SENG 42273

Semantic web and Ontological Engineering

Lecture 2

The world of Semantic Web

- The current Web is made up of many Web documents (pages).
- Any given Web document, in its current form (HTML tags and natural text), only gives the machine instructions about how to present information in a browser for human eyes.
- Therefore, machines have no idea about the meaning of the document they are presenting; in fact, every single document on the Web looks exactly the same to machines.
- Machines have no way to understand the documents and cannot make any intelligent decisions about these documents.
- Developers cannot process the documents on a global scale (and search engines will never deliver satisfactory performance).
- One possible solution is to modify the Web documents, and one such modification is to add some extra data to these documents; the purpose of this extra information is to enable the computers to understand the meaning of these documents.
- Assuming that this modification is feasible, we can then construct tools and agents running on this new Web to process the document on a global scale; and this new Web is now called the Semantic Web.

Suppose you have a relational database with two tables "StudentMaster" and "StudentCourse".

StudentNo	StudentName

StudentMaster

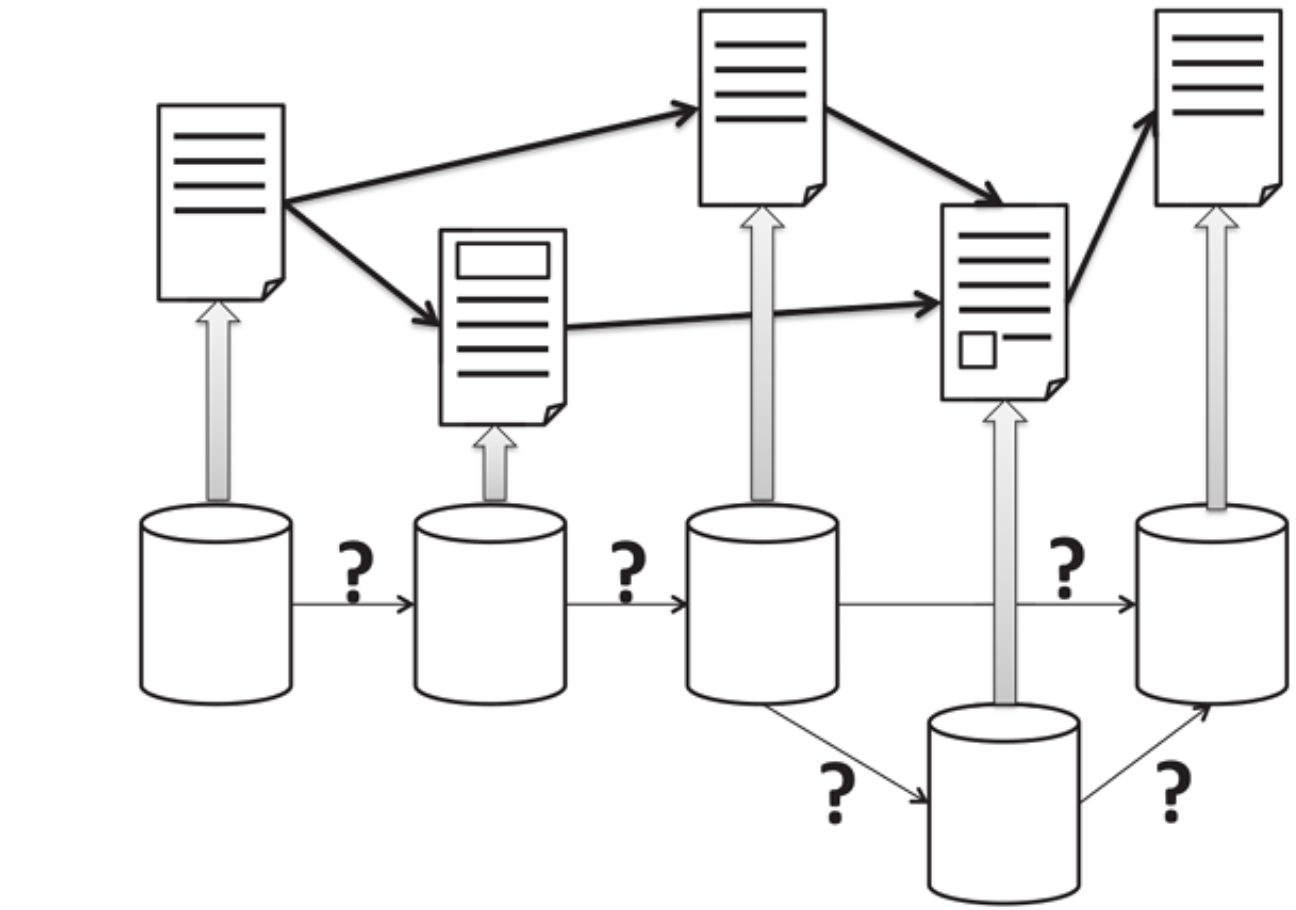
StudentNo	Course	AcademicYear

StudentCourse

How do you get the names of the students registered for a particular course?

- Web search engine should be able to do a similar operation to combine web data sources to produce the desired result for user queries.

Structured and unstructured data on the web



World Wide Web Consortium (W3C)

- There is a dedicated team of people at World Wide Web Consortium (W3C) working to improve, extend, and standardize the system.
  - Having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration, and reuse of data across various applications
  - Building a data model Or in other words, building a machine-readable (machine-understandable)

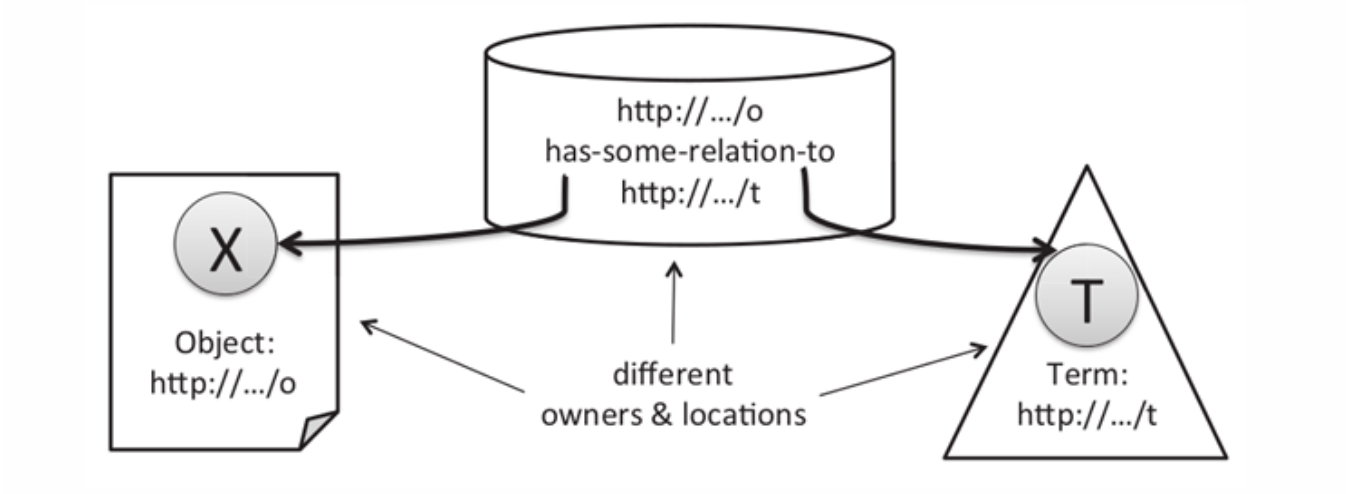
Web For further details: https://www.w3.org/standards/semanticweb/

Design Principles

- The Semantic Web (or The Web of Data, as it is becoming known in recent years) follows different design principles, which can be summarized as follows:
  - Make structured and semi-structured data available in standardized formats on the web;
  - Make not just the datasets, but also the individual data-elements and their relations accessible on the web;
  - Describe the intended semantics of such data in a formalism, so that this intended semantics can be processed by machines.

