Problem Set 2

1. (1) The scores and rankings for the four documents for query “walrus”:

Doc1: d = Doc1, t= “walrus”

We have f (t, d) = 10, o (t) = 2, c = 4, so w (t, d) = , i (t) = and

The similarity = = 0.642.

Doc2: d = Doc2, t= “walrus”

We have f (t, d) = 0, o (t) = 2, c = 4, so w (t, d) = 0, i (t) = and

The similarity =

Doc3: d = Doc3, t= “walrus”

We have f (t, d) = 0, o (t) = 2, c = 4, so w (t, d) = 0, i (t) = and

The similarity =

Doc4: d = Doc4, t= “walrus”

We have f (t, d) = 10, o (t) = 2, c = 4, so w (t, d) = , i (t) = and

The similarity =

The ranking for query “walrus” is Doc1, Doc4, Doc2, Doc3.

(2) The scores and rankings for the four documents for query “walrus carpenter”:

Doc1:

We can know that from (1).

We have f (carpenter, d) = 8, o (t) = 2, c = 4, so w (carpenter, d) = , i (carpenter) = and So

The similarity =

Doc2:

We can know that from (1).

We have f (carpenter, d) = 0, o (t) = 2, c = 4, so w (carpenter, d) = , i (carpenter) = and So

The similarity =

Doc3:

We can know that from (1).

We have f (carpenter, d) = 40, o (t) = 2, c = 4, so w (carpenter, d) = , i (carpenter) = and So

The similarity =

Doc4:

We can know that 8.64 from (1).

We have f (carpenter, d) = 0, o (t) = 2, c = 4, so w (carpenter, d) = 0, i (carpenter) = and So

The similarity =

So the ranking for query “walrus carpenter” is Doc1, Doc4, Doc3, Doc2.

(3) The scores and rankings for the four documents for query “walrus bread butter”:

Doc1:

We can know that from (1).

w(butter, d) = o(butter) = 2, c = 4;

i(butter) =

w(bread, d) = o(bread) = 3, c = 4;

i(bread) =

;

The similarity =

Doc2:

We can know that from (1).

w(butter, d) = o(butter) = 2, c = 4;

i(butter) =

w(bread, d) = o(bread) = 3, c = 4;

i(bread) =

;

The similarity =

Doc3:

We can know that from (1).

w(butter, d) = o(butter) = 2, c = 4;

i(butter) =

w(bread, d) = o(bread) = 3, c = 4;

i(bread) =

The similarity =

Doc4:

We can know that from (1).

w(butter, d) = o(butter) = 2, c = 4;

i(butter) =

w(bread, d) = o(bread) = 3, c = 4;

i(bread) =

;

The similarity =

The ranking for query “walrus bread butter” is Doc1, Doc4, Doc2, Doc3.

1. (1)

Doc1:

= 8.64，

Doc2:

= 0，

Doc1:

= 8.64，

Doc3:

= 0，

Doc1:

= 8.64，

Doc4:

= 8.64，

(2) word “bread”:

word “walrus”:

word “carpenter”:

word “butter”:

1. (1)

|  |  |  |
| --- | --- | --- |
|  | d | e |
| walrus | 10 | 0 |
| bread | 10 | 10 |
| carpenter | 20 | 0 |
| butter | 0 | 10 |

Query: “bread”

Document d and e:

f(bread, d) = f(bread, e) = 1+, i(bread) are both 1+

The similarity

So property A fails to hold in the example above.

(2)

|  |  |  |
| --- | --- | --- |
|  | d | e |
| walrus | 0 | 4 |
| bread | 8 | 4 |
| carpenter | 8 | 4 |
| butter | 0 | 4 |

Query: “bread carpenter”

Form the table above, we know that f(bread, d) = 2\* f(bread, e), f(carpenter, d) = 2\* f(carpenter, e).

Both d and e, i(bread carpenter) =

The similarity =

The similarity =

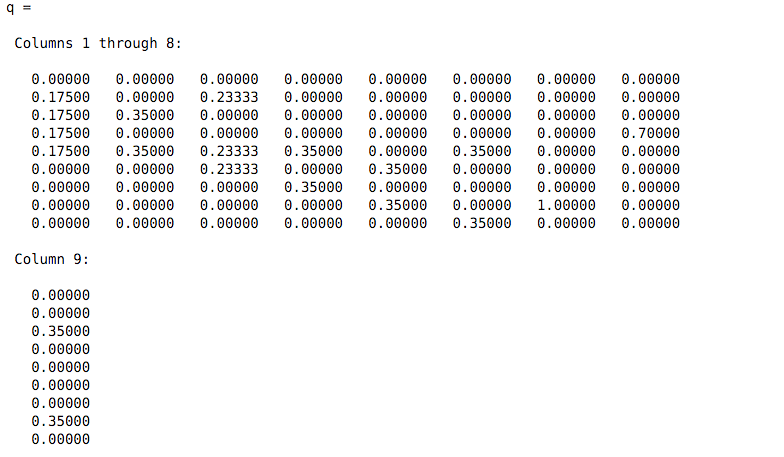
So property B fails to hold in the example above.

