

TagCloudGenerator.java

```
1 import java.util.Comparator;
14
15 /**
16  * Program to take a file of text, count each instance of every word, and
17  * generate a tag cloud with each word and corresponding size
18  *
19  * @author Robert Frenken
20  * @author Bennett Palmer
21  *
22  * NOTE: This code handles all words to lower case
23  */
24 public final class TagCloudGenerator {
25
26     /**
27      * Default constructor--private to prevent instantiation.
28      */
29     private TagCloudGenerator() {
30         // no code needed here
31     }
32
33     /**
34      * Compare {@code String}s in alphabetical order.
35      */
36     private static class StringLT
37         implements Comparator<Map.Pair<String, Integer>> {
38         @Override
39         public int compare(Map.Pair<String, Integer> o1,
40             Map.Pair<String, Integer> o2) {
41             // use to lower case to group both capitalized and non capitalized
42             return o1.key().toLowerCase().compareTo(o2.key().toLowerCase());
43         }
44     }
45
46     /**
47      * Compare {@code Integer}i in decreasing order.
48      */
49     private static class IntegerLT
50         implements Comparator<Map.Pair<String, Integer>> {
51         @Override
52         public int compare(Map.Pair<String, Integer> o1,
53             Map.Pair<String, Integer> o2) {
54
55             return o2.value().compareTo(o1.value());
56         }
57     }
58
59     /**
60      * Generates the set of characters in the given {@code String} into the
61      * given {@code Set}.
62      *
63      * @param str
64      *     the given {@code String}
65      * @param strSet
66      *     the {@code Set} to be replaced
67      * @replaces strSet
68      * @ensures strSet = entries(str)
69      */
69 }
```

TagCloudGenerator.java

```

70 private static void generateElements(String str, Set<Character> strSet) {
71     assert str != null : "Violations of: str is not null";
72     assert strSet != null : "Violation of: strSet is not null";
73
74     for (int i = 0; i < str.length(); i++) {
75         char c = str.charAt(i);
76         if (!strSet.contains(c)) {
77             strSet.add(c);
78         }
79     }
80 }
81
82 /**
83  * Returns the first "word" (maximal length string of characters not in
84  * {@code separators}) or "separator string" (maximal length string of
85  * characters in {@code separators}) in the given {@code text} starting at
86  * the given {@code position}.
87  *
88  * @param text
89  *         the {@code String} from which to get the word or separator
90  *         string
91  * @param position
92  *         the starting index
93  * @param separators
94  *         the {@code Set} of separator characters
95  * @return the first word or separator string found in {@code text} starting
96  *         at index {@code position}
97  * @requires 0 <= position < |text|
98  * @ensures <pre>
99  * nextWordOrSeparator =
100  *   text[position, position + |nextWordOrSeparator|) and
101  *   if entries(text[position, position + 1)) intersection separators = {}
102  * then
103  *   entries(nextWordOrSeparator) intersection separators = {} and
104  *   (position + |nextWordOrSeparator| = |text| or
105  *   entries(text[position, position + |nextWordOrSeparator| + 1))
106  *   intersection separators /= {})
107  * else
108  *   entries(nextWordOrSeparator) is subset of separators and
109  *   (position + |nextWordOrSeparator| = |text| or
110  *   entries(text[position, position + |nextWordOrSeparator| + 1))
111  *   is not subset of separators)
112  * </pre>
113  */
114 private static String nextWordOrSeparator(String text, int position,
115     Set<Character> separators) {
116     assert text != null : "Violation of: text is not null";
117     assert separators != null : "Violation of: separators is not null";
118     assert 0 <= position : "Violation of: 0 <= position";
119     assert position < text.length() : "Violation of: position < |text|";
120
121     int endPosition = position;
122     if (!separators.contains(text.charAt(position))) {
123         // find the length of the word
124         while (endPosition < text.length()
125             && !separators.contains(text.charAt(endPosition))) {
126             endPosition++;

```

```

127     }
128     } else {
129         // find the length of the separator
130         while (endPosition < text.length()
131             && separators.contains(text.charAt(endPosition))) {
132             endPosition++;
133         }
134     }
135
136     return text.substring(position, endPosition);
137 }
138
139 /**
140  * Take file given and build a map, with each word to lower case as the map
141  * key and the occurrence of each word as the map value.
142  *
143  * @param in
144  *         the SimpleReader file
145  * @ensures all words from inFile will be in map, with count of each word
146  * @return Map<String, Integer> of words of the file and their counts
147  */
148 private static Map<String, Integer> readFileToMap(SimpleReader inFile) {
149     assert inFile != null : "Violation of: inFile is not null";
150
151     Map<String, Integer> map = new Map1L<>();
152
153     /*
154      * Define separator characters for test
155      */
156     final String separatorStr = " \\t, .-!?!_@#$$%&*[]()";
157     Set<Character> separatorSet = new Set1L<>();
158     generateElements(separatorStr, separatorSet);
159
160     // Read file, and compile all the words into map
161     while (!inFile.atEOS()) {
162         String oneLine = inFile.nextLine();
163         int i = 0;
164         while (i < oneLine.length()) {
165             String word = nextWordOrSeparator(oneLine, i, separatorSet)
166                 .toLowerCase();
167             boolean isWord = true;
168             for (int j = 0; j < word.length(); j++) {
169                 char c = word.charAt(j);
170                 if (separatorSet.contains(c)) {
171                     isWord = false;
172                 }
173             }
174             if (isWord) {
175                 // add word to map or add one to count
176                 if (map.containsKey(word)) {
177                     int count = map.value(word);
178                     map.replaceValue(word, count + 1);
179                 } else {
180                     map.add(word, 1);
181                 }
182             }
183             i += word.length();

