

HW5

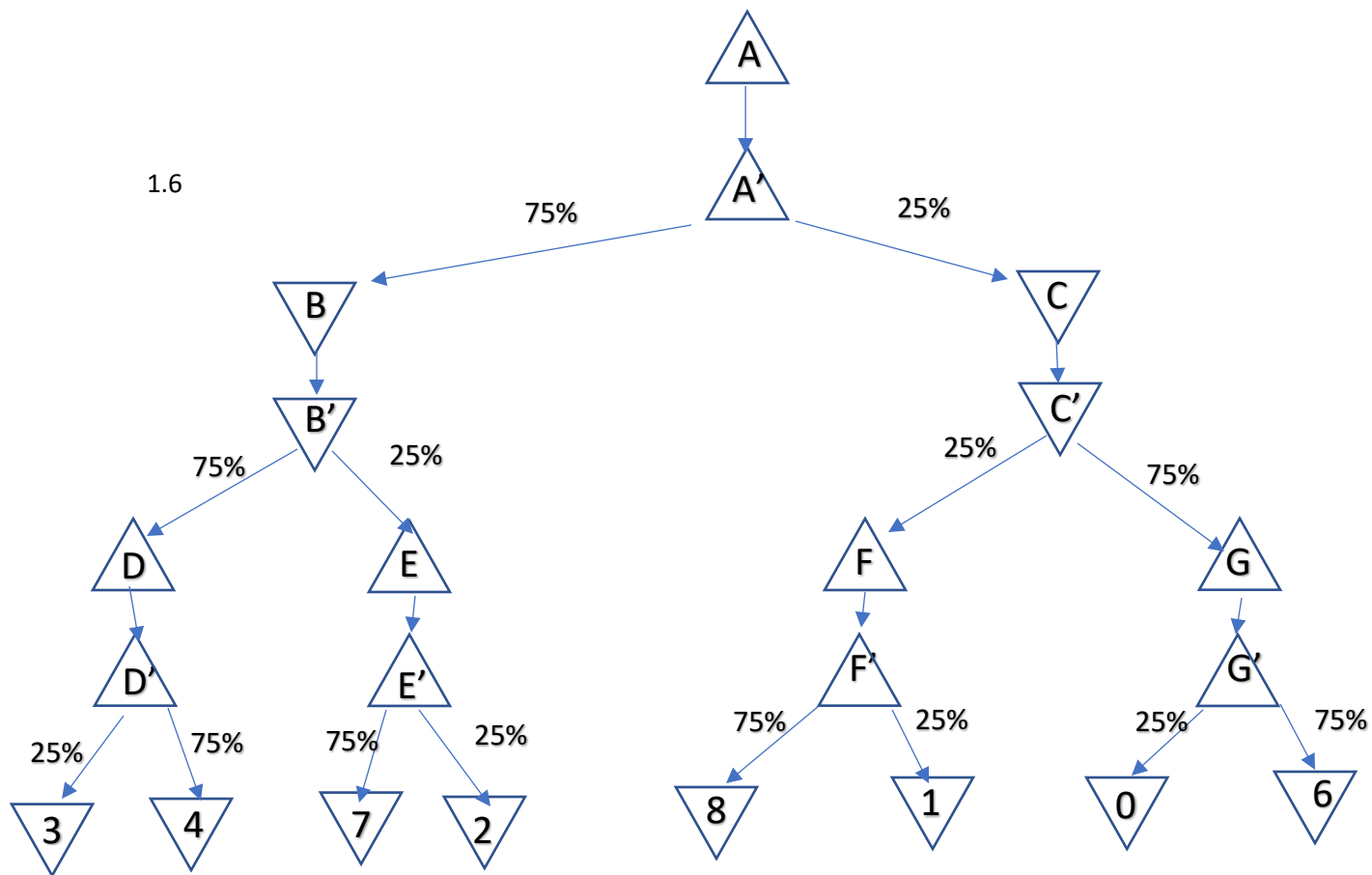
1.1 A= 6
 B= 4
 C= 6
 D= 4
 E= 7
 F= 8
 G=6

1.2 A=6	$\alpha=6$ $\beta=+\infty$
B=4	$\alpha=-\infty$ $\beta=4$
C=6	$\alpha=6$ $\beta=6$
D=4	$\alpha=4$ $\beta=+\infty$
E=7	$\alpha=7$ $\beta=4$
F=8	$\alpha=8$ $\beta=+\infty$
G=6	$\alpha=6$ $\beta=+\infty$

1.3 We would prune the right child of Node E because the left child is already higher than beta.

1.4 We would want to use alpha-beta pruning because by pruning subtrees that are known to not to be optimal for the other player we save memory and time when running our programs.

