Q1:

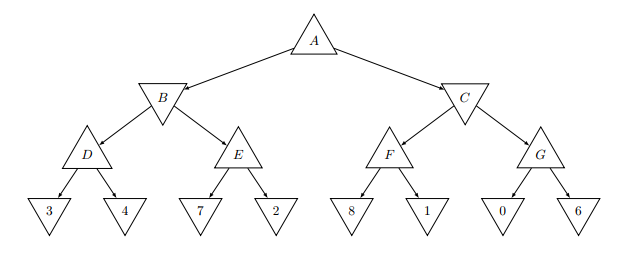
1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| A | B | C | D | E | F | G |
| 4 | 4 | 6 | 4 | 7 | 8 | 6 |

2.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G |
| Minimax | 4 | 4 | 6 | 4 | 7 | 8 | 6 |
| α | 4 | 4 | 4 | 4 | 7 | 8 | 6 |
| β | ∞ | 4 | 6 | ∞ | 4 | ∞ | 8 |

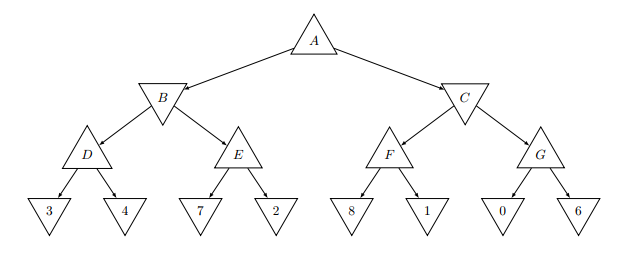
3.



Pruned elements are highlighted in red box

4. It saves time as we prune part of the tree

5.

6. 

0.75

0.25

0.25

0.75

0.25

0.25

0.25

0.25

0.25

0.75

0.75

0.75

0.75

0.75

7.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| A | B | C | D | E | F | G |
| 5.67 | 5.25 | 5.81 | 3.75 | 5.75 | 6.25 | 4.5 |

Q2:

1. We can use java’s scanner library to read from the text files and separate the strings and count the number of words.
2. 307556 occurrences and 15192 unique occurrences

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Word | Count | Word | Count | Word | Count | Word | Count | Word | Count |
| the | 8012 | in | 2911 | be | 2032 | are | 1185 | as | 1088 |
| to | 5069 | a | 2779 | AI | 1926 | it | 1178 | with | 890 |
| of | 4518 | is | 2513 | will | 1543 | on | 1119 | The | 850 |
| and | 3879 | that | 2493 | for | 1209 | not | 1111 | have | 811 |

1. Most used

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| children?s | Employment/ | q | (p.22). | Power) | ?deep | (FDA) | $400. | Yorka??s | “let’s |
| (AI | Trucks.[1] | i | ICD-9 | ‘drivers’ | SET?, | al, | 540. | 35,000 | 5.1% |

1. Least used
2. (Items after 1000th item is trimmed to show the curve of the graph)
4. This shows that only about a few hundred unique words are used in these files, all others are unique words.
5. One issue is that when punctuation marks are used, the system cannot differentia the words from its original form thus count it are two different unique words  
   Another issue is that frequently used words are too few. Only about a hundred or so has a high rate of using.