

INF3580 – Semantic Technologies – Spring 2011

Lecture 1: Introduction

Martin Giese

25th January 2011



DEPARTMENT OF
INFORMATICS



UNIVERSITY OF
OSLO

INF3170 – Logikk

- It has been decided on short notice that
INF3170 – Logikk
will be held!
- Tuesdays, 12:15–14:00 in Lille Aud. (old ifi building)
- Detailed information about theoretical background of Semantic Technologies:
 - Logic
 - Model Semantics
 - Reasoning

Today's Plan

- 1 Practicalities
- 2 Software
- 3 Introduction to Semantic Technologies

Outline

- 1 Practicalities
- 2 Software
- 3 Introduction to Semantic Technologies

When, Where, and Who

When and Where

- Lectures Tuesdays 14:15–16:00 in Prolog (2465).
- No lecture 19. April (Easter break) and 17. May
- Homepage:
<http://www.uio.no/studier/emner/matnat/ifi/INF3580/>

Lecturer



Martin Giese (martingi@ifi.uio.no)

Additional Lecturers



Martin G. Skjæveland
(martige@ifi.uio.no)



Audun Stolpe
(audus@ifi.uio.no)



Kjetil Kjernsmo
(kjekje@ifi.uio.no)

Exercises

Exercises

- Practical exercises every week,
- Fortress (3468), Fridays 12:15–14:00, starting **this week**
- Exercises available on website well in advance. Come prepared!
- Consider bringing your laptop!

Teachers



Håvard M. Ottestad
(haavarot@ifi.uio.no)



Martin G. Skjæveland
(martige@ifi.uio.no)

Mandatory Assignments, Exam

Assignments

- Probably five mandatory assignments
- Corrected by teachers
- Pass/Fail
- Must have passed all assignments in order to attend exam
- First three assignments:
 - Small, one per week (first one published on 1.2.)
 - Automated correction
 - One attempt
- Fourth and Fifth assignment:
 - More substantial, timing will be announced
 - Manual correction
 - Two attempts

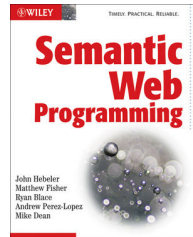
Exam

- Three hours written Exam
- Grades A–F

Reading

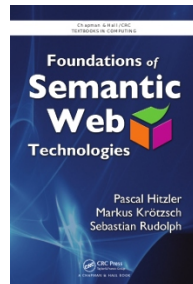
- For practical aspects:

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

- 1 Practicalities
- 2 Software
- 3 Introduction to Semantic Technologies

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 JDK6 from
<http://java.sun.com/>

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 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

So: get Jena 2.6.4 from
<http://jena.sourceforge.net/>

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

So: get Pellet 2.2.2 from
<http://clarkparsia.com/pellet/>

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 4.0.2 from
<http://protege.stanford.edu/>

Alternatives:

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

Outline

- 1 Practicalities
- 2 Software
- 3 Introduction to Semantic Technologies

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.

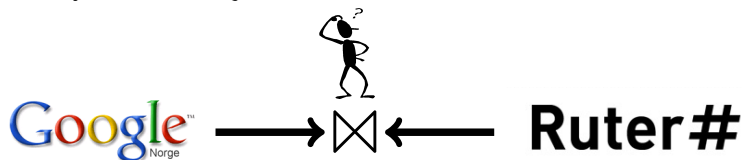


Tim Berners-Lee

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 *Hereafter* 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.trafikanten.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?

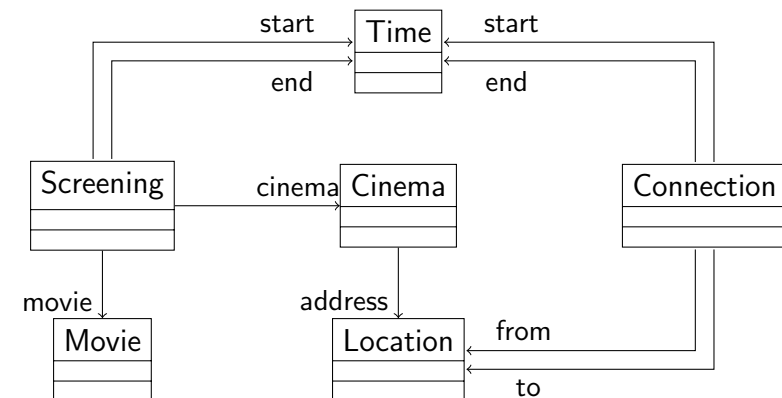
- This sounds like a nice idea, but how can it work?
- There has been a lot of hype around the Semantic Web!
- Visions instantly transformed to promises (and \$\$\$)
- Most of this simply does not work (yet?)
- But then, a lot does!
- Current *partial* solutions build on traditions of
 - Modelling
 -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

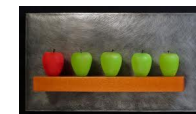
Calculating

- What is calculation?

$$\begin{array}{r} A \text{ owns } x \text{ Bs} \\ A \text{ gets another } y \text{ Bs} \\ \hline A \text{ now owns } (x + y) \text{ Bs} \end{array}$$

e.g.

$$\begin{array}{r} \text{Peter owns } 1 \text{ apple} \\ \text{Peter gets another } 4 \text{ apples} \\ \hline \text{Peter now owns } 5 \text{ apples} \end{array}$$



- 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!

