Outline

• Introduction
• Semantic Web
  o Vision
  o From the Web of Documents to the Web Of Data
  o Technologies
  o Advantages
• Semantic Web and Linguistics
• RDB2RDF
  o Concepts
  o Sparqlify (SML)
• PanLex2RDF
• Demo
• Discussion
About me

- Claus Stadler
- Austria
- PhD Student at the University of Leipzig since 2011
  - In the Agile Knowledge Engineering and Semantic Web (AKSW) research group, headed by Soeren Auer.
- Research Interests: Spatial Data Management, SPARQL-SQL query rewriting and optimization, Data integration.
About me

- Patrick Westphal
- Germany
- Master Student at the University of Leipzig since 2011
  - In the Agile Knowledge Engineering and Semantic Web (AKSW) research group, headed by Soeren Auer.
- Research Interests: Data Quality, SPARQL-SQL query rewriting and optimization, Data integration
Agile Knowledge Engineering and Semantic Web Research Group

- Founded in 2006
- 25+ Researchers
- 3 Sub groups

Goals
- Contributing to the advancement of science in Semantic Web, Knowledge Engineering, Software Engineering
- Cost efficient, high-impact R&D, which proves usefulness at an early stage
- Bridge the gap between research results and applications

Committed to Open Source, Open Access, and Open Knowledge movements
Agile Knowledge Engineering and Semantic Web Research Group

- EU Funded Projects:
  - GeoKnow
  - Linked Open Data 2 (LOD2)
  - LOD Around the Clock (LATC)
  - Open Data Portal (ODP)
  - Semantic Content Management Systems for Enterprise Knowledge Management and News Mining (SCMS)
  - OntoWiki - Semantic Collaboration for Knowledge Management, E-Learning and E-Tourism
Agile Knowledge Engineering and Semantic Web Research Group

- **Further Projects**
  - SlideWiki
    - SlideWiki is a collaboration platform which enables communities to build, share and play online presentations.
  - LinkedGeoData
    - Making OpenStreetMap data available in the Semantic Web
    - Motivation for Sparqlify
  - LIMES
    - Very fast tool for interlinking RDF knowledge bases.
  - DBpedia Live
    - Synchronization of DBpedia with Wikipedia
  - ...

- **Find more at**
  - [http://aksw.org/Projects](http://aksw.org/Projects)
Linked Open Data cloud
From the Web of Documents to the Web of Data

• Web of Documents

[http://www.dailycal.org/section/news/]

CITY

THURSDAY, JULY 25, 2013

Berkeley Food and Housing Project awarded $1 million to fight veteran homelessness

Local non-profit organization Berkeley Food and Housing Project (BFHP) announced on Friday that they have received about $1,000,000 from the U.S. Department of Veteran Affairs as part of a $300 million nationwide effort to eradicate veterans’ homelessness.

CITY

WEDNESDAY, JULY 24, 2013

Postal Service rejects appeal of main Berkeley branch sale

The US Postal Service announced a final decision Thursday to sell Berkeley's main post office despite city and community appeals.

CITY

SUNDAY, JULY 21, 2013

Irish students on work visas gravitate to Berkeley for the summer

Every year, hundreds of Irish students spend their summer in Berkeley. Recognizable to many by
Berkeley Food and Housing Project awarded $1 million to fight veteran homelessness

Local non-profit organization Berkeley Food and Housing Project (BFHP) announced on Friday that they have received about $1,000,000 from the U.S. Department of Veteran Affairs as part of a $300 million nationwide effort to eradicate veterans’ homelessness.

Read More...
Berkeley Food and Housing Project awarded $1 million to fight veteran homelessness

Local non-profit organization Berkeley Food and Housing Project (BFHP) announced on Friday that they have received about $1,000,000 from the U.S. Department of Veteran Affairs as part of a $300 million nationwide effort to eradicate veterans’ homelessness.

Read More...
Berkeley Food and Housing Project awarded $1 million to fight veteran homelessness

Local non-profit organization Berkeley Food and Housing Project (BFHP) announced on Friday that they have received about $1,000,000 from the U.S. Department of Veteran Affairs as part of a $300 million nationwide effort to eradicate veterans’ homelessness.

