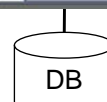
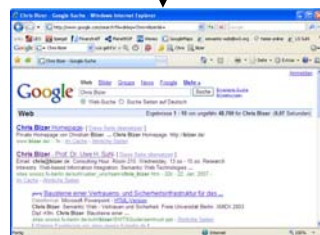


Disputationsvortrag  
14.2.2007

# Turning the Web into a Database

Diplom Kaufmann  
Christian Bizer

## What does the Web offer us today?



Christian Bizer: Turning the Web into a Database (14.2.2007)

## What do we actually want?

Wunschauto finden

Marke

Modell

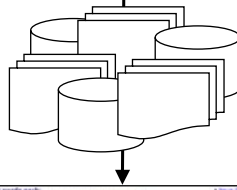
Kraftstoff  Benzin  Diesel

Preis

Leistung von  bis

EZ von  bis

Umkreis



Wunschauto wurde erstellt.

BMW, von € 9.000 bis € 15.000, PLZ: 10965, Umkreis: 20 km

1 bis 22 von 85 Fahrzeugen

Wunschauto	Bezeichnung	Auftraggeber	Wunsch. Preis	Legen Sie
	BMW X5	01-1287 Berlin, 02 20 1986, 4x4-Torque, Silber met., 130.000 km, 87 kW, Klimaanlage	€ 9.200	
	BMW X5 - 4x4-Torque	01-10377 Berlin/PLZ, 02 06 1986, Chromschwarzglanz met., 146.000 km, 210 kW, Automatik, Klimaanlage	€ 8.200	
	BMW X5 - 4x4-Torque	01-10487 Berlin, 02 20 1986, Hands / 100, (Leistungsmessung met., 180.000 km, 142 kW, Automatik, Klimaanlage)	€ 8.200	
	BMW X5	01-10248 Berlin, 02 06 1986, 4x4-Torque, Dekorgrün met., 124.000 km, 142 kW, Klimaanlage	€ 8.200	
	BMW X5 - 4x4-Torque	01-10011 Berlin, 02 20 1986, 4x4-Torque, Schwarzglanz met., 143.000 km, 210 kW, Automatik, Klimaanlage	€ 8.200	

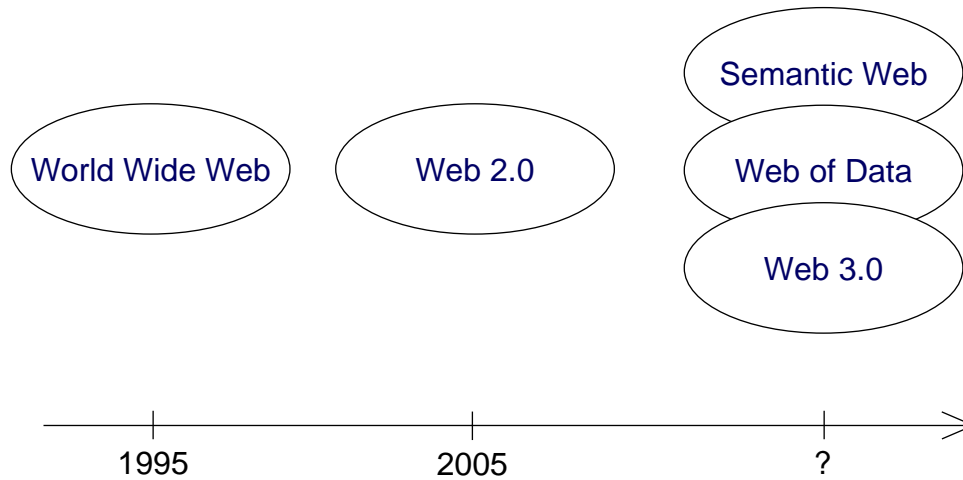
Christian Bizer: Turning the Web into a Database (14.2.2007)

## Outline

1. Development of the Web
2. Overview of some Base Technologies
3. Integrating Relational Databases into the Web
4. Summary and Outlook

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Development of the Web



Christian Bizer: Turning the Web into a Database (14.2.2007)

## Current Stage: Web 2.0

### 1. Read-and-Write Web

- User-generated Content, Weblogs, Social Networks

### 2. Introduction of more structured content formats

- Microformats:  
Small data islands within HTML pages (hCard, hCallender, hReview)
- Really Simple Syndication (RSS):  
Adds explicit timestamp to Web content

**Increased structure enables more sophisticated information processing!**

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Really Simple Syndication (RSS)

```
<?xml version="1.0"?>
<rss version="2.0">
<channel>
  <title>CNN.com</title>
  <language>en-us</language>
  <item>
    <title>Bush: Iraq plan is 'best chance'</title>
    <link>http://edition.cnn.com/2007/POLITICS/...
    <description>President George W. Bush says ...
    <pubDate>Wed, 24 Jan 2007 07:34:44 EST</pubDate>
  </item>
  <item>
    <title>Rape charge for Israeli president</title>
    <link>http://edition.cnn.com/2007/POLITICS/...
    <description>Israel's attorney general has ...
    <pubDate>Wed, 24 Jan 2007 12:44:21 EST</pubDate>
  </item>
  ...

