

BaseX

KIM / HOS Expertenworkshop "Tools für das Metadatenmanagement" an
der SUB Göttingen

12.03.2020

J. Beck - Goethe Universität Frankfurt

Vorstellung des Tools - BaseX

- Open Source **XML-Datenbank** mit XQuery Prozessor
- erlaubt die Speicherung, Abfrage und Manipulation **großer Textdateien** (XML/JSON/CSV/...)
- enthält einen **XQuery Editor**, um XQuery in Echtzeit zu testen
- bietet verschiedenste **Visualisierungsmöglichkeiten** für die Analyse der Daten
- sehr ausführliche **Dokumentation** mit vielen Beispielen
- per **Kommandozeile** oder **GUI** zu bedienen

Für mehr Informationen zu XQuery und BaseX siehe auch Links (z.B. DH-Workshop von 2019)

Einschätzung des Tools (Einsatz innerhalb von GLAM)

- BaseX ist eine XML-Datenbank, kein(!) Metadatenmanagementtool...
- ... aber eignet sich für das Speichern, Manipulation und Validierung von großen* XML-Datenbeständen wie sie in Bibliotheken/Museen/Archiven typischerweise auftreten
- ermöglicht komplexe Abfragen auf großen Datenmengen mit XQuery (wie bei relationaler Datenbank mit SQL)

* <http://docs.basex.org/wiki/Statistics>

Einschätzung des Tools (Einsatz innerhalb von GLAM)

- sehr **mächtiges Tool** mit dem noch viel mehr möglich ist (z.B. Webapplikationen, REST-Schnittstelle, ...)
- für XQuery braucht es zumindest Programmiergrundkenntnisse und ist daher nur bedingt von nicht-technischem Personal einsetzbar

Einschätzung des Tools (Einsatz innerhalb von GLAM)

BaseX im Bibliothekskontext eignet sich vor allem für

- **gezielte Änderungen** von XML-Dokumenten z.B. bei Normalisierung oder Anreicherung...
- **Analyse** des Datenbestands (Ausreißer erkennen, Statistiken aufstellen, Abfrageergebnisse gruppieren und sortieren)
- Analyse von **Volltexten** (Anm.: nicht selbst getestet)
- **Validierung** von XML Daten mittels XML Schema
- **Visualisierung** des Datenbestands (Views: Map, Tree, Plot, Table, Folder, Explorer)

Einschätzung des Tools (Bezug zu den Workshop-Aufgaben)

- für die Bearbeitung der Daten waren weitere Tools nötig (MarcEdit für die Transformation von MARC ISO2709 nach MARCXML)
- Implementierung eines OAI-Harvesters wäre zwar theoretisch möglich, aber der Einsatz existierender Harvester ist sinnvoller
- das Tool eignet sich besonders gut für Daten mit Baumstruktur (v.a. XML, aber auch JSON), es ermöglicht aber auch das Lesen und Schreiben von CSV oder TEXT
- das Tool eignet sich besonders gut für die Analyse der Daten

Einschätzung des Tools (Bezug zu den Workshop-Aufgaben)

- die **Transformation** und **Bereinigung** der Daten ist mittels XQuery oder XSLT möglich. [XSLT kann übrigens in BaseX mittels `xslt:transform(input, stylesheet)` ausgeführt werden. In diesem Fall (Mapping aller Felder) eignet sich XSLT vermutlich besser.]
- eine **Serialisierung** in JSON ist möglich
- **xproc** mit calabash ermöglicht das Bauen von Pipelines (Anm.: nicht selbst getestet)

- BaseX
<http://basex.org/>
- BaseX Dokumentation
http://docs.basex.org/wiki/Main_Page
- Workshop DH zu XML-Manipulation in BaseX
https://github.com/j4lib/talks/blob/master/dh_workshop.pdf

Metropolis beamer theme by M. Vogelsang, cc-by-sa-4.0