



Department of Computer Science - Τμήμα Πληροφορικής
University of Cyprus - Πανεπιστήμιο Κύπρου

Εργαστήρια ΕΠΛ 231

Χειμερινό 2018

Εργασία Αρ. 3
Ημερομηνία έκδοσης: 8/11/2018
Ημερομηνία Παράδοσης: 26/11/2018

Ονοματεπώνυμο: Σωκράτης Γιαννακού
Αριθμός Ταυτότητας: 913240



Φωτογραφία Εδώ

Υπεύθυνος Εργαστηρίου
Πέτρος Παναγή

**ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΛΥΣΗΣ (~1/2 ΣΕΛΙΔΑ):
(30 Μονάδες)**

Αρχικά το πρόγραμμα διαβάζει από ένα αρχείο κειμένου διάφορες ιστοσελίδες. Δημιουργείται Array List (urls) με τα ονόματα των ιστοσελιδών και γράφος του οποίου οι κόμβοι είναι τα ονόματα ιστοσελιδών. Ο γράφος είναι ένα Linked List όπου ο κάθε κόμβος έχει ένα δικό της Linked List. Έτσι παρουσιάζονται οι ακμές του γράφου, οι διασυνδέσεις μεταξύ των σελιδών. Δημιουργείται πίνακας (edges) ο οποίος περιέχει το σύνολο των εξερχόμενων συνδέσμων της κάθε ιστοσελίδας. Η κάθε ιστοσελίδα έχει από 1-30 σύνδεσμους (τυχαίος αριθμός). Αυτός ο αριθμός είναι το σύνολο των συνδέσμων οι οποίοι πάλι επιλέγονται τυχαία. Δημιουργούνται 2 πίνακες, pagerank και temp, το pagerank της προηγούμενης επανάληψης. Ζητείται ο χρήστης να επιλέξει 1 ιστοσελίδα για να δει τα αποτελέσματα της μέσα από τις 3 μεθόδους. Στην μέθοδο Α υπολογίζεται το Pagerank με τον τύπο που μας δόθηκε. Γίνονται επαναλήψεις μέχρι το διάνυσμα των pageranks να είναι μικρότερο από μια μικρή τιμή. Η μέθοδος Β ελέγχει επιπλέον μετά από 5 επαναλήψεις αν κάποιο pagerank είναι μικρότερο από ένα κατώφλι, τότε το μηδενίζει έτσι αποφεύγονται τα spams. Η μέθοδος Γ, εξετάζει αν υπάρχουν ιστοσελίδες με κοινό domain, τότε πολλαπλασιάζει το pagerank με το sibling factor για να μειωθεί. Επίσης ελέγχεται να μην έχουν ιστοσελίδες συνδέσμους ως προς τις ίδιες.

ΕΚΤΥΠΩΣΗ ΚΑΙ ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ (Ο ΓΡΑΦΟΣ ΘΑ ΠΡΕΠΕΙ ΝΑ ΕΙΝΑΙ Ο ΙΔΙΟΣ ΣΕ ΟΛΑ ΤΑ ΠΙΟ ΚΑΤΩ ΠΑΡΑΔΕΙΓΜΑΤΑ)

Υπολογισμός PageRank Μέθοδος Α (2-3 Παραδείγματα)
(5+15 Μονάδες):

```
Select a URL based on number:
9

***** Method A *****

Loops: 78
PageRank for giphy.com/el/index.html: 0.7930572682630037

PageRank for link: go.com/en/index.html: 0.6367017057270618
PageRank for link: rollingstone.com/content.html: 0.8290425569574755
PageRank for link: politico.com/index.html: 1.0324264492810868
PageRank for link: merriam-webster.com/info.html: 1.159271342046261
PageRank for link: goodreads.com/el/index.html: 0.7628232838082968
PageRank for link: twentytwowords.com/index.html: 1.3232131650583567
PageRank for link: usmagazine.com/contact.html: 0.6231127873460761
PageRank for link: ebay.com/en/index.html: 0.5741565504375216
PageRank for link: lifehacker.com/info.html: 0.940843147665534
PageRank for link: giphy.com/el/index.html: 0.7930572682630037
```

```
Number of vertices: 432
Select a URL based on number:
2

***** Method A *****

Loops: 77
PageRank for politico.com/index.html: 1.857158297115325

PageRank for link: go.com/en/index.html: 0.7110441247646299
PageRank for link: rollingstone.com/content.html: 0.965165280298529
PageRank for link: politico.com/index.html: 1.857158297115325
PageRank for link: merriam-webster.com/info.html: 1.1323404117946818
PageRank for link: goodreads.com/el/index.html: 0.5150772238539735
PageRank for link: twentytwowords.com/index.html: 1.29698698719635
PageRank for link: usmagazine.com/contact.html: 0.9015586197451315
PageRank for link: ebay.com/en/index.html: 0.5946605013649872
PageRank for link: lifehacker.com/info.html: 0.9936809175696357
PageRank for link: giphy.com/el/index.html: 1.1265819295498691
PageRank for link: adobe.com/info.html: 0.6954999592138437
PageRank for link: facebook.com/index.html: 0.8510464481103981
PageRank for link: craigslist.org/info.html: 1.1268859568317338
PageRank for link: littlethings.com/info.html: 0.9995373176723507
PageRank for link: wayfair.com/info.html: 0.8653196671082849
PageRank for link: stackexchange.com/en/index.html: 0.6185705375843242
PageRank for link: comcast.net/el/index.html: 1.1773755714438434
PageRank for link: pinterest.com/content.html: 0.7445957908855504
PageRank for link: ups.com/en/index.html: 0.8636570079286897
```

Υπολογισμός pagerank για giphy.com/el/index.html και politico.com/index.html, με τη χρήση του τύπου. Παρουσίαση των συνδέσμων και των pagerank τους.

**Υπολογισμός PageRank Μέθοδος B (2-3 Παραδείγματα στα οποία να φαίνονται οι διαφορές με τις άλλες μεθόδους)
(5+15 Μονάδες):**

```
***** Method B *****

Loops: 6
PageRank for giphy.com/el/index.html: 0.8193667086413974

PageRank for link: go.com/en/index.html: 0.6563645490891724
PageRank for link: rollingstone.com/content.html: 0.8571428523383855
PageRank for link: politico.com/index.html: 1.068077304527258
PageRank for link: merriam-webster.com/info.html: 1.203903103836108
PageRank for link: goodreads.com/el/index.html: 0.7887273148370608
PageRank for link: twentytwowords.com/index.html: 1.3719005512423783
PageRank for link: usmagazine.com/contact.html: 0.6427989177312141
PageRank for link: ebay.com/en/index.html: 0.5907943414454194
PageRank for link: lifehacker.com/info.html: 0.974883557597839
PageRank for link: giphy.com/el/index.html: 0.8193667086413974
```

```
***** Method B *****

Loops: 6
PageRank for politico.com/index.html: 1.9262931127345206

PageRank for link: go.com/en/index.html: 0.7339350189862317
PageRank for link: rollingstone.com/content.html: 0.9961788617408577
PageRank for link: politico.com/index.html: 1.9262931127345206
PageRank for link: merriam-webster.com/info.html: 1.1721802378089377
PageRank for link: goodreads.com/el/index.html: 0.5300465580622382
PageRank for link: twentytwowords.com/index.html: 1.344969188877126
PageRank for link: usmagazine.com/contact.html: 0.9318136021023681
PageRank for link: ebay.com/en/index.html: 0.612148255257625
PageRank for link: lifehacker.com/info.html: 1.027920152542223
PageRank for link: giphy.com/el/index.html: 1.1666365574072737
PageRank for link: adobe.com/info.html: 0.7172853006193384
PageRank for link: facebook.com/index.html: 0.8794358562913172
PageRank for link: craigslist.org/info.html: 1.1682493123811242
PageRank for link: littlethings.com/info.html: 1.0332896980247301
PageRank for link: wayfair.com/info.html: 0.8937246126300247
PageRank for link: stackexchange.com/en/index.html: 0.636938727800092
PageRank for link: comcast.net/el/index.html: 1.2201902360117274
PageRank for link: pinterest.com/content.html: 0.7684210639661625
PageRank for link: ups.com/en/index.html: 0.8914104699940987
```

Υπολογισμός pagerank για giphy.com/el/index.html και politico.com/index.html, με τη χρήση του τύπου. Παρουσίαση των συνδέσμων και των pagerank τους. Επιπρόσθετα χρησιμοποιείται threshold=0.5 για αποφυγή spam.

Υπολογισμός PageRank με sibling factor Μέθοδος Γ (2-3 Παραδείγματα στα οποία να φαίνονται οι διαφορές με τις άλλες μεθόδους) (5+15 Μονάδες):

```
***** Method C *****
```

```
Loops: 48
```

```
PageRank for giphy.com/el/index.html: 0.45868593265724056
```

```
PageRank for link: go.com/en/index.html: 0.38352699333642026
PageRank for link: rollingstone.com/content.html: 0.46762910070100405
PageRank for link: politico.com/index.html: 0.5726522777377513
PageRank for link: merriam-webster.com/info.html: 0.6323402184580409
PageRank for link: goodreads.com/el/index.html: 0.4415298540540531
PageRank for link: twentytwowords.com/index.html: 0.7105762833169281
PageRank for link: usmagazine.com/contact.html: 0.3733082913990242
PageRank for link: ebay.com/en/index.html: 0.35965748390159097
PageRank for link: lifehacker.com/info.html: 0.5223853479046461
PageRank for link: giphy.com/el/index.html: 0.45868593265724056
```

```
***** Method C *****
```

```
Loops: 48
```

```
PageRank for politico.com/index.html: 0.981096814099341
```

```
PageRank for link: go.com/en/index.html: 0.4210316699900284
PageRank for link: rollingstone.com/content.html: 0.559045212863121
PageRank for link: politico.com/index.html: 0.981096814099341
PageRank for link: merriam-webster.com/info.html: 0.6191890532052277
PageRank for link: goodreads.com/el/index.html: 0.3271756236243811
PageRank for link: twentytwowords.com/index.html: 0.6903729624906391
PageRank for link: usmagazine.com/contact.html: 0.5090052593632621
PageRank for link: ebay.com/en/index.html: 0.36625267188977373
PageRank for link: lifehacker.com/info.html: 0.562745374795062
PageRank for link: giphy.com/el/index.html: 0.6087617649520395
PageRank for link: adobe.com/info.html: 0.4114644968527329
PageRank for link: facebook.com/index.html: 0.4843046347969141
PageRank for link: craigslist.org/info.html: 0.611740967822184
PageRank for link: littlethings.com/info.html: 0.5596567709589989
PageRank for link: wayfair.com/info.html: 0.49865592797755875
PageRank for link: stackexchange.com/en/index.html: 0.37735520520062205
PageRank for link: comcast.net/el/index.html: 0.6437376523565427
PageRank for link: pinterest.com/content.html: 0.44082090401948015
PageRank for link: ups.com/en/index.html: 0.5019948704718968
```

