UNIVERSITY OF Southampton
School of Electronics
and Computer Science

COMP6028
Semantic Web Technologies

Dr Nicholas Gibbins
32/3019
nmg@ecs.soton.ac.uk

Course Aims

- Understand the key ideas and history behind the Semantic Web
- Explain the state of the art in Semantic Web technologies
- Gain practical experience of ontology design in OWL
- Understand the future directions of the Semantic Web, and its relationship with other Web developments

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Course Structure

- Two lectures per week
  - Monday 1400-1500 in 54/5025
  - Wednesday 0900-1000 in 02/1039
  - (Thursday 1200-1300 in 35/1005)

Teaching Schedule

Week 18: Introduction to the Semantic Web

Week 19: Ontologies and Description Logic

Week 20: RDF and RDFS

Week 21: OWL
Protege

Week 22: Rules for the Semantic Web
Protege

Week 23: Methodologies and Ontology Design Patterns

Teaching Schedule

Week 28:	SPARQL Publishing on the Semantic Web
Week 29:	Reading Week
Week 30:	SW Application Design SW Application Showcase
Week 31:	Semantic Annotation Semantic Web Services
Week 32:	Semantic Web meets Web2.0
Week 33:	Review

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Assessment

- Examination: 50%
- Ontology design coursework: 20%
  - Specification published in week 20
  - Submission due week 30
  - Feedback due week 33
- Individual report: 30%
  - Subjects published in week 20
  - Submission due week 32

Introduction to the Semantic Web

UNIVERSITY OF Southampton
School of Electronics and Computer Science

History of the Semantic Web

- The World Wide Web was invented by Tim Berners-Lee (amongst others), a physicist working at CERN
- TBL's original vision of the Web was much more ambitious than the reality of the existing (syntactic) Web
- TBL (and others) have since been working towards realising this vision, which has become known as the Semantic Web

UNIVERSITY OF Southampton
School of Electronics and Computer Science

History of the Semantic Web

"... a goal of the Web was that, if the interaction between person and hypertext could be so intuitive that the machine-readable information space gave an accurate representation of the state of people's thoughts, interactions, and work patterns, then machine analysis could become a very powerful management tool, seeing patterns in our work and facilitating our working together through the typical problems which beset the management of large organizations."

T. Berners-Lee, The World Wide Web: Past, Present and Future, 1996

UNIVERSITY OF Southampton
School of Electronics
and Computer Science

What is the Semantic Web?

The Semantic Web is an extension of the current Web in which **information is given a well-defined meaning**, better enabling computers and people to work in cooperation.

It is the idea of having data on the Web defined and linked in a way that it can be used **for more effective discovery, automation, integration and reuse** across various applications.

The Web can reach its full potential if it becomes a place where **data can be processed by automated tools as well as people**.

THE SEMANTIC WEB

W3C Activity Statement

Which Semantic Web? The annotated Web

- Enrich existing web pages with annotations
- Classify web pages
- Use natural language techniques to extract information from web pages

- Annotations enable enhanced browsing and searching
- (but NLP is hard)

Which Semantic Web? The Web of Data

UNIVERSITY OF Southampton
School of Electronics and Computer Science

- Expose existing databases in a common format
- Express database schemas in a machine-understandable form

- Common format allows the integration of data in unexpected ways
- Machine-understandable schemas allow reasoning about data
- (make the most of the structure you already have)

University of Southampton
School of Electronics and Computer Science

Rocket Science (not)

Is this rocket science? Well, not really. The Semantic Web, like the World Wide Web, is just taking well established ideas, and making them work interoperably over the Internet. This is done with standards, which is what the World Wide Web Consortium is all about. **We are not inventing relational models for data, or query systems or rule-based systems. We are just webizing them.** We are just allowing them to work together in a decentralized system - without a human having to custom handcraft every connection.

