



1

INTRODUCTION AU WEB SÉMANTIQUE

Philippe GENOUD – Danielle ZIEBELIN - LIG-Steamer

Prenom.Nom@imag.fr

https://www.w3.org/standards/semanticweb/

In addition to the classic "Web of documents" W3C is helping to build a technology stack to support a "Web of data," the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term "Semantic Web" refers to W3C's vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.

WEB architecture





Web evolution



*d'après le cours Technologie du Web – A. Hombiat - Licence Professionnelle Études Statistiques et Systèmes d'Information Géographique (ESSIG) – UPMF - 2015



Web evolution





From Web of Documents to Web of Data

"The web of human-readable document is being merged with a web of machine understandable data."

- The traditional web (Web of Documents) is for humans
 - based on the HTML markup language
 - HTML describes
 - what information is presented + how it's presented in conjunction with CSS
 - how information is linked
 - but not what the information means
 - meaning (semantics) of information is derived from available information





From Web of Documents to Web of Data

"The web of human-readable document is being merged with a web of machine understandable data."

"The semantic web is an extension of the current web in which information is given welldefined meaning, better enabling computers and people to work in cooperation"

> Tim Berners-Lee, James Hendler and Ora Lassila , *The Semantic Web* Scientific American, 284(5), pp 34-43 (2001)



- The Semantic Web (Web of Data) is for machines
 - meaning of entities is defined explicitly using formal (structured) and standardized knowledge representations (ontologies)
 - ➔ possibility to
 - process the meaning of information automatically
 - automatically deduce implicit information from existing data
 - relate and integrate heterogeneous data





From Web of Documents to Web of Data

Semantic web standards



W3C SEMANTIC WEB ACTIVITY

On this page \rightarrow publications, interviews \circ presentations \circ active groups \circ completed groups \circ past groups

W3C Semantic Web

The Semantic Web is a web of data. There is lots of data we all use every day, and it is not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar?

Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself.

The Semantic Web is about two things. It is about common formals for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.

What is the Semantic Web?

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise. and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF). See also the separate FAQ for further information

Warning: this Activity has been subsumed, in December 2013, by the <u>W3C Data Activity</u>. That activity has a larger scope; new or current Working and Interest Groups related to "traditional" Semantic Web technologies are now part of that Activity.

The current page has been frozen on the 11th December, 2013.

http://www.bnode.org/blog/tag/layer%20cake (Benjamin Nowack)



Linked Data Principles



Tim Berners-Lee

http://www.w3.org/DesignIssues/LinkedData.html

a set of best practices for publishing and interlinking structured data on the Web

Basic idea: to apply the general architecture of the World Wide Web to the task of sharing structured data on global scale.

- 1. Use URIs as names for things.
- 2. Use **HTTP URIs**, so that people can look up those names.
- 3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL...).
- 4. Include **links** to other URIs, so that they **can discover more things**.



Uniform Resource Identifiers (URIs)



http://www.bnode.org/blog/tag/layer%20cake (Benjamin Nowack)













Uniform Resource Identifiers (URIs)



Different URIs (URNs) in different namespaces can represents the same thing





Resources in the Web of data

- Data describe elements (things) for a domain of interest through their ٠ properties and relationships.
- Theses elements can be: ٠



All these elements are resources identified by an URI •



R

Ε

S

0

U

С

Ε

S

Examples of resources about Georges Brassens available in the web





http://fr.wikipedia.org/ wiki/Georges_Brassens









http://culturetheque.org.uk/media/item/

Video (mp4 file) <u>https://www.youtube.com</u> <u>/watch?v=rslShTbqNbo</u>



17545/800/brassens.jpg

HTTP URIS



http://www.bnode.org/blog/tag/layer%20cake (Benjamin Nowack)



HTTP URIS

- Where HTTP URIs identify real-world objects or abstract concepts, it is essential to **not confuse** the objects or concepts themselves with the Web documents that describe them.
 - A real world object or abstract concept can have different representations
 - It allows separate statements to be made about an object and about a document that describes that object.

URI of DBpedia resource representing Georges Brassens http://dbpedia.org/resource/Georges_Brassens







Making URIs Dereferenceable Content negotiation (303 URIs)

• The same URI can be used to retrieve different representations.





Resource Description Framework (RDF)



http://www.bnode.org/blog/tag/layer%20cake (Benjamin Nowack)



Resource Description Framework

- RDF Resource Description Framework
 - a framework for describing resources on the web

"The Resource Description Framework (RDF) is a framework for representing information in the Web." [1]

- is designed to be read and understood by computers
- RDF is a part of the W3C's Semantic Web Activity
 - became a W3C recommendation 10. February 2004
 - Updated February 2014 (RDF 1.1)

http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/ [1]

http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/ [2]



Resource ?

