Seminarpraktikum
Semantic Web Engineering

Thanh Tran, Günter Ladwig, Andreas Wagner
Agenda

Introduction
- Semantic Web
- Pointers

Seminar
- Expectations and Goals
- Topics
- Important Dates

Tools
- Fourty2
- Other software

Discussion
- Form groups and choose topic
- Questions?
INTRODUCTION
Semantic Web – Motivation – BBC Music Site

Eric Clapton
Born 30 March 1945.

Biography
Eric Patrick Clapton, CBE (born 30 March 1945) is an English blues-rock guitarist, singer, songwriter and composer. Clapton has been inducted into the Rock and Roll Hall of Fame as a solo performer, as a member of rock bands; the Yardbirds and Cream. Clapton is the only person ever to be inducted three times. Often viewed by critics and fans alike as one of the most important and influential guitarists of all time, Clapton was ranked fourth in Rolling Stone magazine's list of the "100 Greatest Guitarists of All Time" and #15 on their list of the "100 Greatest Artists of All Time."

Latest Tracks Played On The BBC

- **Promises**
  BBC Radio 2 | Ken Bruce 22/02/2016

- **Bad Love**
  BBC Radio 2 | Alex Lester 22/02/2016

- **Lay Down Sally**
  BBC Radio 2 | Chris Evans Breakfast 18/02/2016

- **I Ain't Gonna Stand For It**
  BBC Radio 2 | Alex Lester 15/02/2016

- **Wonderful Tonight**
  BBC Radio 2 | Ken Bruce 18/02/2016

Audio Previews From Latest Album Review

Me And Mr Johnson
Milkcow's Calf Blues
Come on in My Kitchen
Semantic Web – Motivation – BBC Music Site

BBC – Music – Eric Clapton

http://www.bbc.co.uk/music/artists/61b6900-0618-4f1e-b835-bcc1784294

Read more at Wikipedia...

Wikipedia This entry is from Wikipedia, the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the GNU Free Documentation License. If you find this biography content factually incorrect, defamatory or highly offensive you can edit this article at Wikipedia. Find out more about our use of this data.

Links & Information
- Official homepage at ericclapton.com
- Fanpage at whereseric.com
- Wikipedia article on Eric Clapton
- MySpace at ericclapton.myspace.com
- Last.fm page on Eric Clapton
- Myspace profile of Eric Clapton


Links & information come from MusicBrainz. You can add or edit information about Eric Clapton at musicbrainz.org. Find out more about our use of this data. The BBC is not responsible for the content of external sites.

Latest News Stories

Information displayed about artists played on BBC programmes is incomplete and may not reflect actual play count information.
Semantic Web – Motivation
How to build the BBC Music Site?

1. Attempt

- Site editors roam the Web for new facts
  - may discover further links while roaming
- They update the site **manually**
- And the site gets soon **out-of-date**
Semantic Web – Motivation
How to build the BBC Music Site?

2. Attempt

- Editors roam the Web for new data published on Web sites
- “Scrape” the sites with a program to **extract the information**
  - ie, *write some code* to incorporate the new data
- Easily get **out of date** again 😞
Semantic Web – Motivation
How to build the BBC Music Site?

3. Attempt

- Editors roam the Web for **new data via API-s**
- Understand those…
  - input, output arguments, datatypes used, etc
- **Write some code** to incorporate the new data
- Easily **get out of date** again…
Semantic Web – Motivation
What did the BBC do?

- Use external, public datasets
  - Wikipedia, MusicBrainz, …
- They are available *as data*
  - not API-s or hidden on a Web site
  - data can be extracted using, eg, HTTP requests or standard queries

- Use the Web of Data as a Content Management System
- Use the community at large as content editors
Semantic Web – Motivation Summary

- With increased use of computers more and more data is being stored
  - Organisations rely on data for business decisions
  - Data drives policy decisions in government
  - Individuals rely on data from the Web for information and communication

- Data volumes explode
  - More and more data available on the Web is represented in Semantic Web standards
  - Linking Data initiative [1]

- Semantic Web technologies facilitate the integration of data from multiple sources
  - Combining data from multiple sources enables insights
Semantic Technologies

- Useful for **data publishing, exchange, and integration**
- Insights possible when combining data from multiple sources
- Semantic Web technologies, standardised by the W3C [2], are mature:
  - **RDF** recommendation in 1999, update in 2004
  - RDFa (RDF in HTML) note in 2008
  - **RDFS** recommendation in 2004
  - **SPARQL** recommendation in 2008
  - **OWL** recommendation in 2004, update in 2009
- Linked Data comprises of a few principles for data publishing on the web
Semantic Web Technologies
Resource Description Framework (RDF)

- Directed, labeled graph
- triple(subject, predicate, object)
  - subject: URI or blank node
  - predicate: URI
  - object: URI or blank node or RDF literal

- RDF/XML is the most widely deployed serialization
- Other serializations possible (N-Triples, Turtle, Notation3…)
- Quadruples (or quads) used as internal representation when integrating data
  - quad(subject, predicate, object, context)
    - context: URI (used to store origin of triple)
Resource Description Framework (RDF) Example – We start with a book ...
Resource Description Framework (RDF) Example – A simplified bookstore data

