Ps6: Markov Model

Assignment Description:

This assignment asked us to make a program that outputs similar but not the same text as the input file using a Markov Model. This model takes the input text and breaks it into sub strings called kgrams these kgrams compute the next character based on the probability of that character in the input file. This creates text similar to the input file in vocabulary but is a new creation. The larger the kgram the more accurate the predictions.

Key Concepts and Algorithms:

The key concept in this assignment is a Markov Model. This is a model in probabilistic forecasting that is used to predict future states based on the current state and not previous states. Markov models are applied in many predictive applications but for this assignment we used it to create a predictive text algorithm.

What I learned in this assignment:

During this assignment I learned how to use maps in the standard template library and how to use them to create a table for holding data. This project required this maps instead of other more familiar data structures like vectors or queues. This assignment also taught me how devices like smart phone create predictive text using Markov chains, in order to figure out what you are trying to type.

Ps6: Screen shot

According types: //pess //pess

Ps6 Source Code: Makefile

```
1 CFLAGS = -03 -Wall -Werror -std=c++11 -ansi -pedantic
2 DEPS = -lboost_unit_test_framework
4 all:
           TextGenerator test.o
6 TextGenerator: TextGenerator.o MModel.o
7
       g++ TextGenerator.o MModel.o -o TextGenerator
8
9 TextGenerator.o: TextGenerator.cpp MModel.h
10
       g++ -c TextGenerator.cpp MModel.h $(CFLAGS)
11
12 MModel.o: MModel.cpp MModel.h
       g++ -c MModel.cpp MModel.h $(CFLAGS)
13
14
15 test.o:
       g++ test.cpp MModel.cpp MModel.h -o test.o $(CFLAGS) $(DEPS)
16
17
18 clean:
19
       rm *.o
20
       rm *.gch
21
       rm TextGenerator
```

Ps6 Source Code: TextGenerator.cpp

```
1 // Copyright 2020 Karl Marx
2 #include <string>
3 #include "MModel.h"
   int main(int argc, char* argv[]) {
6
     if (argc != 3) {
       std::cout << "Invalid Number of Arguments. Must == 3";</pre>
7
8
       return -1;
9
     }
10
11
     int k = atoi(argv[1]);
     int L = atoi(argv[2]);
12
13
14
     std::string input;
15
     std::string temp;
16
17
     while (std::cin >> temp) {
       input += " " + temp;
18
```

```
19
        temp = "";
20
21
22
      std::string output;
23
24
      MModel model(input, k);
25
26
      output += "" + model.generate(input.substr(0, k), L);
27
      std::cout << output << std::endl;</pre>
28
29
      return 0;
30 }
```

Ps6 Source Code: test.cpp

```
// Copyright 2020 Karl Marx
1
3 #define BOOST_TEST_DYN_LINK
  #define BOOST_TEST_MODULE Main
4
  #include <boost/test/unit_test.hpp>
  #include "MModel.h"
   BOOST_AUTO_TEST_CASE(Exception) {
8
9
       MModel a("hello", 2);
       BOOST_REQUIRE_THROW(a.freq("h"), std::runtime_error);
10
       BOOST_REQUIRE_THROW(a.freq("hhh"), std::runtime_error);
11
       BOOST_REQUIRE_THROW(a.freq("hhh", 'e'), std::runtime_error);
12
       BOOST_REQUIRE_THROW(a.freq("h", 'e'), std::runtime_error);
13
       BOOST_REQUIRE_THROW(a.generate("hhh", 10), std::runtime_error);
14
15
```

Ps6 Source Code: MModel.h

```
// Copyright 2020 Karl Marx
#ifndef MMODEL_H //NOLINT
#define MMODEL_H
#include <iostream>
#include <algorithm>
#include <string>
#include <map>
#include <stdexcept>
#include <vector>
```

```
10 #include <utility>
11
12 class MModel {
13
   public:
14
        MModel(std::string text, int k);
15
        int kOrder();
16
        int freq(std::string kgram);
        int freq(std::string kgram, char c);
17
18
        char kRand(std::string kgram);
19
        std::string generate(std::string kgram, int L);
20
        friend std::ostream& operator<< (std::ostream &os, MModel &model) {</pre>
21
         os << "Order =" << model.order << std::endl;</pre>
22
         os << "Alphabet = "<< model.alphabet << std::endl;</pre>
23
24
         std::map<std::string, int>::iterator temp;
25
26
          for (temp = model.kgrams.begin(); temp != model.kgrams.end(); temp++) {
27
          os << temp->first << " " << temp->second << std::endl;
28
          }
29
30
     return os;
31
        }
32
33
    private:
34
        std::map <std::string, int> kgrams;
35
        int order;
36
        std::string alphabet;
37
        std::string original;
38 };
39 #endif //NOLINT
```

Ps6 Source Code: MModel.cpp

```
1 // Copyright 2020 Karl Marx
2 #include <string>
3 #include "MModel.h" //NOLINT
   MModel::MModel(std::string text, int k) {
5
6
     srand(time(NULL));
7
     order = k;
8
     original = text;
9
10
     for (unsigned i = 0; i < text.size(); i++)</pre>
11
       if (std::string::npos == alphabet.find(text.at(i)))
```

```
12
          alphabet += text.at(i);
13
      for (unsigned i = 0; i < text.size(); i++) {</pre>
14
15
        std::string temp;
16
        std::string temp2;
17
18
        for (unsigned j = i; j < i + k; j++)</pre>
        temp = (j \ge text.size()) ? temp += text.at(j - text.size()) : temp += text.
19
20
21
        kgrams[temp] = (kgrams.end() == kgrams.find(temp)) ? 1 : kgrams[temp] += 1;
22
23
        for (unsigned j = 0; j < alphabet.size(); j++)</pre>
24
          if (kgrams.end() == kgrams.find(temp + alphabet[j]))
25
            kgrams[temp + alphabet.at(j)] = 0;
26
        for (unsigned j = i; j < i + k + 1; j++)</pre>
27
28
          temp2 = (j \ge text.size()) ? temp2 += (text.at(j - text.size())) : temp2 += (text.at(j - text.size())) :
29
30
        kgrams[temp2] += 1;
31
     }
32
   }
33
34
   int MModel::kOrder() {
35
          return order;
36
37
38
   int MModel::freq(std::string kgram) {
39
     if ((signed)kgram.length() != order) {
40
          throw
41
             std::runtime_error("invalid kgram in freq(std::string)");
42
43
     int return_value = (order == 0) ? (original.size()) : (kgrams[kgram]);
44
      return return_value;
45
   }
46
47
   int MModel::freq(std::string kgram, char c) {
48
      if ((signed)kgram.length() != order) {
49
          throw
50
            std::runtime_error("invalid kgram in freq(std::string, char)");
51
52
      if (order == 0) {
53
        int counter = 0;
54
        for (int i = 0; i < (signed)original.size(); i++) {</pre>
          if (original.at(i) == c) {
55
            counter++;
56
57
          }
```

```
58
59
       return counter;
60
     } else {
61
        return kgrams[kgram + c];
62
63
     return 0;
64 }
65
66
   char MModel::kRand(std::string kgram) {
   if (kgram.length() != (unsigned)order) {
68
       throw std::runtime_error("invalid kgram in kRand");
69
70
  if (kgrams.end() == kgrams.find(kgram)) {
71
       throw std::runtime_error("invalid krgram in kRand");
72 }
73
     double _freq = 0;
74
     double val = 0;
75
     for (unsigned int i = 0; i < alphabet.length(); i++) {</pre>
        _freq = (double)(freq(kgram, alphabet.at(i))) / freq(kgram);//NOLINT
76
77
        if ((double)(rand() % freq(kgram)) / freq(kgram) < _freq + val && _freq != 0</pre>
78
          return alphabet.at(i);
79
       }
80
       val += _freq;
81
82
     return '#';
83
84
85
   std::string MModel::generate(std::string kgram, int L) {
     if (kgram.size() < static_cast<unsigned>(order))
86
87
          throw std::runtime_error("Invalid Kgram");
88
89 L = L - order + 1;
   while(kgram.size() < (unsigned)L)</pre>
91
       kgram += kRand(kgram.substr(kgram.size()-order, order));
92
   return kgram;
93
   }
```