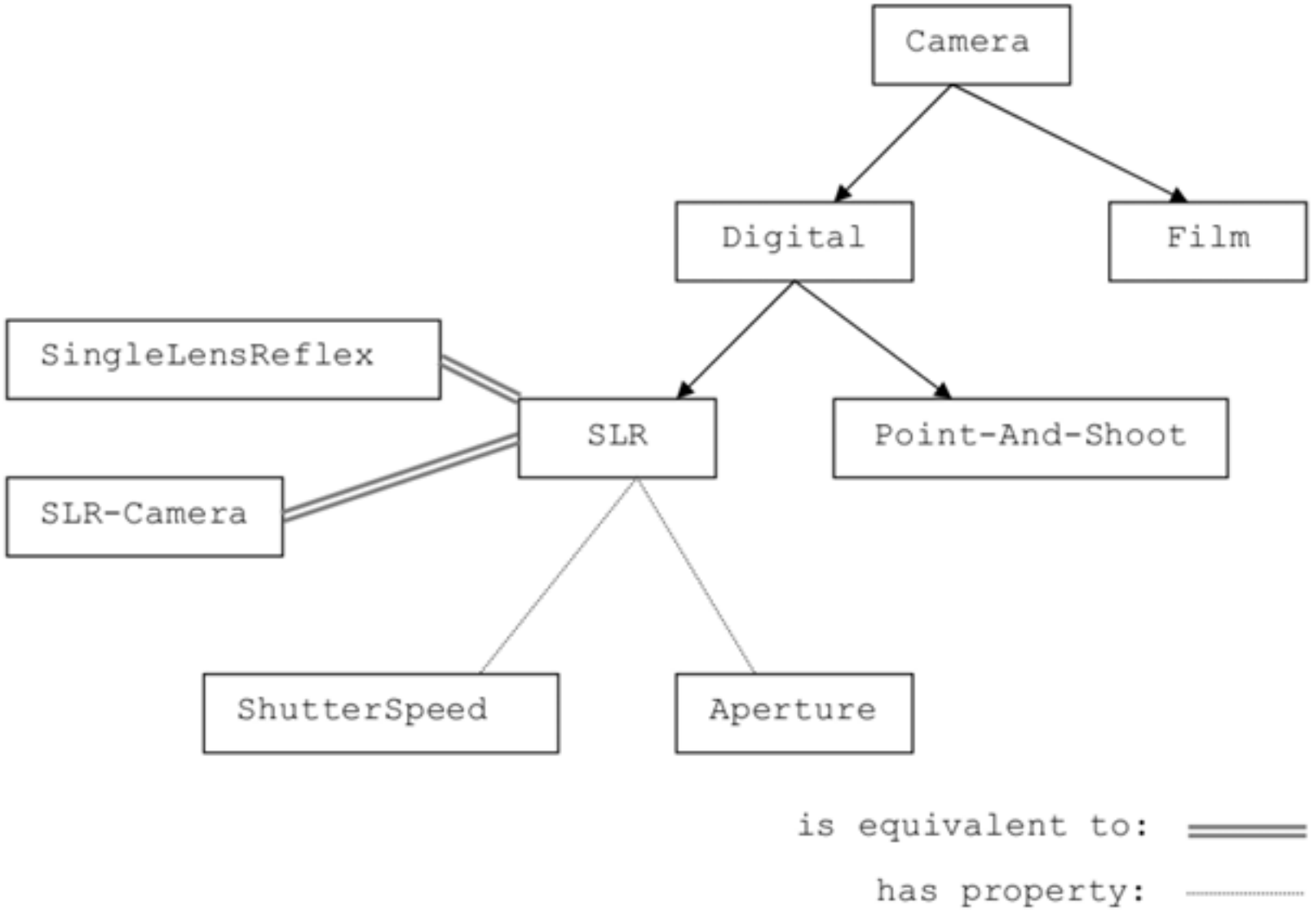
The Architecture of Semantic web

- One can publish a dataset on the web, another can independently publish a vocabulary of terms, and a third party may decide to annotate the object published by the first one with a term published by the second one, without asking for permission from either of them, and in fact without either of these two having to even know about it.