```



Christian Bizer: Turning the Web into a Database (14.2.2007)

## Use Case: Trend Scouting

**Technorati** Search for  in

Home / Search / Results for Kurnaz

3,002 results from  with  in

Customer Support Software  
Web-based customer support solution KB, Live Chat, Ticketing & more [www.SupportMagic.com](http://www.SupportMagic.com)

**IN 24 HOURS**

- "Unschuldiges Folteropfer" Murat Kurnaz doch...**  
 Politically Incorrect · 2 hours ago · [69 blogs link here](#)  
 lieber\_kurnaz.jpg In der Bewertung von Murat Kurnaz (Foto) ... die USA zu kämpfen. Kurnaz' damaliges Flugticket von Frankfurt nach Karatschi wurde mit einer EC ... Kurnaz zum militanten Islam bekehrt. Die Bild berichtet: Tatsache ist, noch 2005 waren sich deutsche
- Schlagzeilen 24. Januar 2007**  
 gedankensplitter... · 2 hours ago · [1 blog links here](#)  
 Bushs Schwänzenesang Bush ist unpopulärer wie nie zuvor Steinmeier und der Fall Kurnaz Haftentlassung früherer RAF-Terroristen weiter umstritten Gnade vor Recht für Mohnhaupt und Klar
- Steinmeier in Not**  
 Der Linksbote · 2 hours ago · [65 blogs link here](#)  
 Der deutsche Außenminister Frank-Walter Steinmeier (SPD) ... Kurnaz frei zulassen. Das brisante dabei, so soll es wohl nicht einmal Bedingungen seitens der ... Murat Kurnaz festgenommen und verschleppt. Die deutsche Bundesregierung hat davon gewusst, zumal die
- "Ich kenne solch ein Angebot nicht"**  
 sebastianbrux|de · 3 hours ago · [1 blog links here](#)  
 Bild: aus der heutigen Süddeutschen Zeitung. Es ist ein Faksimile einer BND-internen E-Mail. Steinmeier hat sich nun geäußert. Kurz, knapp, klar ... dokumentiert wird: "USA sehen die Unschuld von Murat Kurnaz als anwesen an. Er soll in sechs bis acht Wochen ... 2002, Murat Kurnaz freizulassen, nicht angenommen hat." (taz). Eine ausführliche
- Wir machen weiter unsere Propaganda...**  
 Gegenstimme · 3 hours ago · [50 blogs link here](#)  
 ... auch wenn "US- und dt. Geheimdienste [...] anderer Ansicht als die Gegenstimme [sind]" ... Bremen vom Mai 2002, es bestehe Grund zu der Annahme, dass Kurnaz nach Pakistan gereist sei, um von ... verhaftet, drum wird da wohl noch mehr gewesen sein. Aber lesen wir weiter. Kurnaz' Flugticket von

welcome to the social

zune

Advertise on Technorati

**Mentions by Day**

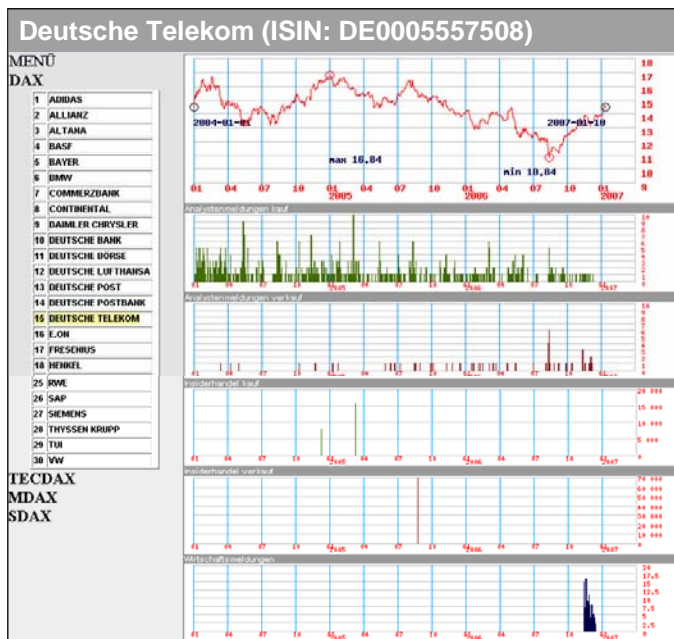
Posts that contain Kurnaz per day for the last 30 days.