1.5 There is a 75% chance of choosing the optimal action at each node. We first take the 50% chance of getting heads and choosing the optimal action then we add the 25% chance of getting the optimal action on the second coin flip.



1.7 $A=4.77$

$$A = C \cdot .75 + B \cdot .25 = 4.94 \cdot .75 + 4.25 \cdot .25$$

$$B = 4.25$$

$$B = D \cdot .75 + E \cdot .25 = 3.75 \cdot .75 + 5.75 \cdot .25$$

$$C = 4.94$$

$$C = G \cdot .75 + F \cdot .25 = 4.5 \cdot .75 + 6.25 \cdot .25$$

$$D = 3.75$$

$$D = 4 \cdot .75 + 3 \cdot .25$$

$$E = 5.75$$

$$E = 7 \cdot .75 + 2 \cdot .25$$

$$F = 6.25$$

$$F = 8 \cdot .75 + 1 \cdot .25$$

$$G = 4.5$$

$$G = 6 * .75 + 0 * .25$$

2.1 The most convenient way to break a corpus into word tokens would be to use scanner and a list ADT. If the token obtained by the is contained in the list increment the count of the word object otherwise add a new word object to the list.

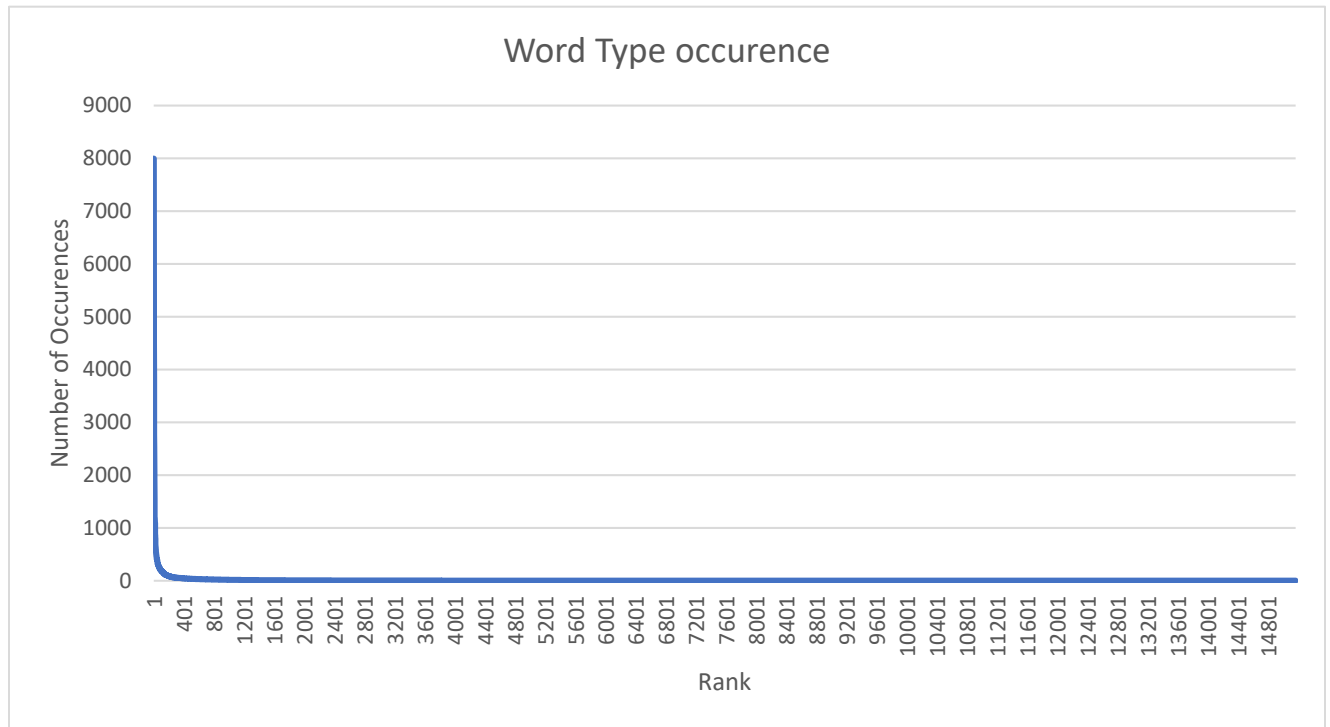
2.2 There are 153,526 word tokens and 15,170 computer word types.