To implement architectural principles

1. Must agree on standard syntax to represent data and metadata
2. Must have sufficient agreement on the metadata vocabularies in order to share intended semantics of the data
3. must publish large volumes of data in the formats of step 1, using the vocabularies of step 2


Camera
Digital
Film
SingleLensReflex
SLR
Point-And-Shoot
SLR-Camera
ShutterSpeed
Aperture
is equivalent to:
has property:

Semantic web search engines

- Building a common language or vocabulary to explain "things" in a particular domain (Eg: laptop computers). The "thing" explained as a connected set of concepts and properties
  - It is a standard way to express the meaning/knowledge of a specific domain
  - It is a common understanding/language/meaning/knowledge shared by different parties over the Web

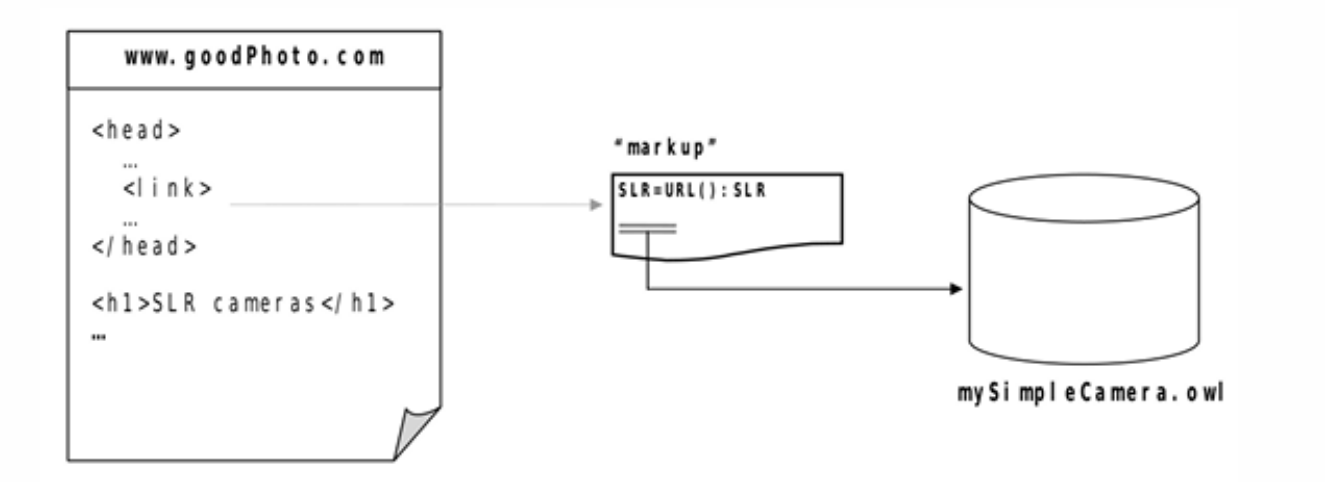
Semantic web search engines

How a Semantic Web search engine would work with the vocabularies

- First, it will scan all the available common vocabularies and find all the ones that define the given concept(key word)
- Then presents the "markup file" which contain the semantic meanings of the given concept (key word)
- Create the index table perform the searching

Markup pages

- How to indicate that a given key word in a particular web page has the semantic meaning defined in the vocabulary?
- In Semantic Web, semantic similarities between the words on the given page and the concepts defined in the vocabulary have to be explicitly expressed.
- Markup pages to add some extra data (meta data) or information to the Web page to describe some specific characteristic of the page.

