**Problem 1**

Query: “walrus”

Doc1

f(t,d) = 10 o(t) = 2 c = 4 w(t,d) = 1 + log2(10) = 4.32 i(t) = 1 + log2(4/2) = 2

dt = 4.32\*2 = 8.64

d = [8.64,0,0,0], q = [1,0,0,0]

**Sim(d,q) = 8.64/8.64 = 1**

Doc2

f(t,d) = 0 w(t,d,) = 0

dt = 0

d = [0,0,0,0] **Sim(doc2,q) = 0**

Doc3

**Sim(d, q) = 0**

Doc4

f(t,d) = 10 o(t) = 2 c = 4 w(t,d) = 4.32 i(t) = 1 + log2(4/2) = 2

dt = 4.32\*2 = 8.64

d = [8,0,0,0], q = [1,0,0,0]

**Sim(d,q) = 1**

Rank: 1) Doc1 2) Doc4 3) Doc2 4) Doc2

Query: “walrus carpenter”

Doc1

q = [1,1,0,0]

dwalrus = 8.64

f(carpenter, d) = 8 o(t) = 2 c=4 w(carpenter,d) = 4 i(carpenter) = 2

dcarpenter = 8

d = [8.64, 8, 0, 0]

**Sim(d,q) = 16.64/16.65 ~ 1**

Doc2

f(walrus,d) = 0

f(carpenter,d) = 0

**Sim(d,q) = 0**

Doc3

f(walrus,d) = 0

dwalrus = 0

f(carpenter,d) = 40 o(t) = 2 c = 4 w(carpenter,d) = 1 + log2(40) = 6.32 i(carpenter) = 2

dcarpenter = 6.32\*2 = 12.64

d = [0,12.64,0,0] **Sim(d,q) = 12.64/sqrt(2)\*12.64 = 1/sqrt(2) = 0.707**

Doc4

f(walrus,d) = 10

dwalrus = 8.64

f(carpenter,d) = 0

dcarpenter = 0

d = [8.64,0,0,0]

**Sim(d,q) = 8.64/sqrt(2)\*8.64 = 0.707**

Rank: 1) Doc1 2) Doc3 3) Doc4 4) Doc2

Query = “walrus bread butter”

q = [1,0,1,1]

Doc1

dwalrus = 8.64

f(bread, d) = 4 o(t) = 3 c = 4 w(bread, d) = 1 + log2(4) = 3 i(bread) = 1.415

dbread = 3 \* 1.415 = 4.245

f(butter, d) = 1 o(t) = 2 c = 4 w(butter, d) = 1 + log2(1) = 1

i(butter) = 1+ log2(4/2) = 3

dbutter = 1\*3 = 3

d = [8.64, 0, 4.245, 3]

**Sim(d,q) = 15.885/(sqrt(3)\*10.08 = 15.885/17.46 = 0.91**

Doc2

dwalrus = 0

f(bread,d) = 24 o(t) = 3 c = 4 w(bread,d ) = 1+ log2(24) = 5.58

i(bread) = 1.415

dbread = 7.8957

f(butter,d) = 16 o(t) = 2 c = 4 w(butter,d) = 1 + 4 = 5 i(butter) = 3

dbutter = 15

d = [0,0,7.8957, 15]

**Sim(d,q) = 22.8957/sqrt(3)\*16.95 = 22.8957/29.36 = 0.7798**

Doc3

d = [0,0,0,0] **Sim(d,q) = 0**

Doc4

dwalrus­ = 8.64

dbutter = 0

f(bread,d) = 20 o(t) = 3 c = 4 w(bread, d) = 5.32 i(bread) = 1.415

dbread = 1.415 \* 5.32

d = [8.64,0,0,7.5278]

**Sim(d,q) = 16.1678/sqrt(3)\*11.46 = 16.1678/19.85 = 0.814**

Rank: 1) Doc1 2) Doc4 3) Doc2 4) Doc3

**Problem 2**

**A.**

Doc1 = [10,8,4,1]

Doc2 = [0,0,24,16]

Doc3 = [0,40,0,0]

Doc4 = [10,0,20,0]

Sim(Doc1, Doc2) = 112/13.45\*28.84 = 112/387.9 = 0.288

Sim(Doc1, Doc3) = 320/40\*13.45 = 0.594

Sim(Doc1, Doc4) = 180/13.45\*22.36 = 0.598

**B.**

walrus = [10,0,0,10]

carpenter = [8,0,40,0]

bread = [4,24,0,20]

butter = [1,16,0,0]

Sim(bread, carpenter) = 32/31.49\*40.79 = 0.025

Sim(bread, walrus) = 240/14.14\*31.49 = 0.539

Sim(bread, butter) = 388/31.49\*16.03 = 0.768

**Problem 3**

**A.** Does work.

Suppose f(t,d) = f(t,e) this means that w(t,d) = w(t,e). Since c and o(t) are constants, then

dt = et. Then d = e.

