

XML Projekt: Web Data Browser

Meilensteinpräsentation

Proxy Server & Triplestore

Grundidee:

- Transparent
- Abfangen von bestimmten URLs
- Weiterleiten an Parser
- HTTP-Server für Datenchronik

Triplestore:

- 4store: RDF-Database, gut skalierbar, SPARQL-Unterstützung
- Python-Wrapper zur Nutzung von 4store

Nächste Schritte

- Einbinden weiterer Parser
- Triplestore einbinden
- Visualisierung der Chronik

Parser

Microdata

- Für Stackoverflow
- Übergabe des Dokumentes vom Proxy
- Durchgehen des Dokumentes Zeile für Zeile
- Pattern Matching auf Microdata Content
- Zusammenbauen des RDF Files
- Rückgabe an Proxy Server

Parser

JSON

- Für Twitter
- Übergabe der URL vom Proxy
- GET-Anfrage an einen Twitter-Feed
→ aktuellster Tweet wird in RDF geparsed
- Rückgabe des RDF-Dokuments an den Proxy
- Noch nicht integriert