CT102 – Assignment 1 Daniel Hannon (19484286)

1.

I. EU citizens have a “Right To Be Forgotten,” meaning that any citizen can request that search engines remove links to pages which are publicly available but where the content on a page is deemed private.

Step 1: Stop word removal – Get rid of any words which hold little to no value in terms of content of the page

EU citizens “Right To Forgotten,” that citizen that search engines links to pages which are available where content on page is deemed.

Step 2: Normailisation – make it lower case

eu citizens “right to forgotten,” that citizen that search engines links to pages which are available where content on page is deemed.

Step 3: Word stemming/ lemmitisation altering the nature of words so plurals have the same weighting as singulars and so on

eu citizen “ right to forgotten ,” that citizen that search engin link to page which are avail where content on page is deem .

II.

Tf of “page” in the above statement is 1/11 as it appears twice and there’s 22 words total.

Idf of the term “page” is log(500/300) which is roughly 0.222

Tf*idf of “page” = 1/11 * 0.511 = roughly 0.0202

2.

< 0.3, 0.25, 0.1, 0.02, 0, 0.11 >

<	0.3,	0.25,	0.1,	0.02,	0,	0.11 >
	t1	t2	t3	t4	t5	t6
<	0.35,	0,	0.3,	0.11,	0.02,	0.2 >

Document Similarity

Using the Cosine similarity formula we can deduce that

$$\frac{(t1_1*t1_2+t2_1*t2_2+t3_1*t3_2+t4_1*t4_2+t5_1*t5_2+t6_1*t6_2)}{\sqrt{(\sum(tx_1)^2)}*\sqrt{(\sum(tx_2)^2)}}=document\ similarity$$

$$\frac{(0.3*0.35+0.25*0.0+0.1*0.3+0.02*0.11+0*0.02+0.11*0.2)}{\sqrt{((0.3)^2+(0.25)^2+(0.1)^2+(0.02)^2+(0.11)^2)}*\sqrt{((0.35)^2+(0.3)^2+(0.11)^2+(0.02)^2+(0.2)^2)}}=\frac{0.1592}{\sqrt{0.265}*\sqrt{0.175}}$$

$$\frac{0.1592}{0.2153}=0.7394$$

The documents have a similarity of 0.7394

3.

