

Benutzerhandbuch

Webanwendung zur Extraktion von Teildatensätzen aus DBpedia

Robert Bielinski Marvin Hofer Till Nestler Hans Angermann
Christian Ernst Dominik Strohscheer Jonas Rebmann

Inhaltsverzeichnis

Einleitung	1
Installation	1
Benutzung	2
Hinweise zur Benutzung des Query-Managers	6

Einleitung

DBpedia ist ein Community-Projekt, welches sich mit der Sammlung strukturierter Daten im RDF Format aus der Web-Enzyklopädie Wikipedia beschäftigt. Die Datensätze wird mit Hilfe des SPARQL-Endpunktes von OpenLink Virtuoso zugänglich gemacht.

Dieses Softwareprodukt stellt eine Weboberfläche zur Verfügung welche auf Benutzerfreundliche Weise die Auswahl, Abfrage, Konsistenzfilterung und Auswertung strukturierter Daten aus DBpedia ermöglicht.

Installation

Für Hinweise zur Installation des Softwareproduktes öffnen Sie bitte das Dokument "Installationsanleitung" im Ordner "Dokumente".

Benutzung

Die Benutzung des Softwareproduktes erfolgt über die Webanwendung, welche man über einen Browser öffnen kann. Die URL zum Aufrufen der Webanwendung erhalten Sie vom ihrem Server-Administrator.

Der Querymanager

Um eigene Sparql-Anfragen an einen gewünschten Sparql-Endpunkt zu verschicken steht auf der Startseite der Webanwendung der Query-Manager zur Verfügung. Die unten zu sehende Grafik beschreibt die Funktionen der einzelnen Elemente und welcher Inhalt jeweils eingetragen werden muss.

The image shows a web form titled "Query". It has two input fields: "Endpoint" containing "http://dbpedia.org/sparql" and "Endpointlimit" containing "10000". Below these is a large empty text area. At the bottom right is a "Senden" button. Numbered callouts 1-4 point to the Endpoint field, the Endpointlimit field, the text area, and the Senden button respectively.

Abbildung 1: Querymanager

1. Im Feld "Endpoint" wird URL des anzufragenden Sparql-Endpunktes eingetragen. Standardmäßig steht an dieser Stelle der DBPedia-Endpunkt.
2. Wenn der Endpunkt ein Limit für die zurückgegebenen Tripel pro Anfrage besitzt, dann muss dieses hier eingetragen werden. Falls der Endpunkt kein Limit besitzt dann kann das Feld freigelassen werden oder eine Zahl >0 eingesetzt werden. Standardmäßig steht hier das Limit des DBPedia-Endpunkts.
3. Textfeld in welches die auszuführende Sparql-Anfrage eingetragen wird. Beachten sie die Hinweise unter dem Punkt "Hinweise zur Benutzung des Query-Managers" falls an dieser Stelle Probleme auftreten.
4. "Senden"-Button zum Abschicken der in "3" eingetragenen Anfrage an den unter "1" angegebenen Endpunkt unter Beachtung des Limits in "2". Das Abschicken der Anfrage leitet sie an die nächste Seite weiter.

Datensatzstatistiken und Filterauswahl

Nachdem die Anfrage abgeschickt wird landen Sie auf der nächsten Seite. Hier wird zunächst der erhaltene Datensatz inklusive einiger Statistiken angegeben und Ihnen die Möglichkeit zur Filterung gegeben. Alle Funktionen sind im unten zusehenden Bildern erklärt.

Dataset information

N-Triples for dataset: 2016-05-30-18-53-44-637985 (1)

from query: prefix dbo: <http://dbpedia.org/ontology/> prefix dbp: <http://dbpedia.org/property/>
select * where { ?film a dbo:Film. ?film dbp:runtime ?runtime.} Limit 100 (2)

Propertys and datatypes

(3)

Filter Datatypes

Property

- <http://dbpedia.org/property/isbn> (4)
- <http://dbpedia.org/property/pages>

Datatype direct input (5)

Remove duplicates (6) (7)

Abbildung 2: Datensatzinformationen und Datentypauswahl

1. Name des erhaltenen Datensatzes und Button zur Anzeige des erhaltenen Datensatzes. Bei Betätigung wird ein neuer Tab aufgerufen. Im Abschnitt “Datensatzanzeige” erfahren Sie mehr dazu.
2. Die im Querymanager ausgeführte Sparql-Anfrage wird in diesem Feld erneut dargestellt.
3. Bei Betätigung dieses Knopfes werden Statistiken zu den Properties angezeigt, mehr dazu in der zweiten Grafik im Abschnitt “Datentypstatistiken und Konsistenzfilterung”
4. Hier kann eine Attribut ausgewählt werden, nachdem gefiltert werden soll.
5. Dieses Feld steht zur direkten Eingabe der Datentypen, nach denen gefiltert werden soll, zur Verfügung.
6. Wenn sie ein Häkchen in dieses Feld setzen, dann werden bei der Filterung der Daten auch noch zusätzlich alle redundanten Datensätze gelöscht.
7. Die Betätigung des “Filter”-Buttons sorgt dafür, dass das in “4” ausgewählte Attribut nach dem in “5” eingetragenen Datentypen gefiltert wird und Sie werden zur Ergebnisseite weitergeleitet.

