INF3580/4580 – Semantic Technologies – Spring 2017

Lecture 1: Introduction

Martin Giese

16th January 2017

Today's Plan

1 Introduction to Semantic Technologies

2 Practicalities

3 Software

Outline

The Vision of a Semantic Web

A vision

I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web—the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize.


Tim Berners-Lee

Quoted from: *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web.* Tim Berners-Lee with Mark Fischetti. Harper San Francisco, 1999.

Let's go to the cinema!

- Kringsjå studentby, 20:00…
- "Let's go to see *My Neighbor Totoro* now!"
- Need to find out which cinema plays the movie tonight, e.g. on `http://www.google.no/movies`
- Need to find out where those cinemas are
- Need to find out which of those cinemas we can reach on time using public transport, e.g. on `http://www.ruter.no/`
- Web user needs to combine information from different sites
- Essentially a database join!
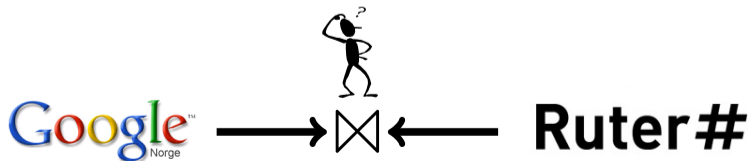
The Solution?

- Wait for Google to produce a Cinema+Public Transport mashup?

Ruter#

- But what about
  - Real estate + public transport?
  - Plane schedules and pricing + weather information?
  - Car rental + tourism?
  - Public information + private information (preferences, calendar, location, etc.)
- Can hardly wait for a separate mashup for each useful combination!
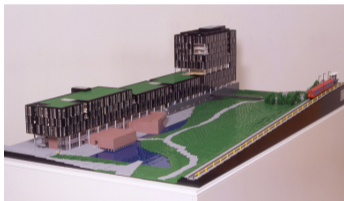
A Web of Data!

Imagine. . .

- All those websites publish their information in a machine-readable format.
- The data published by different sources is *linked*
- Enough domain knowledge is available to machines to make use of the information
- User-agents can find and combine published information in appropriate ways to answer the user's information needs.

But How?

- This sounds like a nice idea, but how can it work?
- There has been a lot of hype around the Semantic Web!
- Visions instantly transformed to promises (and $$$)
- Most of this simply does not work (yet?)
- But then, a lot does!
- Current *partial* solutions build on traditions of
  - Modelling
  - Calculating with Knowledge
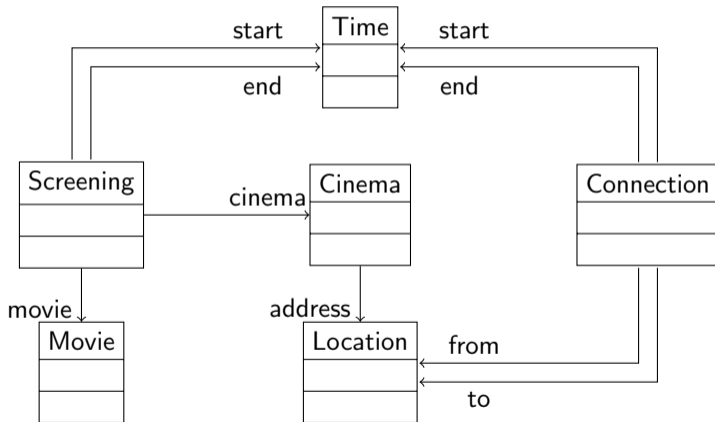 - Information Exchange

Building Models

- A *model* is a simplified representation of certain aspects of the real world.
- Made for
  - understanding
  - structuring
  - predicting
  - communicating

- Can be
  - Taxonomies (e.g. species, genus, family, etc. in biology)
  - Domain models, e.g. in UML
  - Numerical Models (Newtonian mechanics, Quantum mechanics)

A Cinema Transport Model

An example of a UML domain model:

- What is the vocabulary?
- How is it connected?

A Query

What is it we want?

- Screening(s), movie(s, TOTORO)
- cinema(s, k), address(k, l)
- Connection(c), from(c, KRINGSJÅ), to(c, l)
- start(c, cStart), *before*(20:00, cStart)
- end(c, cEnd), start(s, sStart), *before*(cEnd, sStart)

Find s, k, l, c, cStart, cEnd, sStart satisfying this and we have the answer!

