```
    1: // "Copyright 2020 <Greg Kaplowitz>"
    2:
    3: #include "markov.h"
    4: #include <vector>
    5: #include <utility>
    6:
    7: MarkovModel::MarkovModel(std::string text, int k) {
    8:   order = k;  // set order
    9:   // make text circular
   10:   for (int i = 0; i < order-1; i++) {
   11:   text.push_back(text[i]);
   12:   }
   13:   std::cout << text.length() << std::endl;
   14:   symtab.push_back(text.at(0));
   15:   for (unsigned int i = 0; i < text.length(); i++) {
   16:   bool present = false;
   17:
   18:   for (unsigned int j=0; j < symtab.length(); j++) {  // checks if the curre
nt
   19:   //  char of the text is already in or symbol table
   20:   // std::cout<<symtab.length()<<std::endl;
   21:   if (symtab.at(j) == text.at(i)) {  // if yes
   22:   present = true;  // then say so
   23:   }
   24:   }
   25:   if (!present) {  // if it isnt present then add it
   26:   symtab.push_back(text.at(i));
   27:   }
   28:   }
   29:   // we should have a filled out sybol table here
   30:   std::cout << symtab << std::endl;
   31:   for (unsigned int i=0; i < text.length()-1; i++) {  // fill out the kgram
and
   32:   // freq sections of markov chart
   33:   std::string kgram_string = text.substr(i, order);  // gets the kgram
   34:   int freq = countFreq(kgram_string, text);  // gets the freq
   35:   kgram.insert(std::pair<std::string, int>(kgram_string, freq));
   36:   // adds the kgram key with its freq data
   37:   }
   38:   // now we should have a filled out map representing the kgram and freqency
   39:   // for each
   40:   /*for (std::map<std::string, int>::iterator it = kgram.begin();it !=kgram.
end();it++){
   41:   std::cout<< it->first <<"  "<<it->second<<std::endl;// test
   42:   }*/ //this prints out the map
   43:   txt = text;
   44:   }
   45:
   46:
   47: int MarkovModel::k_order() {
   48:         return order;
   49: }
   50:
   51: int MarkovModel::freq(std::string _kgram) {
   52:   if (_kgram.length() != (unsigned)order){
   53:   throw
   54:   std::invalid_argument("kgram not of length k 1");
   55:   }
   56:   if (kgram.find(_kgram) == kgram.end()) {
   57:   return 0;
   58:   }
```

```
   59:   std::map<std::string, int > ::iterator it;
   60:   it = kgram.find(_kgram);
   61:   return it->second;
   62: }
   63:
   64: int MarkovModel::freq(std::string _kgram, char c) {
   65:   if (_kgram.length() != (unsigned) order) {
   66:   throw
   67:   std::invalid_argument("kgram not of length k 2");
   68:   }
   69:   if (kgram.find(_kgram) == kgram.end()) {
   70:   return 0;
   71:   }
   72:   return count_following_freq(_kgram, txt, c);
   73: }
   74:
   75: char MarkovModel::kRand(std::string _kgram) {
   76:   if (_kgram.length() != (unsigned) order) {
   77:   throw
   78:   std::invalid_argument("kgram not of length k 3");
   79:   }
   80:   char result;
   81:   int freq = this->freq(_kgram);
   82:   int RNG = rand() % freq + 1;
   83:   //std::cout<<RNG;
   84:   bool is_zero = false;
   85: //      The way I implement the weighted random sallection here starts a ran
dom
   86: //      integer between
   87: //      1 and the frequency of the kgram then subtract the character freqenc
y
   88: //      so if there are 3 symbols in the symtab a/c/g with 5 kgram frequency

   89: //      we get a random int of 3 and the char freq's are a=3 c=0 g=2
   90: //      we subtract 3 from the freq and check to see if the running count is
 <=0
   91: //      we are at 0 so a is returned
   92:   for (unsigned int i =0; i<symtab.length(); i++) {  // iterate through the
   93:   // symtab
   94:   int diff = this->freq(_kgram, symtab[i]);
   95:   RNG -= diff;
   96:   if (RNG<=0 && !is_zero) {
   97:   result = symtab[i];
   98:   is_zero = true;
   99:   }
  100:   }
  101:   return result;
  102:  }
  103:
  104:   std::string MarkovModel::generate(std::string _kgram, int L) {
  105:   if (_kgram.length() != (unsigned) order){
  106:   throw
  107:   std::invalid_argument("kgram not of length k 4");
  108:   }
  109:   std::string txt = "";
  110:   txt += "" + _kgram;
  111:
  112:   for (int i=0; i<L-order;i++) {
  113:   char x = this->kRand(txt.substr(i, order));
  114:   // std::cout<<x<<std::endl;
  115:   txt.push_back(x);
```

```
  116:   }
  117:   return txt;
  118:   }
  119:
  120:   std::ostream& operator<< (std::ostream &out, MarkovModel &markov) {
  121:   out<<"First the map: "<<std::endl;
  122:   for (std::map<std::string, int>::iterator it = markov.kgram.begin();it !=m
arkov.kgram.end();it++) {// no lint
  123:   out<< it->first <<"  "<<it->second<<std::endl;// print the map;
  124:   }
  125:
  126:   out << "The symtab: "<< std::endl;
  127:   out << markov.symtab << std::endl;
  128:
  129:   out << "The order: "<<std::endl;
  130:   out << markov.order << std::endl;
  131:
  132:   return out;
  133: }
  134:
  135: int countFreq(std::string &pat, std::string &txt) {
  136:   int M = pat.length();
  137:   int N = txt.length();
  138:   int res = 0;
  139:
  140:   /* A loop to slide pat[ ] one by one */
  141:   for (int i = 0; i <= N - M; i++) {
  142:   /* For current index i, check for
  143:   pattern match */
  144:   int j;
  145:   for (j = 0; j < M; j++)
  146:   if (txt[i+j] != pat[j])
  147:   break;
  148:
  149:   // if pat[0...M-1] = txt[i, i+1, ...i+M-1]
  150:   if (j == M) {
  151:   res++;
  152:   j = 0;
  153:   }
  154:   }
  155:   return res;
  156: }
  157:
  158: int count_following_freq(std::string &pat, std::string &txt, char c) {//NOLI
NT
  159:   int M = pat.length();
  160:   int N = txt.length();
  161:   int res = 0;
  162:   /* A loop to slide pat[ ] one by one */
  163:   for (int i = 0; i <= N - M; i++) {
  164:   /* For current index i, check for
  165:   pattern match */
  166:   int j;
  167:   for (j = 0; j < M; j++)
  168:   if (txt[i+j] != pat[j])
  169:   break;
  170:   // if pat[0...M-1] = txt[i, i+1, ...i+M-1]
  171:   if (j == M) {
  172:   if (txt[i + j] == c) {
  173:   res++;
  174:   }
```

```
  175:   j = 0;
  176:   }
  177:   }
  178:   return res;
  179: }
```