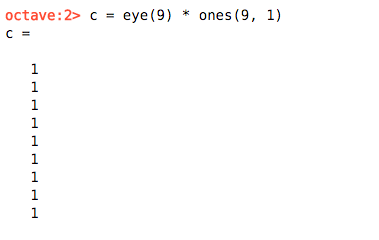
(3) Property C is correct.

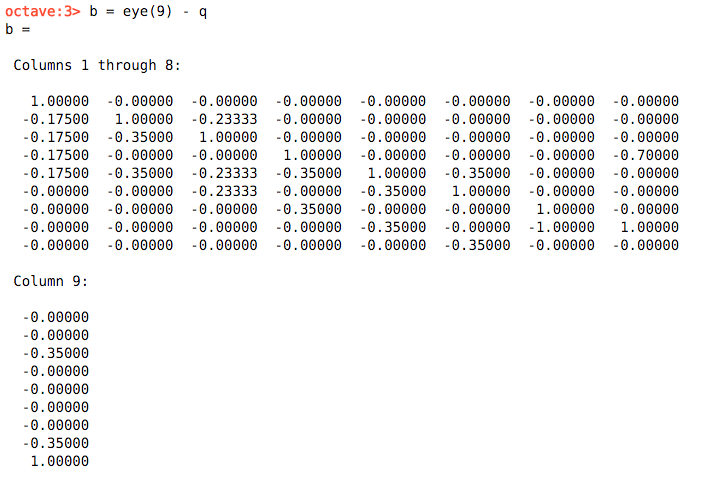
We suppose that d is ranked higher than e in collection b. Since everything in the formulation is unchanged in different collections (b and c), except i(t). However, i(t) are the same for d and e in the same collection. So the rank result will not change the order, d is still ranked higher than e in collection c.

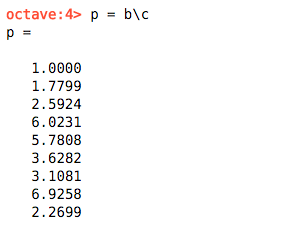
1. (1) Linear equations for PageRank (e = 0.3):

