The Semantic Data Web: A Networked Knowledge Ecosystem

Sören Auer

UNIVERSITÄT LEIPZIG
**Problem:** Try to search for these things on the current Web:
- Apartments near German-Russian bilingual childcare in Berlin.
- ERP service providers with offices in Vienna and London.
- Researchers working on breast cancer topics in Eastern Europe.

Information is available on the Web, but opaque to current search.

**Solution:** complement text on Web pages with structured linked open data & intelligently combine/integrate such structured information from different sources:
Web 1.0

Many Web sites containing unstructured, textual content

Web 2.0

Few large Web sites are specialized on specific content types

Web 3.0

Many Web sites containing & semantically syndicating arbitrarily structured content
The Long Tail of Information Domains

Currently supported structured content types:
- Pictures
- Recipes
- News
- Video
- Calendar

SemWeb supported structured content:
- Requirements-Engineering
- Talent management
- Special interest communities
- Itinerary of King George
- Gene sequences

Not or insufficiently supported content types:
- ...
1. Uses RDF Data Model

Klopotek organizes PubFor2012

PubFor2012 starts "20120423"^^xsd:date

PubFor2012 takesPlaceIn Berlin

2. Is serialised in triples:

Klopotek organizes PubFor2012
PubFor2012 starts "20120423"^^xsd:date
PubFor2012 takesPlaceAt Berlin

3. Uses Content-negotiation
The emerging Web of Data

Virtouso
SemMF
SILK
poolparty
DL-Learner
Sindice
Sigma
ORE
OntoWiki
MonetDB
dxx Engine
WiQA
repair
interlink
fuse
classify
enrich
create
Linked Data Lifecycle

- Extraction
- Search/Browsing/Exploration
- Storage/Querying
- Manual revision/authoring
- Inter-linking/Fusing
- Classification/Enrichment
- Quality Analysis
- Evolution/Repair
Creating Knowledge out of Interlinked Data

Sören Auer

– DBpedia and the Emerging Web of Linked Data

http://lod2.eu

Extraction
Extraction

From **unstructured** sources
- NLP, text mining, annotation

From **semi-structured** sources
- DBpedia, LinkedGeoData, SCOVO/DataCube

From **structured** sources
- RDB2RDF
extract structured information from Wikipedia & make this information available on the Web as LOD:

- ask sophisticated queries against Wikipedia (e.g. universities in brandenburg, mayors of elevated towns, soccer players),
- link other data sets on the Web to Wikipedia data
- Represents a community consensus

Recently launched DBpedia Live transforms Wikipedia into a structured knowledge base
Structure in Wikipedia

- Title
- Abstract
- Infoboxes
- Geo-coordinates
- Categories
- Images
- Links
  - other language versions
  - other Wikipedia pages
  - To the Web
  - Redirects
  - Disambiguations
Infobox templates

Wikitext-Syntax

{{Infobox Korean settlement
| title       = Busan Metropolitan City
| img         = Busan.jpg
| imgcaption  = A view of the [[Geumjeong]] district in Busan
| hangul      = 부산 광역시
... 
| area_km2    = 763.46
| pop         = 3635389
| popyear     = 2006
| mayor       = Hur Nam-sik
| divs        = 15 wards (Gu), 1 county (Gun)
| region      = [[Yeongnam]]
| dialect     = [[Gyeongsang]]
}}

http://dbpedia.org/resource/Busan

RDF representation

dbp:Busan    dbpp:title    "Busan Metropolitan City"
dbp:Busan    dbpp:hangul   "부산 광역시"@Hang
dbp:Busan    dbpp:area_km2  "763.46"^xsd:float
dbp:Busan    dbpp:pop      "3635389"^xsd:int
dbp:Busan    dbpp:region   dbp:Yeongnam
dbp:Busan    dbpp:dialect  dbp:Gyeongsang
...

Revised Romanization     Busan
McCune-Reischauer       Pusan
Kwanyŏkshi

Short name
Hangul    부산
Hanja     釜山

Korean name
Hangul    부산 광역시
Hanja     釜山廣域市

Region
Yeongnam

Statistics
Area 763.46 km² (295 sq mi)
Population density 4,762/km² (12,334/sq mi)
Government Metropolitan City
Mayor Hur Nam-sik
Administrative divisions 15 wards (Gu), 1 county (Gun)
...
A vast multi-lingual, multi-domain knowledge base

DBpedia extraction results in:

- **descriptions of ca. 3.4 million things** (1.5 million classified in a consistent ontology, including 312,000 persons, 413,000 places, 94,000 music albums, 49,000 films, 15,000 video games, 140,000 organizations, 146,000 species, 4,600 diseases)

- labels and abstracts for these 3.2 million things in up to 92 different languages; 1,460,000 links to images and 5,543,000 links to external web pages; 4,887,000 external links into other RDF datasets, 565,000 Wikipedia categories, and 75,000 YAGO categories

- altogether **over 1 billion pieces of information** (i.e. RDF triples): 257M from English edition, 766M from other language editions

- **DBpedia Live** (http://live.dbpedia.org/sparql/) & **Mappings Wiki** (http://mappings.dbpedia.org) integrate the community into a refinement cycle
DBpedia SPARQL Endpoint

```
  ?person skos:subject dbp:Cat:German_musicians .
  ?person foaf:name ?name .
  FILTER (LANG(?description) = 'en') .
} ORDER BY ?name
```
Tapping the intelligence of the crowd for text annotation

Pop star Michael Jackson has died in Los Angeles, aged 50.