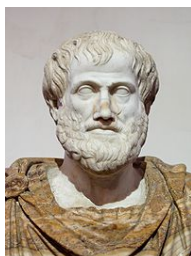
Calculating with Knowledge

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

$$\begin{array}{r} \text{All } A \text{ are } B \\ \text{All } B \text{ are } C \\ \hline \text{All } A \text{ are } C \end{array}$$

e.g.

$$\begin{array}{r} \text{All } \text{Greeks} \text{ are } \text{men} \\ \text{All } \text{men} \text{ are } \text{mortal} \\ \hline \text{All } \text{Greeks} \text{ are } \text{mortal} \end{array}$$



- Algorithmic manipulation of *knowledge*. . .
- . . . where the *meaning* of the words is not needed!
- Also an abstraction!
- The topic of *formal logic*
- a.k.a. INF3170!

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 Clint Eastwood does not direct documentaries
 - ④ Clint Eastwood directed *Hereafter*
 - ⑤ There is a screening of *Hereafter* at 19:00.
 - . . .
- Let us calculate. . .
 - ⑥ From 3 and 4: *Hereafter* is not a documentary
 - ⑦ From 6 and 2: A screening of *Hereafter* 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.

Bringing it together

- RDF as common knowledge format:
 - `movie:Hereafter movie:director people:ce.`
 - `people:ce people:name "Clint Eastwood".`
- URLs to avoid naming conflicts:
 - `http://heim.ifi.uio.no/martingi/movies#Hereafter`
- 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:Hereafter movie:director people:ce.`
- Saga Kino: `movie:Hereafter movie:shownAt oslokino:Saga.`
- VG: `movie:Hereafter vg:terningkast 5.`
- Three statements from three sources about the same subject
`movie:Hereafter!`
- My homepage: `movie:Hereafter movie:director mg:myself.`

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



Problems with the Semantic Web

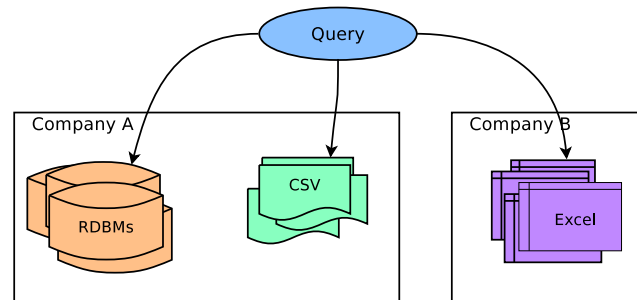
- Relies on ontologies
 - Have to agree on and communicate ontologies
 - Have to agree on the precise meaning of ontologies
- Anyone can say Anything about Anything
 - Good, simple, necessary
 - Difficult to locate relevant information
 - Difficult to trust data sources
 - Have to deal with unreliable, 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

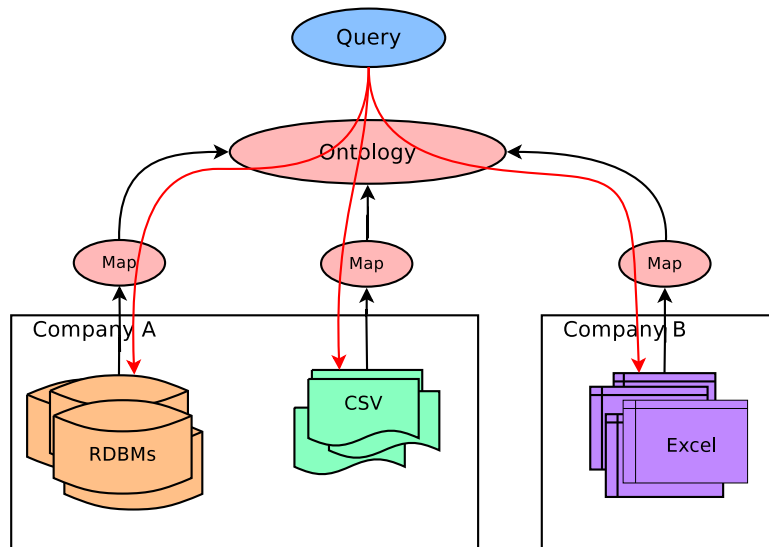
- One of the foremost problems in industry today
 - within one organization
 - between organizations
- Enormous amounts of data gathered over the last decades
 - different formats, different data models
 - specialists needed to find, access, convert data when it is needed
 - large need for automated,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

Semantic Technologies at ifi

- Currently 1 professor, 2 post-docs, 3 PhD-students, ~6 MSc students directly concerned with semantic technologies in OMS group.
- Semicolon
 - Data exchange between public sector institutions in Norway
 - Publication and interlinking of public data.
 - User partners: Brønnøysundregistrene, Helsedirektoratet, Skattedirektoratet, Statistisk sentralbyrå, ...
- IOHN (Integrated Operations in the High North)
 - Partners include two oil companies, major software vendors like IBM, SAP, Siemens
 - Data exchange and integration for the oil industry
- Great opportunities for both practically and theoretically oriented MSc theses, PhD work, ... with strong connections to industry and public sector!