- Maybe not the easiest way to ask, but it's a start.
- Models are an important part of a Web of Data!
- Need to connect models from different domains.

Nothing But Questions?

- Tim Berners-Lee talks about "intelligent agents"
- More than just question answering.
- "Agents" can act!
- Make a doctor's appointment:
    - Find and commit to a time that fits agenda and public transport
    - Notify the employer
    - Possibly reschedule conflicting meetings
    - . . .
- Queries over distributed information are at the centre of all this.

Calculating

- What is calculation?

$$\frac{A \text{ owns } x \ Bs \qquad A \text{ gets another } y \ Bs}{A \text{ now owns } (x + y) \ Bs}$$

e.g.

$$\frac{\text{Peter owns } 1 \text{ apple} \qquad \text{Peter gets another } 4 \text{ apples}}{\text{Peter now owns } 5 \text{ apples}}$$

- Calculation is algorithmic manipulation of numbers. . .
- . . . where the *meaning* of the numbers is not needed
- Can calculate $1 + 4 = 5$ without knowing what is counted
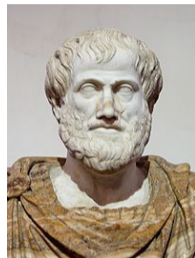- Abstraction!

Calculating with Knowledge

- Can be traced back to Aristotle (384–322 BC)
- Modus Barbara:

$$\frac{\text{All } A \text{ are } B \quad \text{All } B \text{ are } C}{\text{All } A \text{ are } C}$$

  e.g.

$$\frac{\text{All Greeks are men} \quad \text{All men are mortal}}{\text{All Greeks are mortal}}$$

- Algorithmic manipulation of *knowledge*...
- ...where the *meaning* of the words is not needed!
- Also an abstraction!
- The topic of *formal logic*

Computing with Knowledge About Movies

- Query: find a *fun event* we can reach by public transport
- Knowledge base:
    1. A movie screening is an event
    2. A movie screening is fun if the movie being shown is not a documentary
    3. Hayao Miyazaki does not direct documentaries
    4. Hayao Miyazaki directed *My Neighbor Totoro*
    5. There is a screening of *My Neighbor Totoro* at 19:00.
       . . .
- Let us calculate. . .
    6. From 3 and 4: *My Neighbor Totoro* is not a documentary
    7. From 6 and 2: A screening of *My Neighbor Totoro* is fun
    8. From 1, 5, 7: there is a fun event at 19:00
       . . .
- Computing with Knowledge is an important part of a Web of Data!

Exchanging Information
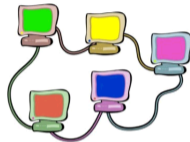
- 1974: The Internet: Global network. Unified network addresses. TCP/IP protocol.
- 1990: The WWW: HTTP protocol. HTML markup. URLs.
- 1996: XML: more data-oriented markup.

- All these (and more) are obviously ingredients for a Web of Data!

- Semantic Web standards are being managed by W3C.

The "Home" of the Semantic Web

See the W3C pages for the Semantic Web effort:

```
http://www.w3.org/2001/sw/
```

For standards (RDF, OWL, SPARQL, etc.), see:

```
http://www.w3.org/2001/sw/wiki/Main_Page
```

Bringing it together

- RDF as common knowledge format:
  - `movie:totoro movie:director people:hm.`
  - `people:hm people:name "Hayao Miyazaki".`
- URIs to avoid naming conflicts:
  - `http://heim.ifi.uio.no/martingi/movies#totoro`
- existing protocols to move data:
  - Use HTTP for queries to a semantic web server
  - Use XML for answers, to encode RDF, etc.
- OWL to express ontologies
  - Somewhat like UML class diagrams but better for Sem. Web
- Reasoners to infer new knowledge
  - Hidden from other tools by standardized interfaces

The AAA slogan

Anyone can say Anything about Anything.

- IMDB: `movie:totoro movie:director people:hm.`
- Saga Kino: `movie:totoro movie:shownAt oslokino:Saga.`
- VG: `movie:totoro vg:terningkast 6.`
- Three statements from three sources about the same subject `movie:totoro`!
- My homepage: `movie:totoro movie:director mg:myself.`

Problems with the Semantic Web

- Relies on ontologies
  - Have to agree on and communicate ontologies
  - Have to agree on the precise meaning of ontologies

- Anyone can say Anything about Anything
  - Good, simple, necessary
  - Difficult to locate relevant information
  - Difficult to trust data sources
  - Have to deal with unreliable, inconsistent data
  - Have to deal with enormous amounts of data

- . . .

