topKMatches("notarealword", 32) - 9.1191795E-4
Time for topKMatches("notarealword", 64) - 9.1805506E-4
Time for topKMatches("notarealword", 128) - 8.6484614E-4
Time for topKMatches("notarealword", 256) - 9.1804318E-4

BinarySearchAutocomplete

Found 228488 words
Time to initialize - 0.018118753
Time for topMatch("") - 1.93685233E-4
Time for topMatch("aenk") - 1.262375E-6
Time for topMatch("a") - 1.462523E-5
Time for topMatch("ae") - 2.256471E-6
Time for topMatch("notarealword") - 1.227668E-6
Time for topKMatches("", 1) - 0.00252638915
Time for topKMatches("", 2) - 0.00294522348
Time for topKMatches("", 4) - 0.00508374688
Time for topKMatches("", 8) - 0.00602682892
Time for topKMatches("", 16) - 0.00711736396
Time for topKMatches("", 32) - 0.00832230167
Time for topKMatches("", 64) - 0.00964181898
Time for topKMatches("", 128) - 0.01118838903
Time for topKMatches("", 256) - 0.0133964975
Time for topKMatches("aenk", 1) - 2.28278E-6
Time for topKMatches("aenk", 2) - 9.883E-7

Time for topKMatches("aenk", 4) - 9.7319E-7
Time for topKMatches("aenk", 8) - 9.2095E-7
Time for topKMatches("aenk", 16) - 9.8241E-7
Time for topKMatches("aenk", 32) - 8.358E-7
Time for topKMatches("aenk", 64) - 9.0779E-7
Time for topKMatches("aenk", 128) - 1.00517E-6
Time for topKMatches("aenk", 256) - 1.26332E-6
Time for topKMatches("a", 1) - 1.9480776E-4
Time for topKMatches("a", 2) - 2.4926395E-4
Time for topKMatches("a", 4) - 3.3044545E-4
Time for topKMatches("a", 8) - 4.375304E-4
Time for topKMatches("a", 16) - 5.7439521E-4
Time for topKMatches("a", 32) - 6.955714E-4
Time for topKMatches("a", 64) - 7.9679851E-4
Time for topKMatches("a", 128) - 9.3981874E-4
Time for topKMatches("a", 256) - 0.00112091299
Time for topKMatches("ae", 1) - 9.25313E-6
Time for topKMatches("ae", 2) - 1.029526E-5
Time for topKMatches("ae", 4) - 1.65762E-5
Time for topKMatches("ae", 8) - 2.161207E-5
Time for topKMatches("ae", 16) - 2.185133E-5
Time for topKMatches("ae", 32) - 1.3525027E-4
Time for topKMatches("ae", 64) - 2.850959E-5
Time for topKMatches("ae", 128) - 3.178494E-5
Time for topKMatches("ae", 256) - 3.998929E-5
Time for topKMatches("notarealword", 1) - 5.5693E-7
Time for topKMatches("notarealword", 2) - 4.1136E-7
Time for topKMatches("notarealword", 4) - 5.7072E-7
Time for topKMatches("notarealword", 8) - 3.0751E-7
Time for topKMatches("notarealword", 16) - 4.5573E-7
Time for topKMatches("notarealword", 32) - 2.9447E-7
Time for topKMatches("notarealword", 64) - 2.4406E-7
Time for topKMatches("notarealword", 128) - 2.5315E-7
Time for topKMatches("notarealword", 256) - 1.8739E-7

TrieAutocomplete

Found 228488 words
Time to initialize - 0.068766415
Created 237628 nodes
Time for topMatch("") - 4.002144E-6
Time for topMatch("aenk") - 2.51842E-7
Time for topMatch("a") - 1.064524E-6
Time for topMatch("ae") - 2.105929E-6
Time for topMatch("notarealword") - 1.18626E-7
Time for topKMatches("", 1) - 4.322019E-5
Time for topKMatches("", 2) - 4.065537E-5
Time for topKMatches("", 4) - 5.596916E-5
Time for topKMatches("", 8) - 3.350776E-5
Time for topKMatches("", 16) - 5.861338E-5
Time for topKMatches("", 32) - 1.0832626E-4
Time for topKMatches("", 64) - 1.8948587E-4
Time for topKMatches("", 128) - 3.8685277E-4
Time for topKMatches("", 256) - 7.7659208E-4
Time for topKMatches("aenk", 1) - 1.8251E-6
Time for topKMatches("aenk", 2) - 1.1474E-6
Time for topKMatches("aenk", 4) - 2.41959E-6
Time for topKMatches("aenk", 8) - 1.37005E-6
Time for topKMatches("aenk", 16) - 1.93987E-6
Time for topKMatches("aenk", 32) - 2.15104E-6
Time for topKMatches("aenk", 64) - 4.1895E-7
Time for topKMatches("aenk", 128) - 4.6243E-7

Time for topKMatches("aenk", 256) - 5.0766E-7
Time for topKMatches("a", 1) - 2.82384E-6
Time for topKMatches("a", 2) - 4.80967E-6
Time for topKMatches("a", 4) - 7.6565E-6
Time for topKMatches("a", 8) - 1.494857E-5
Time for topKMatches("a", 16) - 3.687335E-5
Time for topKMatches("a", 32) - 6.34804E-5
Time for topKMatches("a", 64) - 1.3287502E-4
Time for topKMatches("a", 128) - 2.0361715E-4
Time for topKMatches("a", 256) - 2.8149123E-4
Time for topKMatches("ae", 1) - 9.1688E-7
Time for topKMatches("ae", 2) - 1.35618E-6
Time for topKMatches("ae", 4) - 2.0756E-6
Time for topKMatches("ae", 8) - 4.01388E-6
Time for topKMatches("ae", 16) - 6.50306E-6
Time for topKMatches("ae", 32) - 1.367249E-5
Time for topKMatches("ae", 64) - 1.969847E-5
Time for topKMatches("ae", 128) - 2.810853E-5
Time for topKMatches("ae", 256) - 6.019537E-5
Time for topKMatches("notarealword", 1) - 1.19406E-6
Time for topKMatches("notarealword", 2) - 1.5574E-7
Time for topKMatches("notarealword", 4) - 1.4661E-7
Time for topKMatches("notarealword", 8) - 1.4956E-7
Time for topKMatches("notarealword", 16) - 2.129E-7
Time for topKMatches("notarealword", 32) - 1.512E-7
Time for topKMatches("notarealword", 64) - 1.4157E-7
Time for topKMatches("notarealword", 128) - 1.4462E-7
Time for topKMatches("notarealword", 256) - 1.54E-7