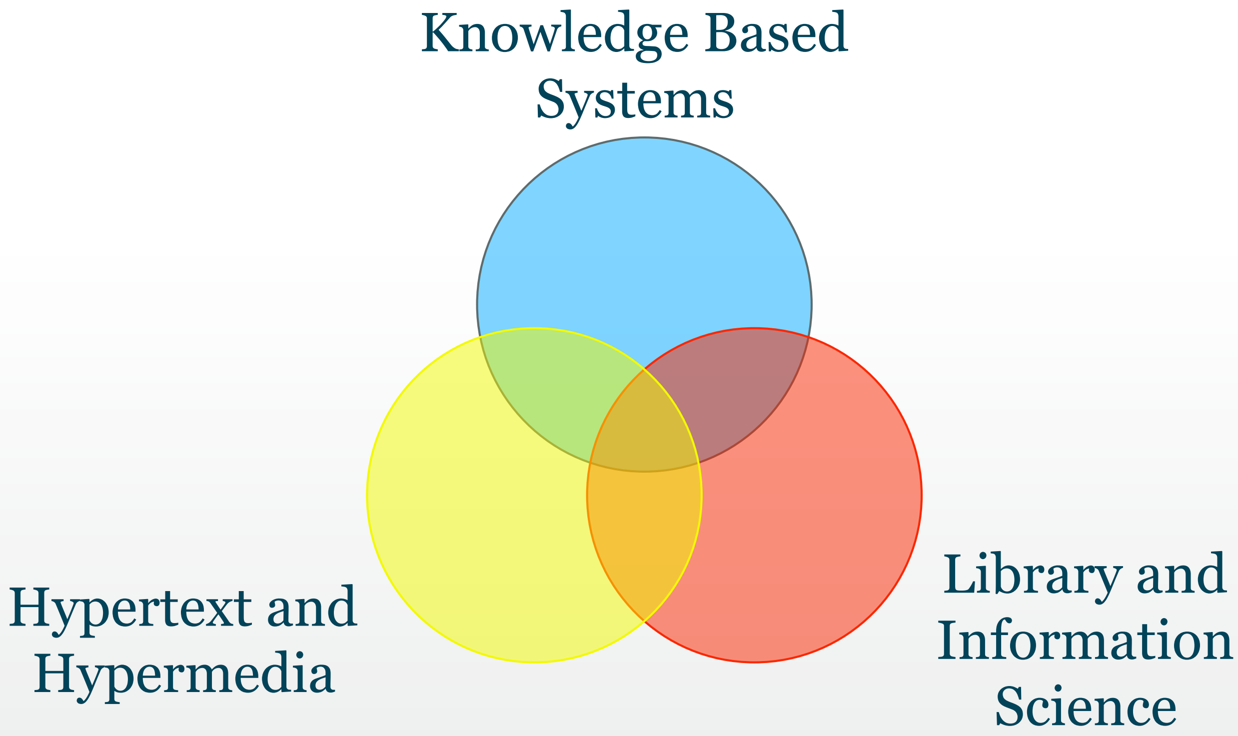
Tim Berners-Lee, Business Case for the Semantic Web,
http://www.w3.org/DesignIssues/Business

UNIVERSITY OF Southampton

School of Electronics
and Computer Science

The Origins of the Semantic Web

Interwoven themes



Metadata

- The origins of the Semantic Web lie in metadata
- Metadata is data about data
  - A webpage is data
  - A description of the webpage is metadata
  - Metadata for a webpage could include
    - author
    - date of publication
    - file size
    - …
- Library cataloguing = metadata

Beyond metadata

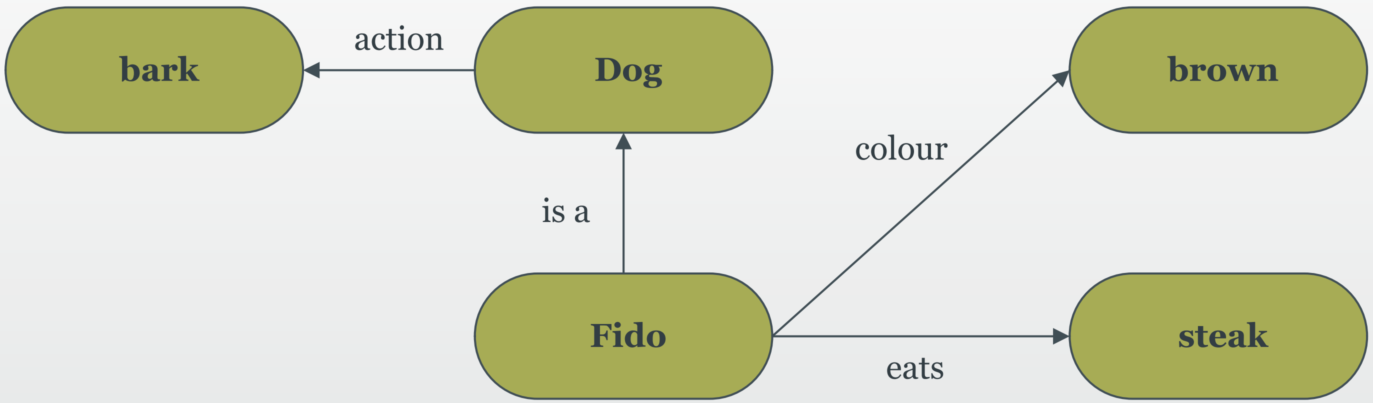
- The scope of the modern Semantic Web goes beyond bibliographic metadata for webpages
- Metadata is still just data
- If we have an infrastructure for metadata, we can use it for data in general

Knowledge representation

- Long-standing discipline within Artificial Intelligence
  - (the Semantic Web has a strong heritage!)
- Knowledge representation languages should:
  - Handle qualitative knowledge
  - Allow new knowledge to be inferred
  - Represent both the general and the specific
  - Capture complex meaning
  - Allow meta-level reasoning
- RDF, RDF Schema and OWL are knowledge representation languages