- Crawls 63.2 million RSS feeds

- Provides keyword search and trend analysis

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Use Case: Financial Information Analysis



Christian Bizer: Turning the Web into a Database (14.2.2007)

- Diploma Thesis
- Aggregates RSS feeds from:
  - 296 analyst houses
  - 700 newspapers
  - 1 feed about insider trades

## Structure enables Information Processing

- Tim Berners-Lee (MIT/W3C)
- Basic Ideas
  - Publish pure data in addition to HTML pages on the Web
  - Set links between data items within different data sources
- Realization
  - Pushed by W3C with several standards
  - World wide research effort
  - Industry is getting involved (Oracle, HP, IBM)
- Base Technologies
  - Resource Description Framework (RDF)
  - Query Language for RDF (SPARQL)

Christian Bizer: Turning the Web into a Database. Freie Universität Berlin, 14.2.2007

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Resource Description Framework (RDF)

- RDF provides a simple data model for publishing data on the Web.
- Information is represented in the form of triples.

### Triple: Chris knows Richard

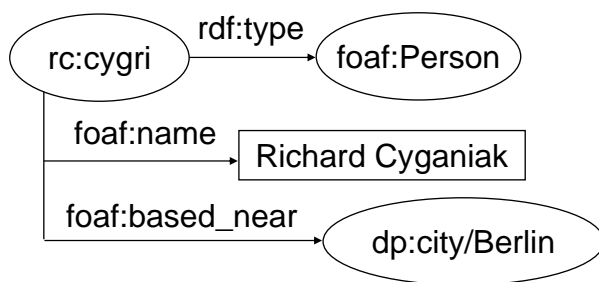
**Subject:** `http://www.bizer.de#chris`  
**Predicate:** `http://xmlns.com/foaf/0.1/knows`  
**Object:** `http://richard.cyganiak.de/foaf.rdf#cygri`

### Triple: Chris was born 1973-09-20

**Subject:** `http://www.bizer.de#chris`  
**Predicate:** `http://www.w3.org/2001/vcard-rdf/3.0#BDAY`  
**Object:** `"1973-09-20"^^http://www.w3.org/2001/XMLSchema#date`

Christian Bizer: Turning the Web into a Database (14.2.2007)