2.3	1. the	7996
	2. to	5068
	3. of	4503
	4. and	3866
	5. in	2898
	6. a	2778
	7. is	2513
	8. that	2498
	9. be	2029
	10. Al	1931
	11. will	1544
	12. for	1204
	13. are	1186
	14. it	1177
	15. on	1117
	16. not	1106
	17. as	1088
	18 with	889
	19. The	852
	20. have	811

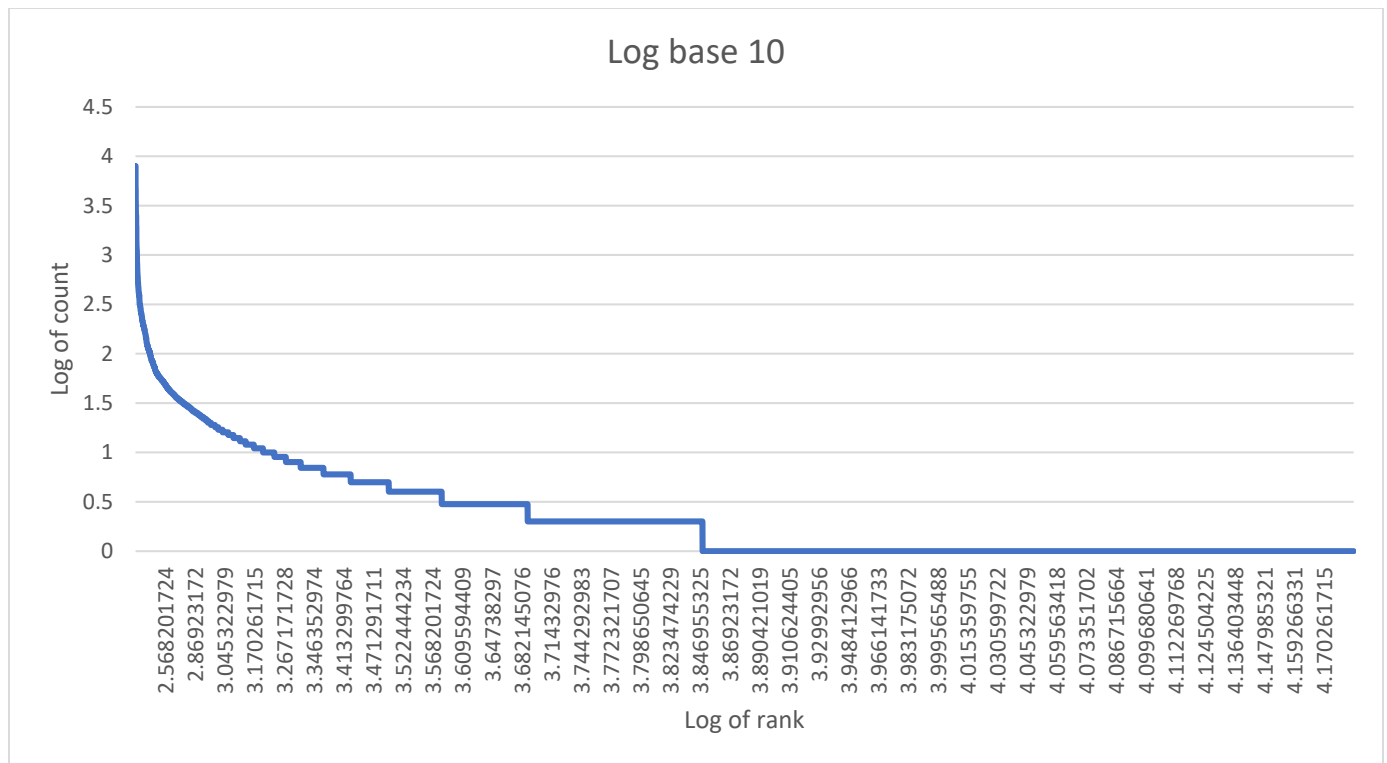
2.4	1. name.	1
	2. cyberterrorism	1
	3. provoke	1
	4. township	1
	5. diverting	1
	6. breaching	1
	7. outlets,	1
	8. senators	1
	9. Amerithrax	1
	10. small-scale	1
	11. posed;	1
	12. FBI	1
	13. warped	1
	14. fine,	1
	15. budgetary	1
	16. 9000,	1
	17. (Hal	1

18. genre 1
19.witout 1
20.attacks; 1

2.5



2.6



2.7 These two curves represent the rank and occurrence of a particular word. The first shows the exact rank and occurrence and the second shows the log base 10 of each rank and count. The shape of both curves shows the drastic drop from the first few most used words. The curve then flattens out towards the least used words because of the large amount of unique words.

2.8 Two potential major issues with my computer words for the use of natural language processing are punctuation and capitalization. My program creates a unique word based on space separation which will add existing words that have punctuation attached to it. My program also does not ignore capitalization so existing words with different letter casing will also be added.