Network knowledge representation

- "Traditional" knowledge representation is formal logic
- Network knowledge representation originated in 1960s with psychologists and linguists
- Knowledge is represented as a graph
  - Nodes are objects or concepts
  - Edges are relations or associations



Vocabularies and ontologies

- A knowledge representation language by itself is of little use
- We need to be able to tailor the language to our application domain
  - The bibliographic domain needs to be able to talk about works and authors
  - The e-commerce domain needs to be able to talk about orders and prices
  - …
- We need domain-specific vocabularies and ontologies

Hypertext and hypermedia

- Non-linear writing
  - Interlinked texts
  - Multiple pathways, multiple reading sequences
  - Multiple media: video, audio, images, emails, databases, spreadsheets
- Annotation and commentary
- Association of ideas

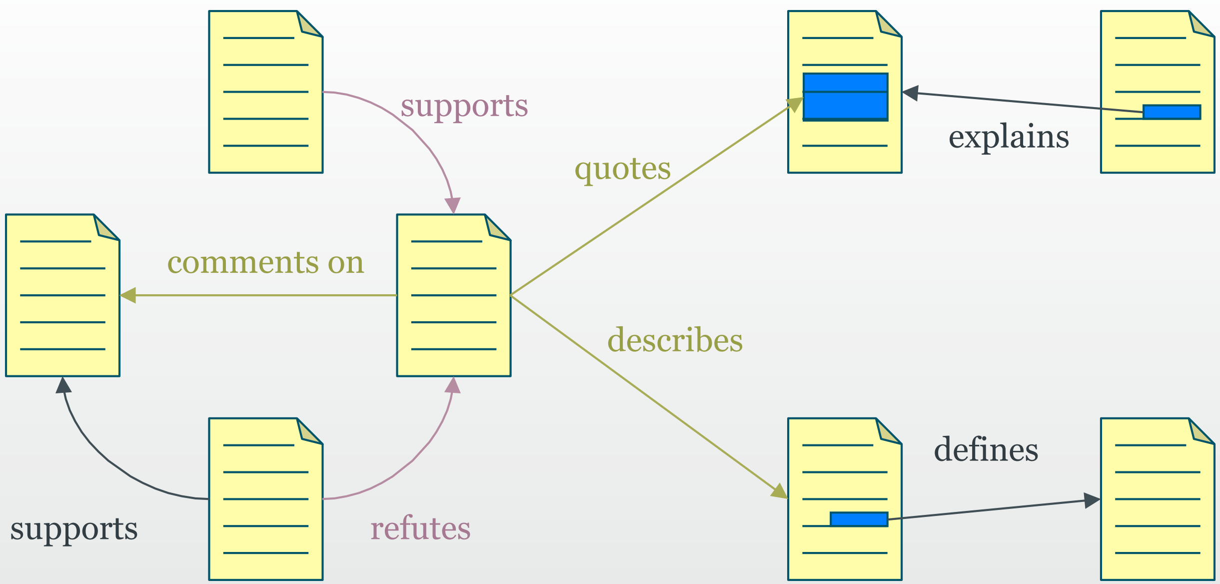
Links

- Essence of hypermedia is connections
  - Relationships in an abstract domain
  - Implemented as navigable links
- Many kinds of relationships
  - Author-of, homepage-of, see-also, background-info, definition, more-detail
  - Typed links
- Links are complex structures
  - Multivalent, rich metadata
  - Not just simple GOTOs

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Open Hypermedia

- Links should be
  - first-class objects
  - manipulated independently



Open hypermedia versus network KR

UNIVERSITY OF Southampton
School of Electronics and Computer Science

- Open hypermedia makes links between different bits of knowledge
  - Knowledge is expressed as text, images, etc
- Network knowledge representation makes links that are knowledge
- Are typed hypermedia links knowledge?
- Is a set of hypermedia link types an ontology?

UNIVERSITY OF Southampton

School of Electronics and Computer Science

Basic Concepts

The World Wide Web versus The Semantic Web

UNIVERSITY OF
Southampton
School of Electronics
and Computer Science

- The World Wide Web is the Web for people
  - Information is predominantly textual
  - Technologies include URI, HTTP, XML, HTML
- The Semantic Web is the Web for machines
  - Information needs to be structured
  - Technologies include RDF, RDFS, OWL (in addition to those for the Web)

Machine readable versus machine understandable

UNIVERSITY OF Southampton
School of Electronics and Computer Science

- On the World Wide Web, information needs humans to give it interpretation
  - Information is predominantly natural language
  - Difficult to mediate by software agents
- On the Semantic Web, information is structured so that it can be interpreted by machines
  - Humans need not interact directly with Semantic Web information – mediation through agents
- Formal meaning is critical to understanding