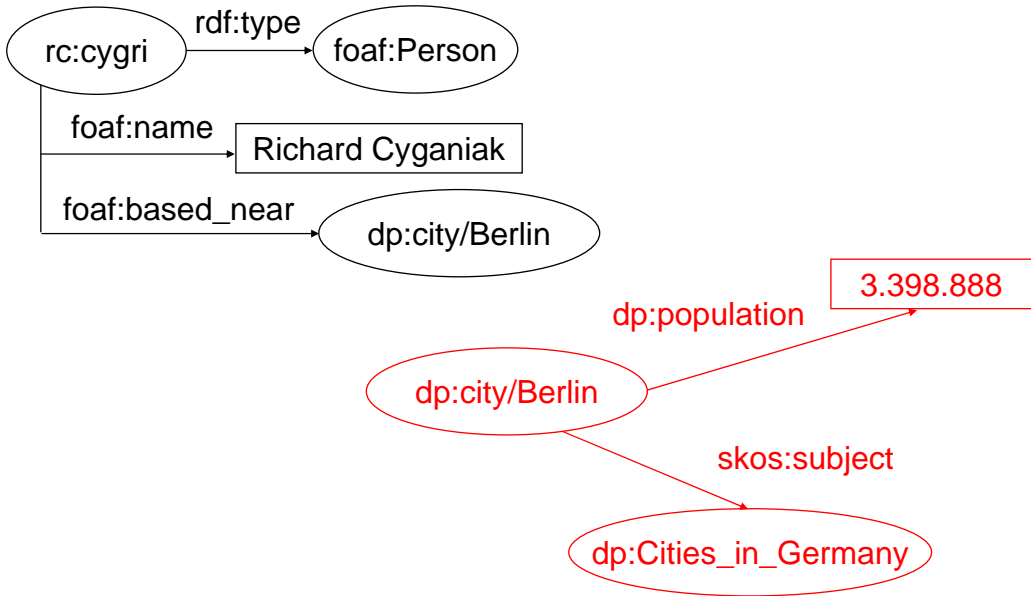
## Every Triple is a Hyperlink



**GET /city/Berlin HTTP/1.0**  
**Accept: application/rdf+xml**

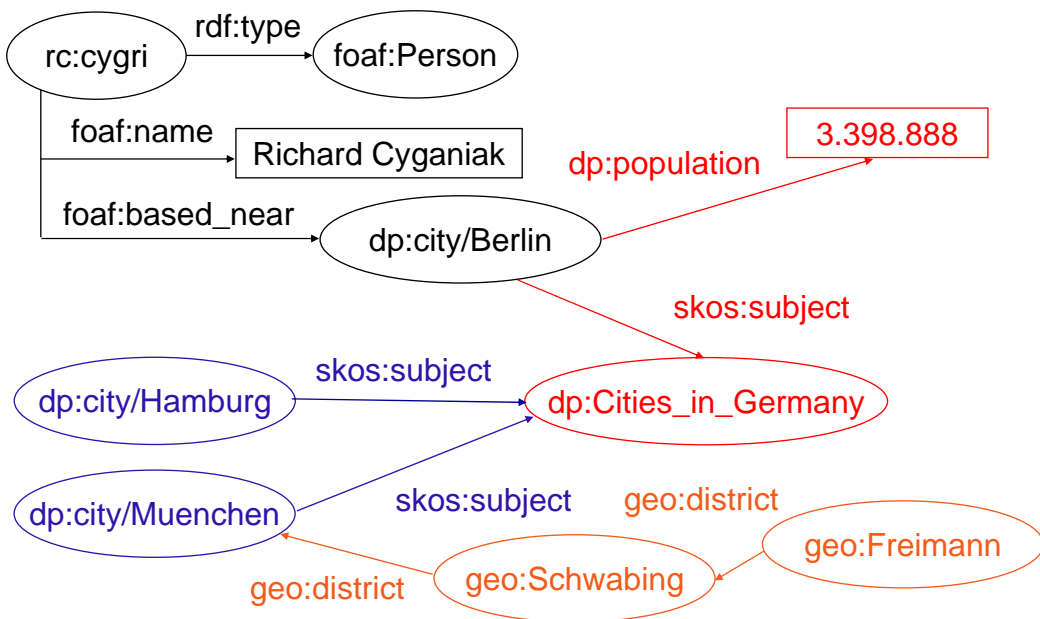
Christian Bizer: Turning the Web into a Database (14.2.2007)

## Every Triple is a Hyperlink



Christian Bizer: Turning the Web into a Database (14.2.2007)

## Every Triple is a Hyperlink



Christian Bizer: Turning the Web into a Database (14.2.2007)

## Richard Cyganiak

URI: 

Property	Value	Sources
event	...	<a href="#">G2</a>
type	<a href="http://xmlns.com/foaf/0.1/Person">http://xmlns.com/foaf/0.1/Person</a>	<a href="#">G1</a> <a href="#">G2</a> <a href="#">G3</a> <a href="#">G4</a>
seeAlso	<a href="http://richard.cyganiak.de/cygri.rdf">http://richard.cyganiak.de/cygri.rdf</a>	<a href="#">G2</a>
seeAlso	<a href="http://richard.cyganiak.de/foaf.rdf">http://richard.cyganiak.de/foaf.rdf</a>	<a href="#">G3</a>
nearest airport	...	<a href="#">G1</a>
phone	tel:+49-175-5630408	<a href="#">G1</a>
sameAs	Richard Cyganiak	<a href="#">G1</a>
based_near	...	<a href="#">G1</a>
based_near	Berlin	<a href="#">G1</a>
based_near	<a href="http://sws.geonames.org/2950159/">http://sws.geonames.org/2950159/</a>	<a href="#">G1</a>
currentProject	<a href="http://page.mi.fu-berlin.de/~cyganiak/foaf.rdf#StatCvs">http://page.mi.fu-berlin.de/~cyganiak/foaf.rdf#StatCvs</a>	<a href="#">G3</a>
currentProject	<a href="http://www.wiwiss.fu-berlin.de/suhl/bizer#d2rq">http://www.wiwiss.fu-berlin.de/suhl/bizer#d2rq</a>	<a href="#">G3</a>
depiction		<a href="#">G4</a>
gender	male	<a href="#">G1</a>

## Berlin

URI: 

Property	Value	Sources
population	3398888	<a href="#">G2</a>
type	<a href="http://dbpedia.org/City">http://dbpedia.org/City</a>	<a href="#">G2</a>
comment	Berlin is the capital city and one of the sixteen Federal States of Germany. It is the country's largest city in area and population, and the second most populous city in the European Union.	<a href="#">G2</a>
comment	Berlin ist die deutsche Bundeshauptstadt und als Stadtstaat ein eigenständiges Land der Bundesrepublik Deutschland. Berlin ist die bevölkerungsreichste und flächengrößte Stadt Deutschlands und nach Einwohnern die zweitgrößte Stadt der EU.	<a href="#">G2</a>
label	Berlin	<a href="#">G2</a>
sameAs	<a href="http://sws.geonames.org/2950159/">http://sws.geonames.org/2950159/</a>	<a href="#">G2</a>
subject	<a href="http://dbpedia.org/resource/category/Berlin">http://dbpedia.org/resource/category/Berlin</a>	<a href="#">G2</a>
subject	<a href="http://dbpedia.org/resource/category/Capitals_in_Europe">http://dbpedia.org/resource/category/Capitals_in_Europe</a>	<a href="#">G2</a>
subject	<a href="http://dbpedia.org/resource/category/Cities_in_Germany">http://dbpedia.org/resource/category/Cities_in_Germany</a>	<a href="#">G2</a>
subject	<a href="http://dbpedia.org/resource/category/German_state_capitals">http://dbpedia.org/resource/category/German_state_capitals</a>	<a href="#">G2</a>
subject	<a href="http://dbpedia.org/resource/category/Host_cities_of_the_Summer_Olympic_Games">http://dbpedia.org/resource/category/Host_cities_of_the_Summer_Olympic_Games</a>	<a href="#">G2</a>
subject	<a href="http://dbpedia.org/resource/category/States_of_Germany">http://dbpedia.org/resource/category/States_of_Germany</a>	<a href="#">G2</a>
sourceURL	<a href="#">Berlin</a>	<a href="#">G1</a>
depiction		<a href="#">G2</a>
page	<a href="http://en.wikipedia.org/wiki/Berlin">http://en.wikipedia.org/wiki/Berlin</a>	<a href="#">G2</a>
is birthplace of	<a href="#">Adolf von Baeyer</a>	<a href="#">G2</a>



## The SPARQL Query Language

- W3C standard query language for RDF data
- Example query

Give me the name and birthday of all people that were born after January 1st, 1970.

```
SELECT ?name ?birth
WHERE { ?person foaf:name ?name .
        ?person vcard:BDAY ?birth .
FILTER (?birth > "1970-01-01"^^xsd:date) }
```

?name	?birth
"Chris"	"1973-09-20"^^xsd:date
"Andy"	"1980-03-12"^^xsd:date
...	...

Christian Bizer: Turning the Web into a Database (14.2.2007)

## What to query?

1. Querying a Single Data Source
2. Querying Multiple Data Sources
3. Querying the complete Semantic Web

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Querying a Single Data Source

- Data source: [dbpedia.org](http://dbpedia.org)
- 19 million triples extracted from Wikipedia info boxes.
- Example Query

Give me all persons from Wikipedia that were born in Berlin before 1900.

```
SELECT ?name ?birth ?description
WHERE {
  ?person dbpedia:birthplace
    dbpedia:resource/city/Berlin .
  ?person dbpedia:birth ?birth .
  ?person foaf:name ?name .
FILTER (?birth < "1900-01-01"^^xsd:date) }
```

Landesflagge	Landeswappen
	
(Details)	
Basisdaten	
Fläche:	891,82 km² (14.)
Einwohner:	3.402.312 (8.) (30. September 2006)
Bevölkerungsdichte:	3.815 Einw. je km² (1.) als Bundesland, (2.) als Stadt
BIP:	79,60 Mrd. € (2005)
Höhe:	34–115 m
Geographische Lage:	52° 31' N, 13° 25' O 
Postleitzahlen:	10001–14199
Vorwahl:	030
Kfz-Kennzeichen:	B
Gemeineschlüssel:	11 0 00 000
ISO 3166-2:	DE-BE
UNLOCODE:	DE BER
Website:	<a href="http://www.berlin.de">www.berlin.de</a> 

Christian Bizer: Turning the Web into a Database (14.2.2007)

## SPARQL Explorer for <http://dbpedia.org/sparql>

### SPARQL:

```
PREFIX db: <http://dbpedia.org/resource/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX map: <file:///Users/richard/dbpedia/dbpedia-mapping.n3#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX link: <http://richard.cyganiak.de/2006/link#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
```

```
SELECT ?name ?birth ?description WHERE {
  ?person dbpedia:birthplace <http://dbpedia.org/resource/city/Berlin> .
  ?person dbpedia:birth ?birth .
  ?person foaf:name ?name .
  ?person rdfs:comment ?description .
  FILTER (?birth < "1900-01-01"^^xsd:date) .
}
```

ORDER BY ?name

Results:

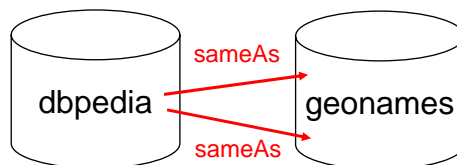
### Browse:

- [Classes](#)
- [Properties](#)

### SPARQL results:

name	birth	description
"Adalbert of Prussia"@en	"1811-10-29"^^xsd:date	"Heinrich Wilhelm Adalbert Prince of Prussia (October 29, 1811 in Berlin – June 6, 1873 in Karlsbad) was a Prussian naval theorist and admiral. A son of Prince William, the youngest brother of King Frederick William III, Adalbert was instrumental during the Revolutions of 1848 in founding the first unified German fleet. During the 1850s he helped establish the Prussian Navy."@en
"Adolf von Baeyer"@en	"1835-10-31"^^xsd:date	"Johann Friedrich Wilhelm Adolf von Baeyer (October 31, 1835 - August 20, 1917) was a German chemist who synthesized indigo, and was the 1905 recipient of the Nobel Prize in Chemistry . Born in Berlin, he initially studied mathematics and physics at Berlin University before moving to Heidelberg to study chemistry with Robert Bunsen. There he worked primarily in August Kekulé's laboratory, earning his doctorate (from Berlin) in 1858."@en
"Albert Lortzing"@en	"1801-10-23"^^xsd:date	"Gustav Albert Lortzing (October 23, 1801 - January 21, 1851) was a German composer."@en
"Alexander Gottlieb Baumgarten"@en	"1714-07-17"^^xsd:date	"Alexander Gottlieb Baumgarten (July 17, 1714 &ndash; May 26, 1762) was a German philosopher. He was a follower of Leibniz and Christian Wolff, and gave the term aesthetics its modern meaning aesthetics."@en
"Alexander Mitscherlich"@en	"1836-05-28"^^xsd:date	".This article is about the chemist. Go to Alexander Mitscherlich (Psychology) for the psychologist."@en

## Querying Multiple Data Sources



### Three Options:

#### 1. Virtual Integration

- Query Routing: Split query and ask data sources for the things that they can answer.
- HU Berlin: DARQ
- Complicated and slow.

#### 2. Materialized Integration

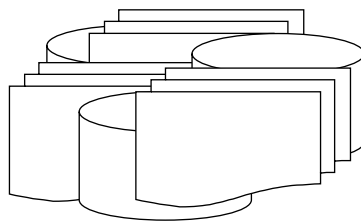
- Use Links between data items to crawl all data into a single repository.
- Crawlers under development: Zitgist, OpenLink USA and SWSE, DERI Irland
- Fast, but requires huge RDF repositories.

#### 3. Materialization On-the-Fly

- Crawl only data that is needed while answering the query.
- FU Berlin: Semantic Web Client Library
- Works, but is really slow.

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Querying the complete Semantic Web



### Two Options:

#### 1. Materialized Integration

- Use Links between data items to crawl all data into a single repository.
- Worked for HTML (Google) and RSS (Technorati)

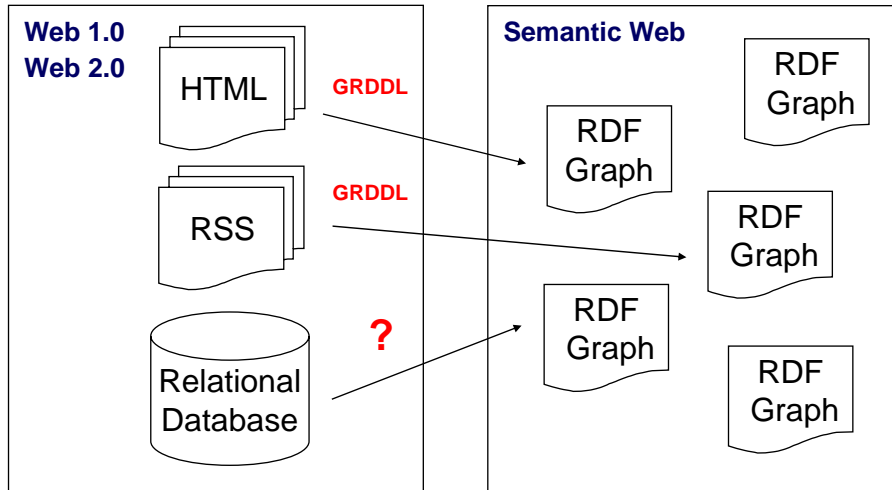
#### 2. Materialization On-the-Fly

- Crawl only data that is needed while answering the query.
- Works, but is really slow.

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Viewing the Web as RDF

- Now we have a data model and a query language.
- What is still needed is an RDF view on the complete Web.



Christian Bizer: Turning the Web into a Database (14.2.2007)

## Integrating Relational Databases into the Semantic Web

- What is needed?
  - A way to map data between the relational data model and RDF
  - Some software to make the resulting RDF data Web-accessible
- We developed two artifacts:
  - The D2RQ Mapping Language
  - D2R Server

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Mapping Relational Data to RDF

### People

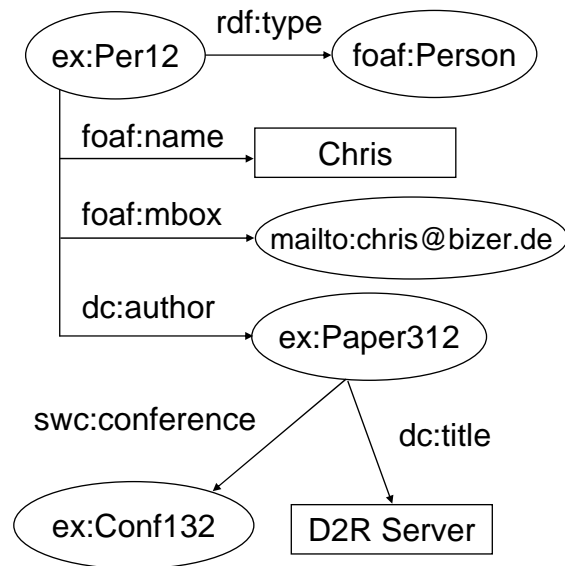
<u>ID</u>	name	email
12	Chris	chris@bizer.de