- Extent of these problems is in stark contrast to the visions that have been stated and the promises that have been made.

- Hype has brought some amount of discredit to the Semantic Web effort.

Semantic technologies

- If Tim Berners-Lee's vision of a Semantic Web is still far away, then what is this course about?
- Let's have a look at what we do have:
    - W3C standards: RDF, SPARQL, OWL, some more
    - Technology like reasoners, ontology editors
    - Interfacing to relational databases, etc.
    - Existing ontologies for applications in medicine, industry, some of them with over 1M concepts
- Possible, and a lot easier, to use Semantic Web technologies for more closed, controlled applications
- We talk about "semantic technologies" since they make sense independent of the Web

Data integration

- One of the foremost problems in industry today
  - within one organization
  - between organizations
- Enormous amounts of data gathered over the last decades
  - different formats, different data models
  - specialists needed to find, access, convert data when it is needed
  - large need for automated, unified data access
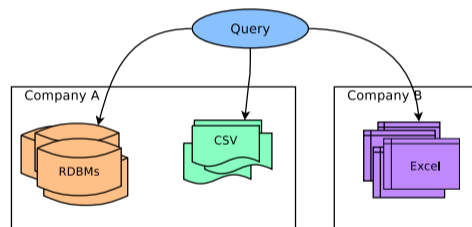
Ontology-based data access

- Use ontology to define common vocabulary
- Possibly by connecting ontologies for different sources using mediating ontologies
- Create mappings between the common vocabulary and what is in the data sources.
- Access data using queries expressed using the common vocabulary
- Background machinery gives answers as if data had always been stored according to a common data model

Ontology-based data access (cont.)
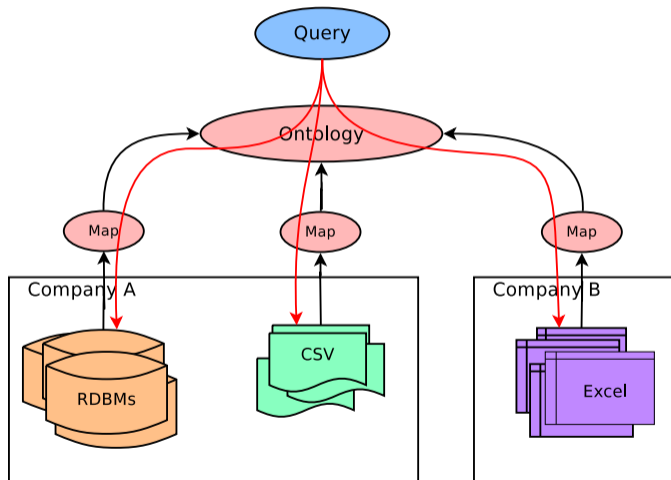
This course

The aim of this course is to teach you...

- ...enough of the semantics in semantic technologies (logic, reasoning) for you to get an idea of what this is all about, what can and cannot be done.
- ...enough of the technology in semantic technologies (standards, languages, programming interfaces) for you to be able to use them in practice.
- ...enough overview for you to know where to look and what to read when you need a deeper understanding of either side.

If you want to learn more:

- Contact us for possible MSc degree topics

The LogID group – Logic and Intelligent Data

- Resarch in semantic technologies, mostly around Ontology-based Data Access.
- Optique http://www.optique-project.eu/
    - 4 year EU project (just finished), led by LogID
    - Ontology Based Data-Access
    - Industry: Siemens, Statoil, DNV, fluid Ops
    - Universities: Oxford, Hamburg, Bolzano, Rome, Athens
- Sirius http://www.sirius-labs.no/
    - Center for Scalable Data Access in the Oil&Gas Domain
    - 8 years funding, 7 left
    - UiO, NTNU, Statoil, Oracle, IBM, Computas, Numascale . . .
- BigMed: personalised medicine
- Great opportunities for both practically and theoretically oriented MSc theses, PhD work,. . . with strong connections to industry and public sector!