Machine readable versus machine understandable

UNIVERSITY OF Southampton
School of Electronics and Computer Science

XML is a machine readable format

- It can be parsed to give an unambiguous document structure

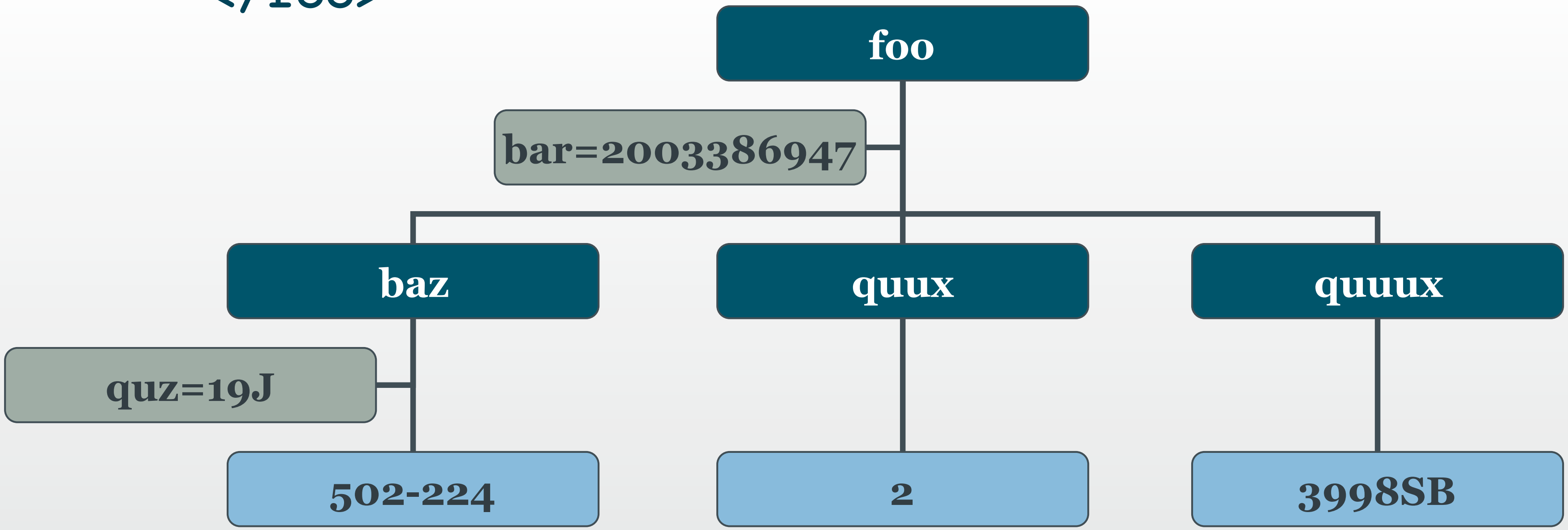
but

- It has no formal meaning
- Meanings of XML interchange formats must be explicitly agreed

UNIVERSITY OF Southampton
School of Electronics
and Computer Science

Machine readable: XML

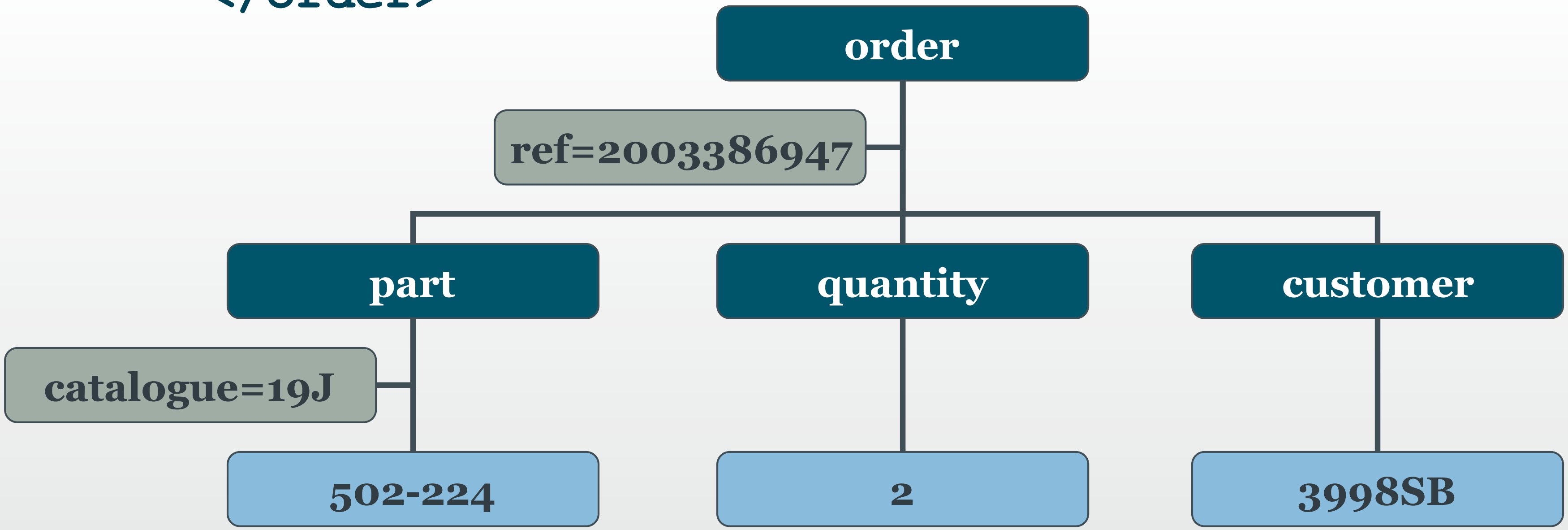
```
<foo bar=“2003386947”>
  <baz qux=“19J”>502-224</baz>
  <quux>2</quux>
  <quuux>3998SB</quuux>
</foo>
```



