

## Gruppe 3 / 5

# Tutorübung zu Einsatz und Realisierung von Datenbanksystemen (SS 17)

Michael Schwarz

Institut für Informatik  
Technische Universität München

26.07 / 27.07.2017

## Zentralübung

27.07 statt der Vorlesung

# Information Retrieval

- Informationsexplosion im Internet
- Ranking von Dokumenten um relevante Information zu finden
- Ähnlichkeit von Dokumenten (Dissertationen) zu erkennen

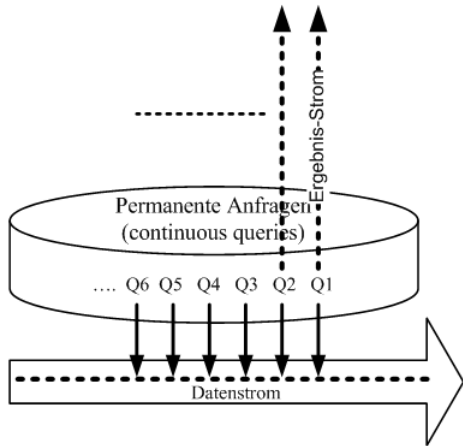
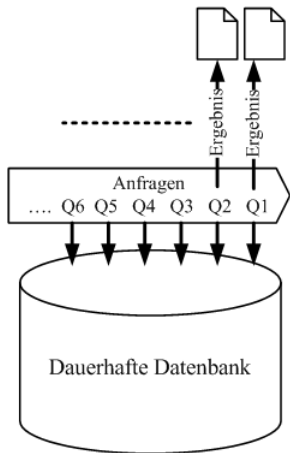
# TF-IDF: Term Frequency – Inverse Document Frequency

$$TF_{ij} = f_{ij} / \sum_{i=1 \dots |V|} f_{ij}$$

$$IDF_i = \log(N/n_i)$$

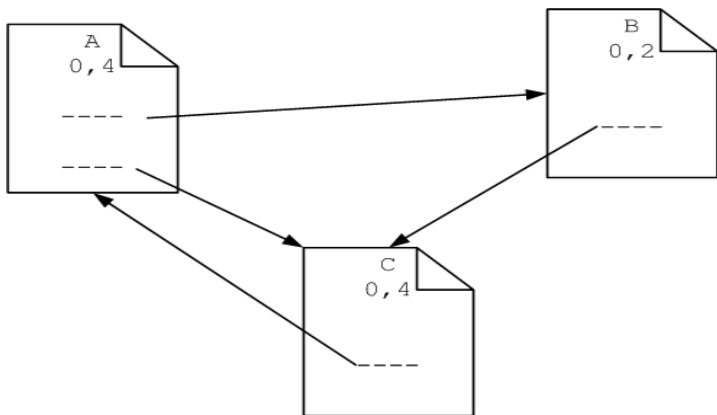
$$rel(D_j, Q) = \sum_{i \in Q} TF_{ij} * IDF_i$$

# Datenbank versus Datenstrom



# Page Rank: Grundidee

$$r(A) = \frac{\alpha}{N} + (1 - \alpha) \left( \frac{r(B_1)}{|B_1|} + \dots + \frac{r(B_n)}{|B_n|} \right)$$