<table>
<thead>
<tr>
<th>ID</th>
<th>Author</th>
<th>Title</th>
<th>Publisher</th>
<th>Year</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Homepage</th>
</tr>
</thead>
<tbody>
<tr>
<td>id_xyz</td>
<td>Ghosh, Amitav</td>
<td><a href="http://www.amitavghosh.com">http://www.amitavghosh.com</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Publisher’s name</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>id_qpr</td>
<td>Harper Collins</td>
<td>London</td>
</tr>
</tbody>
</table>
Resource Description Framework (RDF) Example

- RDF triples form a directed, labeled graph

```
<rdf:Description rdf:about="http://.../isbn/2020386682">
  <f:year>2000</f:year>
</rdf:Description>
```

Note: namespaces are used to simplify the URI-s
Resource Description Framework (RDF) Example

- “Internal” nodes (aka blank nodes)
  - Consider the following statement:
    - “the publisher is a «thing» that has a name and an address”
  - Until now, nodes were identified with a URI. But…
  - …what is the URI of «thing»?

```
<rdf:Description rdf:about="http://.../isbn/000651409X">
  <a:publisher rdf:nodeID="A234"/>
</rdf:Description>

<rdf:Description rdf:nodeID="A234">
  <a:p_name>HarpersCollins</a:p_name>
  <a:city>HarpersCollins</a:city>
</rdf:Description>
```

Note: Internal = these resources are not visible outside
Resource Description Framework Schema

- First step towards the “extra knowledge”:
  - define the terms we can use
  - what restrictions apply
  - what extra relationships are there?

- RDFS defines resources and classes:
  - everything in RDF is a “resource”
  - “classes” are also resources, but…
  - …they are also a collection of possible resources (i.e., “individuals”)

- Relationships are defined among resources:
  - “typing”: an individual belongs to a specific class
    - “The Glass Palace” is a novel
  - “subclasseing”: all instances of one are also the instances of the other
Resource Description Framework Schema

- Property is a special class (rdf:Property)
  - properties are also resources identified by URI-s
- There is also a possibility for a “sub-property”
  - all resources bound by the “sub” are also bound by the other
- Range and domain of properties can be specified
  - i.e., what type of resources serve as object and subject
Resource Description Framework Schema Example

- Classes, resources in RDF(S)

RDFS defines the meaning of these terms (these are all special URI-s, we just use the namespace abbreviation)
Points

- **Books**
  - *Foundations of semantic web technologies*
    - Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph
  - **Linked Data**
    - Tom Heath and Chris Bizer
      - [http://linkeddatabook.com/editions/1.0/](http://linkeddatabook.com/editions/1.0/)
  - **Handbook on Ontologies**
    - Steffen Staab

- **Slides, Talks etc**
  - [http://www.w3.org/2001/sw/](http://www.w3.org/2001/sw/)

- **Tools**
  - Fourty2 Platform
System *Fourty2* – Demo
Expectations

- Have **system running** at the end
- Keep the **usability** in mind → system must be understandable to anybody
- Keep **reaction times** in mind → a running system must be fast

- What tools are used is up to you. However, we support our system **Fourty2**
- Work as a **team**
  - Talk to each other
  - Organize your work
  - Implement and test your ideas together
Expectations

- Employ **waterfall model (or similar model)**

See also: Royce, Winston (1970), "Managing the Development of Large Software Systems" [3]
Expectations – Keep track of time …

25/ Apr/  5/ Mai/  15/ Mai/  25/ Mai/  4/ Jun/  14/ Jun/  24/ Jun/  4/ Jul/  14/ Jul/

Entwicklung  3
Realisierung  4
Planung  6
Ideenpräsentation
Aufgaben definieren und verteilen
Kick-Off Programmierung
HTML Code
Google Maps API
Homepage
JavaScript
CSS Code
DBpedia
Zwischenpräsentation
Foaf
Revyu
RDF Book Mashup
Routenberechnung
Routenberechnung
Datensätze
Computer
Mobil
Abschlusspräsentation

Requirements
Design
Implementation
Verification
Topics …

- Short overview of predefined topics
  - Hybrid search *(text and structured constraints)*
    - Integrate existing hybrid search code in frontend
    - Make hybrid search interface more usable, e.g., a result preview
  - Linked data source selection
    - Describe data source in intuitive manner
    - Allow user to select data source for her queries
  - Pivotsearch
    - Allow users to „jump“ from one result to another
    - For instance, a result may contain the U.S. presidents; now users may want to see the universities, which the presidents attended
  - Context-aware search
    - Allow users to issue query using „context“ information
    - For instance, find all tire manufacturer trusted by ADAC
Topics …

- Short overview of predefined topics
  - Hybrid content authoring (text and structured data)
    - Create an intuitive interface for modifying structured and unstructured data
  - Result set visualisation
    - Allow a flexible result presentation (not only lists), depending on the contents of the current result set
  - Natural Syntax for SPARQL
    - More natural and easier syntax for basic SPARQL expressions
    - Parse new syntax and translate to SPARQL
  - SPARQL query builder
    - Create an intuitive SPARQL query builder
    - Enable (for instance) a result preview or warn users, if a query may lead to an empty result
- Other topics are also possible → discussion at the end
Getting a good grade …

