The RDF Book Mashup

From Web APIs to a Web of Data

Chris Bizer, Freie Universität Berlin
Richard Cyganiak, Freie Universität Berlin
Tobias Gauß, Freie Universität Berlin

Outline

1. Web APIs and Mashups
2. The RDF Book Mashup
3. Use Cases
   1. Annotate HTML Pages with Book Mashup Data
   2. Enrich other Data Sources with Book Mashup Data
   3. Form a Part of the Emerging Web of Data
Web APIs

- A growing number of Web data sources provide query APIs.
- ProgrammableWeb.com lists 445 Web APIs
- Most popular Web APIs

![Pie chart showing popularity of Web APIs]

ProgrammableWeb.com 06/02/07

Mashups

- Mashups combine and visualize data from multiple sources.
## Limitations of Web APIs

- They provide diverse interfaces using
  - REST or SOAP web services
  - Return XML, JSON or ATOM feeds
- They use proprietary schemata to represent information
- Therefore,
  - Web APIs can not be accessed using generic data browsers
  - their content can not be crawled by search engines
  - Mashups must be implemented against a fixed number of data sources

## Approach to Overcome these Limitations

**Integrate Web APIs into the Semantic Web via Linked Data wrappers**

- **Linked Data Principles**  
  (Tim Berners-Lee, 2007)
  1. Use URIs as names for things
  2. Use HTTP URIs so that people can look up those names
  3. When someone looks up a URI, provide useful RDF information
  4. Include RDF statements that link to other URIs so that they can discover related things
Benefits of the Linked Data Approach

- Data is accessible to generic data browsers
- Data is accessible to generic data crawlers
- It is possible to interlink data items between data sources
- Clients can work against an un-bound set of data sources
- It is easier to map between different schemata and to use mappings on term level

The RDF Book Mashup

- is a small PHP script (300 lines of code)
- demonstrates how Web 2.0 data sources like Amazon, Google and Yahoo can be integrated into the Semantic Web
- applies the Linked Data principles to two Web APIs (Amazon and Google Base) which supply information about books.
The RDF Book Mashup

- assigns HTTP URIs to books, authors, reviews, online bookstores and purchase offers
- makes these URIs dereferenceable by querying
  - Google Base API
  - Amazon API

Architecture of the RDF Book Mashup

Description of a Book

```
<http://www4.wiwiss.fu-berlin.de/bookmashup/books/006251587X>

<scom:Book rdf:about="/bookmashup/books/006251587X">
  <dc:title>Weaving the Web</dc:title>
  <dc:creator rdf:resource="/bookmashup/persons/Tim+Berners-Lee"/>
  <dc:date>2000-11-01</dc:date>
  <dc:format>Paperback</dc:format>
  <dc:publisher>Collins</dc:publisher>
  <rev:hasReview rdf:resource="/bookmashup/reviews/006251587X_EditorialReview1"/>
  <scom:hasOffer rdf:resource="/bookmashup/offers/006251587XamazonOffer"/>
</scom:Book>
```
Use Cases

1. Annotate HTML Pages with Book Mashup URIs
2. Enrich other Data Sources with Book Mashup Data
3. Form a Part of the Emerging Web of Data
**UC1: Annotate HTML Pages with Book Mashup URIs**

- **Textbooks**

  - **2007**

  - **2006**

**UC2: Enrich other Data Sources Mashup Data**

- **Example: DBpedia**
  - Community effort to extract structured information from Wikipedia
  - Make this information available on the Web under an open license
  - Interlink the DBpedia dataset with other datasets on the Web

- **Contributors**
  - Freie Universität Berlin (Germany)
  - Universität Leipzig (Germany)
  - OpenLink Software (UK)
  - Linking Open Data Community (W3C SWEO)
Extracting Infobox Data

http://en.wikipedia.org/wiki/Calgary

<http://dbpedia.org/resource/Calgary>
  dbpedia:native_name "Calgary";
dbpedia:altitude "1048";
dbpedia:population_city "988193";
dbpedia:populationMetro "1079310";
mayor_name
  dbpedia:Dave_Bronconnier;
governing_body
  dbpedia:Calgary_City_Council;
...

- Altogether 9,100,000 RDF triples extracted from 754,000 infoboxes

Multi-Lingual Abstracts

- The dataset contains a short and a long abstract for each concept.

  - Short abstracts
    - English: 1,637,622
    - German: 246,791
    - French: 206,085
    - Dutch: 133,746
    - Polish: 118,874
    - Italian: 113,950
    - Spanish: 112,417
    - Japanese: 106,610
    - Portuguese: 104,842
    - Swedish: 100,267
    - Chinese: 54,991
The DBpedia Dataset

- 1,600,000 concepts
- including
  - 58,000 persons
  - 70,000 places
  - 35,000 music albums
  - 12,000 films
  - 9,000 books
- described by 93 million triples
- using 8,141 different properties.

DBpedia is interlinked with the Book Mashup

- RDF Link within the DBpedia dataset
  
  `<http://dbpedia.org/resource/Harry_Potter_and_the_Half-Blood_Prince>`
  `owl:sameAs`
  `<http://www4.wiwiss.fu-berlin.de/bookmashup/books/0747581088>`

- Altogether 9000 RDF links from DBpedia to the Book Mashup.
- Links are generated automatically using ISBN numbers from within Wikipedia.
Data from DBpedia and from the RDF Book Mashup
UC3: Form a Part of the Emerging Web of Data

- Over one billion RDF triples served as Linked Data
- Around 120,000 RDF links between data sources

Accessing the Web of Data

- Semantic Web Browsers, like
  - DISCO Hyperdata Browser
  - Tabulator Browser
  - OpenLink RDF Browser

- Semantic Web Crawlers, like
  - SWSE (DERI, Ireland)
  - Zitgist (Zitgist LLC, USA)
  - Swoogle (UMBC, USA)
Participating in the Linking Open Data Project

- Publish more RDF datasets as Linked Data
- Interlink your datasets with other datasets

Wiki Page
- http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

Mailing List
- http://simile.mit.edu/mailman/listinfo/linking-open-data

Participating in the project
- Put your name on the Wiki page
- Subscribe to the mailing list

Related Work

- OpenLink SPONGER and /proxy REST service
  - General framework to wrap Web APIs as Linked Data
    - Extractors, Mappers
  - Already wrapped Data Sources
    - Google Base
    - Flickr
    - HTML Web (page meta information/GRDDL??)
  - http://docs.openlinksw.com/virtuoso/rdfsparqlprotocolendpoint.html#rdfproxyservice

- D2R Server
  - Solution for publishing relational databases as Linked Data
    - Declarative Mapping Language
  - Already wrapped Data Sources
    - DBLP Bibliography, Project Gutenberg, CIA World Fact Book, DBtune
  - http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/
Thanks!

**RDF Book Mashup Website**
http://sites.wiwiss.fu-berlin.de/suhl/bizer/bookmashup/

**Linking Open Data Project Website**
http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData