



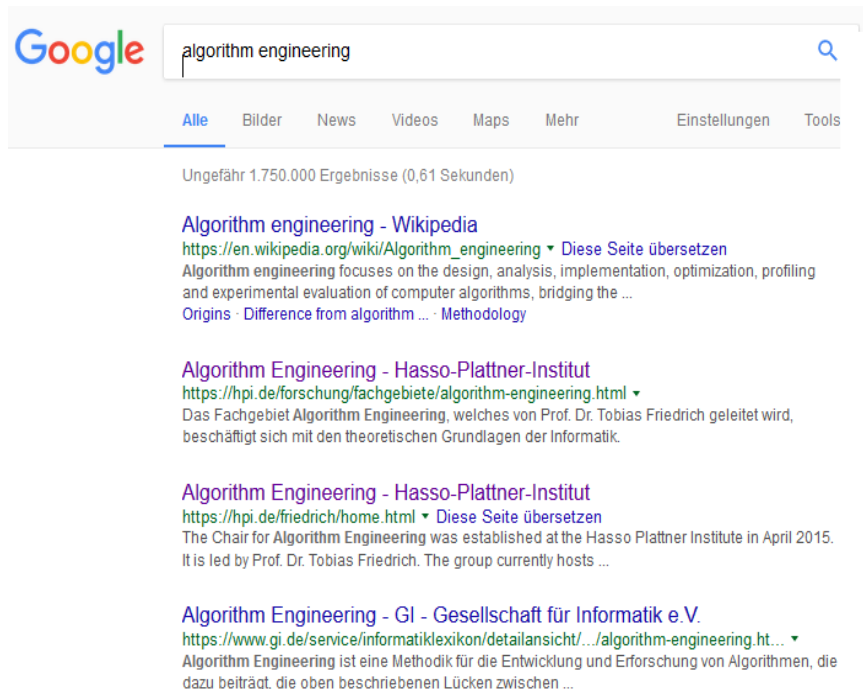
openHPI  
Einführung in die Mathematik der Algorithmik  
– PageRank-Intro

**Thomas Bläsius, Karen Seidel**

Hasso-Plattner-Institut

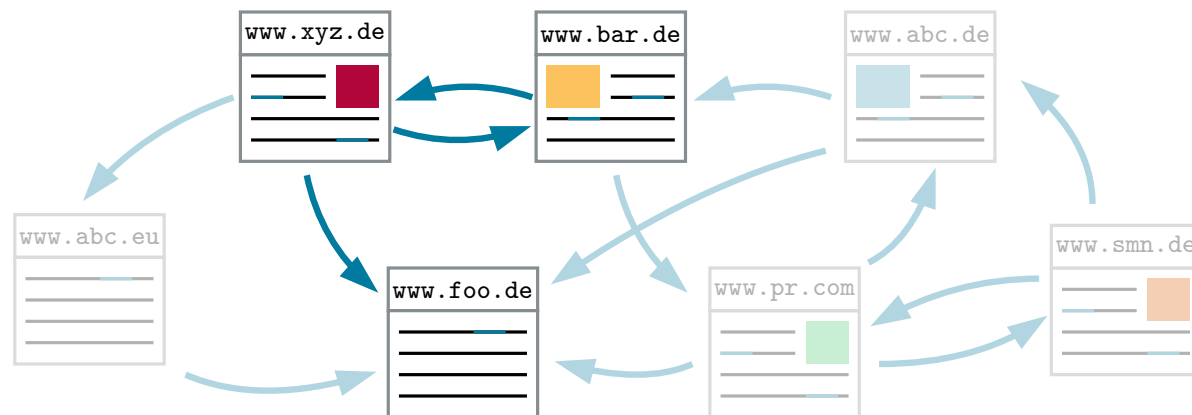
Universität Potsdam

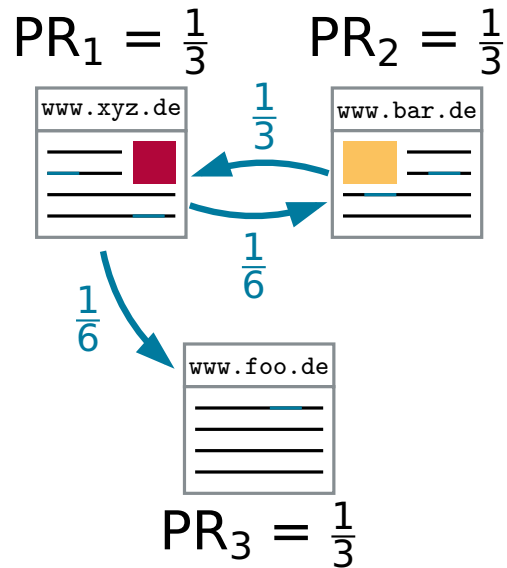
- **Informationen** (wieder)zu**finden** ist schon immer eine Herausforderung
- erste computergestützte Systeme zur Verwaltung von **großen Datenmengen** wurden ausschließlich von speziell dafür ausgebildeten Personen bedient
- Eine **Internetrecherche** wirft viele Fragen auf, beispielsweise:



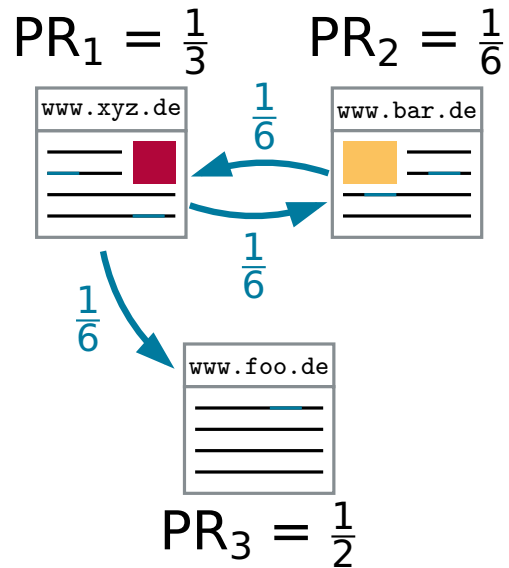
- Wie gestaltet die suchende Person im Idealfall ihre **Suchanfrage**?
- Welche der Milliarden Webseiten sind **passend** für die Suchanfrage?
- In welcher **Reihenfolge** sollen die inhaltlich passenden Webseiten angezeigt werden?

- **Googles Kernalgorithmus** wurde von Sergey Brin und Lawrence Page and der Universität Stanford entwickelt und 1998 veröffentlicht
- Googles aktueller Suchalgorithmus ist ein Betriebsgeheimnis und enthält neben PageRank **auch andere Bestandteile**
- Beispiele für weitere Suchalgorithmen sind Hubs and Authorities (HITS), **Hilltop** und TrustRank
- Grundidee von PageRank ist es, die Wichtigkeit einer Seite allein aufgrund der **Hyperlink-Struktur des WWW** zu ermitteln

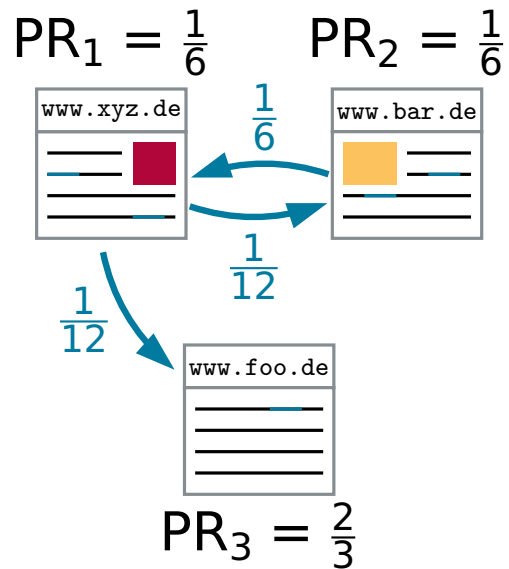




|   | $PR_1$ | $PR_2$ | $PR_3$ |
|---|--------|--------|--------|
| 0 | 0,3333 | 0,3333 | 0,3333 |

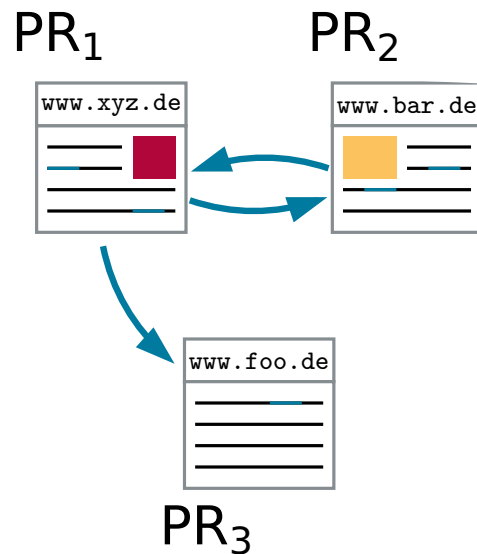


|   | $PR_1$ | $PR_2$ | $PR_3$ |
|---|--------|--------|--------|
| 0 | 0,3333 | 0,3333 | 0,3333 |
| 1 | 0,3333 | 0,1667 | 0,5000 |



|   | $PR_1$ | $PR_2$ | $PR_3$ |
|---|--------|--------|--------|
| 0 | 0,3333 | 0,3333 | 0,3333 |
| 1 | 0,3333 | 0,1667 | 0,5000 |
| 2 | 0,1667 | 0,1667 | 0,6667 |