```

TagCloudGenerator.java

```

184     }
185 }
186
187     return map;
188
189 }
190
191 /**
192  * Takes map and integer value, and first sorts the map with sortingMachine
193  * by occurrence of each word in decreasing order. Then makes a second
194  * sortingMachine, and sorts alphabetically. numberDisplay determines the
195  * number of words to be put in second sortingMachine.
196  *
197  * @param map
198  *     map of all of the words and their counts
199  * @param numberDisplay
200  *     the amount of words that will be in the first sortingMachine
201  * @requires map is not null
202  * @ensures all Map.Pairs is sorted in decreasing order by their values
203  * @return SortingMachine<Map.Pair<String, Integer>> of words and their
204  *     counts in alphabetical order
205  *
206  */
207 public static SortingMachine<Map.Pair<String, Integer>> mapToSortingMachineAlphabet(
208     Map<String, Integer> map, Integer numberDisplay) {
209     assert map != null : "Violation of: words is not null";
210
211     Comparator<Map.Pair<String, Integer>> sortByCount = new IntegerLT();
212     Comparator<Map.Pair<String, Integer>> sortByAlpha = new StringLT();
213     SortingMachine<Map.Pair<String, Integer>> sortCount = new SortingMachine1L<>(
214         sortByCount);
215     SortingMachine<Map.Pair<String, Integer>> sortAlpha = new SortingMachine1L<>(
216         sortByAlpha);
217
218     // build the sorting machine with the map
219     while (map.size() > 0) {
220         Map.Pair<String, Integer> temp = map.removeAny();
221         sortCount.add(temp);
222     }
223
224     sortCount.changeToExtractionMode();
225
226     // build sorting machine in alphabetical order
227     for (int i = 0; i < numberDisplay; i++) {
228         sortAlpha.add(sortCount.removeFirst());
229     }
230
231     sortAlpha.changeToExtractionMode();
232
233     return sortAlpha;
234 }
235
236 /**
237  * Outputs the tag cloud given the words in SortingMachie
238  *
239  * @param sortAlpha
240  */

```

TagCloudGenerator.java

```

241     *           the map sorted alphabetically
242     * @param out
243     *           the output stream
244     * @param title
245     *           the string of the file name
246     * @updates out.content
247     * @requires out.is_open
248     * @ensures out.content = #out.content * [the HTML tags]
249     */
250     private static void outputhtml(
251         SortingMachine<Map.Pair<String, Integer>> sortAlpha,
252         SimpleWriter out, String title) {
253         assert out.isOpen() : "Violation of: out.is_open";
254         out.println("<html>");
255         out.println("<head>");
256         out.println("<title>" + title
257             + "</title><link href=\"http://web.cse.ohio-state.edu/software/2231/web-
sw2/assignments/projects/tag-cloud-generator/data/tagcloud.css\" rel=\"stylesheet\" type=
\"text/css\">")
258             + "</head><body><h2>" + "Top " + sortAlpha.size() + " words in "
259             + title + "</h2><hr>");
260
261         out.println("<div class=\"cdiv\">");
262         out.println("<p class =\" + '\"' + \"cbox\" + '\"' + >");
263         int maxCount = 0;
264         for (Map.Pair<String, Integer> p : sortAlpha) {
265             if (p.value() > maxCount) {
266                 maxCount = p.value();
267             }
268         }
269
270         while (sortAlpha.size() > 0) {
271             Map.Pair<String, Integer> temp = sortAlpha.removeFirst();
272
273             int fontSize = 37 * temp.value() / maxCount + 11;
274             out.println("<span style=\"cursor:default\" class=\"f\" + fontSize
275                 + \"\" title=\"count:\" + temp.value() + \"\">\" + temp.key()
276                 + "</span>");
277         }
278         out.println("</p>");
279         out.println("</div>");
280         out.println("</body>");
281         out.println("</html>");
282     }
283
284     /**
285     * Main method.
286     *
287     * @param args
288     *           the command line arguments; unused here
289     */
290     public static void main(String[] args) {
291         SimpleReader in = new SimpleReader1L();
292         SimpleWriter out = new SimpleWriter1L();
293
294         out.print("Enter file that will be used to obtain the words: ");
295         String textFile = in.nextLine();

```

TagCloudGenerator.java

```
296     SimpleReader inFile = new SimpleReader1L(textFile);
297
298     out.print("Enter the name of a of the html page to write to: ");
299     String htmlpage = in.nextLine();
300     SimpleWriter outputhtml = new SimpleWriter1L(htmlpage);
301
302     out.print(
303         "Enter the number of words that will be displayed (integer): ");
304     String numberToDisplay = in.nextLine();
305     // checks if number is positive integer
306     while (!FormatChecker.canParseInt(numberToDisplay)
307         || !(Integer.parseInt(numberToDisplay) > 0)) {
308         out.print(
309             "ERROR: Not Positive Integer \nPlease enter the number of words that will
be displayed (integer): ");
310         numberToDisplay = in.nextLine();
311     }
312     int numberDisplay = Integer.parseInt(numberToDisplay);
313
314     Map<String, Integer> map = readFileToMap(inFile);
315
316     SortingMachine<Map.Pair<String, Integer>> sortAlpha = mapToSortingMachineAlphabet(
317         map, numberDisplay);
318
319     outputhtml(sortAlpha, outputhtml, textFile);
320
321     inFile.close();
322     out.close();
323     in.close();
324 }
325
326 }
327
```