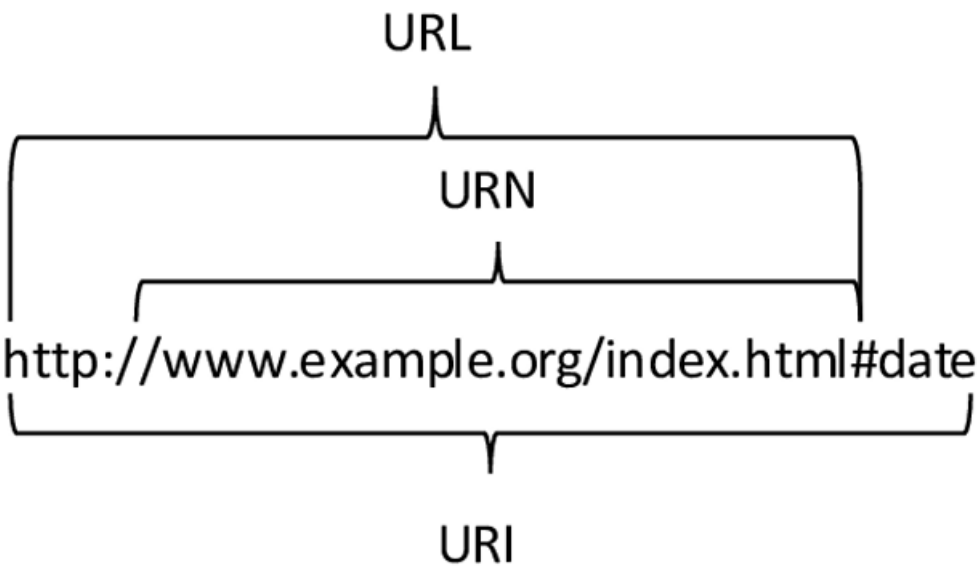


Basic Technology for the Semantic web

- Use labeled graphs as the data model for objects and their relations, with objects as nodes in the graph, and the edges in the graph depicting the relations between these objects. “Resource Description Framework” (RDF) is used as the formalism to represent such graphs.
- Use web identifiers (Uniform Resource Identifiers- URI) to identify the individual data-items and their relations that appear in the datasets.
- Use ontologies (briefly: hierarchical vocabularies of types and relations) as the data model to formally represent the intended semantics of the data.

What is a web resource?

- Any identifiable resource (digital, physical, or abstract) present on or connected to the WWW (Wikipedia)
- Resources are identified using Uniform Resource Identifiers (URI)

