AKSW Group

- 3 PostDocs, 15 PhDs, 3 Subgroups (Emergent Semantics, MOLE, SIMBA)
- centered around Semantic Web tools and applications
- [http://aksw.org](http://aksw.org)
- [http://www.slideshare.net/soeren1611/overview-ag-aksw](http://www.slideshare.net/soeren1611/overview-ag-aksw)
AKSW Projects

Active Funded Projects:
- LOD2
- LATC
- SCMS
- LE4SW

Research/Software Projects:
- DBpedia
- LinkedGeoData
- OntoWiki
- DL-Learner
- BorderFlow
- LIMES
- SAIM
- NLP2RDF
- Triplify
- . . .

http://lod2.eu
Personal Background

- Studied Computer Science at TU Dresden 2001-2006
- PhD Uni Leipzig 2006-2010
- PostDoc AKSW / Uni Leipzig ab 2010
- Head of Machine Learning and Ontology Engineering Group (MOLE)
- EU-Projects: LOD2, LATC
- Software/Research-Projects: DBpedia, DL-Learner, ORE, LinkedGeoData
- Interests: Semantic Web, Machine Learning, Logic
Why Semantic Web?
Problem: Try to search for these things on the current Web:
- Apartments near German-Russian bilingual childcare in Leipzig.
**Problem:** Try to search for these things on the current Web:

- Apartments near German-Russian bilingual childcare in Leipzig.
- ERP service providers with offices in Vienna and London.
Problem: Try to search for these things on the current Web:
• Apartments near German-Russian bilingual childcare in Leipzig.
• ERP service providers with offices in Vienna and London.
• Researchers working on multimedia topics in Eastern Europe.
Information is available on the Web, but opaque to current Web search.
**Problem:** Try to search for these things on the current Web:
- Apartments near German-Russian bilingual childcare in Leipzig.
- ERP service providers with offices in Vienna and London.
- Researchers working on multimedia topics in Eastern Europe.

Information is available on the Web, but opaque to current Web search.
**Problem:** Try to search for these things on the current Web:
- Apartments near German-Russian bilingual childcare in Leipzig.
- ERP service providers with offices in Vienna and London.
- Researchers working on multimedia topics in Eastern Europe.

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

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

- *leipzig.de*  
  Has everything about childcare in Leipzig.
- *Immobilien scout.de*  
  Knows all about real estate offers in Germany.
**Problem:** Try to search for these things on the current Web:
- Apartments near German-Russian bilingual childcare in Leipzig.
- ERP service providers with offices in Vienna and London.
- Researchers working on multimedia topics in Eastern Europe.

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

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

- **leipzig.de**
  - Has everything about childcare in Leipzig.

- **Immobilienscout.de**
  - Knows all about real estate offers in Germany.
### Semantic Web - Standards

#### Standardization Semantic Web

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>First public presentation of idea</td>
</tr>
<tr>
<td>1998</td>
<td>Start of standardization of RDF and RDFS</td>
</tr>
<tr>
<td>2000</td>
<td>Large research projects in US and Europe</td>
</tr>
<tr>
<td>2002</td>
<td>Start of OWL standardisation</td>
</tr>
<tr>
<td>2004</td>
<td>RDF, OWL</td>
</tr>
<tr>
<td>2006</td>
<td>RDFa</td>
</tr>
<tr>
<td>2008</td>
<td>SPARQL</td>
</tr>
<tr>
<td>2009</td>
<td>OWL 2</td>
</tr>
<tr>
<td>2010</td>
<td>RIF</td>
</tr>
</tbody>
</table>
1. (usually) uses RDF as data model

AKSW organizes LSWT2011

LSWT2011 takesPlaceAt 5.5.2011

LSWT2011 takesPlaceIn Leipzig

AKSW
1. (usually) uses RDF as data model

Serialised in Triples:

- AKSW organizes LSWT2011
- LSWT2011 takesPlaceAt "20110505"^^xsd:date
- LSWT2011 takesPlaceIn Leipzig