Read More…
From the Web of Documents to the Web of Data

- Web of Documents

<h2>Berkeley Food and Housing Project awarded $1 million to fight veteran homelessness</h2>
<div>
<p>Local non-profit organization Berkeley Food and Housing Project (BFHP) announced on Friday that they have received about $1,000,000 from the U.S. Department of Veteran Affairs as part of a $300 million nationwide effort to eradicate veterans’ homelessness. <a href="http://...">Read More…</a></p>
</div>
From the Web of Documents to the Web of Data

- Web of Documents

<h2>BERKELEY Food and Housing Project awarded $1 million to fight veteran homelessness</h2>

http://en.wikipedia.org/wiki/Berkeley,_New_South_Wales
http://en.wikipedia.org/wiki/Berkeley,_California
http://en.wikipedia.org/wiki/Berkeley,_Missouri
http://en.wikipedia.org/wiki/Berkeley,_Albemarle_County,_Virginia
http://en.wikipedia.org/wiki/Berkeley,_Charles_City_County,_Virginia

<a href="http://...">Read More…</a>
From the Web of Documents to the Web of Data

- Web of Documents

<h2>Berkeley Food and Housing Project awarded $1 million to fight veteran homelessness</h2>

Local non-profit organization Berkeley Food and Housing Project (BFHP) announced on Friday that they have received about $1,000,000 from the U.S. Department of Veteran Affairs as part of a $300 million nationwide effort to eradicate veterans’ homelessness.

<a href="http://...">Read More…</a>
From the Web of Documents to the Web of Data

- **Web of Documents**

  <h2>Berkeley Food and Housing Project awarded $1 million to fight veteran homelessness</h2>
  <div>
    <p>Local non-profit organization Berkeley Food and Housing Project (BFHP) announced on Friday that they have <strong>received about $1,000,000</strong> from the U.S. Department of Veteran Affairs as part of a $300 million nationwide effort to eradicate veterans’ homelessness. <a href="http://...">Read More…</a></p>
  </div>
• Web of Data


the $1,000,000 are a funding.

the $1,000,000 fund <http://bfhp.org/>.
From the Web of Documents to the Web of Data

• Web of Data

<http://bfhp.org/>
  <http://ex.org/ontology/isA>

<http://bfhp.org/>
  <http://ex.org/ontology/isLocatedIn>
    <http://cityofberkeley.info> .

<http://ex.org/funding2342>
  <http://ex.org/ontology/isA>
    <http://ex.org/ontology/Funding> .

<http://ex.org/funding2342>
  <http://ex.org/ontology/funds>
  <http://bfhp.org/> .
From the Web of Documents to the Web of Data

• Web of Data

<http://bfhp.org/>
  <http://ex.org/ontology/isA>

<http://bfhp.org/>
  <http://ex.org/ontology/isLocatedIn>
  <http://cityofberkeley.info> .

<http://ex.org/funding2342>
  <http://ex.org/ontology/isA>
  <http://ex.org/ontology/Funding> .

<http://ex.org/funding2342>
  <http://ex.org/ontology/funds>
  <http://bfhp.org/> .
From the Web of Documents to the Web of Data

Web of Data

<http://bfhp.org/>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://xmlns.com/foaf/0.1/Organization> .

<http://bfhp.org/>
  <http://dbpedia.org/ontology/}
  <http://cityofberkeley.info> .

<http://ex.org/funding2342>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
    <http://vocab.ox.ac.uk/projectfunding#Funding> .

<http://ex.org/funding2342>
  <http://vocab.ox.ac.uk/projectfunding#funds>
    <http://bfhp.org/> .
From the Web of Documents to the Web of Data

• Web of Data

<http://bfhp.org/>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://xmlns.com/foaf/0.1/Organization> .

<http://bfhp.org/>
  <http://dbpedia.org/ontology/locationCity>
  <http://cityofberkeley.info> .

<http://ex.org/funding2342>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://vocab.ox.ac.uk/projectfunding#Funding> .