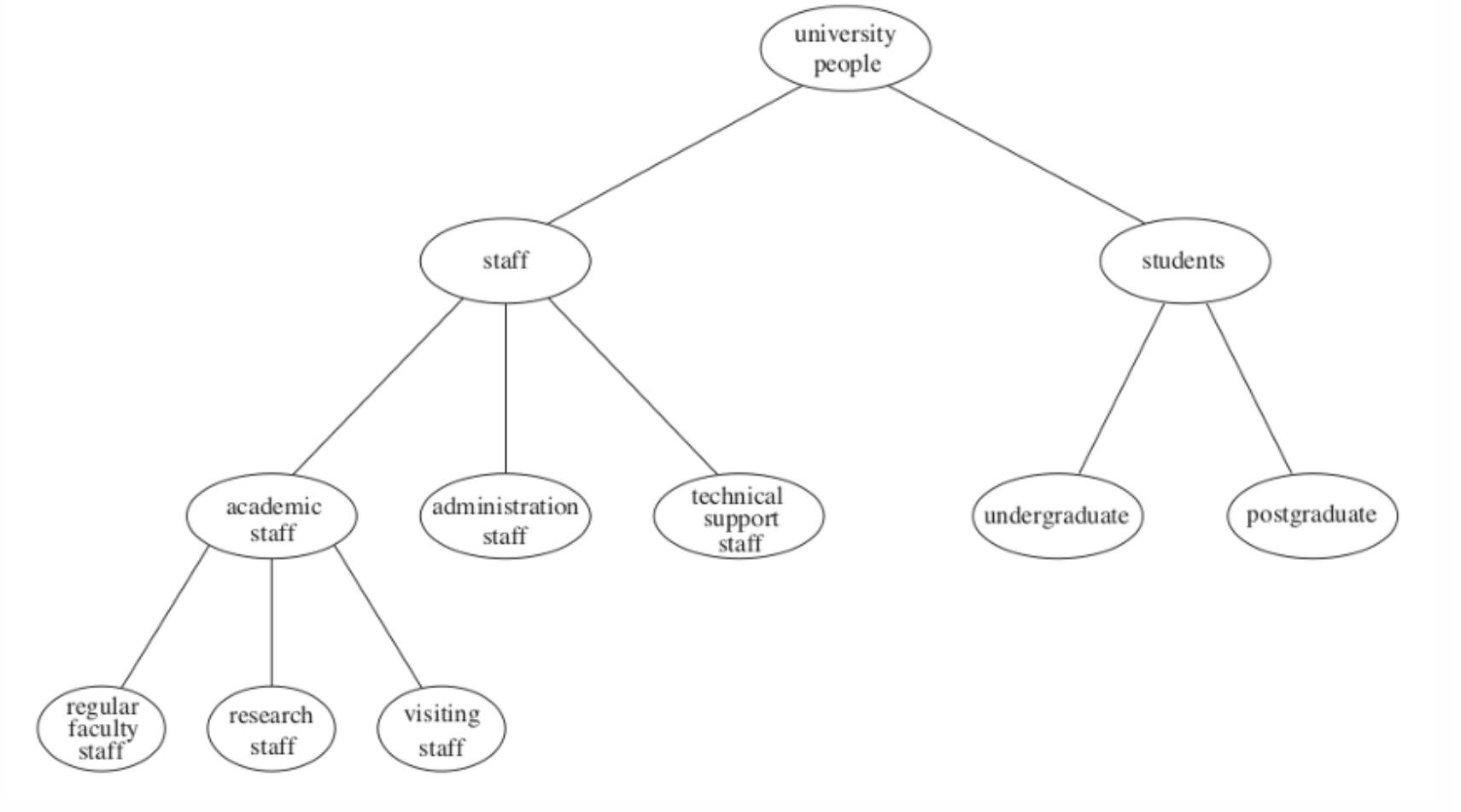


RDF (Resource Description Framework)	Key concepts: Resource, Property, Statement
RDFS (Resource Description Framework Schema)	Lightweight knowledge representation language
OWL(Web Ontology Language)	Lightweight knowledge representation language, richer than RDFS

Ontologies

- Consists of a finite list of terms and the relationships between these terms (describes formally a domain of discourse). The terms denote important concepts (classes of objects) of the domain. For example, in a university setting, staff members, students, courses, lecture theaters, and disciplines are some important concepts.
- The relationships typically include hierarchies of classes. A hierarchy specifies a class C to be a subclass of another class C if every object in C is also included in C.

A hierarchy


university people
staff
students
academic staff
administration staff
technical support staff
undergraduate
postgraduate
regular faculty staff
research staff
visiting staff

Ontologies

- Provide a shared understanding of a domain in the context of web (Eg: one application’s zip code may be the same as another application’s postcode)
- Useful for the organization and navigation of websites. Many web-sites today expose on the left-hand side of the page the top levels of a concept hierarchy of terms.

Resource Definition Framework (RDF)

- In this section, we will learn the following facts about RDF:
  - RDF is a language recommended by W3C, and it is all about metadata
  - RDF is a graph data model
  - RDF is capable of describing any fact (resource) independent of any domain
  - RDF provides a basis for coding, exchanging, and reusing structured metadata
  - RDF is structured; i.e., it is machine-understandable. Machines can do useful operations with the knowledge expressed in RDF
  - RDF allows interoperability among applications exchanging machine-understandable information on the Web

  A model to represent/assemble knowledge

Resource Definition Framework (RDF)

- RDF is language used to add meta data to web resources. RDF:
  - Has a vocabulary
  - Has syntax
  - Use to build graph data model (Subject, Predicate, Object)
  - A domain independent data model (Ex: an application should be able to interpret the concept "friend" in social network description and and concept "city" in geographical description)

RDF described as a language as well as a data model. It is more of a framework for describe resources in the web.

Resource Definition Framework (RDF)

- Resource Definition Framework use to create Markup file. Thus, RDF is the language we use to construct these metadata files.
  - RDF is the basic building block for supporting the Semantic Web
  - RDF is to the Semantic Web what HTML has been to the Web (assume that you have learnt web development and HTML)
- RDF is an XML-based language for describing information contained in a Web resource. A Web resource can be a:
  - Web page, an entire Web site
  - Any item on the Web that contains information in some form

Resource Definition Framework (RDF)

- Perhaps, you are wondering about the URIs and URLs in a RDF statement. Where do they come from?
  - Linked Open Data Cloud (LOD)
    - Wikipedia, Dbpedia, GeoNames, IMDb, ...
  - Four design principles of linked data
    - Use URIs as names for things
    - Use HTTP URIs so that people can look up these names (eg: "http://www.setu.kln.ac.lk/courses/seng42273/notes")
    - When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
    - Include links to other URIs so that they can discover more things

Assignment 01

- Visit the Linked Open Data (LOD) Cloud website and identify at least three datasets available in the LOD cloud.
- For each dataset you identified, summarize its purpose and the type of data it provides.

Deadline: 10th November 2024

Q&A