(2) The Matrix Q is





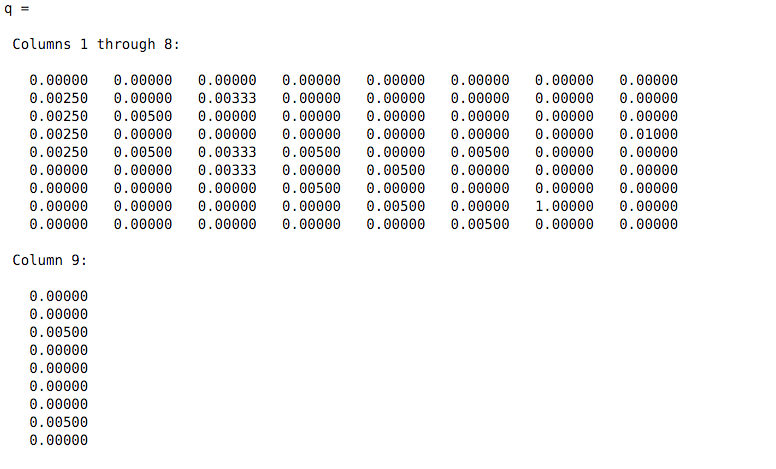


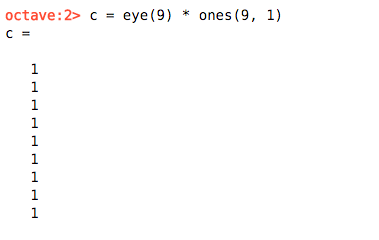


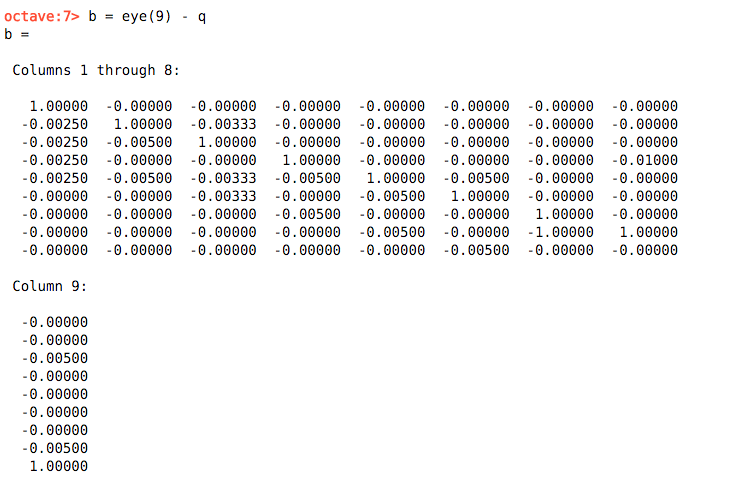
The solution of (1) is p as shown above. And the order of pages is H, D, E, F, G, C, I, B, A.

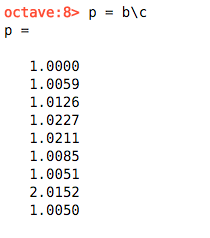
1. (1) Linear equations for PageRank (e = 0.99):

The Matrix Q is





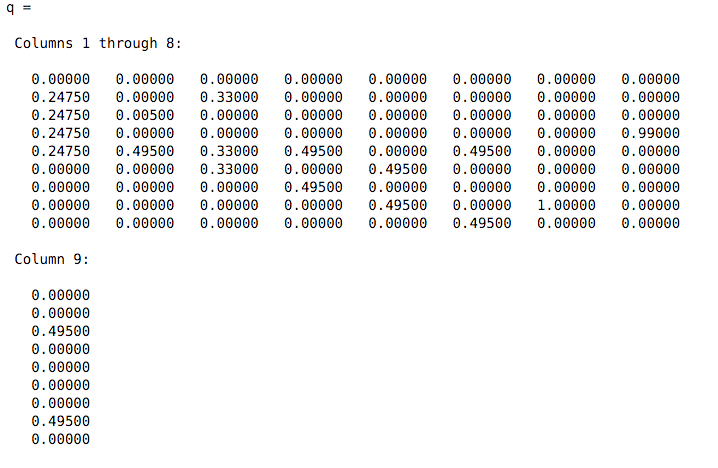


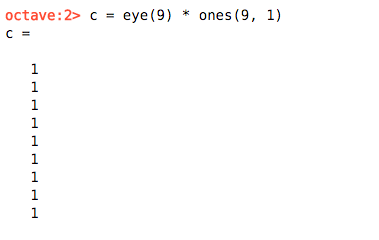


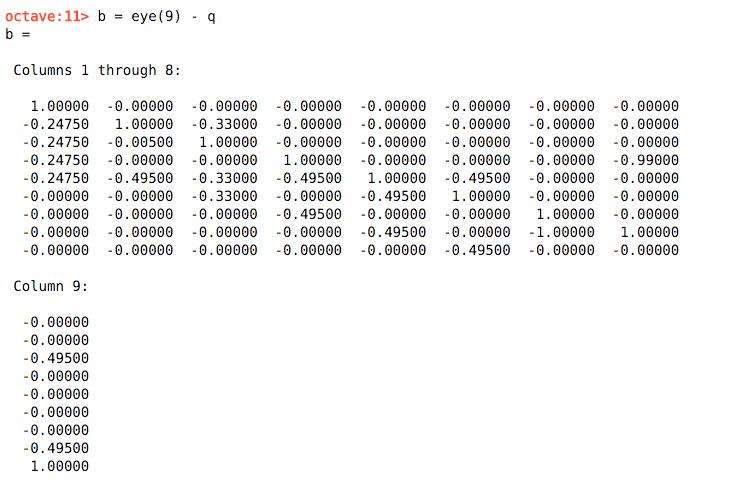
The solution of (1) is p as shown above. And the order of pages is H, D, E, C, F, B, G, I, A.

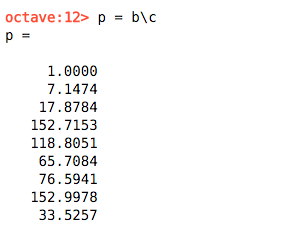
1. (2) Linear equations for PageRank (e = 0.01):

The Matrix Q is









The solution of (1) is p as shown above. And the order of pages is H, D, E, G, F, I, C, B, A.