<http://ex.org/funding2342>
  <http://vocab.ox.ac.uk/projectfunding#funds>
  <http://bfhp.org/> .
Semantic Web Technology Stack

User Interface & Applications

Trust

Proof

Unifying Logic

Query: SPARQL

Ontology: OWL

Rule: RIF

RDFS

Data interchange: RDF

XML

URI/IRI

Crypto
Semantic Web Technology Stack

URI/IRI

- **Uniform Resource Identifier**
  - http://panlex.org
  - foo://example.com:8042/over/there?name=ferret#nose

- **Internationalized Resource Identifier**
  - http://ld.panlex.org/expression/कममजलदकर

Semantic Web Technology Stack

RDF (*Resource Description Framework*)
- W3C specification
- "for conceptual description or modeling of information"

[RDF documents (files)
- graphs
  - triples
    - resources | blank nodes | literals]
Semantic Web Technology Stack

RDF (Resource Description Framework)

- resources
  - sth. that has an identity
  - identified by IRI
  - often URLs used
    - "owned" by one person/organization → one authority
    - Web documents can be placed on the target location to provide further information about the resource
Semantic Web Technology Stack

RDF (Resource Description Framework)

- blank nodes (*anonymous resources*)
  - to be able to make statements about a resource with a certain identity without naming it

  e.g. "John has a friend born on 21st of April"

  [http://en.wikipedia.org/wiki/Blank_node]

  - represented by
    - an underscore (with certain id) "_:23"
    - or "[]"
Semantic Web Technology Stack

RDF (Resource Description Framework)

• literals
  o can be without a type, but having a language tag (according to RFC 5646, BCP 47)
    → plain literal
    e.g. "apple"@en, "Apfel"@de
  o can have a type: string, integer number, date, ...
    → typed literal
    e.g. "42"^^xsd:int,"2005-01-01"^^xsd:date
Semantic Web Technology Stack

RDF *(Resource Description Framework)*

- **triples**
  - subject
  - predicate
  - object
  - resources/blank nodes
  - resources
  - resources/blank nodes/literals

- **graphs**
  - each triple is associated with a certain graph (or a default graph)
  - extension of triples to quads:
    - graph
    - subject
    - predicate
    - object
    - resource
    - resources/blank nodes
    - resources
    - resources/blank nodes/literals
Semantic Web Technology Stack

RDF (Resource Description Framework)
- example

http://www.w3.org/1999/02/22-rdf-syntax-ns#type

http://xmlns.com/foaf/0.1/Organization

http://dbpedia.org/ontology/locationCity

http://cityofberkeley.info

http://vocab.ox.ac.uk/projectfunding#funds

Berkeley

http://ex.org/funding2342

http://www.w3.org/2000/01/rdf-schema#label
Semantic Web Technology Stack

RDF (Resource Description Framework)

- abbreviations for URIs: prefixes

```xml
@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>

```

```xml
@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>

```
Semantic Web Technology Stack

RDF (Resource Description Framework)
• abbreviations for URIs: prefixes

```
@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.

```
Semantic Web Technology Stack

User Interface & Applications

Trust

Proof

Unifying Logic

Query: SPARQL

Ontology: OWL

Rule: RIF

RDFS

Crypto

Data interchange: RDF

XML

URI/IRI
Semantic Web Technology Stack

RDFS (RDF Schema)

- additional features for
  - classifications:
    - classes
    - resources
    - properties
    - literals
  - domain and range restrictions
  - definitions of sub-classes and sub-properties
  - comments and labels of resources
  - links to other resources providing additional informations
Semantic Web Technology Stack

OWL (Web Ontology Language)

- ontology modeling language with further features, e.g. for the statement that
  - a resource represents a concept ("class") or an instance ("entity", "individual")
  - two resource identifiers represent (not) the same entity
  - two classes of entities are disjoint (e.g. the class car and the class person)
  - two classes are the same
  - a certain property ("predicate") has cardinality constraints
  - one property is the inverse of another property

...
Semantic Web Technology Stack

SPARQL *(SPARQL Protocol and RDF Query Language)*
- language to query RDF documents or databases
- based on pattern matching and graph traversal concepts
- current versions: 1.0 and 1.1
- main template:

```
SELECT <variables that should appear in result>
WHERE {
  <query patterns>
  FILTER (<additional constraints>)
}
```
Semantic Web Technology Stack