AKSW organizes LSWT2011

Leipzig

5.5.2011
1. (usually) uses RDF as data model

Serialised in Triples:

- AKSW organizes LSWT2011
- LSWT2011 takesPlaceAt "20110505"^^xsd:date
- LSWT2011 takesPlaceIn Leipzig

2. Uses dereferencable HTTP-URIs, e.g. Content-negotiation
1. (usually) uses RDF as data model

Serialised in Triples:

AKSW organizes LSWT2011
LSWT2011 takesPlaceAt “20110505”^^xsd:date
LSWT2011 takesPlaceIn Leipzig

2. Uses dereferencable HTTP-URIs, e.g. Content-negotiation

3. Links to Resources (in external knowledge bases)
LOD achievements and challenges

- Web - a global, distributed platform for data, information and knowledge integration
- exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF
LOD achievements and challenges

- Web - a global, distributed platform for data, information and knowledge integration
- exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF

Achievements

1. Extension of the Web with a data commons (25B facts)
2. vibrant, global RTD community
3. Industrial uptake begins (e.g. BBC, Thomson Reuters, Eli Lilly, NY Times, Facebook, Google, Yahoo)
4. Emerging governmental adoption in sight
5. Establishing Linked Data as a deployment path for the Semantic Web.
LOD achievements and challenges

• Web - a global, distributed platform for data, information and knowledge integration
• exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF

Achievements

1. Extension of the Web with a data commons (25B facts)
2. vibrant, global RTD community
3. Industrial uptake begins (e.g. BBC, Thomson Reuters, Eli Lilly, NY Times, Facebook, Google, Yahoo)
4. Emerging governmental adoption in sight
5. Establishing Linked Data as a deployment path for the Semantic Web.

Challenges

1. Coherence: Relatively few, expensively maintained links
2. Quality: partly low quality data and inconsistencies
3. Performance: Still substantial penalties compared to relational
4. Data consumption: large-scale processing, schema mapping and data fusion still in its infancy
5. Usability: Missing direct end-user tools and network effect

These issues are closely related and should ultimately lead to an ecosystem of interlinked knowledge!

July 2007

(a) July 2007
(b) April 2008
(c) September 2008
(d) July 2009
Linked Data Lifecycle

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

Jens Lehmann – Linked Data Lifecycle
7. Juni 2011

http://lod2.eu
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
Structure in Wikipedia

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

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

Busan

From Wikipedia, the free encyclopedia

"Pusan" redirects here. For the Vedic Hindu god, see Pushan.

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

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

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

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

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

Categories: Busan | Cities in South Korea | Coastal cities
Structure in Wikipedia