Sollten Sie den Button in “3” betätigt haben, dann werden Sie folgende veränderte Anzeige der Attribute und Datentypen erhalten.

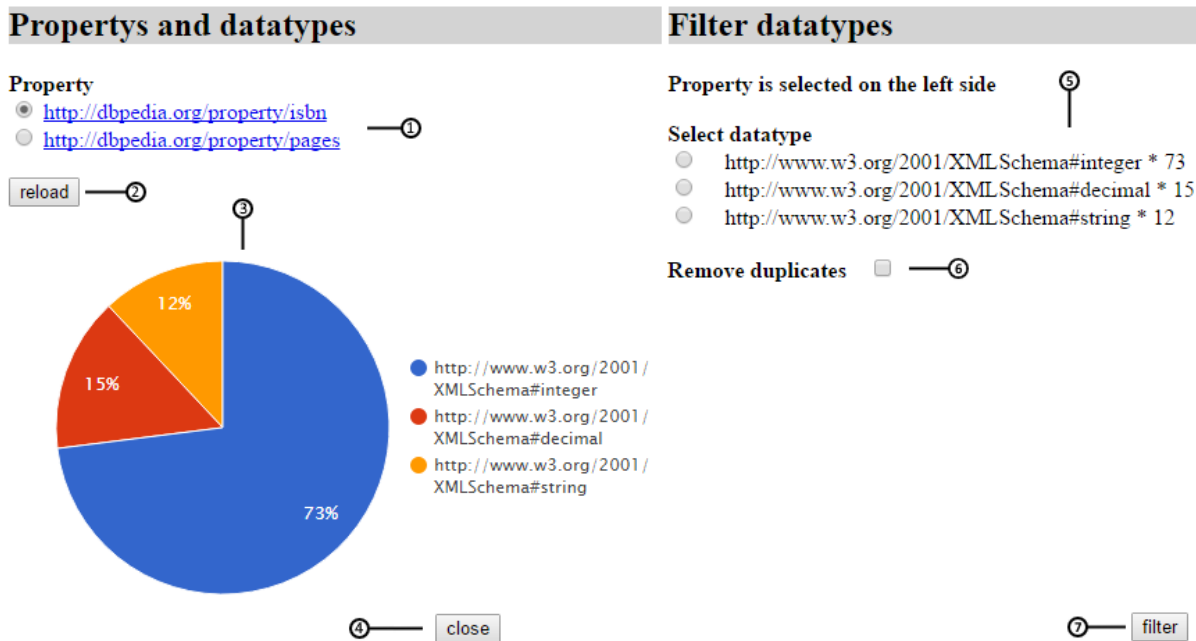


Abbildung 3: Statistik und Filterung

- Hier wird das Attribut gewählt, zu welchem alle auftretenden Datentypen angezeigt werden sollen und nach dem gefiltert werden soll.
- Um die Statistiken zu einem anderen Attribut anzeigen zu lassen muss dieser “Reload”-Button betätigt werden, damit sich die Anzeige aktualisiert.
- In dem Kuchendiagramm werden alle Datentypen zum in “1” gewählten Datentyp angezeigt, nachdem der Button in “2” betätigt wurde.
- Sie können den “Close”-Button betätigen um das Anzeige der Statistiken wieder zu schließen und kehren zur ersten Ansicht zurück.
- Hier werden alle auftretenden Datentypen inklusive deren Quantität des in “1” ausgewählten Attributs angezeigt, nachdem der Button in “2” betätigt wurde.
- Dieser Button hat die gleiche Funktion wie in der ersten Ansicht. Wenn sie ein Häkchen in dieses Feld setzen, dann werden bei der Filterung der Daten auch noch zusätzlich alle redundanten Datensätze gelöscht.
- Die Betätigung des “Filter”-Buttons sorgt hier dafür, dass das in “1” ausgewählte Attribut nach dem in “5” ausgewählten Datentyp gefiltert wird und Sie werden zur Ergebnisseite weitergeleitet.

Datensatzanzeige

Wenn Sie sich vor oder nach der Filterung nach Datentypen den Datensatz anzeigen lassen, dann werden die Daten auf folgender Seite präsentiert.

Show N-Triples

from 1 to 200 ①

Index	Subject	Predicate	Object
0000001:	http://dbpedia.org/resource/Into_the_River	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/Book
0000002:	http://dbpedia.org/resource/Into_the_River	http://dbpedia.org/property/isbn	9780473205089.0
0000003:	http://dbpedia.org/resource/Into_the_River	http://dbpedia.org/property/pages	279
0000004:	http://dbpedia.org/resource/Hopper's_Destiny:_Book_Two	http://dbpedia.org/property/pages	352
0000005:	http://dbpedia.org/resource/Hopper's_Destiny:_Book_Two	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/Book
0000006:	http://dbpedia.org/resource/Hopper's_Destiny:_Book_Two	http://dbpedia.org/property/isbn	9781481420891.0
0000007:	http://dbpedia.org/resource/Rogue_Lawyer	http://dbpedia.org/property/isbn	978
0000008:	http://dbpedia.org/resource/Rogue_Lawyer	http://dbpedia.org/property/pages	352
0000009:	http://dbpedia.org/resource/Rogue_Lawyer	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/Book

Abbildung 4: Datensatzanzeige

1. Hier können sie die Indizes der aus dem Datensatz angezeigten Tripel auswählen und mit dem “Load”-Button nur die Datensätze im eingegebenen Bereich anzeigen lassen.
2. Hier werden die Datensätze nach dem Index sortiert im Format Subjekt, Prädikat und Objekt angezeigt.

