COMP318

Ontologies and Semantic Web

http://www.csc.liv.ac.uk/~valli/Comp318.html

UNIVERSITY OF LIVERPOOL

Dr Valentina Tamma
Room: Ashton 2.12
Dept of computer science
University of Liverpool
V.Tamma@liverpool.ac.uk

Motivation

- The current Web is made of text, pictures, music, movies ….
  - Great for People!
    - We can make sense of what we see, and select, combine and integrate information from these various sources
  - Useless for Computer Automation
    - Computers can't easily do "intelligent" tasks, which requires understanding this information

- But what if the Web was made of text, pictures, music, movies etc that were also readable by machines?
  - What would be the impact on tasks like search, query answering and knowledge aggregation?

Representing facts

- The success of the WWW is based on the use of standard mechanisms to exchange and communicate data, information and knowledge to different stakeholders:
  - HTML:primarily for human consumption
  - XML: data exchange mechanism
  - …

```
<html>
<head><title>John Smith</title></head>
<body bgcolor="white">
<b>John Smith</b><br>
is a Lecturer at the
<i>University of Liverpool</i><br>
John teaches <i>Semantic Web
technologies</i><br>
…
</body>
</html>
```

XML: data exchange mechanism

- Exchange languages components
  - Syntax: how to write the data
  - Data model: how to structure or organise the data
  - Semantics: how to interpret (meaning) this data

```
<lecturer name John Smith>
<university> University of Liverpool</university>
<course>SemWeb Technologies</course>
<url>http://www.csc.liv.ac.uk/~jsmth</url>
</lecturer>
```

- Semantics clarifies the meaning, specifies assumptions and allows deductions

```
<course name SemWeb Technologies>
<university> University of Liverpool</university>
<lecturer>John Smith</lecturer>
<url>http://www.csc.liv.ac.uk/~Comp356</url>
</course>
```

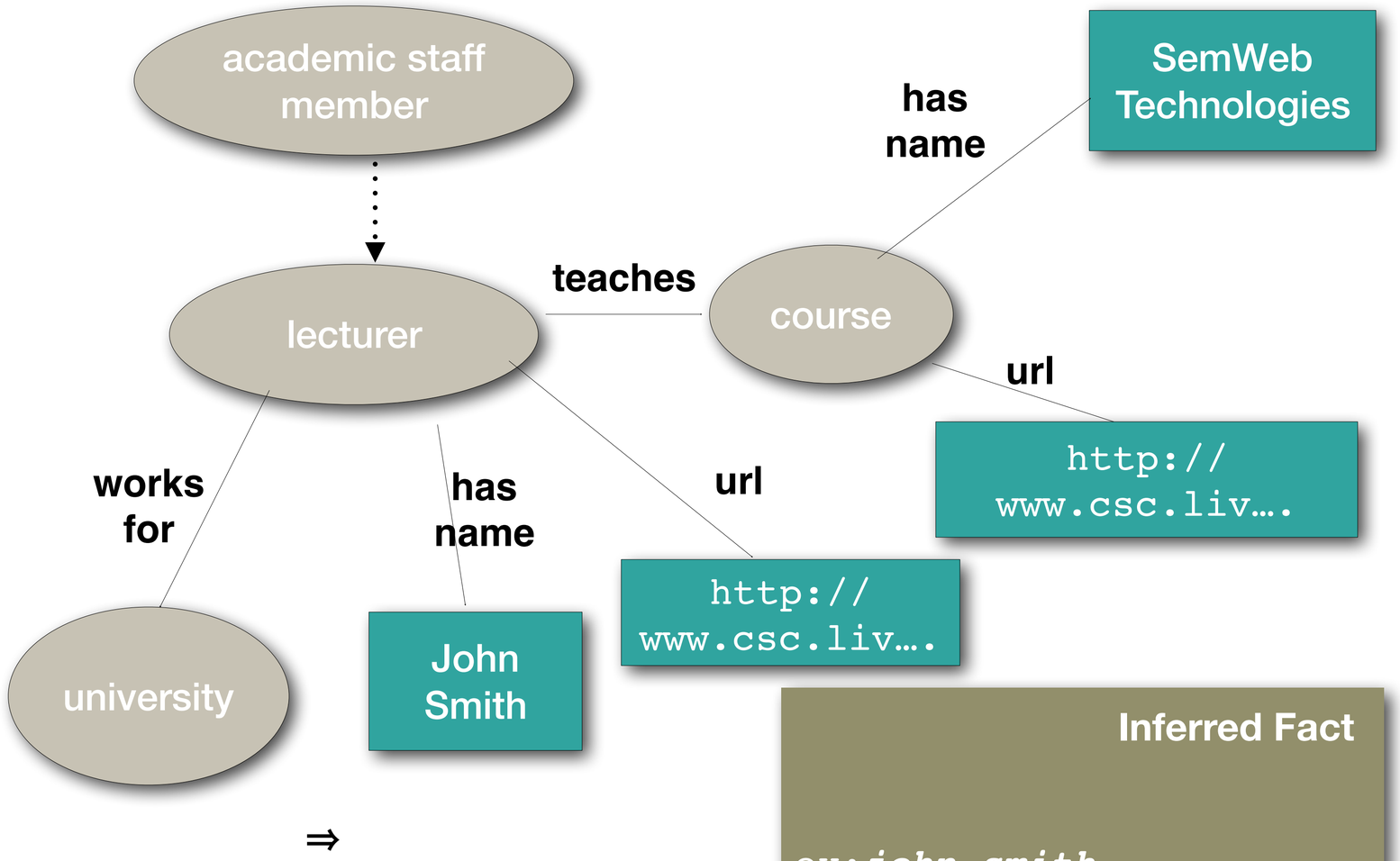
Adding Semantics

```
ex:Lecturer
  rdfs:subClassOf
  ex: AcademicStaff
  ex: hasName
  ex:teaches
  ex:url
…
```