Title
Abstract
Infoboxes
Geo-coordinates
Categories
Images
Links
other language versions
other Wikipedia pages
To the Web
Redirects
Disambiguations
{{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
...
```
• **descriptions of ca. 3.4 million things** (1.5 million classified in a consistent ontology), including 312K *persons*, 413K *places*, 94K *music albums*, 49K *films*, 15K *video games*, 140K *organizations*, 146K *species*, 5K *diseases*

• labels & abstracts in up to 92 different languages

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

  • **substantial impact** in science, technology and society
  • became a **central interlinking hub** on the Data Web
  • Scientific publications attracted more than **500 citations**
  • More than 15,000 monthly visits on DBpedia.org, numerous press articles, blog posts …
  • Ecosystem of commercial and community applications: ThomsonReuters, BBC, Neofonie, Openlink, Faviki …
OpenStreetMap is an incredible source of spatial data. LinkedGeoData makes this data available as Linked Data:

- Mappings to DBpedia, Geonames
- SPARQL Endpoint
- REST API – Linked Data Services

<table>
<thead>
<tr>
<th>Users</th>
<th>309465</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS points</td>
<td>1993482813</td>
</tr>
<tr>
<td>Nodes</td>
<td>795615514</td>
</tr>
<tr>
<td>Ways</td>
<td>66102033</td>
</tr>
<tr>
<td>Relations</td>
<td>751638</td>
</tr>
</tbody>
</table>
Many different approaches (D2R, Virtuoso RDF Views, Triplify, …)
No agreement on a formal semantics of RDF2RDF mapping
• LOD readiness,
  SPARQL-SQL translation
W3C RDB2RDF WG

<table>
<thead>
<tr>
<th>Tool</th>
<th>Triplify</th>
<th>D2RQ</th>
<th>Virtuoso RDF Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>Scripting languages (PHP)</td>
<td>Java</td>
<td>Whole middleware solution</td>
</tr>
<tr>
<td>SPARQL endpoint</td>
<td>-</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mapping language</td>
<td>SQL</td>
<td>RDF based</td>
<td>RDF based</td>
</tr>
<tr>
<td>Scalability</td>
<td>Medium-high (but no SPARQL)</td>
<td>Medium</td>
<td>High</td>
</tr>
</tbody>
</table>
Extraction Challenges

From **unstructured** sources
- Deploy existing NLP approaches (OpenCalais, Ontos API)
- Develop standardized, LOD enabled interfaces between NLP tools (NLP2RDF)

From **semi-structured** sources
- Efficient bi-directional synchronization

From **structured** sources
- Declarative syntax and semantics of data model transformations (W3C WG RDB2RDF)

Orthogonal challenges
- Using LOD as background knowledge
- Provenance
Storage and Querying
Still by a factor 5-50 slower than relational data management
Performance increases steadily
Comprehensive, well-supported open-source and commercial implementations are available:

- OpenLink’s Virtuoso (os+commercial)
- Big OWLIM (commercial), Swift OWLIM (os)
- Talis (hosted)
- Bigdata (distributed)
- Allegrograph (commercial)
- Mulgara (os)
Adaptive SPARQL Query Cache

RDF Graph:

<table>
<thead>
<tr>
<th></th>
<th>rdf:type</th>
<th>foaf:person</th>
</tr>
</thead>
<tbody>
<tr>
<td>JensLehmann</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JensLehmann</td>
<td>foaf:knows</td>
<td>SoerenAuer</td>
</tr>
<tr>
<td>SebastianTramp</td>
<td>rdf:type</td>
<td>foaf:person</td>
</tr>
<tr>
<td>SebastianTramp</td>
<td>foaf:knows</td>
<td>SoerenAuer</td>
</tr>
</tbody>
</table>

...
Adaptive SPARQL Query Cache

RDF Graph:

Graph pattern (SPARQL core):

```sparql
?person rdf:type foaf:person
?person foaf:knows http://aksw.org/SoerenAuer
```
Adaptive SPARQL Query Cache

**RDF Graph:**

<table>
<thead>
<tr>
<th>...</th>
<th>rdf:type</th>
<th>foaf:person</th>
</tr>
</thead>
<tbody>
<tr>
<td>JensLehmann</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JensLehmann</td>
<td>foaf:knows</td>
<td>SoerenAuer</td>
</tr>
<tr>
<td>SebastianTramp</td>
<td>rdf:type</td>
<td>foaf:person</td>
</tr>
<tr>
<td>SebastianTramp</td>
<td>foaf:knows</td>
<td>SoerenAuer</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Graph pattern (SPARQL core):**

```sparql
?person rdf:type foaf:person
?person foaf:knows http://aksw.org/SoerenAuer
```
Adaptive SPARQL Query Cache

**RDF Graph:**

<table>
<thead>
<tr>
<th>...</th>
<th>rdf:type</th>
<th>foaf:person</th>
</tr>
</thead>
<tbody>
<tr>
<td>JensLehmann</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JensLehmann</td>
<td>foaf:knows</td>
<td>SoerenAuer</td>
</tr>
<tr>
<td>SebastianTramp</td>
<td>rdf:type</td>
<td>foaf:person</td>
</tr>
<tr>
<td>SebastianTramp</td>
<td>foaf:knows</td>
<td>SoerenAuer</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Graph pattern (SPARQL core):**

