



PageRank

Wie Google die Welt eroberte

Gliederung

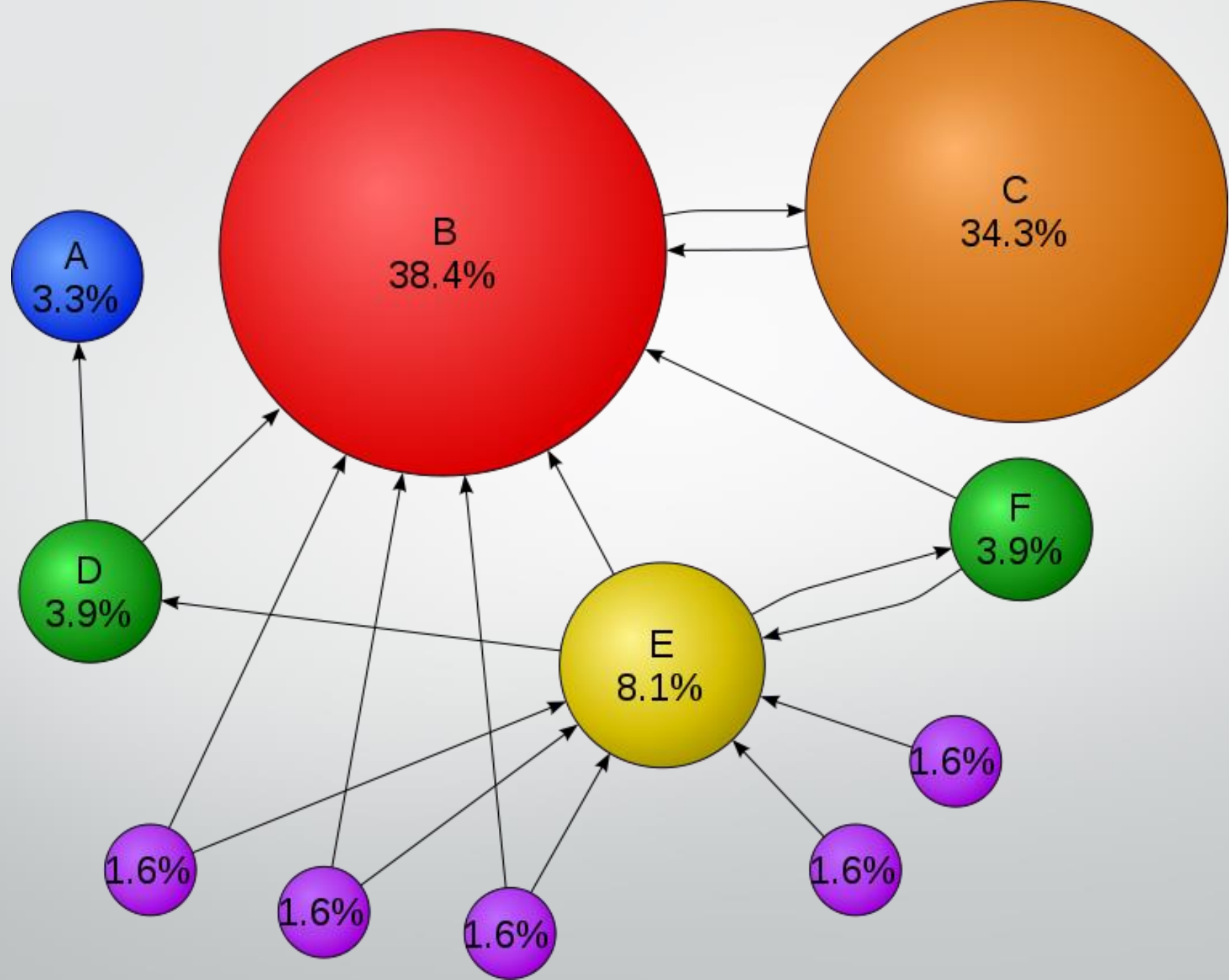
- Historisches
- Was ist PageRank?
- Algorithmus
- Random Surfer Model
- Verwendung außerhalb von Suchmaschinen
- Google Bomb

Historisches

- Entwickelt von Larry Page (-> PageRank) und Sergey Brin
- Idee 1996: Suchergebnisse nach Anzahl und Qualität von Hyperlinks zu einer Website zu ordnen
- 1998: Veröffentlichung Suchmaschine Google und PageRank

Was ist PageRank?