### Papers

<u>ID</u>	title	confID
312	D2R Server	132

### Rel\_People\_Papers

<u>PersonID</u>	<u>PaperID</u>
12	312



Christian Bizer: Turning the Web into a Database (14.2.2007)

## The mapping has to bridge:

### ■ Data Model Heterogeneity

- relations versus a graph
- primary keys versus URIs

### ■ Schema Heterogeneity

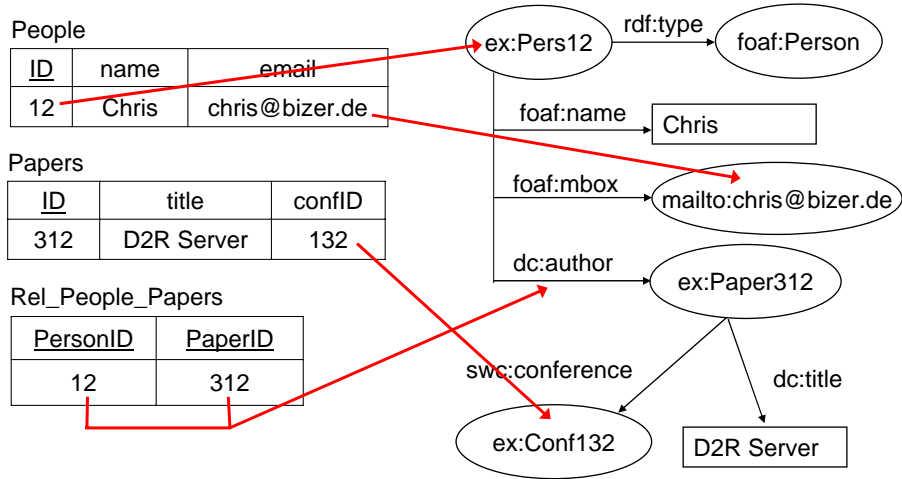
- name versus given name, surname

### ■ Syntactic Heterogeneity

- timestamp versus "2007-02-14"^^xsd:data

Christian Bizer: Turning the Web into a Database (14.2.2007)

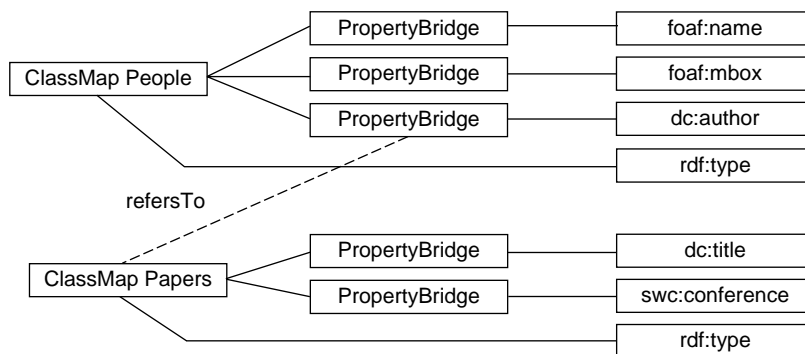
## Value Correspondences



■ D2RQ is a language for writing these correspondences down.

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Example D2RQ Mapping



Christian Bizer: Turning the Web into a Database (14.2.2007)

## Class Map

### People

<u>ID</u>	name	email
12	Chris	chris@bizer.de

```
:PeopleClassMap a d2rq:ClassMap;  
  d2rq:class foaf:Person;  
  d2rq:uriPattern  
    "http://example.org/person@@People.ID@@".
```

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Property Bridge

### People

<u>ID</u>	name	email
12	Chris	chris@bizer.de

```
:PeopleEmailProperty a d2rq:PropertyBridge;  
  d2rq:belongsToClassMap :PeopleClassMap;  
  d2rq:property foaf:mbox;  
  d2rq:uriPattern "mailto:@@People.email@@".
```