Υπολογισμός pagerank για giphy.com/el/index.html και politico.com/index.html, με τη χρήση του τύπου. Παρουσίαση των συνδέσμων και των pagerank τους. Επιπλέον πολλαπλασιασμός με siblingFactor=0.85 όπου υπάρχει κοινός domain και αποφυγή σύνδεσμου σε εαυτό.

ΚΩΔΙΚΑΣ (ΜΟΝΟ ΣΗΜΑΝΤΙΚΑ ΑΠΟΣΠΑΣΜΑΤΑ ΑΠΟ ΤΟΝ ΚΩΔΙΚΑ Π.Χ. ΥΠΟΛΟΓΙΣΜΟΙ): (10 Μονάδες):

(Κώδικας + ΣΧΟΛΙΑ)

Fonts: Courier New 10

```
for (int i = 0; i < V; i++) {
    pagerank[i] = 1;
    temp[i] = 1;
}
double result = 0;
int counter = 0;
do {
    counter++;
    for (int i = 0; i < V; i++) {
        temp[i] = pagerank[i];
    }

    for (int i = 0; i < V; i++) {
        double rank = 0;
        for (int j = 0; j < V; j++) {
            if (graph.adjListArray[j].contains(i)) {
                rank += 0.85 * temp[j] / edges[j];
            }
        }
        pagerank[i] = 0.15 + rank;
    }

    double distance = 0;
    for (int i = 0; i < V; i++) {
        distance += Math.abs(temp[i] - pagerank[i]);
    }
    result = distance;
} while (result > 0.00001);

for (int i = 0; i < V; i++) {
    pagerank[i] = 1;
    temp[i] = 1;
}
double threshold = 0.5;
result = 0;
counter = 0;
do {
    counter++;
    for (int i = 0; i < V; i++) {
        temp[i] = pagerank[i];
    }

    if (counter > 5) {
        for (int i = 0; i < V; i++) {
            if (pagerank[i] < threshold) {
                pagerank[i] = 0;
                temp[i] = 0;
            }
        }
    }
}
```

```
        else {
            for (int i = 0; i < V; i++) {
                if (pagerank[i] != 0) {
                    double rank = 0;
                    for (int j = 0; j < V; j++) {
                        if (graph.adjListArray[j].contains(i)) {
                            rank += 0.85 * temp[j] / edges[j];
                        }
                    }
                    pagerank[i] = rank + 0.15;
                }
            }
        }

        double distance = 0;
        for (int i = 0; i < V; i++) {
            distance += Math.abs(temp[i] - pagerank[i]);
        }
        result = distance;
    } while (result > 0.00001);

    for (int i = 0; i < V; i++) {
        pagerank[i] = 1;
        temp[i] = 1;
    }
    double siblingFactor = 0.85;
    result = 0;
    counter = 0;
    do {
        counter++;
        for (int i = 0; i < V; i++) {
            temp[i] = pagerank[i];
        }

        for (int i = 0; i < V; i++) {
            int c = urls.get(i).indexOf("/");
            String s = urls.get(i).substring(0, c);
            double rank = 0;
            for (int j = 0; j < V; j++) {
                if (graph.adjListArray[j].contains(i) && i!=j) {
                    int c2 = urls.get(j).indexOf("/");
                    String s2 = urls.get(j).substring(0, c2);
                    if (s2.equals(s))
                        rank += 0.85 * temp[j] / edges[j];
                    else
                        rank += siblingFactor * 0.85 * temp[j] /
edges[j];
                }
            }
            pagerank[i] = 0.15 + rank;
        }
    }
```

```
double distance = 0;
for (int i = 0; i < V; i++) {
    distance += Math.abs(temp[i] - pagerank[i]);
}
result = distance;
} while (result > 0.00001);
}
```


BONUS (20 ΜΟΝΑΔΕΣ)