**B.** Doesn’t work.

Suppose f(t,d) = p\*f(t,e) then:

w(t,d) = 1 + log2(f(t,d)) = 1 + log2(p\*f(t,e)) = 1+log2(p) + log2(f(t,e)) = log2(p) + w(t,e)

So w(t,d) = log2(p) + w(t,e) = C + w(t,e) (C being a constant)

Then dt = w(t,d)\*i(t) = (C+w(t,e))\*i(t) = C‘ + w(t,e)\*i(t)

So dt = C‘ + de­. Since Sim(d,q) is not linear, then Sim(d,q) != Sim(e,q)

E.g.

If d = [1,2] and C’ = 1 and e = [2,3] and q = [1,1]

So Sim(d,q) = 3/sqrt(2)sqrt(5) = 0.9486

But Sim(e,q) = 5/\*sqrt(2)\*sqrt(13) = 0.9805

**C.** True

Using one collection or another will only change the constants o(t) and c.

Let Collb(e) denote e in the context of collection b

For every t, Collc(et)= C\*Collb(dt)

So if Rank(Collb(d)) > Rank(Collb(e))

Then Rank(Collc(d)) = Rank(C\*Collb(e))

And Rank(Coll­c(e)) = Rank(C\*Collb(e))

So Rank(Collc(d)) > Rank(Coll­c(e))

**Problem 4**

**A.**

N = 9

f = 0.7

e = 0.3

E = e/N = 0.3/9 = 0.0333

A = 0.0333

B = 0.0333 + 0.7\*(A/4 + C/3)

C = 0.0333 + 0.7\*(A/4 + B/2 + I/2)

D = 0.0333 + 0.7\*(A/4 + H)

E = 0.0333 + 0.7\*(A/4 + B/2 + C/3 + F/2 + D/2)

F = 0.0333 + 0.7\*(C/3 + E/2)

G = 0.0333 + 0.7\*(D/2)

H = 0.0333 + 0.7\*(I/2 + G + E/2)

I = 0.0333 + 0.7\*(F/2)

**B.**

Solution

Using Matlab:

a = zeros(9,1); c = 0.0333\*ones(9,1); q = matrix rep of part **A**

for i = 1:50

a = c+q\*a

end

Solution Vector:

0.0333

0.0591

0.0857

0.1702

0.1801

0.1163

0.0929

0.1873

0.0740

**Problem 5**

**e = 0.99**

f = 0.01

E = e/N = 0.99/9 = 0.11

A = 0.11

B = 0.11+ 0.01\*(A/4 + C/3)

C = 0.11+ 0.01\*(A/4 + B/2 + I/2)

D = 0.11+ 0.01\*(A/4 + H)

E = 0.11+ 0.01\*(A/4 + B/2 + C/3 + F/2 + D/2)

F = 0.11+ 0.01\*(C/3 + E/2)

G = 0.11+ 0.01\*(D/2)

H = 0.11+ 0.01\*(I/2 + G + E/2)

I = 0.11+ 0.01\*(F/2)

Using Matlab:

a = zeros(9,1); c = 0.11\*ones(9,1); q = matrix rep of above system of equations

for i = 1:50

a = c+q\*a

end

Solution Vector:

0.1100

0.1106

0.1114

0.1114

0.1123

0.1109

0.1106

0.1122

0.1106

**e = 0.01**

f = 0.99

E = e/N = 0.01/9 = 0.0011

Using Matlab:

a = zeros(9,1); c = 0.11\*ones(9,1); q = matrix

for i = 1:50

a = c+q\*a

end

Solution Vector:

0.0011

0.0063

0.0151

0.0916

0.0746

0.0424

0.0457

0.0926

0.0218