```sparql
?p person rdf:type foaf:person
```

**Question:** Which query results are affected by which updates?
Adaptive SPARQL Query Cache

**RDF Graph:**

<table>
<thead>
<tr>
<th>...</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>JensLehmann</td>
<td>rdf:type</td>
</tr>
<tr>
<td>JensLehmann</td>
<td>foaf:knows</td>
</tr>
<tr>
<td>SebastianTramp</td>
<td>rdf:type</td>
</tr>
<tr>
<td>SebastianTramp</td>
<td>foaf:knows</td>
</tr>
</tbody>
</table>
| ... | ...

**Graph pattern (SPARQL core):**

```sparql
?person rdf:type foaf:person
?person foaf:knows http://aksw.org/SoerenAuer
```

**Question:** Which query results are affected by which updates?

**Theorem:** Query results only change, when updated triples match one of the triple patterns. Formal proof wrt. SPARQL algebra in [1].

---

Adaptive SPARQL Query Cache

Graph pattern (SPARQL core):

```
?person rdf:type foaf:person
?person foaf:knows http://aksw.org/SoerenAuer
```

Question: Which query results are affected by which updates?

Theorem: Query results only change, when updated triples match one of the triple patterns. Formal proof wrt. SPARQL algebra in [1].

Work with OpenLink Ltd. (UK) on integrating the approach into the Virtuoso triple store and with CWI (NL) into MonetDB.

Storage and Querying Challenges

- Reduce the performance gap between relational and RDF data management
- SPARQL Query extensions
- Spatial/semantic/temporal data management
- More advanced query result caching
- View maintenance / adaptive reorganization based on common access patterns
- More realistic benchmarks
Creating Knowledge out of Interlinked Data

Jens Lehmann – Linked Data Lifecycle

7. Juni 2011

http://lod2.eu

Authoring
Two Kinds of Semantic Wikis

1. Semantic (Text) Wikis
   - Authoring of semantically annotated texts

2. Semantic Data Wikis
   - Direct authoring of structured information
     (i.e. RDF, RDF-Schema, OWL)
Two Kinds of Semantic Wikis

1. Semantic (Text) Wikis
   - Authoring of semantically annotated texts

2. Semantic Data Wikis
   - Direct authoring of structured information (i.e. RDF, RDF-Schema, OWL)
Two Kinds of Semantic Wikis

1. Semantic (Text) Wikis
   - Authoring of semantically annotated texts

2. Semantic Data Wikis
   - Direct authoring of structured information (i.e. RDF, RDF-Schema, OWL)
OntoWiki – a semantic data wiki

Versatile domain-independent tool
Serves as Linked Data / SPARQL endpoint on the Data Web
Open-source project hosted at Google code
Not just a Wiki UI, but a whole framework for the development of Semantic Web applications
Developed in PHP based on the Zend framework
Very active developer and user community
More than 500 downloads monthly
Large number of use cases
Management of multimedia content for a record label in the UK
<table>
<thead>
<tr>
<th></th>
<th>Harpactea golovatchi</th>
<th>published in</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Dunin (1992)</td>
<td>Shvaridzor</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Harpactea nenilini</td>
<td>Dunin (1992)</td>
<td>Kuris</td>
</tr>
<tr>
<td>Property</td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ns1:to</td>
<td><a href="http://philipp.frischmuth24.de/pingback/ping/">http://philipp.frischmuth24.de/pingback/ping/</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type</td>
<td>Person</td>
<td></td>
<td></td>
</tr>
<tr>
<td>seeAlso</td>
<td><a href="http://feeds.delicious.com/rss/pfrischmuth">http://feeds.delicious.com/rss/pfrischmuth</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sameAs</td>
<td>id490bd762-0352-40e8-8559-3726c31bbd8a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>birthday</td>
<td>1984-10-09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>current project</td>
<td><a href="http://pingback.aksw.org/">http://pingback.aksw.org/</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>depiction</td>
