

XML-Projekt SoSe 2012

GPSies.org, DBPedia.org, Twitter.com, XML/XSD/XSLT

Gruppe 11

Eike Cochu, Samer El-Safadi, Hannes Geist,
Cenk Gündogan, Michael Pluhatsch

7. Juli 2012

Inhaltsverzeichnis

1	Aufgabenstellung und Organisation	2
1.1	Aufgabenstellung	2
1.2	Aufgabenverteilung und -bewältigung	2
1.2.1	Hannes Geist	3
1.2.2	Cenk Gündogan	4
1.2.3	Eike Cochu	5
1.2.4	Samer El-Safadi	6
1.2.5	Michael Pluhatsch	7
2	Datenbank-Info	8

1 Aufgabenstellung und Organisation

Gruppenleiter: Hannes Geist

Dokumentierer: Eike Cochu

1.1 Aufgabenstellung

Durch den XML-Endpoints von gpsies.com 100.000 Wanderstrecken laden, per XSLT-Schema in ein eigens entwickeltes XSD-Schema transformieren und in einer geeigneten XML-Datenbank speichern. Zu diesen Strecken soll per SPARQL der Endpoint von dbpedia.org abgefragt und alle auf (oder an) der Strecke liegenden Sehenswürdigkeiten abgefragt werden können. Ein HTML-Formular (oder ähnliches) entwickeln, mit dem die XML-Datenbank abgefragt werden kann. Das Ergebnis der Abfrage soll wiederum per XSLT in HTML transformiert und zusammen mit den vorhandenen Sehenswürdigkeiten und den zu diesen Sehenswürdigkeiten vorhandenen Tweets von Twitter.com angezeigt werden.

1.2 Aufgabenverteilung und -bewältigung

- Hannes Geist: XSLT-Transformierungen und Füllen der XML-Datenbank
- Cenk Gündogan: Abfrage der Strecken von gpsies.org
- Eike Cochu: Erstellung des XSD-Schemas, Bereitstellung eines Servers, Dokumentation
- Samer El-Safadi: SPARQL-Abfrage der Sehenswürdigkeiten von dbpedia.org
- Michael Pluhatsch: HTML-Formular mit PHP, Abfrage der XML-Datenbank

1.2.1 Hannes Geist

1.2.2 Cenk Gündogan

1.2.3 Eike Cochu

Aufgabenbereich: Erstellung eines XSD-Schemas für die XML-Datenbank, Bereitstellung des Datenbank- und Applikationsservers sowie Installation der Datenbank, Einfügen der Inhalte in die Datenbank und Installation der Webseite, kleine kosmetische Änderungen an der Webseite und Formatierung.

Zum XSD-Schema: Nachdem wir uns mit einem kurzen Testlauf einige Anschauungsdaten von gpsies.org beschafft hatten, konnten wir basierend darauf grob die Anforderungen an das XSD-Schema entwerfen. Die Idee des Schemas sollte es sein, mehr mit Attributen zu arbeiten und so die Anzahl an Elementen zu verringern, damit die Speichergröße der resultierenden Dateien minimal gehalten wird. Dazu haben wir uns entschlossen, einige weniger wichtige Datenteile der crawl-Daten gar nicht erst mit in die Datenbank zu speichern, im Schema waren diese dann auch nicht vorgesehen. Die meisten Elemente sind mittels minOccurs=0 ebenfalls optional gehalten, Attribute sind per default optional.

Wir haben uns für BaseX als Datenbank entschieden, da dieses auf Java basiert und somit systemunabhängig installierbar und auch leicht zu handhaben ist. Die Datenbank wurde auf einem dedizierten Ubuntu 12.04 Server installiert, als Webserver wurde Apache + PHP 5 gewählt.

1.2.4 Samer El-Safadi

1.2.5 Michael Pluhatsch

Organisatorisches: Zur Gruppenkoordinierung hat sich unsere Gruppe jede Woche Sonntags beim Gruppenleiter Hannes Geist getroffen und das weitere Vorgehen besprochen.

2 Datenbank-Info

Verwendete Datenbanksoftware: BaseX XML-Datenbank Anzahl der vorhandenen Datensätze:
100.000