



Übung zur Vorlesung *Einsatz und Realisierung von Datenbanken im SoSe25*

Alice Rey, Maximilian Reif, Tobias Goetz (i3erdb@in.tum.de)

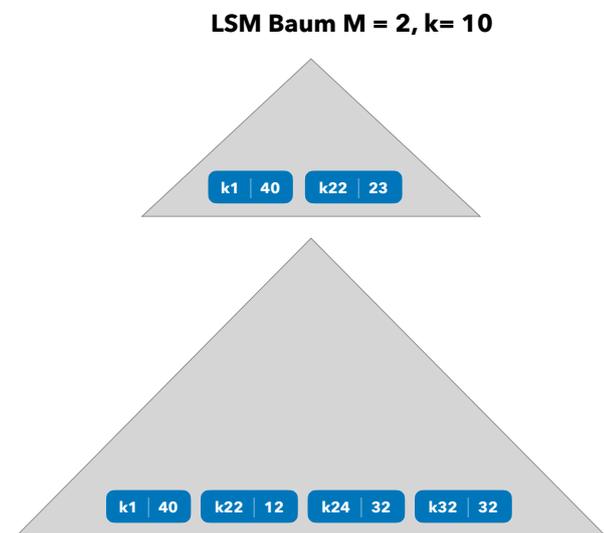
<http://db.in.tum.de/teaching/ss25/impldb/>

Blatt Nr. 11

Hinweise Für die aktive Teilnahme an der Vorlesung am 24. Juli benötigen Sie Spark auf Ihrem Rechner installiert. Eine Anleitung finden Sie unter https://db.in.tum.de/teaching/ss24/impldb/Spark_preparation.pdf. Falls Sie Schwierigkeiten haben, wenden Sie sich bitte in dieser Übungswoche an Ihren Tutor.

Hausaufgabe 1

Erklären Sie, wie LSM-Bäume das Problem der Write-Amplification beheben:



- Lesen sie k22 aus dem abgebildeten LSM-Baum
- Setzen sie k22 auf 33
- Setzen sie k34 auf 19
- Setzen sie k1 auf 20

Hausaufgabe 2

In dem in Abbildung 1 gezeigten Netzwerk von Web-Seiten wird ein kleines Beispiel für einen Webgraphen gezeigt. Lösen Sie folgende Aufgaben.

1. Berechnen Sie, für das in Abbildung gezeigte Netzwerk, den PageRank, sowie die HITS-Werte nach 2 Iterationen. Nutzen Sie $1/|V|$ als Anfangswert für den PageRank und 1 für HITS. $a = 0.1$

2. Formulieren sie eine Iteration des Pagerank Algorithmus in SQL. Der Graph ist dabei in der Tabelle $edges(src,dst)$ gespeichert, die aktuelle PageRank Gewichtung in der Tabelle $pagerank(node,pr)$.
3. Formulieren Sie die SQL Anfrage nun als rekursive SQL Anfrage (100 Iterationen) um.

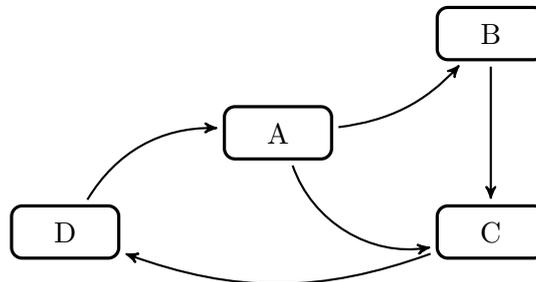


Abbildung 1: Ein kleiner Webgraph.

Hausaufgabe 3

Berechnen Sie für folgende drei Dokumente die TF-IDF-Werte:

1. „Beim Fußball dauert ein Spiel neunzig Minuten – und am Ende gewinnen die Deutschen“
2. „Beim Fußball muss das Runde (der Ball) in das Eckige (das Tor)“
3. „Nie war ein Tor so wertvoll wie jetzt“

Welches Ranking ergibt sich gemäß der Relevanzwerte für die Anfrage: „Fußball“ \wedge „Tor“. Zur Ermittlung des TF Wertes gehen sie davon aus, dass alle Wörter eines Dokuments *interessant* sind?

Hausaufgabe 4

Vervollständigen Sie die untere Anfrage um die Namen der Freunde von Personen mit dem Vornamen *Sokrates* zu finden, die älter als 30 Jahre sind. Die *foaf* Ontology is unter <http://xmlns.com/foaf/spec/> beschrieben. Nutzen Sie <https://rdf.db.in.tum.de/> für Ihre Abfrage.

```

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name2
WHERE {
    ....
}
  
```

Hausaufgabe 5

```

@prefix ex: <http://example.org>.
ex:Rapunzel ex:hatAutor ex:Sokrates.
ex:Rapunzel ex:erschiene 2006.
ex:Aschenputtel ex:hatAutor ex:Archimedes.
ex:Aschenputtel ex:hatAutor ex:Platon.
ex:Schneewittchen ex:hatAutor ex:Platon.
ex:Schneewittchen ex:erschiene 2004.
  
```

Drücken Sie die folgenden Anfragen in SPARQL aus:

1. Geben Sie alle Bücher aus, für die sowohl der Autor als auch das Erscheinungsjahr in der Datenbank enthalten sind.
2. Geben Sie die gemeinsamen Autoren der beiden Bücher Aschenputtel und Schneewittchen aus.
3. Geben Sie die Namen aller Autoren (ohne Duplikate) von Büchern mit einem Erscheinungsjahr nach 2004 aus.

Hausaufgabe 6

wikidata.org ist ein Projekt, das strukturierte Informationen für Wikimedia-Schwesterprojekte bereitstellt. Informationen über das Datenmodell finden Sie unter <https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer>. Praktischerweise bietet es auch eine SPARQL-Schnittstelle für Ihre Erkundungen unter query.wikidata.org.

Schreiben Sie SPARQL-Abfragen, um die folgenden Fragen zu beantworten:

1. Listen Sie alles auf, was München als Objekt verwendet. Wikidata hat den URI <http://www.wikidata.org/entity/Q1726> für München vergeben. Daher können Sie unter Verwendung einer Präfixdefinition auf München verweisen, indem Sie `wd:Q1726` verwenden.
2. Welche Prädikat wird am häufigsten verwendet?
3. Welche der Städte in der Datenbank hat die früheste schriftliche Aufzeichnung?
4. Führen Sie die Unterklassen von Sport (Q349) und ihre Bezeichnungen auf, falls es eine gibt.
5. Listen Sie die transitiven Unterklassen von Sport auf.