- Methode zur Platzierung von Suchergebnissen nach ihrer Wichtigkeit
- Gemessen an Anzahl und Qualität von Hyperlinks zu einer Website
- Je populärer eine Website ist, umso mehr Links führen zu ihr und umso größer ist ihr PageRank
- $PR(A)$: PageRank von Website A



Random Surfer Model

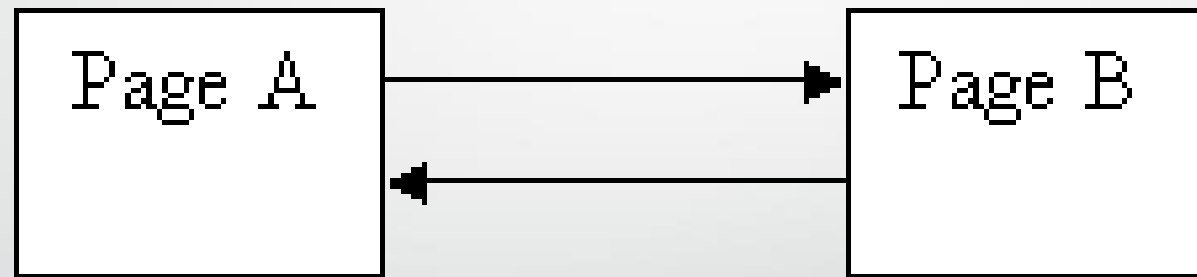
- Besucher einer Website folgt einem beliebigen Link ohne Berücksichtigung des Inhalts
- Langweilt sich nach kurzer Zeit
- Aufrufen einer beliebige Website

Algorithmus

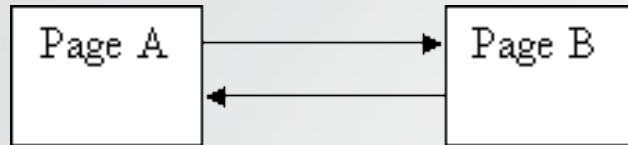
$$PR(A) = (1 - d) + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right)$$

- $PR(T_i)$ PageRank der Seiten T_i , die einen Link zu Seite A haben
- $C(T_i)$ Anzahl der Links von der Website T_i
- d Dämpfungsfaktor, es gilt $0 \leq d \leq 1$
- $(1 - d)$ Verhalten des Random Surfers

Berechnung des PageRank

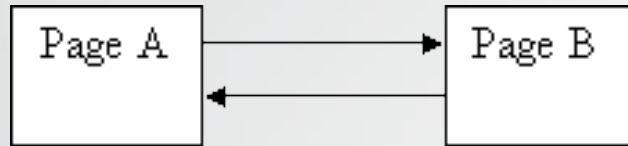


Berechnung des PageRank



$$PR(A) = (1 - d) + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right)$$

Berechnung des PageRank

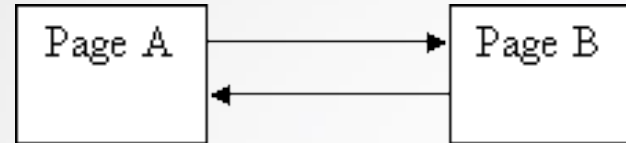


$$PR(A) = (1 - d) + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right)$$

$$PR(A) = (1 - d) + d \left(\frac{PR(B)}{1} \right)$$

$$PR(B) = (1 - d) + d \left(\frac{PR(A)}{1} \right)$$

Berechnung des PageRank

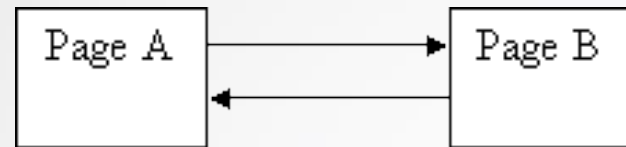


$$d = 0,85$$

$$PR(A) = (1 - 0,85) + 0,85 \left(\frac{PR(B)}{1} \right)$$

$$PR(B) = (1 - 0,85) + 0,85 \left(\frac{PR(A)}{1} \right)$$

Berechnung des PageRank

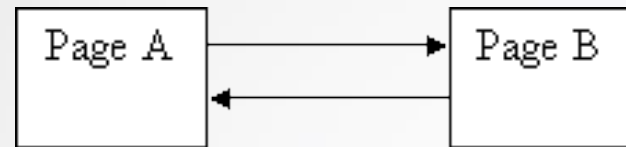


$$d = 0,85$$
$$PR(A) = 0$$

$$PR(A) = (1 - 0,85) + 0,85 \left(\frac{PR(B)}{1} \right)$$

$$PR(B) = (1 - 0,85) + 0,85 \left(\frac{PR(A)}{1} \right)$$

Berechnung des PageRank



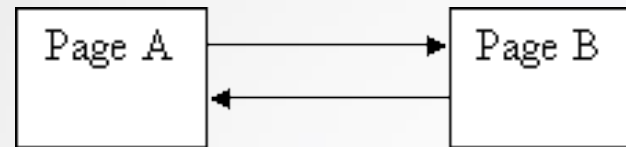
$$d = 0,85$$

$$PR(A) = 0$$

$$PR(A) = (1 - 0,85) + 0,85 * (0)$$

$$PR(B) = (1 - 0,85) + 0,85 * \left(\frac{PR(A)}{1} \right)$$

Berechnung des PageRank



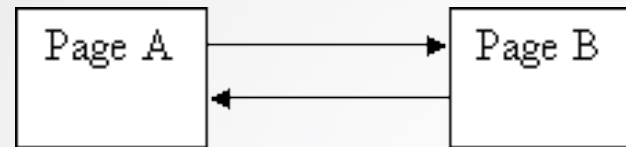
$$d = 0,85$$

$$PR(A) = 0$$

$$PR(A) = (1 - 0,85) + 0,85 * (0) = 0,15$$

$$PR(B) = (1 - 0,85) + 0,85 * \left(\frac{PR(A)}{1} \right)$$

Berechnung des PageRank

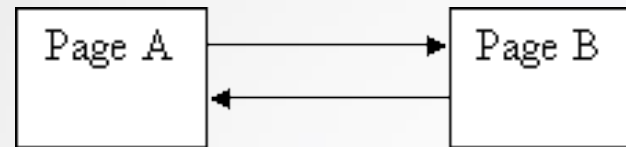


$$d = 0,85$$

$$PR(A) = (1 - 0,85) + 0,85 * (0) = 0,15$$

$$PR(B) = (1 - 0,85) + 0,85 * (0,15) = 0,2775$$

Berechnung des PageRank



$$d = 0,85$$

$$PR(B) = 0,2775$$

$$PR(A) = (1 - 0,85) + 0,85 * (0,2775) = 0,385875$$

$$PR(B) = (1 - 0,85) + 0,85 * (0,385875) = 0,47799375$$

Erweiterter Algorithmus

$$PR(A) = \frac{(1 - d)}{N} + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right)$$

- $PR(T_i)$ PageRank der Seiten T_i , die einen Link zu Seite A haben
- $C(T_i)$ Anzahl der Links von der Website T_i
- d Dämpfungsfaktor, es gilt $0 \leq d \leq 1$
- N Anzahl aller Websites

Vergleich

Algorithmus:

- Erwartungswert für die Auswahl einer Website

Erweiterter Algorithmus:

- Tatsächliche Wahrscheinlichkeit

Verwendung außerhalb von Suchmaschinen

- Biologie:
 - Analyse von Proteinnetzwerken
 - Artensterben
- Twitter: Freundschaftsvorschläge

Google Bomb

- Manipulation des PageRanks
- Google-Suche bewertet Seiten auch anhand von Texten von Links, die auf die Seite verweisen.
- In den Suchergebnissen erscheinen auch Seiten die mit einem Text verlinkt wurden, der der Suchanfrage entspricht.
- 1999: 1. Google Bombe
- "more evil than satan himself" wurde mit Microsoft verlinkt
- Viele Google Bomben wurden manuell entschärft

Diskussion

- Was könnte Google dazu bewegt haben den Suchalgorithmus zu ändern?
- Was könnten weitere Kriterien sein um Suchergebnisse zu verbessern?
- Sind Google Bomben eurer Meinung nach legitim?
- Inwiefern hat sich dieser Algorithmus auf die Gesellschaft ausgewirkt?

Quellen

- <https://en.wikipedia.org/wiki/PageRank> (Grafik Folie 5)
- [https://en.wikipedia.org/wiki/Ranking_\(information_retrieval\)](https://en.wikipedia.org/wiki/Ranking_(information_retrieval))
- <http://pr.efactory.de/e-pagerank-algorithm.shtml>
- <https://de.wikipedia.org/wiki/Google-Bombe>
- <http://nlp.stanford.edu/IR-book/html/htmledition/pagerank-1.html>
- <https://www.youtube.com/watch?v=ly5zHY7DI7o>
- <https://web.eecs.umich.edu/~michjc/eecs584/notes/lecture19-pagerank.ppt>
- <http://www.cs.princeton.edu/~chazelle/courses/BIB/pagerank.htm> (Grafik Folie 8)