<td>![Image]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>firstName</td>
<td>Philipp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>holds account</td>
<td>pfrischmuth</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pfrischmuth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>homepage</td>
<td><a href="http://www.frischmuth24.de">http://www.frischmuth24.de</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>knows</td>
<td>me</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sebastian Tramp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>personal mailbox</td>
<td><a href="mailto:frischmuth@informatik.uni-leipzig.de">mailto:frischmuth@informatik.uni-leipzig.de</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="mailto:philipp@frischmuth24.de">mailto:philipp@frischmuth24.de</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>name</td>
<td>Philipp Frischmuth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nickname</td>
<td>Phil</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
RDFauthor in OntoWiki
OntoWiki: Supporting Requirements engineering

To each lab in the repository users can write comments. To accomplish that a simple text field is required and a function which saves the comments related to the specific resource.

This functionality provides a communication channel between users and developers as well as among users. It allows them to give hints for improvements, to criticize, to leave compliments for the authors. To avoid spam it is necessary to implement methods like Captcha or math story problems.
Linking
Manual:
- SILK
- LIMES

Semi-Automatic:
- SAIM (soon)
Only 5% of the information on the Data Web is actually linked

• Make sense of work in the de-duplication/record linkage literature
• Consider the open world nature of Linked Data
• Use LOD background knowledge
• Zero-configuration linking
• Explore active learning approaches, which integrate users in a feedback loop
• Maintain a 24/7 linking service: Linked Open Data Around-The-Clock project ([LATC-project.eu)](http://lod2.eu/LATC-project.eu)
Enrichment
ORE (Ontology Repair and Enrichment) tool allows to improve an OWL ontology by fixing inconsistencies & making suggestions for adding further axioms.

- **Ontology Debugging:** OWL reasoning to detect inconsistencies and satisfiable classes + detect the most likely sources for the problems. User can create a repair plan, while maintaining full control.

- **Ontology Enrichment:** uses the DL-Learner framework to suggest definitions & super classes for existing classes in the KB. Works if instance data is available for harmonising schema and data.

http://aksw.org/Projects/ORE
**Given:**
- **Background knowledge** base
- **Positive and negative examples**
  (example = individual in ontology)

**Goal:**
Find an OWL **Class Expression** / DL concept which
- **covers** as many positive examples as possible
- **covers** as few negative examples as possible

Concept C covers example a \( \Rightarrow \)
a is instance of C
Analogous problem can be defined for logic programs \( \Rightarrow \) Inductive Logic Programming
**Supervised** Machine Learning Task

**Diagram:**
- **Background knowledge** (OWL-DL ontology, DL knowledge base)
- **Positive and negative examples**
- **Concept generator**, e.g., refinement operator based, genetic programming
- **Reasoner**, e.g., Pellet, FaCT, KAON2, Racer Pro
- **Query results**
- **Instance checks**
- **Return quality**
- **Test concept**

**URL:**[http://lod2.eu](http://lod2.eu)
Application areas:

- **Ontology Engineering**: learn axioms for existing classes, PlugIns for Protégé and OntoWiki
- recommendation/navigation, e.g. music based on last songs heard or navigation suggestions in large knowledge bases (Dbpedia navigation)