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Joins

### People

<u>ID</u>	name	email
12	Chris	chris@bizer.de

### Papers

<u>ID</u>	title	confID
312	D2R Server	132

### Rel\_People\_Papers

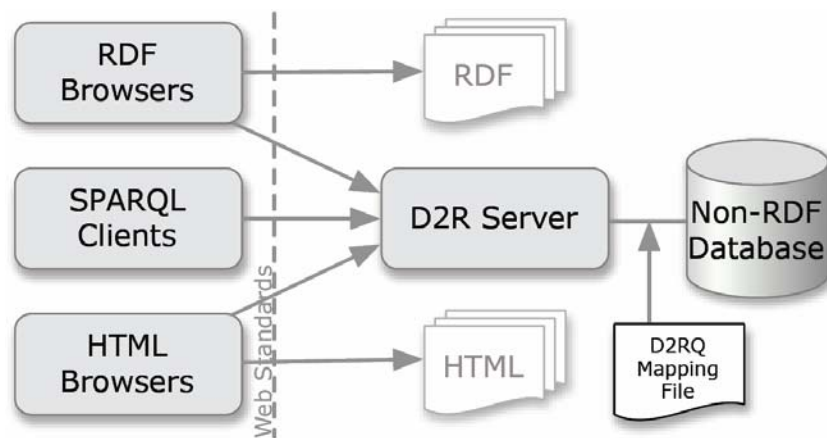
<u>PersonID</u>	<u>PaperID</u>
12	312

```
:PeoplePaperRelation a d2rq:PropertyBridge;  
  d2rq:belongsToClassMap :PeopleClassMap;  
  d2rq:property dc:author;  
  d2rq:refersToClassMap :PapersClassMap;  
  d2rq:join "People.ID=Rel_People_Papers.PersonID";  
  d2rq:join "Rel_People_Papers.PersonID=Papers.ID";
```

Christian Bizer: Turning the Web into a Database (14.2.2007)

## D2R Server

- open source project (GNU General Public License)
- around 1500 downloads (138 in January 2007)

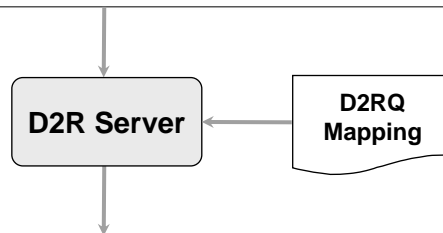


Christian Bizer: Turning the Web into a Database (14.2.2007)



## Example Rewriting from SPARQL to SQL

```
SELECT ?person ?mbox
WHERE { ?person foaf:name "Chris" .
        ?person foaf:mbox ?mbox . }
```



```
SELECT People.ID, People.email
FROM People
WHERE People.name = "Chris";
```

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Usage of D2RQ and D2R Server

### ■ Public D2R Servers on the Web

- dbpedia, FU Berlin and Universität Leipzig
- DBLP Bibliography, FU Berlin
- Information Systems Group, FU Berlin
- Roller, Weblog Server, Sun Microsystems
- Images of Fruitfly Embryogenesis, Berkeley Drosophila Genome Project

### ■ OEM Distribution

- Part of the TopBraid Composer Ontology Editor
- TopBraid is a startup company in Mountain View, California
- Customers: NASA, US Military

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Publications and Outreach

### ■ Publications

- **Christian Bizer, Andy Seaborne: D2RQ -Treating Non-RDF Databases as Virtual RDF Graphs.** Poster at 3rd International Semantic Web Conference, Hiroshima, Japan, November 2004.
- **Christian Bizer, Richard Cyganiak: D2R-Server - Publishing Relational Databases on the Web as SPARQL-Endpoints.** 15th International World Wide Web Conference, Edinburgh, Scotland, May 2006.

### ■ Outreach

- **Ivan Herman: A short Introduction into the Semantic Web.**  
Talks at multiple occasions, December 2006 and January 2007
- **Tim Berners-Lee: Semantic Web URIs.**  
Talk at MIT, Cambridge, USA, January 2007
- **Bob DuCharme: Relational Database Integration with RDF/OWL.**  
XML 2006 Conference, Boston, USA, December 2006
- **Alois Reibauer: Tool-Workshop Triple Stores, SPARQL und D2RQ.**  
Semantic Web School Austria, Wien, January 2007 (340 €excl. USt.)

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Summary

- Gave an overview about the development of the Web towards a database
- Introduced RDF and SPARQL
- Described the D2RQ Mapping Language
- Described D2R Server

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Outlook

- **Web 2.0 kick-started Semantic Web roll-out.**
- **There are more and more RDF data sources on the Web.**
  - FOAF, DOAP, SIOC Communities
  - dbpedia
  - geonames
  - RDF Book Mashup
  - WordNet
  - MusicBrainz
- **Semantic Web Search Engines are getting into place.**
  - Zitgist, Openlink USA
  - SWSE, DERI Irland
  - Swoogle, University of Maryland USA
- **Things will stay exiting in the years to come.**

Christian Bizer: Turning the Web into a Database (14.2.2007)

## Thanks!

**This slides are online at**

**<http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/disp-bizer.pdf>**

Christian Bizer: Turning the Web into a Database (14.2.2007)