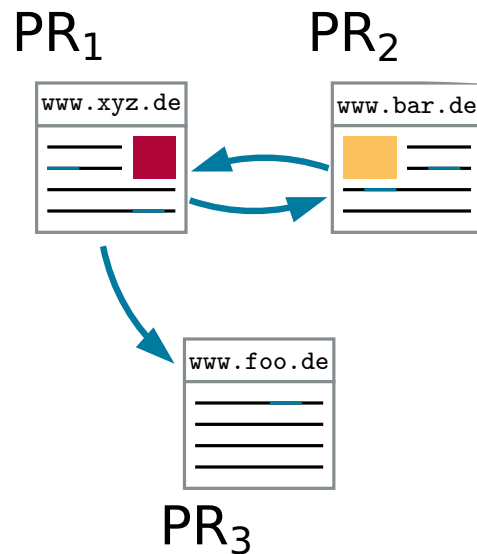




|          | $PR_1$   | $PR_2$   | $PR_3$   |
|----------|----------|----------|----------|
| 0        | 0,3333   | 0,3333   | 0,3333   |
| 1        | 0,3333   | 0,1667   | 0,5000   |
| 2        | 0,1667   | 0,1667   | 0,6667   |
| 3        | 0,1667   | 0,0833   | 0,7500   |
| 4        | 0,0833   | 0,0833   | 0,8333   |
| 5        | 0,0833   | 0,0471   | 0,8750   |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 28       | 0,0000   | 0,0000   | 1,0000   |

## PageRank-Basisalgorithmus

- Weise jeder der  $n$  Webseiten den **Anfangswert**  $\frac{1}{n}$  zu.
- Wende wiederholt folgende **Aktualisierungsvorschrift** an:
  - **Verteile** für jede Webseite ihren **aktuellen PageRank-Wert gleichmäßig** auf alle Webseiten, auf die sie verlinkt.  
(Webseiten ohne ausgehende Links behalten ihren aktuellen PageRank-Wert.)
  - **Addiere** für jede Webseite alle erhaltenen **Werte**.



|          | $PR_1$   | $PR_2$   | $PR_3$   |
|----------|----------|----------|----------|
| 0        | 0,3333   | 0,3333   | 0,3333   |
| 1        | 0,3333   | 0,1917   | 0,4750   |
| 2        | 0,2129   | 0,1917   | 0,5954   |
| 3        | 0,2129   | 0,1405   | 0,6466   |
| 4        | 0,1694   | 0,1405   | 0,6901   |
| 5        | 0,1694   | 0,1220   | 0,7086   |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 18       | 0,1448   | 0,1116   | 0,7436   |

- **Idee:** Gewichte das Ergebnis des Basisalgorithmus mit 85% und verteile die verbleibenden 15% gleichmäßig auf alle Seiten.

- **Beispielrechnung:**

- $PR_1 = \frac{1}{3} \cdot 0,85 + \frac{0,15}{3} = \frac{1}{3} \approx 0,3333$

- $PR_2 = \frac{1}{6} \cdot 0,85 + \frac{0,15}{3} = \frac{23}{120} \approx 0,1917$

- $PR_3 = \left(\frac{1}{6} + \frac{1}{3}\right) \cdot 0,85 + \frac{0,15}{3} = \frac{19}{40} \approx 0,4750$



## Skalierter Basisalgorithmus mit Dämpfung $d = 0,85$

- Weise jeder der  $n$  Webseiten den Anfangswert  $\frac{1}{n}$  zu.
- Wende wiederholt folgende Aktualisierungsvorschrift an:
  - Verteile für jede Webseite ihren aktuellen PageRank-Wert gleichmäßig auf alle Webseiten, auf die sie verlinkt.  
(Webseiten ohne ausgehende Links behalten ihren aktuellen PageRank-Wert.)
  - Addiere für jede Webseite alle erhaltenen Werte und multipliziere das Ergebnis mit dem Dämpfungsfaktor  $d$ .
  - Verteile  $1 - d = 0,15$  gleichmäßig auf alle Webseiten des Netzwerks.

## Warum führt der Algorithmus zum Ziel?