SPARQL

- variable definition via leading '?', e.g. "?someVar"
- query pattern definition via triple pattern:
  `<subject> <predicate> <object> .`
  → every position can contain a certain resource (or literal in case of the object position) or a variable
- example: get all resources of type Organisation

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?org
WHERE {
}
```
Semantic Web Technology Stack

SPARQL

• example: get all resources of type Organisation from Berkeley

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dbo: <http://dbpedia.org/ontology/>

SELECT ?org
WHERE {
}
Semantic Web Technology Stack

SPARQL

example: get all resources of type City having a label starting with "Berk"

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>

SELECT ?city
WHERE {
    ?city rdf:type dbo:City .
    FILTER(regex(?label, "^Berk"))
}

<http://cityofberkeley.info>
Outline

• Introduction
• Semantic Web
  o Vision
  o From the Web of Documents to the Web Of Data
  o Technologies
  o Advantages
• Semantic Web and Linguistics
• RDB2RDF
  o Concepts
  o Sparqlify (SML)
• PanLex2RDF
• Demo
• Discussion
Linguistic Linked Open Data

- Working Group on Open Data in Linguistics (OKFN)
  - bring the idea of "openness" defined on http://opendefinition.org/ to the linguistic world
  - one project: Linguistic Linked Open Data cloud [http://linguistics.okfn.org/resources/llod/]

  - main focus: enhance the LLOD cloud
Linguistic Linked Open Data

Snapshot of the Linguistic Linked Open Data Cloud as a result of the MLODE workshop Sept. 2012 in Leipzig
http://sabre2012.infai.org/mlode

Open Linguistics Working Group
http://linguistics.okfn.org/llod
Standards using Semantic Web Technologies

NIF (NLP Interchange Format)

- makes certain parts of a text available via URIs
- "Welcome to Dublin in Ireland!"