Ergebnisseite nach Datentypfilterung

Nachdem Sie einen Datensatz nach Datentyp gefiltert haben wird folgende Seite angezeigt, welche die Möglichkeit zum Download des Datensatzes und zur Weiterverarbeitung mit RDFUnit bietet.

Filtered_dataset

New N-Triple dataset: 2016-05-31-03-20-53-335993 ①

Download

N-Triples ②

③

RDF-Unit

Schema (empty for automatic detection) ④

Extra CLI arguments (refer to RDFUnit help) ⑤

⑥

Abbildung 5: Ergebnisseite

1. Hier wird der neue Name für den erhaltenen Datensatz nach der ausgeführten Filterung angezeigt. Mit dem “View”-Button können Sie sich den gefilterten Datensatz anzeigen lassen.
2. Mit Hilfe dieses Buttons können sie den Datensatz nach der Filterung downloaden.
3. Hier können Sie das gewünschte Format für den Download ihres gefilterten Datensatzes auswählen.
4. Hier können Sie die URL für ein Schema eingeben, auf welches der Datensatz überprüft werden soll. Falls Sie dieses Feld leer lassen wird RDFUnit automatisch Tests auf Grundlage des zu verarbeitenden Datensatzes generieren.
5. In diesem Feld können zusätzliche Commands zur Ausführung von RDFUnit angegeben werden.
6. Mit der Betätigung dieses Knopfes wird RDFUnit mit den unter “4” und “5” angegebenen Werten gestartet.

Ergebnisseite nach RDFUnit Tests

Nach Beendigung der Tests durch RDFUnit können Sie auf dieser Seite bequem den Testoutput und den erneut gefilterten Datensatz downloaden.

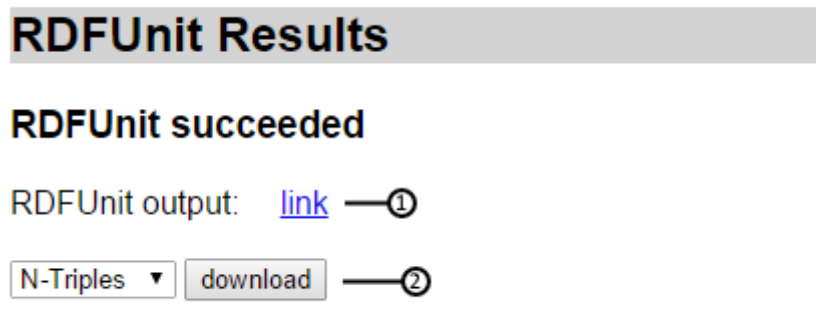


Abbildung 6: RDFUnit

1. Wenn Sie auf diesen Link klicken wird eine Ntriples-Datei mit den RDFUnit-Testergebnissen heruntergeladen.
2. Hier steht Ihnen der endgültige Download des Datensatzes nach allen Filterungen im gewünschten Format zur Verfügung.

Hinweise zur Benutzung des Query-Managers

Das Softwareprodukt arbeitet mit RDF-Daten. Um Daten im RDF-Format vom Endpunkt ihrer Wahl zu erhalten muss eine “Construct”-Anfrage gesendet werden. Für erleichterte Benutzung werden an dieser Stelle auch einfache “Select”-Anfragen unterstützt,

welche in “Construct”-Anfragen umgewandelt werden. Allerdings werden durch die Umwandlung einige Funktionen nicht mehr unterstützt.

CONSTRUCT: Jede “Construct”-Anfrage wird unterstützt. Aufgrund eines vom Softwareprodukt genutzten Methodenpaketes ist es nicht möglich in “Construct”-Anfragen das “CONSTRUCT” komplett in Großbuchstaben zu schreiben. Bitte nutzen Sie dafür “construct” in Kleinschreibung.

SELECT: Jede Anfrage, die dem folgenden Schema entspricht, wird vom Softwareprodukt unterstützt.

```
select * where { [subjekt prädikat objekt.] [FILTER( ... )] [OPTIONAL { ...
}.] [select ... ] } LIMIT X OFFSET Y
```

Anmerkungen: UNION Funktion wird *nicht* unterstützt. Unterabfragen werden nur *teilweise* unterstützt. Select-item-list hat keinen Einfluss auf die Ergebnisse. Tripel Abkürzungen (z.B.: ‘subject predicate object; predicate2 object2.’) werden *nicht* unterstützt.

DESCRIBE Describe-Anfragen werden nicht unterstützt.

ASK Ask-Anfragen werden nicht unterstützt.