Outline

1 Introduction to Semantic Technologies

2 Practicalities

3 Software

When, Where, and Who

When and Where

- Lectures Mondays 14:15–16:00 in OJD 2453, Smalltalk.
- No lecture 10 and 17 April (Easter break), and 1 May
- Guest Lecture: Veronika Heimsbakk, Acando, 24 April

Lecturers

Martin Giese
(martingi@ifi.uio.no)

Leif Harald Karlsen
(leifhka@ifi.uio.no)

Ernesto Jiménez-Ruiz
(ernestoj@ifi.uio.no)

Exercises

Exercises

- Practical exercises every week,
- Limbo (3418), Thursdays 10:15–12:00, starting **this** week
- Exercises available on website well in advance. Come prepared!
- First session: help with setting up software. Bring your laptop!
- In general: part repetition of lectures, part exercises

Teacher

Ole Jørgen Brønner
(olebr@student.matnat.uio.no)

Mandatory Assignments

Assignments

- Six mandatory assignments
- Corrected by teachers. **Tell us if you don't get feedback!**
- Pass/Fail
- Must have passed all assignments in order to attend exam
- First four assignments:
    - Small, about one per week (first one published on 23.1.)
    - (semi-)automated correction
    - One attempt
- Fifth and Sixth assignment:
    - More substantial, timing will be announced
    - Manual correction
    - Two attempts
- For INF4580:
    - more substantial assignments five and six

Piazza

Exam

- Four hours written Exam
- Same exam for INF3580 and INF4580
- Grades A–F
- Probably 15 June – Check semester page!
- Probably 1 May – Check semester page!

Reading

- For practical aspects:

  Semantic Web Programming.
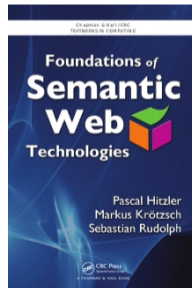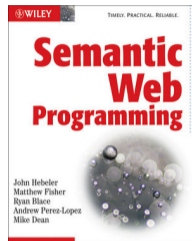 Hebeler, Fisher, Blace, Perez-Lopez.
  Wiley 2009

- For theoretical aspects:

  Foundations of Semantic Web Technologies.
  Hitzler, Krötzsch, Rudolph.
  CRC Press 2009

- Can buy both in Akademika
- Slides available on course homepage

Outline

Software

- Programming-oriented course.
- With non-trivial theoretical components.
- Various off-the-shelf software required to work on exercises.
- Installation help in weekly exercises and exercise sessions.
- Most software already installed on ifi machines.

Software: Java

In principle, any programming language can be used for semantic web programming, but. . .

- Will explain Sem. Web programming using Java libraries
- The textbook concentrates on Java
- Exercises are built around Java

So: get JDK 8 from

`http://www.oracle.com/technetwork/java/javase/downloads/index.html`

Software: Eclipse

In principle, you can use any environment to develop Java programs, but. . .

- The Eclipse IDE is free, open source software
- It is particularly suited for Java development
- We will use the Eclipse IDE for demonstrations
- We will be able to help you with Eclipse problems

So: get the Eclipse Neon IDE from

http://www.eclipse.org/

Software: Jena

There are various Java libraries for Sem. Web programming out there, but...

- The textbook uses Jena
- It is one of the most used and mature Java libraries for Sem. Web
- It is powerful enough for our purposes

> Download Jena 3.1.1 from:
> `http://jena.apache.org/`

Alternatives:

- Sesame, `http://www.openrdf.org/`
- OWL API, `http://owlapi.sourceforge.net/`
- Redland RDF Libraries (C), `http://librdf.org/`
- etc., Google for "RDF library"...

Software: Pellet

There are several reasoning systems around, but. . .

- The textbook uses Pellet
- It is open source software
- It has a direct interface to Jena
- It is one of the more mature and comprehensive reasoners
- It is powerful enough for our purposes

Pellet sources are available from:

`https://github.com/complexible/pellet`

But wait a bit. . . maybe we can offer a precompiled package. Alternatives:

- FaCT++, `http://owl.man.ac.uk/factplusplus/`
- RacerPro, `http://www.racer-systems.com/`
- Hermit, `http://hermit-reasoner.com/`
- etc., `http://en.wikipedia.org/wiki/Semantic_reasoner`

Software: Protégé

There are several ontology editors available, but. . .

- The textbook uses Protégé
- It is open source software
- It is the most widely used ontology editor
- Probably the best non-commercial one

So: get Protégé 5.1 from

http://protege.stanford.edu/

Alternatives:

- see http://en.wikipedia.org/wiki/Ontology_editor

Next weeks. . .

- RDF – knowledge representation – Leif Harald
- Jena – Java API for RDF – Martin
- SPARQL – Query Language – Ernesto
- Maths & Logic – Martin
- . . . reasoning and semantics