Schema

```
ex:john_smith
  rdf:type
  ex: Lecturer
  ex: hasName value “John
  Smith”
  ex: url value “http://www…”
  ex: teaches comp_356
  ex: comp_356 hasName
       “SemWeb Technologies”
  …
```

Assertion



Semantics clarifies the meaning, specifies assumptions and allows deductions.

Inferred Fact

```
ex:john_smith
rdf:type
ex: AcademicStaffMember
```

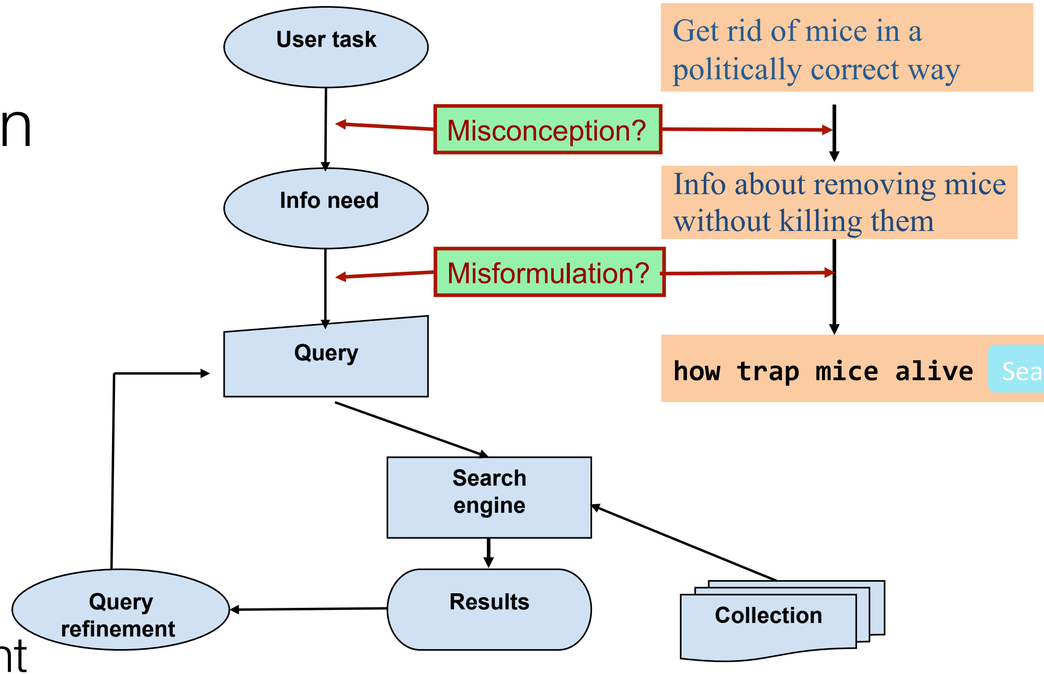
Search engines allow the web to scale

- The success of the web is due to effective search engines
  - No incentive in creating content unless it can be easily found
    - other methods for locating content haven't kept pace (taxonomies, bookmarks, etc)

- The web is both a technological artefact and a social environment
  - Search engines make aggregation of interests possible:
    - Create incentives for very specialised niche players:
      - Economical: specialised stores, providers, etc
      - Social: narrow interests, specialised communities, etc

Is it all working smoothly?