- **Grading**
  - Code ~ 50%
  - Documentation (waterfall or similar model) ~ 40%
  - Final presentation ~ 10%
Goals beyond a good grade ...
Elsevier Apps for Science Challenge

- http://appsforscience.com/
- Deadline: July 31st
- Goal: science applications on top of Fourty2
Goals beyond a good grade ... 
Open Data Challenge

- http://opendatachallenge.org
- Deadline: **June 5th**
- Goal: Running application employing open government data

€20,000 to win 
48 days left to enter

**What is it?**

- What are the prizes?
- How can I enter?
- Who is behind it?
- What are the rules?
- Where can I find open data?

**European public bodies produce thousands upon thousands of datasets every year - about everything from how our tax money is spent to the quality of the air we breathe.**

We are challenging designers, developers, journalists, researchers and the general public to come up with something useful, valuable or interesting using open public data.

There are four main strands to the competition:

- **Ideas** – Anyone can suggest an idea for projects which reuse public information to do something interesting or useful.
Goals ... What does that mean for me?

Getting your system running and usable may be a lot of work

You can learn a lot of things (e.g., working in a group or developing software)
You can get a reward beyond „some“ grade
Important Dates

- **Kick-Off** today: Think about topic, group etc.
- Until **May 4th, 5.15 pm (room 226)**. From Groups, choose system (and have it up and running) and topic

![Work in group](arrow) ![Support from AIFB](arrow)

- **June 8th, 5.15 pm (room 226)**. Present and discuss intermediate results and problems

![Work in group](arrow) ![Support from AIFB](arrow)

- **Final Presentation**, beginning of July: Turn in code and documentation

![Work in group](arrow) ![Support from AIFB](arrow)
Final Presentation

- Present your running application
  - Is it usable?
  - Does it scale w.r.t. time and data size?
  - Live demo …

- Turn in documentation
  - ~ 10 pages, use LaTeX and write in English
    - Employ waterfall model
      - What are the requirements?
      - How is your system designed (UML)?
      - How is it implemented (most important classes, other software packages etc)?
      - How does your software fulfill the requirements (testing)?
  - Source code of your implementation
TOOLS
System *Fourty2*

*Fourty2* is a state-of-the-art **search system developed at AIFB**

- ... is a layered and extendable system
- ... can handle structured and unstructured search
- ... can handle keyword, structured and hybrid queries
- ... uses a standard interface (SAIL interface [4]) for communication between the layers
- ... comes with a GUI framework on top
- ... comes with documentation on how to set it up and run it
System *Fourty2* – Architecture

- *Fourty2* is a state-of-the-art **search system developed at AIFB**
- ... is a layered and extendable system

Most topics will have focus in these layers
The information workbench [5] (IWB) is an open source framework for easily creating user interface components.

```java
FTableModel tm = new FTableModel();
tm.addColumn("Subject");
tm.addColumn("Predicate");
tm.addColumn("Object");

for (Statement st : res) {
    String[] row = new String[3];
    row[0] = st.getSubject().stringValue();
    row[1] = st.getPredicate().stringValue();
    row[2] = st.getObject().stringValue();
    tm.addRow(row);
}

FContainer container = getFacetContainer(pc);
FTable tbl = new FTable("table", ((KeywordQueryResult)
                       pc.resultSet).tm);
...
```
SAIL interface [4] is used by major RDF stores
- The interface defines (in particular) a data model, how query may be issued and what the result looks like

```java
try {
    RepositoryConnection con = myRepository.getConnection();

    try {
        TupleQuery tupleQuery =
            con.prepareTupleQuery(QueryLanguage.SPARQL,
                      queryString);
        TupleQueryResult result = tupleQuery.evaluate();

        try {
            .... // do something with the result
        } ....
    } ...
} ...
```

Other Software

- Software useful for working in groups ...
  - Wikis or Groups for communication
    - Wikis: [www.wikispaces.com](http://www.wikispaces.com) ...
    - Groups: [https://groups.google.com/](https://groups.google.com/) or [http://groups.yahoo.com/](http://groups.yahoo.com/) ...

- Subversion Server for code and document versioning
  - SVN hosting: [http://www.assembla.com/](http://www.assembla.com/) ...
  - SVN clients:
    - [http://tortoisesvn.tigris.org/](http://tortoisesvn.tigris.org/) (Win OS, Explorer)
    - [http://subclipse.tigris.org/](http://subclipse.tigris.org/) (Eclipse)

- IDE

Everything comes free
DISCUSSION
Discussion

- Form groups
- Think about topic for your group
- Questions?

1. Hybrid search (text and structured constraints)
2. Linked data source selection for query answering
3. Pivotsearch
4. Context-aware search
5. Context-aware navigation in complex web data
6. Hybrid content authoring (text and structured data)
7. Result set visualisation
8. Natural Syntax for SPARQL
9. SPARQL query builder with result preview
REFERENCES