```r
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix nifstr: <http://nlp2rdf.lod2.eu/schema/string/> .
@prefix doc: <http://example.com/examplodoc.html#> .

```
Standards using Semantic Web Technologies

lemon (*Lexicon Model for Ontologies*)

- modeling lexicons and machine-readable dictionaries and link them to the Semantic Web and the Linked Data cloud
Standards using Semantic Web Technologies

GOLD (General Ontology for Linguistic Description)

[http://linguistics-ontology.org/]

- ontology for descriptive linguistics
Outline

• Semantic Web
  o Vision
  o From the Web of Documents to the Web Of Data
  o Technologies
  o Advantages
• Semantic Web and Linguistics
• RDB2RDF
  o Concepts
  o Sparqlify (SML)
• PanLex2RDF
• Discussion
RDB2RDF

Relational DB containing public data:
- PanLex
- OSM
- ...

SPARQL server

Mappings
RDB2RDF Mapping Tools

- Sparqlify
  - SPARQL -> SQL Rewriter

- ontop

- D2RQ
  - Accessing Relational Databases as Virtual RDF Graphs
Sparqlify

- project of the AKSW group
  (University of Leipzig)
- http://sparqlify.org
- source code: https://github.com/AKSW/Sparqlify
- datasets run by Sparqlify
  - PanLex (http://ld.panlex.org)
  - LinkedGeoData (http://linkedgeodata.org)
  - Deutscher Wortschatz (offline due to maintenance)
  - LODStats (http://stats.lod2.eu)
  - Converted CSV files
Sparqlification Mapping Language (SML)

- SML is an independent effort for an RDB2RDF mapping language:
  - Combines syntactic elements from SQL and SPARQL
  - Aimed at being as human readable as possible
  - "Core" Expressivity equivalent to RDB to RDF Mapping Language (R2RML)
    - [http://www.w3.org/TR/r2rml/](http://www.w3.org/TR/r2rml/)
  - Some differences in "advanced" features exist:
    - R2RML supports inverse expressions
    - SML supports constraints
    - However, future adoptions of additional features possible in both languages.
SML Structure

<Prefixes>
Create View <ViewName> As
    Construct
        <Quads>
    With
        <VariableDefinitions>
    From
        <LogicalTable>
SML Structure

<Prefixes>
Create View <ViewName> As
    Construct
        <Quads>  // The quads to create
    With
        <VariableDefinitions>  // What RDF terms to create
    From
        <LogicalTable>  // Query, Table or View
Example: Creating an SML View

- example:
  lemon
Example: Creating an SML View 1/5

Prefix plx: <http://panlex.org>
Prefix lemon: <http://www.lemon-model.net/lemon#>

<table>
<thead>
<tr>
<th>mn (meaning)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mn</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
Prefix plx: <http://panlex.org>
Prefix lemon: <http://www.lemon-model.net/lemon#>

Create View Lemon_LexicalSense As

<table>
<thead>
<tr>
<th>mn (meaning)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mn</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>
Example: Creating an SML View 3/5

Prefix plx: <http://panlex.org>
Prefix lemon: <http://www.lemon-model.net/lemon#>

Create View Lemon_LexicalSense As

Construct {
  ?lexSense a lemon:LexicalSense.
}

<table>
<thead>
<tr>
<th>mn (meaning)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mn</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>
Example: Creating an SML View 4/5

Prefix plx: <http://panlex.org>
Prefix lemon: <http://www.lemon-model.net/lemon#>

Create View Lemon_LexicalSense As

Construct {
    ?lexSense a lemon:LexicalSense.
}

With

?lexSense = uri(plx:meaning, '/', ?mn)

<table>
<thead>
<tr>
<th>mn (meaning)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mn</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>
Example: Creating an SML View 5/5

Prefix plx: <http://panlex.org>
Prefix lemon: <http://www.lemon-model.net/lemon#>

Create View Lemon_LexicalSense As

Construct { ?lexSense a lemon:LexicalSense. }

With

?lexSense = uri(plx:meaning, '/', ?mn)

From

mn

<table>
<thead>
<tr>
<th>mn (meaning)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>
Mapping to different RDF Terms

?s = uri('http://ex.org/', ?id)

?la = plainLiteral(?name)
?lb = plainLiteral(?name, 'en')

?d = typedLiteral(?age, xsd:int)

?f = bNode(?id)
PanLex2RDF

- modelled the PanLex structures
- created PanLex vocabulary
- re-used established vocabularies
- data accessible under http://ld.panlex.org/rdf.html
- web viewer: http://ld.panlex.org/vsnorql/
- deployment notes: http://sparqlify.org/wiki/Deployment_PanLex
PanLex2RDF
PanLex2RDF

• integration of
  o the lemon model
  o the linguistic sign model (from GOLD)

→ make PanLex data available in other model structures
→ better interoperability
Demo

...
Thanks!

- https://github.com/AKSW/PanLex-2-RDF
- http://www.semantic-web-journal.net/content/countering-language-attrition-panlex-and-web-data-1
- http://aksw.org
- cstadler@informatik.uni-leipzig.de
- pwestphal@informatik.uni-leipzig.de
Advantages

Try to search for these things on the current Web:

• Apartments near German-French bilingual childcare in Leipzig.
• ERP service providers with offices in Vienna and Berlin.
• Researchers working on DB related topics in south-east Asia.

Information to answer such search queries is available on the Web, but opaque to current Web search. (Semantic) Data Web allows to complement text on Web pages with structured data and to intelligently combine and integrate such structured information from different sources.
Outline

• Semantic Web [30min]
  o Vision
  o From the Web of Documents to the Web Of Data
  o Technology stack
    ▪ RDF Datenmodell
    ▪ SPARQL
    ▪ Ontology Languages (RDF / RDFs / OWL2)
      • Promote Web Data Lectures in SlideWiki!
  o Advantages
    ▪ Faceted Search (DBpedia Faceted Browser)
    ▪ Paradigm: "Think about the things you want to model, rather than how to physically store the data)
    ▪ Interlinkability
  o Demo von Queries @ev Facete

• Linguistic Resources in the Semantic Web [10min]
  o Patrick’s Folien
  o MLODE

• RDB2RDF [@35, 15MIN]
  o Concepts (How does it work, ???)
  o Sparqlify (SML)
    ▪ Architektur Mapping Language
  o Mapping of PanLex (Best practices)
  o Related Work (Mapping Languages, Other Tools)
    ▪ Integration von Lemon (GOLD falls nicht zu kompliziert)
    ▪ Erstellung der DBpedia Links
  o Demo

• Discussion [@55]