TDT 44 – Semantic Web

Starts at 13.15
TDT 44 – Semantic Web

Özlem Özgöbek

ozlem.ozgobek@ntnu.no
Let's meet!

Go to menti.com and use the code 82 29 19 8

or use this link:
https://www.menti.com/q64cuijdsf
About the course

**Goal:** The students get an orientation of the theoretical, methodological and technological background and of the ongoing standardization work that support semantic Web.

You will gain:

- Background theoretical knowledge
- Insight to relevant problems within Web information resources and services
- Insight to how to use semantic web technology to solve problems in different domains

Mostly self-driven.
Learning Material

BOOKS

ACADEMIC PAPERS

PRESENTATIONS

EXTERNAL LINKS

Follow the course wiki page: https://www.ntnu.no/wiki/display/idiemner/TDT44+-+Semantic+Web
Schedule

24.09.2020    Introductory meeting
    Presentation of semantic web concepts
22.10.2020    Mid-term follow-up meeting - 1 with short presentations from students
05.11.2020    Mid-term follow-up meeting - 2 with short presentations from students
25.11.2020    Report deadline
XX.11.2020 (TBA)    Oral exam – Presentation of reports

Mandatory activities: Peer review, project report, final presentation.

Final grade: Project report and final presentation.

The details will be available on the wiki page soon.
About the project

First deadline: 9 October 2020
Brief Introduction to Semantic Web
Semantic Web

Machines understanding the meaning
or
Meaning is machine readable.

Why do we need it?
Too much information, too little structure.
Meaning (semantics) of data is only human readable.
Connected data is smarter data.
Semantic Web

Set of standards

W3C - World Wide Web Consortium
https://www.w3.org/
Semantic Data

Representation of data

Distributed and connected data
Application areas

- Information retrieval
- Knowledge management
- Natural language processing

“Smart systems of any type”

- Search engines
  - E-learning
  - E-commerce
- Recommender systems
Open World

The world is dynamic.

Information changes all the time.

The Web is an Open World, we treat it using the “Open World Assumption”.

It is very important when we want to draw conclusions based on distributed data.
Semantic Modeling
“to make sense of the world around us”

Models are abstract descriptions that hides certain details while illuminating others.

Models:
Help people communicate
Explain and make predictions
Mediate among multiple viewpoints

The Semantic Web standards enable people to collaborate on models that organizes the information. Common collection of knowledge.
A semantic web model must be able to organize **variations** in a meaningful and manageable way.

In semantic web context, modeling is an ongoing process.
How to do it?

Frameworks, descriptions, structured representation of data, links, inference etc.

Ontology
An ontology describes a formal specification of a certain domain. The types of entities and relationships are defined in ontologies.

There are different ontology/semantic web modeling languages: XML, RDF, OWL ...
An example:
What is the usefulness of an ontology?

- To make domain assumptions explicit
  - Ontological analysis
    - clarifies the structure of knowledge
    - allow domain knowledge to be explicitly defined and described
- Enrich software applications with the additional semantics
- To facilitate communications among systems without semantic ambiguity. I.e. to
  - achieve inter-operability
- Thus, practically, improving: computer-computer, computer-human, and human-
  - human communication
- To provide foundations to build other ontologies (reuse)
- To save time and effort in building similar knowledge systems (sharing)
Challenges

- Scalability
- Logical structure

“The challenge of the Semantic Web is to find a representation language powerful enough to support automated reasoning but simple enough to be usable” [AKT 2003]
However...

“The Semantic Web is not about getting everyone agree on a single ontology, but rather about coping in a world where not everyone will agree, achieving some degree of interoperability.”

Book: Semantic Web for the Working Ontologist
The Semantic Web Stack

- **XML**
  Surface syntax, no semantics
- **XML Schema**
  Describes structure of XML documents
- **RDF**
  Datamodel for “relations” between “things”
- **RDF Schema**
  RDF Vocabulary Definition Language
- **OWL**
  A more expressive Vocabulary Definition Language

XML

XML is a labeled tree. XML is a markup language that enables creation of documents composed of semi-structured data.

- **node** = label + attr/values + contents

```xml
<course date="...">
  <title>...</title>
  <teacher>...</teacher>
  <name>...</name>
  <http>...</http>
  <students>...</students>
</course>
```

- **XML Schema:** grammars for describing legal trees and datatypes
RDF → the standard for representing knowledge graphs
RDFS and OWL → two standards for defining ontologies
SPARQL → the standard for querying knowledge graphs
RDF (Resource Description Framework)

RDF is a standard to describe entities/resources. A resource can be anything we can identify, such as a person, a homepage or great dragons in the Game of Thrones.

Resources in RDF are described as triplets:

[subject, predicate, object]

[lemon, is, yellow]
[Ann, knows, Barbara]
[Mary, likes, pizza]
RDF

RDF models the relations between things.

- It has more interoperability,
- Provides better understanding of the data.
RDFS
Resource Description Framework Schema

RDF defines the structure of the data
RDFS defines semantic relationships

org:Organization rdf:type rdfs:Class .
org:Start-up a rdfs:Class .
org:hasHomePage rdf:type rdfs:Property .
org:Start-up rdfs:subClassOf org:Organization .
OWL

Web Ontology Language

https://www.w3.org/2001/sw/wiki/OWL

“It is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things.”
SPARQL

Standard query language for RDF and OWL

- Turtle syntax
- SQL like
- Based on pattern-matching mechanism
Knowledge Graphs

“A knowledge graph consists of a set of interconnected typed entities and their attributes.”

Popularized by Google in 2012.

Now also used by Facebook, LinkedIn and Microsoft.
<table>
<thead>
<tr>
<th>Features</th>
<th>Pure RDF datasets</th>
<th>Linked Data</th>
<th>Knowledge Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine readability</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Human readability</td>
<td>NN</td>
<td>NN</td>
<td>Y</td>
</tr>
<tr>
<td>Data distribution</td>
<td>N</td>
<td>Y</td>
<td>NN</td>
</tr>
<tr>
<td>Inter-dataset linkage</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Data integration</td>
<td>NN</td>
<td>NN</td>
<td>Y</td>
</tr>
<tr>
<td>Data consistency</td>
<td>NN</td>
<td>NN</td>
<td>Y</td>
</tr>
<tr>
<td>Reliability</td>
<td>NN</td>
<td>NN</td>
<td>Y</td>
</tr>
<tr>
<td>High quality</td>
<td>NN</td>
<td>NN</td>
<td>Y</td>
</tr>
</tbody>
</table>

Y = Yes, NN = Not Necessarily, N = No, L = Limited
Some Standard Ontologies

- Friend-of-a-Friend (FOAF)
  https://en.wikipedia.org/wiki/FOAF_(ontology)

- schema.org