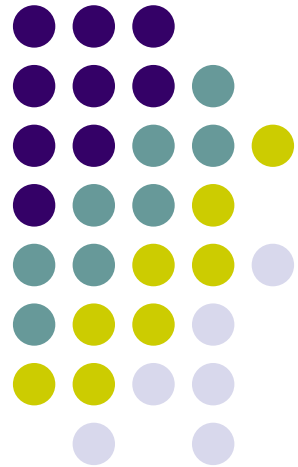
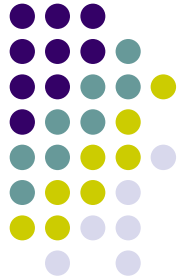
(Team-)Projekt: Entwicklung eines RDF Stores

SS 08

Betreuer: Thomas Hornung, Michael Schmidt



(Team-)Projekt-Anforderungen

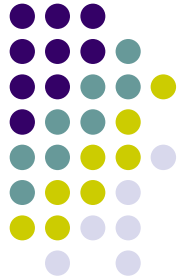


- Bachelor (6 ECTS) [180 Stunden i.S. ~ 12.5 h/Woche]
  - Softwareentwicklung (Lasten-,Pflichtenheft,..)
  - Gruppengröße: 2 – 5
  - Protokoll, Referat, Rechnerpräsentation

- Master/Diplom (16 ECTS/15 KP) [480 Stunden i.S. ~ 34h/Woche]
  - Wissenschaftliche Methoden zur Problemlösung
  - Eigenständige Themeneinarbeitung anhand von Fachliteratur
  - Gruppengröße: 2 – 5
  - Schriftlich je Gruppenmitglied: 15 – 25 Seiten
  - Mündliche Präsentation: je Mitglied 15 Minuten

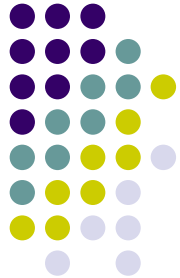
ANMELDUNG DER PROJEKTTEILNAHME
BEIM PRÜFUNGSAMT IST
ERFORDERLICH UND VERBINDLICH

Organisation



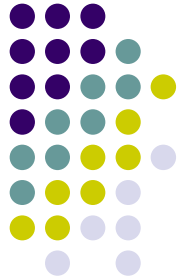
- Zeit & Ort:
  - Dienstag 14 – 18 Uhr (s.t.)
  - Raum 00-006, Geb. 051
- 2-wöchige Präsentation des Fortschritts
- (Selbständige) Einteilung des Projektes in Milestones
- Abschließende Evaluierung m.H. öffentlicher und „geheimer“ Test Cases

Aufgabenstellung



- Implementierung eines RDF Stores
  - Einlesen von RDF Daten
    - Vorgegebene Indizes/physikalische Sortierung
  - Statistiken
  - Anfrage API (Datenzugriff)

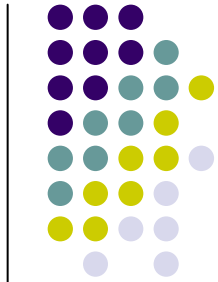
Einlesen von RDF Daten



- Speichern in Berkeley DB
- Effizientes Anlegen der Indizes
- Dictionary Encoding
- [Kompression]

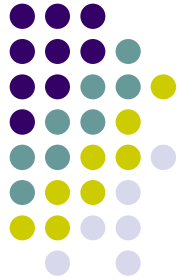
!!! MEMORY BOUNDED !!!

Vorgegebene Indizes



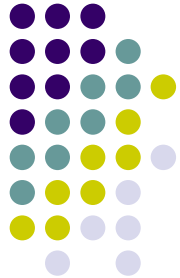
- 6fach redundantes Speichern der RDF-Tripel
  - Subject – Predicate – Object
  - Subject – Object – Predicate
  - Predicate – Subject – Object
  - Predicate – Object – Subject
  - Object – Subject – Predicate
  - Object – Predicate– Subject
- ***Vorteil: effiziente Unterstützung verschiedenster Zugriffsmethoden***

Statistiken



- Global
  - #dist(URI)
  - #dist(Literals)
  - #dist(Blanks)
- URI-Selektivität
- Weitere RDF-spezifische Statistiken (z.B. rdf:type)
- **Wahl geeigneter Zugriffsmethoden anhand vorhandener Statistiken**

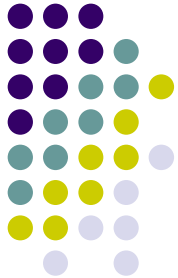
Anfrage API (Datenzugriff)



- Triple Patterns mit Variablen
- AND: Joins von Ergebnissen
- UNION: Vereinigung von Ergebnissen
- FILTER (String, Int; <, <=, etc.)
- [OPTIONAL]

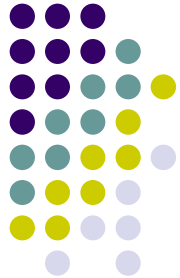
!!! MEMORY BOUNDED !!!

Teamarbeit



- **ENTSCHEIDEND** für das Gelingen des Projekts
- Einsatzbereitschaft
  - Vertrauen + Zuverlässigkeit
  - Organisation
    - Organisationsstruktur: Aufgabenverteilung und Verantwortungsbereiche abgrenzen
    - Regelmässige Treffen (Meetings)
    - Regelmässiger Austausch
    - Gegenseitige Hilfe/Planung bei kritischen Aufgaben

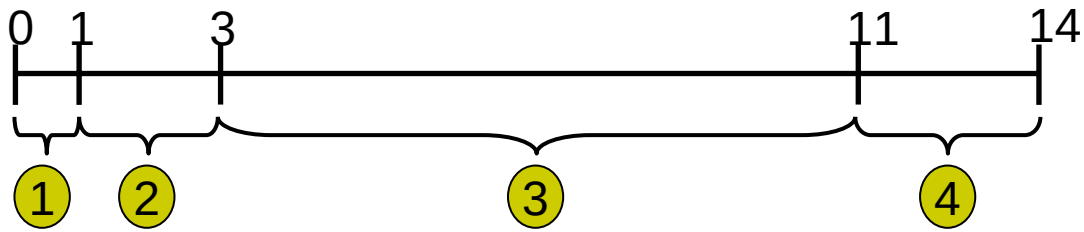
Effiziente Besprechungen



- Moderatoren bestimmen
- Pünktlich anfangen
- Tagesordnung
  - Stand der Dinge
  - Ziele
  - Umsetzung
  - Nächste Schritte
  - Verschiedenes

Zeitplanung

1. Interne Aufteilung
   - Wer macht was?
   - Projektleiter?
   - Einlesen

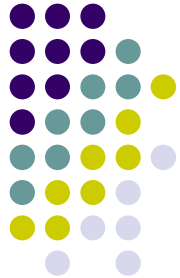


2. Lasten-/Pflichtenheft
   - Funktionalitätsübersicht
   - Klassendiagramm
   - Beschreibung Zusammenhänge
   - Selbständiges Festlegen der Milestones für Periode 3

3. Selbständige Arbeit und Präsentation der Milestones, Ende der Periode: Erste Korrekheitstests, Abgabe einer ersten voll-funktionsfähigen Version

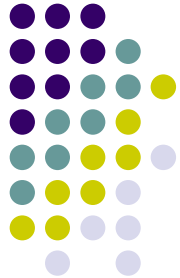
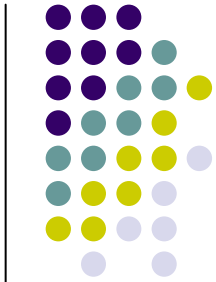
4. Zeit für Optimierung ☺

Material (Links s. Homepage)



- SVN Repository
- Oracle Berkeley DB (Java Edition)
- RDF Parser
  - Format: RDF/XML, N3
- Literatur
  - YARS
  - SPARQL Tutorial
  - Triple Pattern Semantik

Bis nächsten Dienstag…



1. Interne Aufteilung
   - Wer macht was? (Grobeinteilung)
   - Projektleiter?
2. Lesen /aller/ verlinkten Dokumente auf der Homepage