UNIVERSITY OF Southampton
School of Electronics
and Computer Science

Machine readable: XML

```
<order ref=“2003386947”>
  <part catalogue=“19J”>502-224</part>
  <quantity>2</quantity>
  <customer>3998SB</customer>
</order>
```



Machine readable versus machine understandable

UNIVERSITY OF Southampton
School of Electronics and Computer Science

RDF is a machine understandable format

- The structures generated by an RDF parser have a formal meaning
- RDF is a framework for interchange formats that provides a base level of common understanding
- RDF provides basic notions of classes and properties
- RDF enables simple inference
  - RDF permits certain types of deduction to be made from existing knowledge

UNIVERSITY OF Southampton
School of Electronics and Computer Science

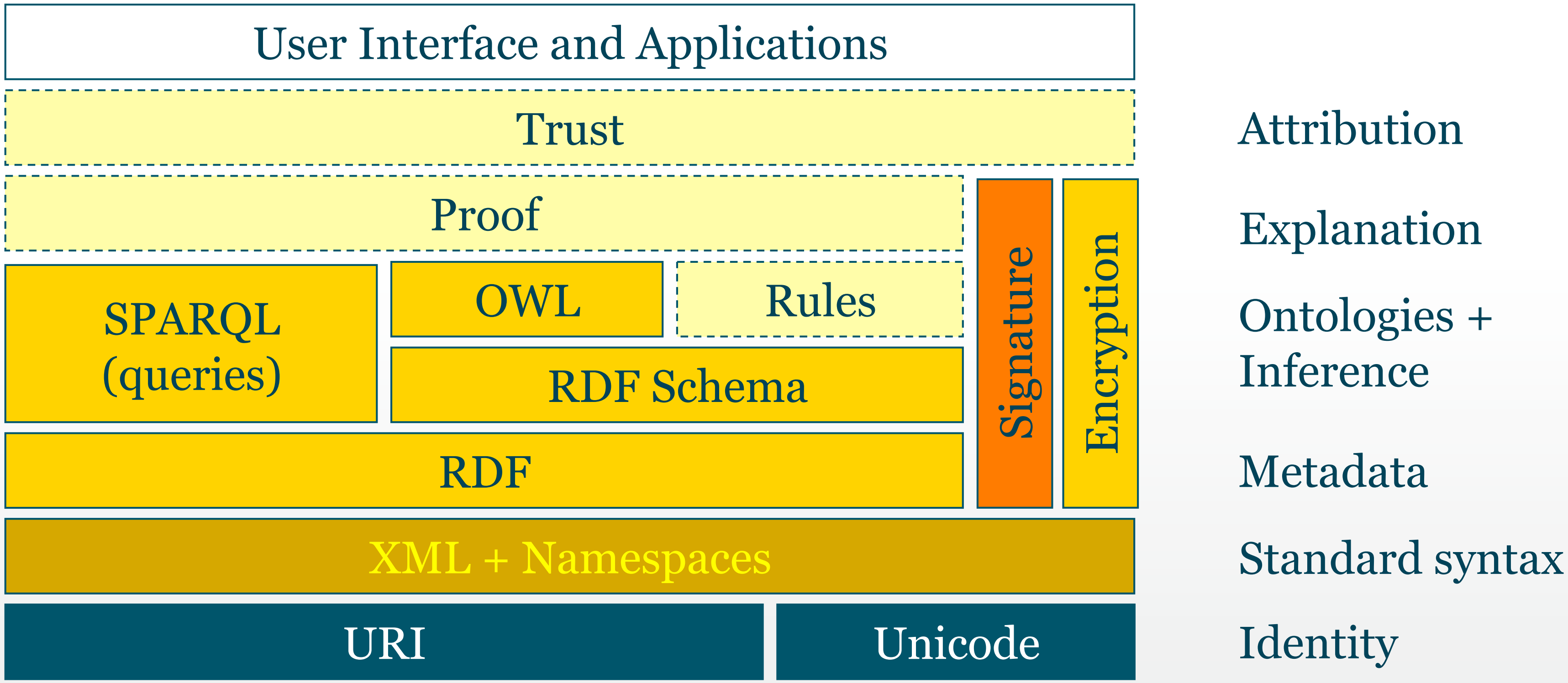
Semantic Web Technical Architecture

Fundamental Principles

- Anyone can make assertions about anything
- Entities are referred to using Uniform Resource Identifiers
- Based on XML technologies
- Formal semantics

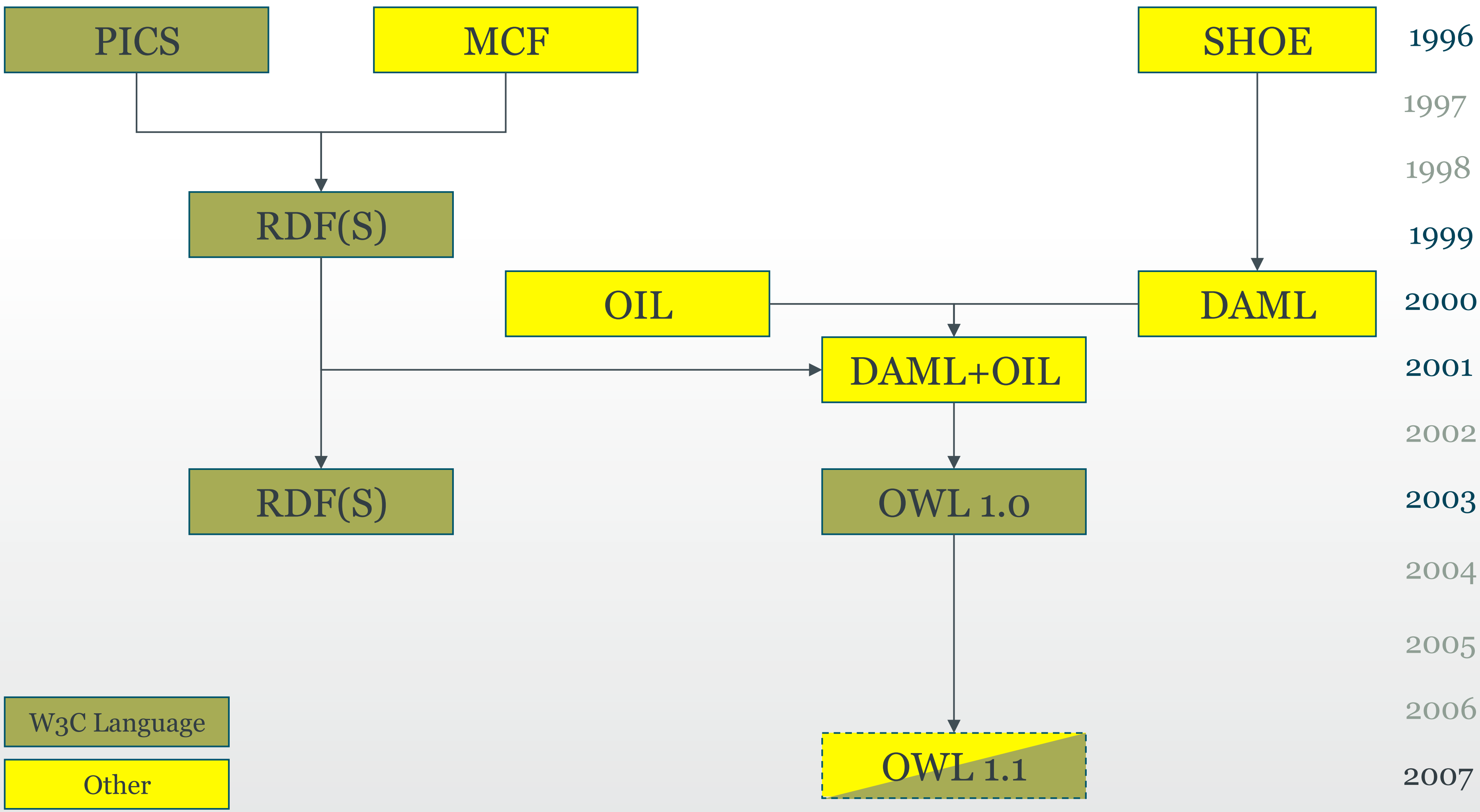
UNIVERSITY OF Southampton
School of Electronics and Computer Science