Paramedics were called to his Beverly Hills home at about midday on Thursday after he stopped breathing. He was pronounced dead two hours later at the UCLA medical centre. Jackson’s brother, Jermaine, said he was believed to have suffered a cardiac arrest.
DBpedia Applications: Relfinder

http://www.visualdataweb.org/relfinder.php
AKSW adventures to Sächsische Schweiz!

On the 3rd of April 2011, a beautiful sunny day in Saxony, the AKSW members decided to visit the beautiful Sächsische Schweiz along with their respective family and friends.

Path:

Stadt Wehlen → Leipzig Hbf → Dresden Hbf → River Elbe → Sächsische Schweiz → Bastei → Saxony

Related Articles:

- AKSW adventures to Sächsische Schweiz
- Guide To Choosing Prams For Sale
- Choosing Bugaboo Prams For Your Baby
DBpedia Applications: Faceted-Browser

DBpedia

search powered by neofonie

Your Filters
artist The Beach Boys × released in year 1980 to 1990 × item type Album ×

The Beach Boys (album)
The Beach Boys is the twenty-fifth studio album by The Beach Boys, released in 1985. This album is notable for several reasons: it was the band's first album to be recorded after the death of Dennis Wilson. It was the first of the band's albums to be released on CD. It was the band's last album for CBS Records. It was the first album for which Brian Wilson's Svengali, Eugene Landy, was noted as a collaborating songwriter with Brian. It was the band's last album with all new compositions.

Keepin' the Summer Alive
Keepin' the Summer Alive is the twenty-fourth studio album by The Beach Boys, their first released in the 1980s, and their second release under their contract with CBS Records. After L.A. (Light Album) failed to live up to expectations, the executives at CBS expected Brian Wilson's full involvement, and thus, there was another attempt to lure him back into the producer's chair.

Still Cruisin'
Still Cruisin' is the twenty-sixth studio album by The Beach Boys, their thirty-fifth official album (counting compilations and live packages), and their last release of the 1980s. It is also the last album of new material released during a brief return to Capitol Records. After "Kokomo" (when released as a single from the Cocktail film soundtrack) gave the Beach Boys their first number-one hit in the U.S.

Ten Years of Harmony
Ten Years of Harmony is an official double album compilation album released...
Creating Knowledge out of Interlinked Data

Sören Auer

– DBpedia and the Emerging Web of Linked Data

http://lod2.eu
1. Semantic (Text) Wikis
   
   • Authoring of semantically annotated texts

2. Semantic Data Wikis
   
   • Direct authoring of structured information (i.e. RDF, RDF-Schema, OWL)
Dynamic views on knowledge bases
RDF triples on resource details page
Dynamische Vorschläge aus dem Daten Web
Prof. Dr. phil. Levin Ludwig Schücking

Lebensdaten

geb. 29.05.1878 in Burgsteifurt
gest. 12.10.1964 in Farchant

Lebenslauf

Studium

- 1897-1901 Studium: Neuere Sprachen und Kunstgeschichte in Freiburg, Göttingen, Berlin und München

Qualifikation

- 1904 Habilitation für Englische Philologie an der Universität Göttingen
  Titel der Arbeit: Grundzüge der Satzverknüpfung im Beowulf.
- 1901 Promotion zum Dr. phil. in Englischer Philologie an der Universität Göttingen
  Titel der Arbeit: Stoffliche Beziehungen der englischen Komödie zur italienischen bis Lilly.
Messi’s Goals Worthy of Praise

Barcelona’s Lionel Messi, left, celebrates after scoring his second goal against Real Madrid.

By ROB HUCHER

Published: April 27, 2011

Thank the world for Lionel Messi. Without him, without the two goals he scored for Barcelona to beat Real Madrid, 2-0, in Madrid’s Bernabeu stadium Wednesday night, soccer would now be lamenting the ugly face of the millionaires’ game. As it turned out, Messi’s goals, the second when he took on the entire Madrid defense on his own, turned brutishness into beauty.

Path: div
Creating Knowledge out of Interlinked Data

Sören Auer – DBpedia and the Emerging Web of Linked Data

http://lod2.eu
Integrating various NLP APIs
LOD Lifecycle

supported by

Debian based

LOD2 Stack

(http://stack.lod2.eu)
Publisher Opportunities

Publishers have vast repositories of textual (and semi-structured) content

Publishing metadata will raise awareness about publishing products

Semantic annotation of content will enable
• the repurposing, repackaging, tailoring of content
• Facilitate better content search (e.g. faceted-browsing)
• Increase interoperability
• Optimize publishing workflows
Publishers in the Linked Data Web

- Become linking hubs for the Data Web
- Inter-linking/Fusing
- Classifying/Enrichment
- Quality Analysis
- Evolution/Repair
- Manual revision/authoring
- Storage/Querying
- Extraction
- Hosting & maintenance of exploration tools
- Search/Browsing/Exploration

Support facilities for knowledge based authoring & collaboration

Provide storage facilities for Linked Data

Extract and publish structured (meta-)data for publishing content

Authorative Linked Data for quality assessment

Publisher is valuable background knowledge for KB enrichment & repair.

Be the “lighthouse” for the LOD ocean.

Be the "lighthouse" for the LOD ocean.

Publisher is valuable background knowledge for KB enrichment & repair.
Thanks for your attention!

Sören Auer
auer@uni-leipzig.de

UNIVERSITÄT LEIPZIG