

Ο αλγόριθμος PageRank

Φανταστείτε ότι ψάχνετε για την πιο δημοφιλή και αξιόπιστη σελίδα στο διαδίκτυο. Ο αλγόριθμος PageRank, βοηθά να βρούμε ποιες σελίδες είναι πιο «σημαντικές».

Πώς Λειτουργεί;

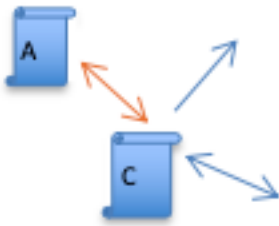
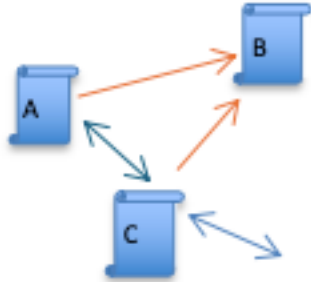
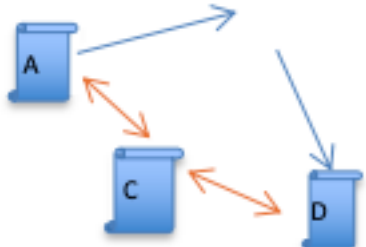
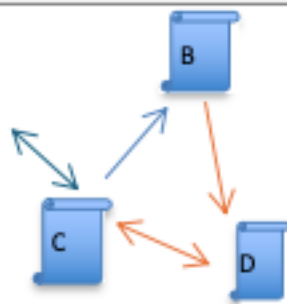
Ο αλγόριθμος στηρίζεται στις παρακάτω παραδοχές:

1. **Κάθε σελίδα στο διαδίκτυο μπορεί να συνδέεται με άλλες σελίδες.** Αυτοί οι σύνδεσμοι είναι σαν ψήφοι εμπιστοσύνης.
2. **Όσο περισσότερες σελίδες συνδέονται με μία συγκεκριμένη σελίδα, τόσο πιο σημαντική θεωρείται.**
3. **Όμως, δεν έχουν όλοι οι ψήφοι την ίδια αξία!** Αν μια σελίδα που ήδη θεωρείται πολύ σημαντική συνδέεται με μια άλλη, τότε της δίνει μεγαλύτερη βαρύτητα.

Ένα Απλό Παράδειγμα

Ας προσπαθήσουμε να κατανοήσουμε την λειτουργία του με τη βοήθεια του παρακάτω παραδείγματος:

Έστω ότι υπάρχουν 4 ιστοσελίδες: A, B, C και D.

Κόμβοι	Συνδεσιμότητα	1 ^{ος} γύρος	2 ^{ος} γύρος	3 ^{ος} γύρος	Βαθμός ιεράρχησης
A		$\frac{1}{4}$	$(\frac{1}{4})/3 = 1/12$ Ο Α έχει εισερχόμενο σύνδεσμο μόνο από τον C και ο C έχει 3 εξερχόμενους συνδέσμους	$(\frac{4,5}{12})/3 = 1,5/12$	1
B		$\frac{1}{4}$	$(\frac{1}{4})/2 + (\frac{1}{4})/3 = 2,5/12$ Ο Β έχει εισερχόμενους συνδέσμους από τους Α και C. Ο Α έχει 2 εξερχόμενους. Ο C έχει 3 εξερχόμενους	$(\frac{1}{12})/2 + (\frac{4,5}{12})/3 = 2/12$	2
C		$\frac{1}{4}$	$(\frac{1}{4})/2 + (\frac{1}{4})/1 = 4,5/12$ Ο C έχει 2 εισερχόμενους συνδέσμους από τους Α, D. Ο Α έχει 2 εξερχόμενους. Ο D έχει 1 εξερχόμενο.	$(\frac{1}{12})/2 + (\frac{4}{12})/1 = 4,5/12$	4
D		$\frac{1}{4}$	$(\frac{1}{4})/3 + (\frac{1}{4})/1 = 4/12$ Ο D έχει 2 εισερχόμενους συνδέσμους από τους Β, C. Ο C έχει 3 εξερχόμενους. Ο Β έχει 1 εξερχόμενο.	$(\frac{4,5}{12})/3 + (\frac{2,5}{12})/1 = 4/12$	3

- Η **A** έχει συνδέσμους προς τις **B** και **C**.
- Η **B** έχει σύνδεσμο προς την **D**.
- Η **C** έχει συνδέσμους προς τις **A**, **B**, **D**.
- Η **D** έχει σύνδεσμο προς την **C**.

Αρχικά, ο PageRank θεωρεί ότι όλες οι σελίδες έχουν την ίδια σημασία. Για παράδειγμα, αν υπάρχουν 4 σελίδες, η κάθε μία ξεκινά με βαθμολογία **1/4 (25%)**.

Πώς υπολογίζεται η νέα βαθμολογία;

Ο PageRank υπολογίζει τη σημασία μιας σελίδας με βάση:

1. Πόσες άλλες σελίδες συνδέονται προς αυτήν (εισερχόμενοι σύνδεσμοι).
2. Πόσο σημαντικές είναι αυτές οι σελίδες.
3. Πόσους εξερχόμενους συνδέσμους έχει κάθε σελίδα (αν μια σελίδα συνδέεται με πολλές άλλες, ο κάθε σύνδεσμος της δίνει μικρότερη βαρύτητα).

Κάθε σελίδα **μοιράζει** το PageRank της στις σελίδες προς τις οποίες συνδέεται.

1ος Γύρος Υπολογισμού

- Όλες οι σελίδες ξεκινούν με **1/4 (ή 0.25)**.
- Η **A** μοιράζει το PageRank της στις δύο σελίδες που συνδέεται (**B** και **C**). Άρα κάθε μία παίρνει **$0.25/2 = 0.125$** από την A.
- Η **B** έχει σύνδεσμο μόνο προς την **D**, άρα όλο το PageRank της (0.25) πάει στη **D**.
- Η **C** έχει συνδέσμους προς **A**, **B**, **D**. Άρα κάθε μία από αυτές παίρνει **$0.25/3 = 0.0833$** από την C.
- Η **D** έχει σύνδεσμο μόνο προς την **C**, άρα όλο το PageRank της (0.25) πάει στη **C**.

Μετά τον 1ο γύρο οι νέες βαθμολογίες είναι:

- **A = 0.0833**
- **B = $0.125 + 0.0833 = 0.2083$**
- **C = $0.125 + 0.25 = 0.375$**

$$D = 0.25 + 0.0833 = 0.3333$$

2ος Γύρος Υπολογισμού

Τώρα επαναλαμβάνουμε τη διαδικασία με τις νέες τιμές.

- **A** λαμβάνει PageRank από **C** (που είχε 0.375). Επειδή η C έχει 3 συνδέσμους, δίνει στην A **$0.375/3 = 0.125$** .

- **B** λαμβάνει από **A** ($0.0833/2 = 0.0417$) και από **C** ($0.375/3 = 0.125$). Άρα νέο PageRank της B: **0.1667**.
- **C** λαμβάνει από **D** ($0.3333/1 = 0.3333$) και από **A** ($0.0833/2 = 0.0417$). Άρα νέο PageRank της C: **0.375**.
- **D** λαμβάνει από **B** (**0.1667**) και από **C** ($0.375/3 = 0.125$). Άρα νέο PageRank της D: **0.2917**.

Τελική Κατάταξη των Σελίδων

Μετά από αρκετές επαναλήψεις, οι βαθμολογίες σταθεροποιούνται:

- **C = 0.40**
- **D = 0.30**
- **B = 0.20**
- **A = 0.10**

Αυτό σημαίνει ότι η **C είναι η πιο σημαντική σελίδα**, επειδή έχει πολλούς εισερχόμενους συνδέσμους από σελίδες που θεωρούνται επίσης σημαντικές.

Τι καταλαβαίνουμε;

Για να είναι μια σελίδα σημαντική:

- ✓ Πρέπει να τη συστήνουν πολλές άλλες σελίδες.
- ✓ Πρέπει να τη συστήνουν σελίδες που είναι ήδη δημοφιλείς.
- ✓ Πρέπει και η ίδια να συστήνει άλλες σελίδες.

Έτσι, με τη χρήση του PageRank καθορίζεται ποιες σελίδες θα εμφανίζονται πρώτες στα αποτελέσματα αναζήτησης