- The simple fact is that a generic web search returns too many pages!
  - Fine grained interpretation and / or aggregation of results is left to the user
  - Search engines try to contextualise the aim of the query;
  - It is simply difficult to distinguish the meaning between these two sentences:
    - “I am a lecturer of computer science” vs “I am an assistant professor of computer science”
    - Jaguar vs jaguar



The problem's with today's web

Human Centered:

- **Information has to be processed by users;**
- **Direct questions are not understood:**
  - **Ask: "*How large is the web?*"**
  - **Answer:**

About 8,040,000,000 results (0.68 sec)

Requires active involvement:

- **Browsing and searching is keyword based;**
- **Laborious and often ineffective:**
  - **Ask: "*How large is the web today?*"**
  - **Answer:**

About 2 results (0.27 sec)

Do you *trust* what you find?

- **Where is the information coming from?**
- **Is it up to date?**

The problem's with today's web

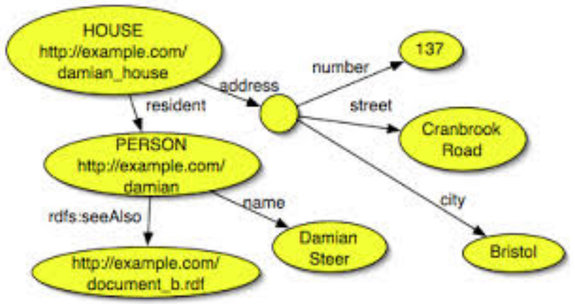
Information needs to be integrated manually:

- Users have to read a number of pages / documents;
- Each of these can be relevant / contain part of the answer;
- Users need to search and integrate information that is:
- Disperse;
- Heterogeneous;

Test:
Cost of a camera Canon EOS T3i

Results:
- **Price VAT ex;**
- **Price VAT inc;**
- **Shipping included ih the price;**
- **Price in Euros;**
- **...**

Syllabus



- In this module we:
  - Introduce the next generation web (Semantic Web): separation between presentation and content
  - Introduce the notion of explicitly representing content through models:
    - different markup languages: RDF and OWL, and query languages: SPARQL
  - Study the principles for modelling expressive schemas (Ontologies) to be used to support different applications requiring knowledge sharing
  - Look at various applications of these principles and technologies:
    - Knowledge graphs
    - Linked open data
  - Assignments use industry tools, i.e. Jena java API and Protege ontology editor

Learning outcomes

- Have an understanding of the basic formal methods and techniques for designing and implementing advanced web applications

- Have an appreciation for Artificial Intelligence and Semantic Web research related to advanced web technology applications

- Be able to apply specific methods and techniques in the design and development of an application of advanced web technology for a case study

Lecturer and demonstrators

Dr Valentina Tamma
Room: Ashton 2.12
Ashton Bldg
V.Tamma@liv.ac.uk

Office hours
Tue 14 – 15
Fri 12 – 13

Joshua Alcock
sgjalcoc@student.liverpool.ac.uk

Reham Alharbi
R.Alharbi@liverpool.ac.uk

Make an appointment by email!

What to study

- Lecture slides:
  - Slides are available on VITAL;
- Text books:
  - There isn't a single book, but rather a collection of book chapters. All of these are available from the library:
    - Antoniou, Groth, Hoekstra, van Harmelen: A Semantic Web Primer. MIT Press, 2012
    - Uschold: Demistifying OWL for the enterprise. Morgan & Claypool 2018.
    - Heath, Bizer: Linked Data - Evolving the Web into a Global Data Space. Morgan & Claypool 2011.
    - Euzenat, Shvaiko: Ontology matching. Springer, 2012
    - Gomez-Perez, Fernandez-Lopez, Corcho: Ontological Engineering. Springer-Verlag, 2003
- Useful readings:
  - Papers of interest and web resources will be indicated during the lectures, and are available on the course web pages. They will allow you to get a better understanding of the topic, and ultimately better marks!

A word on self study

- Part of the course consists in becoming familiar with markup languages: (XML), RDF, OWL
- There is so much one can explain during lectures, but…
  - You need to study their syntax in your own time, and practice in the lab
  - Use the lab sessions to check your progress!
- XML is assumed, but some notes are available for you to refresh your memory...

Web page

- All the material (slides, appendices, exercises, etc) are published on VITAL;
- All announcements are made during lectures and on VITAL;
- First point of reference on the course
- Consult the course page on VITAL!

Assessment

- 80% Exam:
  - Past exam papers available at:
  - http://www.csc.liv.ac.uk/student/exampapers/
  - Careful, though, every year I refresh the content to keep into account the latest advances in the area!

- 20% Coursework (1 assignment, 1 class test):
  - You need to work on your own, this course is about learning the fundamental principles!

Organisation of the course

- 30 lectures (3 lectures a week)
  - But extra “surgery” lectures from week 11 if needed
- Lectures
  - Tue 11 - 12 (NICH-LT)
  - Thu 11 - 12 (CHEM-GOS)
  - Fri 11 - 12 (NICH-LT)
- Labs - two groups:
  - Wed 10 - 11 (GHolt H116/117 - Lab2)
  - Fri 09 - 10 (GHolt H105 - Lab 3)
  - Check on Orbit which group you belong to
- First Lab on Wed 05/02/2020

Practicals

- A set of exercises to be completed in the lab in preparation for the assignment and the class test.