The Semantic Web layer cake



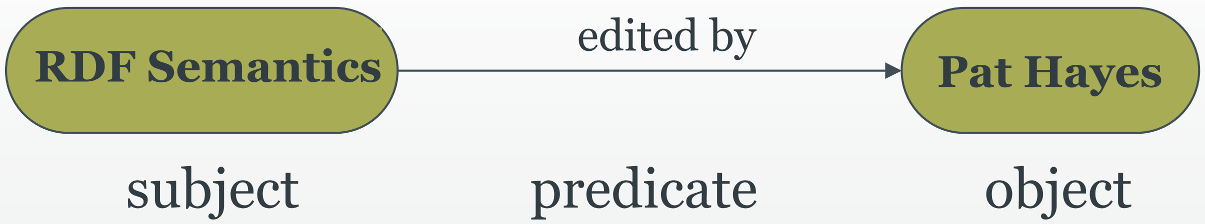
Languages of the Semantic Web

UNIVERSITY OF Southampton
School of Electronics and Computer Science



The triple

- Underlying model of **triples** used to describe the relations between entities in the Semantic Web
- (*subject, predicate, object*)
  - e.g. “RDF Semantics”, “edited by”, “Pat Hayes”



- Network knowledge representation
  - Labelled, directed graph
  - Entities as nodes, relations as edges

Example

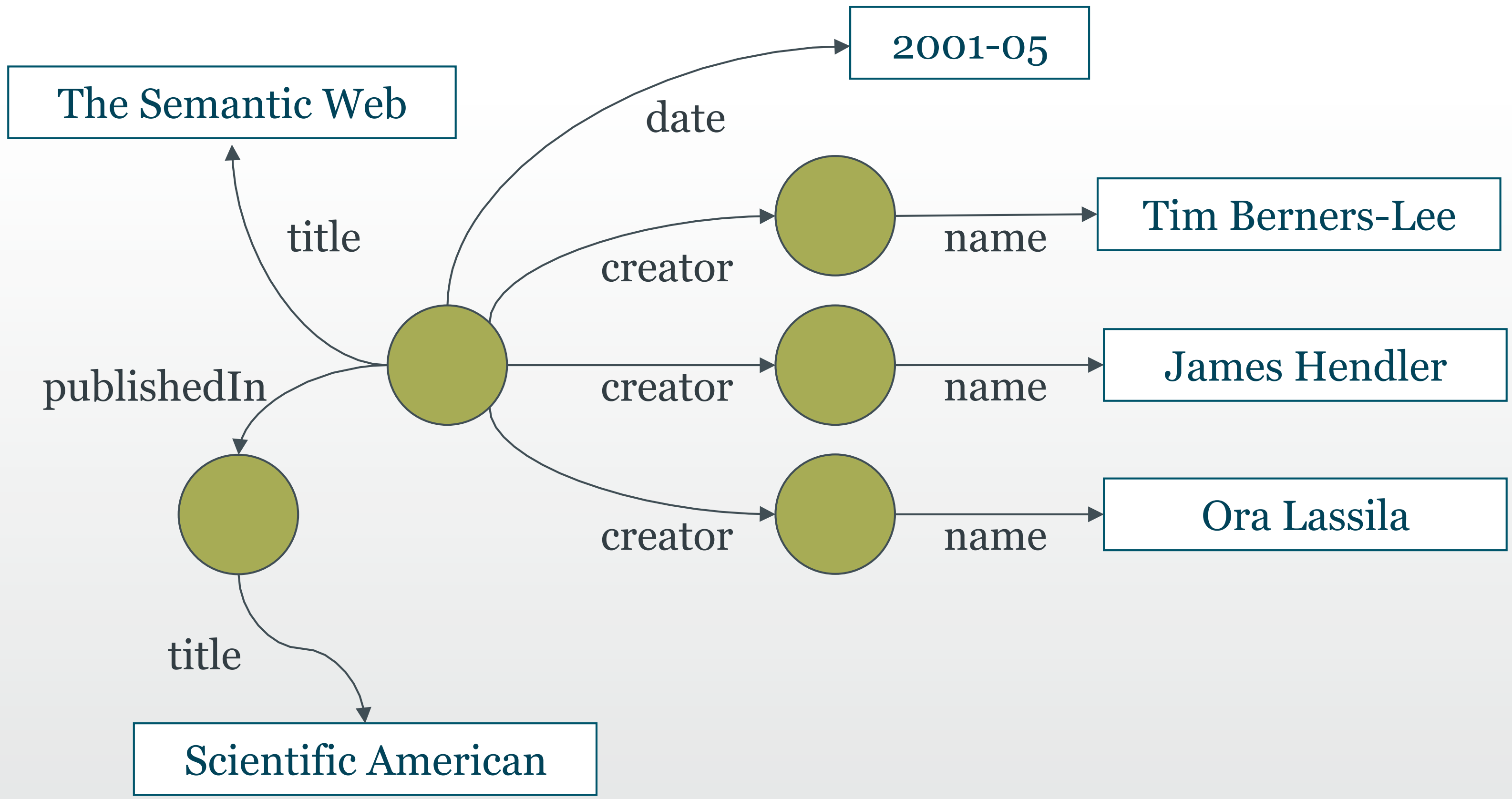
- Take a citation:

  Tim Berners-Lee, James Hendler and Ora Lassila. The Semantic Web. *Scientific American*, May 2001

- We can identify a number of distinct statements in this citation:
  - There is an article titled “The Semantic Web”
  - One of its authors is a person named “Tim Berners-Lee” (etc)
  - It appeared in a publication titled “Scientific American”
  - It was published in May 2001

Example

- We can represent these statements graphically



Example

- There are two types of node in this graph:
  - **Literals**, which have a value but no identity (a string, a number, a date)

Scientific American

  - **Resources**, which represent objects with identity (a web page, a person, a journal)

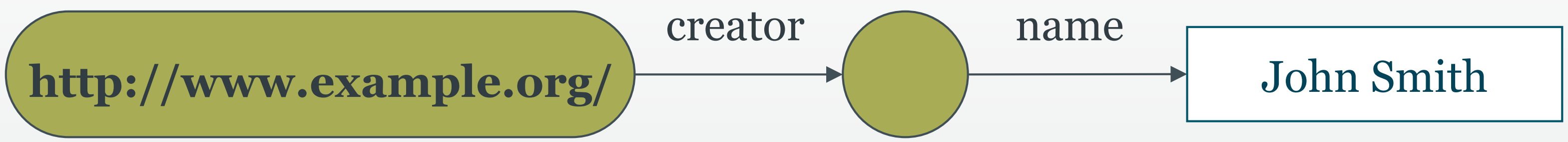
Example

- Resources are identified by URIs
- Property labels are also identified by URIs, and are drawn from a vocabulary or ontology

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Blank nodes (bNodes)

- Sometimes we have resources which we do not wish to identify with a URI
- These are **blank nodes** or **anonymous resources**



UNIVERSITY OF Southampton
School of Electronics and Computer Science

Resource Description Framework

Resource Description Framework

- RDF is a language for representing information about resources on the World Wide Web and beyond
- RDF uses Uniform Resource Identifiers to identify things and the relations between them
- XML-based standard syntax for RDF

UNIVERSITY OF Southampton
School of Electronics and Computer Science

RDF Vocabulary Description Language (RDF Schema)

Limitations of RDF

- RDF lets us make assertions about resources using a given vocabulary
- RDF does not let us define these domain vocabularies by itself
- RDF Schema is an RDF vocabulary which we can use to define other vocabularies

Defining an RDF vocabulary

- Define classes of objects and their relationship with other classes
  - "there is a class called Employee which is a subclass of the class called Person"
- Define properties that relate objects together and their characteristics
  - "there is a property called worksFor which relates objects of class Employee to objects of class Company"

Using RDF to define RDFS

- RDFS is an RDF vocabulary which contains:
  - Classes for defining classes and properties
  - Properties for defining basic characteristics of classes and properties
  - Ancillary properties to aid curation

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Web Ontology Language (OWL)

Rationale

- RDF Schema is not expressive enough for many applications
  - Only supports explicit class/property hierarchies
  - Only supports global range and domain constraints

OWL features

- Property restrictions
  - Local range/cardinality/value constraints
- Equivalence and identity relations
- Property characteristics
  - Transitive/Symmetric/Functional
- Complex classes
  - Set operators
  - Enumerated classes
  - Disjoint classes

OWL integration with RDF(S)

- OWL syntax is RDF-based
  - Existing RDF tools can use OWL-defined vocabulary
  - OWL tools can use RDF-defined knowledge
- OWL semantics needs to be compatible with RDF(S)
  - Meaning of equivalent OWL and RDF(S) constructs must be equivalent, given different model theories for RDF(S) and OWL

UNIVERSITY OF Southampton
School of Electronics and Computer Science

OWL species

- OWL provides three sublanguages of increasing expressivity (and complexity)
  - OWL Lite
  - OWL DL
  - OWL Full

UNIVERSITY OF
Southampton
School of Electronics
and Computer Science

SPARQL Protocol and RDF Query Language

RDF Query Languages

- No query language was defined in the RDF or OWL specifications
- Several non-W3C query languages for RDF had been proposed:
  - SerQL
  - RDQL
  - RQL
  - SquishQL
  - …
- W3C Data Access Working Group tasked with designing a standard query language for RDF

UNIVERSITY OF Southampton
School of Electronics and Computer Science

SPARQL

- The SPARQL Protocol and RDF Query Language
  - Became a W3C Recommendation in January 2008
- Simple SQL-like language for querying RDF systems
- REST-ful protocol defined in WSDL
- Currently supported by Jena, Sesame and 3store (among others)