- “Classical” Machine Learning, e.g. predicting Carcinogenesis

Works on OWL Files and SPARQL endpoints
Supports different reasoner interfaces
Accessible via command-line, GUI, web service
Implemented in Java, available as open-source,

---


Linked Data Quality Analysis

Quality on the Data Web is varying a lot

- Hand crafted or expensively curated knowledge base (e.g. DBLP, UMLS) vs. extracted from text or Web 2.0 sources (DBpedia)

Research Challenge

- Establish measures for assessing the authority, provenance, reliability of Data Web resources
EvoPat – Pattern based KB Evolution

- unified method, for both data evolution and ontology refactoring.
- modularized, declarative definition of evolution patterns is relatively simple compared to an imperative description of evolution
  - allows domain experts and knowledge engineers to amend the ontology structure and modify data with just a few clicks
- Combined with RDF representation of evolution patterns and their exposure on the Linked Data Web, EvoPat facilitates the development of an evolution pattern ecosystem
  - patterns can be shared and reused on the Data Web.
- declarative definition of **bad smells and corresponding evolution patterns** promotes the (semi-)automatic improvement of information quality.
Evolution Patterns

compound pattern (CP)  
- description  
- title  

basic pattern (BP)  
- description  
- title  

functional attributes  
$V_i \quad CP_i \quad 0 < i \leq n$  

SPARQL query  

variable *  

update query  
$U_S \quad U_P \quad U_O$  

functional attributes  
Typ  
Namen  

SPARQL query Variable to Pattern Variable Mapping  

bad smell  

SPARQL query
Exploration
Prof. Dr. phil. Levin Ludwig Schücking

Lebensdaten

geb. 29.05.1878 in Burgsteinfurt
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.
Properties for Levin Ludwig Schücking

Nachname: Schücking
Voller akademischer Titel: Prof. Dr. phil.
Eine Person hat: Levin
Vorname: Ludwig
Person ID: 144
Geburtsdatum: 1878-05-29
Geburtsort: Burgsteinfurt
Geburtsland: Preußen
Todesdatum: 1964-10-12
Todesort: Parchant

Bildherkunft: Privatbesitz
Links: http://de.wikipedia.org/wik/Levin_Ludwig_Sch%20C%20Cocking

Abitur Stadt: Münster
Abitur Land: Preußen
Professor hat: 1897-1901 Studium: Neuere Sprachen und Kunstgeschichte in Freiburg.
The CPL Model

[Diagram of the CPL Model with arrows and nodes representing relationships and entities such as Person, Organization, Country, City, AdministrativeDistrict, etc.]
CPL Authoring Activity

- Any Activity
- Statements Added
- Statements Removed
- Property Changed

- Professors Total
- Professors Added
- Editors
- Access

Jens Lehmann – Linked
Levin Ludwig Schücking

Geburtsstadt: Burgsteinfurt
Geburtsort: Preußen
Todesort: Farchant
Todeszeit: 1964-10-12
Sterbeort: Prof. Dr. phil.

1946-1952 Lehrstuhlvertretung an der Universität Erlangen

Münster

1925-1944 oProf für Anglistik an der Philologisch-Historischen Abteilung der Philosophischen Fakultät der Universität Leipzig

1910-1916 aoProf für Englische Sprache und Literatur an der Universität Jena

1916-1925 oProf für Englische Sprache und Literatur an der Universität Breslau

1952 oProf für Englische Sprache und Literatur an der Universität München

Mitglied der American Academy of Arts and Sciences

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

von 1897
Mitgliedschaft: Universität München
Fach: Kunstgeschichte
bis 1901-01-01
mentions: Studium, Data Creation, Data Item, Data Creating Service
request type: Request for <http://catalogus-professorum.org/lipsiensis/Study_110>
label: 1897-1901 Studium: Neuere Sprachen und Kunstgeschichte in Freiburg, Göttingen, Berlin und München
is-period-of: Levin Ludwig Schücking
is-used-by: This Session

Find all
Visual Query Builder

CPL Navigation

OntoWiki (Admin)

Navigation

Edit  View  Type

Search in Navigation