- The Programming assignment has to be completed in JAVA
  - Use the industry standard API Apache Jena
  - Assignments marked on the grounds of problem solving ability, rather than mere JAVA skills
  - ***The aim of the assignment is to understand the principles rather than demonstrate programming ability***

Tips for passing!

- Attend the lectures, actively!
  - Ask questions, stop me if I'm not clear;
  - Warning: If some topics are to be read in your own time, that doesn't mean they won't appear in the exam;
  - We'll have class exercises, answers will be given in groups, you have no excuse not to participate!

Semantic Web and Linked (Open) data

- Evolution of the current web towards a web of data:
  - **From information to knowledge based**
    - Semantic web
      - www.semanticweb.org

- Delegate machines to “understand” what is the need behind a query...
  - **Represent knowledge on the Web in a form that is more easily machine-processable:**
    - Improve retrieval;
  - **Use intelligent techniques to take advantage of these representations:**
    - Delegate autonomous software components (agents)

But what is the Semantic Web?

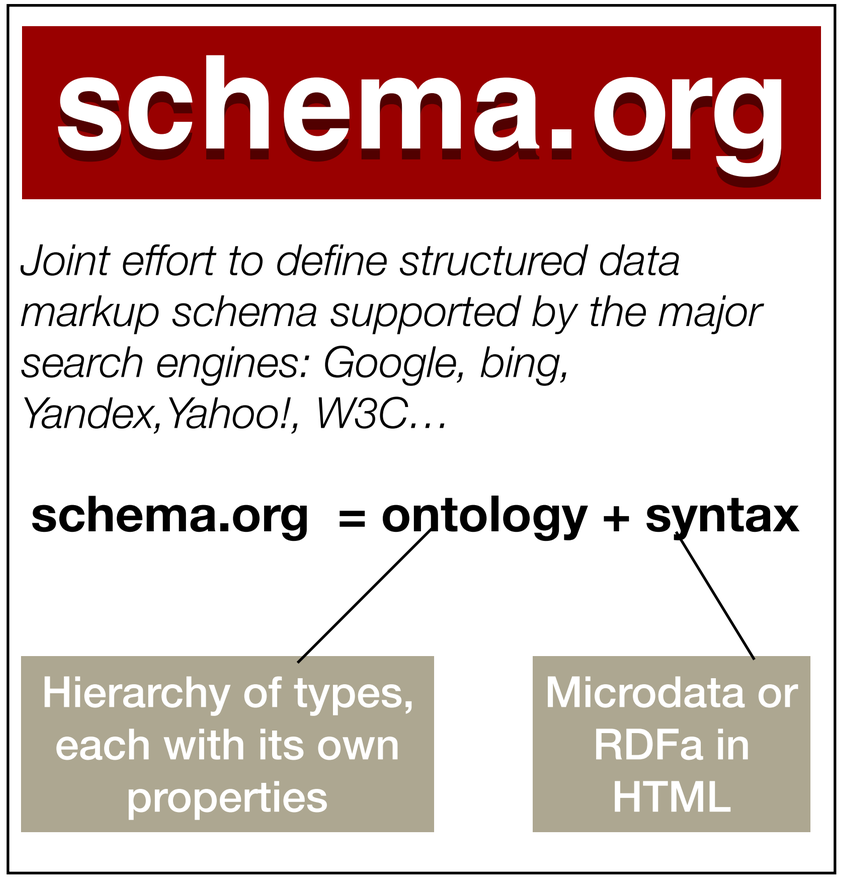
- It NOT just another centralised Knowledge Base framework !!!
  - The Semantic Web is a web of (partially connected) ontologies and knowledge bases

- Facts scattered in many locations…
  - Breaking the “file” paradigm
    - Related instances & ontology fragments can be asserted anywhere
    - Scope & association is managed through namespaces
    - References to ontologies and instances managed through URI

  - Dynamic knowledge source that is typically:
    - Inconsistent / chaotic
    - Incomplete (open-world)
    - Composed of Several Ontologies
    - Generated by novices and machines, as well as experts

A richer data model

- Semantic Web: **beyond machine readable** to **machine understandable.**
  - data model:
    - **data model that can be used by multiple applications**
      - not only for describing documents
      - for people to describe application-specific information
    - **data model that is domain independent**
      - any application can use it to describe information
  - semantics:
    - mechanism to interpret the data model
    - describes the interpretations of the data items wrt the domain
  - syntax:
    - standardised exchange mechanism



Schema.org, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=51534848

Recap

- Problems with today's web
- Introduction to the Semantic Web