	A	B	C	D
A	0	1	0	1
B	0	0	1	0
C	0	1	0	1
D	1	1	1	0

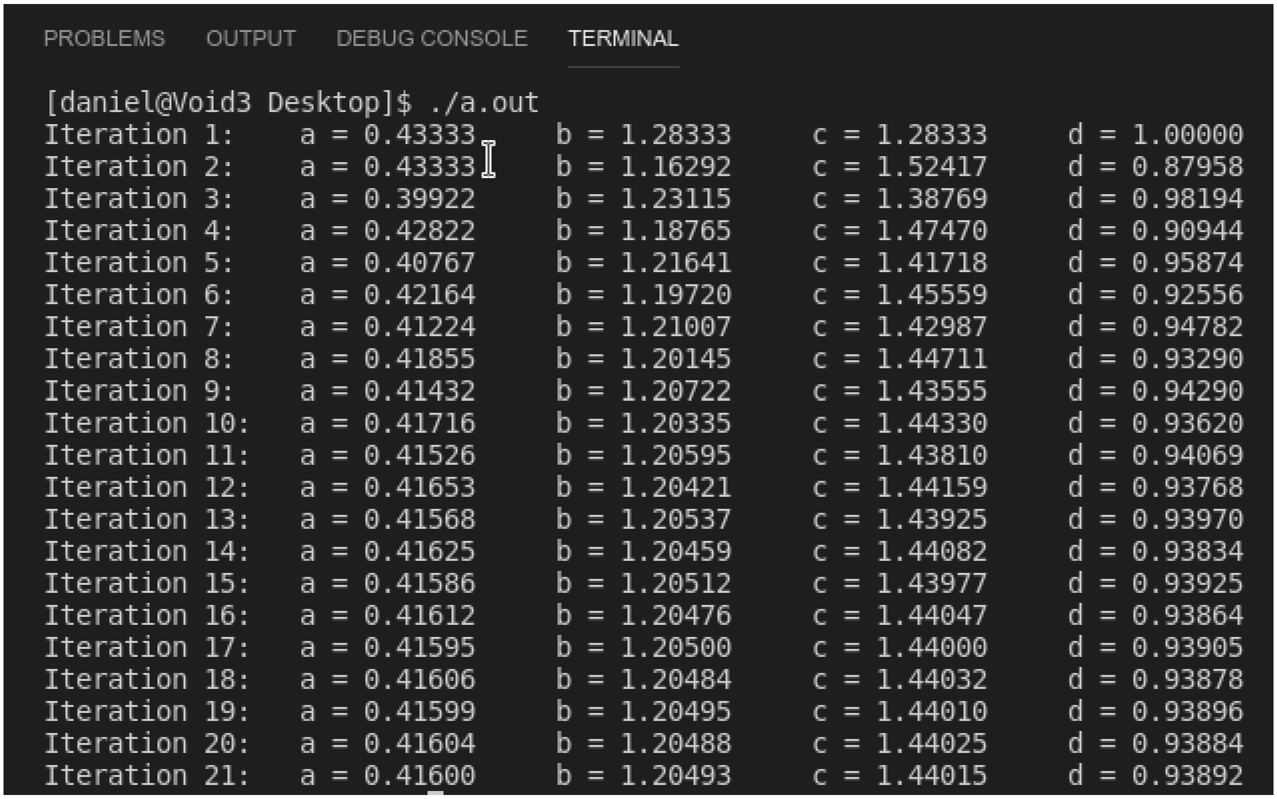
Outgoing Links

A = 2
B = 1
C = 2
D = 3

Formulas for Page Ranks
Pr(A) = 0.15 + 0.85(Pr(D)/C(D))
Pr(B) = 0.15 + 0.85(Pr(A)/C(A) + Pr(C)/C(C) + Pr(D)/C(D))
Pr(C) = 0.15 + 0.85(Pr(B)/C(B) + Pr(D)/C(D))
Pr(D) = 0.15 + 0.85(Pr(A)/C(A) + Pr(C)/C(C))

C Code to perform said calculation

```c
#include <stdio.h>
int main() { /*Page rank Script */
    double dVal = 0.85;
    double a = 1.0, b = 1.0, c = 1.0, d = 1.0, a_pr, b_pr, c_pr, d_pr;
    int links_a = 2, links_b = 1, links_c = 2, links_d = 3, count = 1; /*Outgoing Links and count, 0th Iteration all equal 1*/
    while (1 == 1) {
        a_pr = a;
        b_pr = b;
        c_pr = c;
        d_pr = d;
        a = 0.15 + (dVal * (d_pr/links_d));
        b = 0.15 + (dVal * (a_pr/links_a + c_pr/links_c + d_pr/links_d));
        c = 0.15 + (dVal * (b_pr/links_b + d_pr/links_d));
        d = 0.15 + (dVal * (a_pr/links_a + c_pr/links_c));
        printf("Iteration %d:\ta = %0.5lf\tb = %0.5lf\tc = %0.5lf\td = %0.5lf\n",count,a,b,c,d);
        if ((a_pr - a <= 0.0001 && a_pr - a >= -0.0001)&&(b_pr - b <= 0.0001 && b_pr - b >= -0.0001)&&(c_pr - c <= 0.0001 && c_pr - c >= -0.0001)&&(d_pr - d <= 0.0001 && d_pr - d >= -0.0001)) break;
        count++;
    }
}
```



```
[daniel@Void3 Desktop]$ ./a.out
Iteration 1:    a = 0.43333     b = 1.28333     c = 1.28333     d = 1.00000
Iteration 2:    a = 0.43333     b = 1.16292     c = 1.52417     d = 0.87958
Iteration 3:    a = 0.39922     b = 1.23115     c = 1.38769     d = 0.98194
Iteration 4:    a = 0.42822     b = 1.18765     c = 1.47470     d = 0.90944
Iteration 5:    a = 0.40767     b = 1.21641     c = 1.41718     d = 0.95874
Iteration 6:    a = 0.42164     b = 1.19720     c = 1.45559     d = 0.92556
Iteration 7:    a = 0.41224     b = 1.21007     c = 1.42987     d = 0.94782
Iteration 8:    a = 0.41855     b = 1.20145     c = 1.44711     d = 0.93290
Iteration 9:    a = 0.41432     b = 1.20722     c = 1.43555     d = 0.94290
Iteration 10:   a = 0.41716     b = 1.20335     c = 1.44330     d = 0.93620
Iteration 11:   a = 0.41526     b = 1.20595     c = 1.43810     d = 0.94069
Iteration 12:   a = 0.41653     b = 1.20421     c = 1.44159     d = 0.93768
Iteration 13:   a = 0.41568     b = 1.20537     c = 1.43925     d = 0.93970
Iteration 14:   a = 0.41625     b = 1.20459     c = 1.44082     d = 0.93834
Iteration 15:   a = 0.41586     b = 1.20512     c = 1.43977     d = 0.93925
Iteration 16:   a = 0.41612     b = 1.20476     c = 1.44047     d = 0.93864
Iteration 17:   a = 0.41595     b = 1.20500     c = 1.44000     d = 0.93905
Iteration 18:   a = 0.41606     b = 1.20484     c = 1.44032     d = 0.93878
Iteration 19:   a = 0.41599     b = 1.20495     c = 1.44010     d = 0.93896
Iteration 20:   a = 0.41604     b = 1.20488     c = 1.44025     d = 0.93884
Iteration 21:   a = 0.41600     b = 1.20493     c = 1.44015     d = 0.93892
```

Output of program (Results converge on 21st Iteration) Final Rankings in Order C,B,D,A

“I am aware of what plagiarism is and include this here to confirm that this work is my own”