Root

professor
archive
forename
qualification paper
person
study
career
social role
faculty
academy

Knowledge Bases

Edit  View

Catalogus Professorum Model (CPM)
professors catalogue Leipzig
Professors born in Dresden and related (subset of CPL)

Saved Queries  Query Editor  Graphical Query Builder

Display results  Show Query  en  de

Save Query

Properties

is located at City
is located at Country
related body
type of body
is located at State

Restrictions:

AND

name of the institution contains "Leipzig"

available links:

• institution consists of (body)
• an institution is part of (body)
• body place (not found)
selected links:
• none

<table>
<thead>
<tr>
<th>#</th>
<th>professor1</th>
<th>qualificationpaper1</th>
<th>academy1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cpl:Anger_1210</td>
<td>cpl:Promotion_1477</td>
<td>cpl:Hochschule-1</td>
</tr>
<tr>
<td>2</td>
<td>cpl:Anger_1210</td>
<td>cpl:Promotion_1476</td>
<td>cpl:Hochschule-1</td>
</tr>
</tbody>
</table>

Relationship Finder in CPL
Linked Data Lifecycle

- Interlinking/Fusing
- Classification/Enrichment
- Storage/Querying
- Extraction
- Evolution/Repair
- Search/Browsing/Exploration
- Manual revision/authoring
- Quality Analysis

Interlinking/Fusing

Classification/Enrichment

Storage/Querying

Extraction

Evolution/Repair

Search/Browsing/Exploration

Manual revision/authoring

Quality Analysis
Make the Web a Linked Data Washing Machine
Make the Web a Linked Data Washing Machine
AKSW Data Web Building Blocks

**RDB2RDF**
Mapping relational data to RDF

**DL-Learner**
Machine Learning for Ontologies

**ORE**
Ontology Enrichment & Repair

**LIMES**
Link Discovery Framework for metric spaces

**RDF Query Subsumption & View Maintenance**
Scaling triple stores

**Foundations**
Marrying databases with RDF and ontologies
AKSW Data Web Building Blocks

**Foundations**

- RDB2RDF: Mapping relational data to RDF
- DL-Learner: Machine Learning for Ontologies
- ORE: Ontology Enrichment & Repair
- LIMES: Link Discovery Framework for metric spaces
- RDF Query Subsumption & View Maintenance: Scaling triple stores

**Tools**

- OntoWiki: Collaborative creation of explicit knowledge via Semantic Wikis
- DBpedia: “Semantification” of Wikipedia
- LinkedGeoData: “Semantification” of OpenStreetMaps
- Triplify: “Semantification” of (small) Web Applications
- LESS: Semantification Syndication

**Semantification**

- DBpedia
- OntoWiki
- LinkedGeoData
- Triplify
- LESS

**Resources**

- LOD2: http://lod2.eu
AKSW Data Web Building Blocks

**Foundations**
Marrying databases with RDF and ontologies

**Tools**

- **RDB2RDF**
  Mapping relational data to RDF

- **DL-Learner**
  Machine Learning for Ontologies

- **ORE**
  Ontology Enrichment & Repair

- **LIMES**
  Link Discovery Framework for metric spaces

- **RDF Query Subsumption & View Maintenance**
  Scaling triple stores

**OntoWiki**
Collaborative creation of explicit knowledge via Semantic Wikis

**LinkedGeoData**
“Semantification” of OpenStreetMaps

**Triplify**
“Semantification” of (small) Web Applications

**LESS**
Semantification Syndication

**Applications**
Bringing the Data Web to end users

**Vakantieland**
Building Data Web applications

**SoftWiki**
Distributed, stakeholder driven Requirements Engineering

**OpenResearch.org**
A semantic Wiki for the sciences

**xOperator**
Combining Instant Messaging with the Data Web

**Catalogus Professorum**
Prosopographical knowledge base

**DBpedia**
“Semantification” of Wikipedia

**OntoWiki**
“Semantification” of Wikipedia

**LinkedGeoData**
“Semantification” of OpenStreetMaps

**Triplify**
“Semantification” of (small) Web Applications

**LESS**
Semantification Syndication

**Foundations**
Marrying databases with RDF and ontologies
Thanks for your attention!

Jens Lehmann

lehmann@informatik.uni-leipzig.de