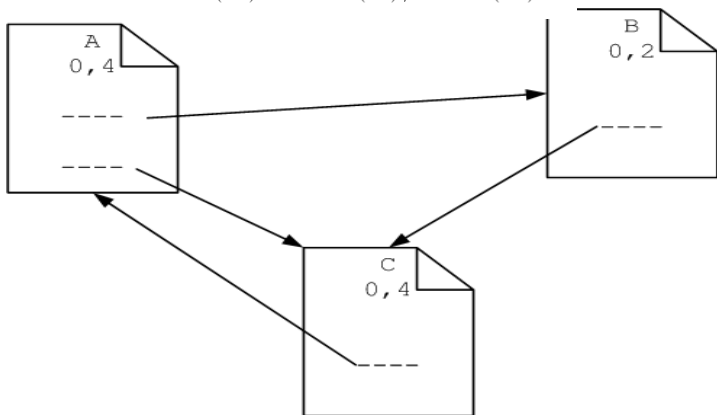


# Page Rank: Grundidee

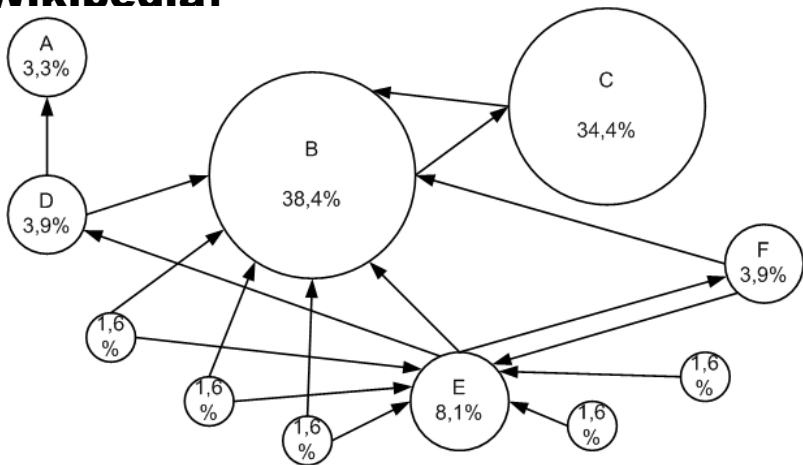
$$r(A) = r(C)/1$$

$$r(B) = r(A)/2$$

$$r(C) = r(A)/2 + r(B)$$



# PageRank für größeren Graph [aus Wikipedia]





## Idee

- Hubs: Verlinken viele inhaltlich wertvolle Dokumente
- Authorities: Seiten mit wertvollem Inhalt

Bewerte jede Seite nach beiden Kategorien.

## Idee

Der Algorithmus nimmt an:

- Gute Hubs haben Links zu vielen Authorities
- Gute Authorities sind von vielen Hubs verlinkt

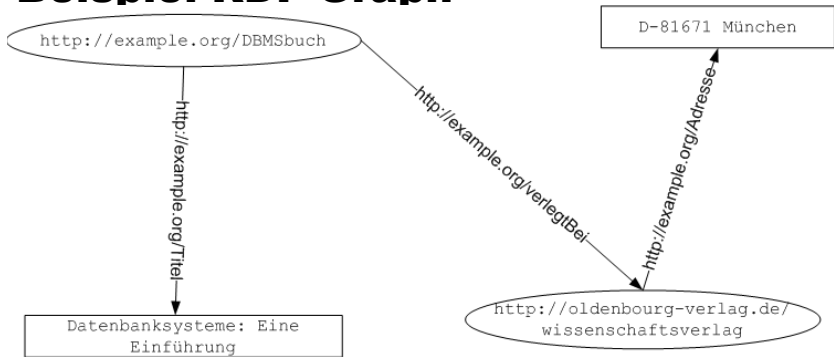
# Algorithmus

1. Berechne die Hub-Werte jeder Seite  $q$  indem man die Summe der Autoritätswerte aller Seiten  $r$  ermittelt, auf die  $q$  verweist.
2. Berechne die Autorität der Seite  $p$  durch Summierung der Hub-Werte der Seiten  $q$ , die auf  $p$  verweisen
3. Normalisiere die so erhaltenen Autoritätswerte indem man sie mit  $\lambda = 1/\max$  multipliziert, wobei  $\max$  den Maximalwert aller gerade neu berechneten Autoritätswerte darstellt. Diese Normalisierung ist nötig, um die Werte nicht ins Unermessliche steigen zu lassen.

# Semantic Web: Resource Description Framework (RDF)

- Triple-Datenmodell
  - (Subjekt, Prädikat, Objekt)
- Meist graphische Visualisierung
  - Subjekte und Objekte sind Knoten
  - Prädikate sind gerichtete Kanten
    - Von Subjekt-Knoten nach Objekt Knoten

# Beispiel-RDF-Graph



# SPARQL: Die RDF Anfragesprache

```
SELECT ?Var1 ?Var2 ... $VarN  
WHERE {Muster1. Muster2. ... MusterM. }
```

```
PREFIX ex: <http://www.example.org>
```

```
SELECT ?AutorenDesOldenbourgVerlags WHERE  
{ ?buch ex:Autor ?a.  
  ?a ex:NachName ?AutorenDesOldenbourgVerlags.  
  ?buch ex:verlegtBei <http://oldenbourg-verlag.de/  
  wissenschaftsverlag>.  
}
```

---

## Feedback

- Was war gut?
- Was sollte ich das nächste Mal anders machen?
- Anregungen, Kommentare, ...

Einfach anonym auf die Zettel

## Fragen

- Entweder jetzt gleich
- Oder: Per E-Mail ([m.schwarz@tum.de](mailto:m.schwarz@tum.de))  
(Deadline für garantierte Antwort: 04.08.2017 23:30h)

# Feuer frei!