 "The Resource Description Framework (RDF) is a language for representing information about resources in the World Wide Web. It is particularly intended for representing metadata about Web resources, such as the title, author, and modification date of a Web page, copyright and licensing information about a Web document, or the availability schedule for some shared resource. However, by generalizing the concept of a "Web resource", RDF can also be used to represent information about things that can be identified on the Web, even when they cannot be directly retrieved on the Web."

http://www.w3.org/TR/rdf-primer/

"To publish data on the Web, the items in a domain of interest must first be identified. These are the things whose properties and relationships will be described in the data, and may include Web documents as well as real-world entities and abstract concepts. As Linked Data builds directly on Web architecture, the Web architecture term resource is used to refer to these things of interest, which are, in turn, identified by HTTP URIs."

Tom Heath, Christian Bizer : *Linked Data: Evolving the Web into a Global DataSpace* <u>http://linkeddatabook.com/editions/1.0/</u>



RDF Data Model

- With RDF, knowledge is represented by a set of assertions (statements)
- All RDF statements follow a simple structure composed of three parts :
 - the thing the statement describes
 - the properties of the thing the statement describes
 - the values of those properties the statement describe





RDF Data Model

• RDF Statements are *triples*

SubjectPredicateObjectURIURIURI/Literal

- the subject and the predicate are resources : RDF uses **URIs** (Universal Resource Identifiers) for **uniquely identifying** them
- object can be a **resource** (URI) or a **literal** (constants that don't have other attributes that describe them)





RDF Data Model

• RDF Statements are *triples*



- RDF data can be viewed as a directed labeled graph
 - subjects and objects are nodes (vertices)
 - predicates are oriented edges (arcs)





RDF Datasets

 In practice, RDF tool builders and data managers needed a mechanism to talk about subsets of a collection of triples → RDF 1.1 a mechanism to group RDF statements in multiple graphs and associate such graphs with an IRI.



RDF and **Data** Integration





Examples of Links between datasets

Prefixes

dbpedia: "http://dbpedia.org/resource/"
dbo: "http://dbpedia.org/ontology/"
geonames: "http://sws.geonames.org/"
gn: "http://www.geonames.org/ontology#"

insee: inseePop: inseeDef: "http://id.insee.fr/geo/commune/"
http://id.insee.fr/demo/populationLegale/commune/
http://rdf.insee.fr/def/demo#populationTotale

Different URIs (URNs) in different namespaces can represents the same thing



http://dbpedia.org/resource/Toulouse



"449328"^^xsd:integer

GeoNames

http://sws.geonames.org/2972315/

inseeDef:populationTotal



Web Sémantique – Ontologies P. Genoud D. Ziébelin

RDF limitations

• RDF provides a standard way to express simple statements about resources, using named properties and values...

But you can't express knowledge about the properties and types of the resources

- what are the types allowed for resources ?
- what are the properties allowed for a given type of resource ?
- what are the allowed values for a given property ?
- what are the relations between types of resources (generalization/specialization) ?

— ...



Vocabularies / taxonomies/ ontologies





What is a vocabulary ?

• A person's **vocabulary** is the set of words within a language that are familiar to that person.

(Wikipedia)

• = "all the words known and used by a particular person"

(Cambridge Advanced Learners Dictionary)

On the Semantic Web, **vocabularies** define the **concepts and relationships** (also referred to as "terms") used to describe and represent an area of concern.

Vocabularies are used

to classify the terms that can be used in a particular application,

characterize possible relationships,

and define possible constraints on using those terms.

In practice, vocabularies can be very complex (with several thousands of terms) or very simple (describing one or two concepts only).

http://www.w3.org/standards/semanticweb/ontology



RDF Schema - RDFS

- RDFS extends RDF with a schema vocabulary (Resource, CLass, Property, subCLassOf, subPropertyOf, range, domain ...) that allows to define basic vocabulary terms and the relations between them.
- A well-defined **semantics** gives "extra meaning" to these particular RDF predicates and resources
 - specifies how RDFS terms should be interpreted
 - allows to draw simple inferences (*entailments*)
 - Ex : type propagation



- W3C recommendation RDF Vocabulary Description Language –RDF 1.1 (Feb. 2014) http://www.w3.org/TR/rdf-schema/ (previous rec. RDF 1.0 Feb. 2004)
- the RDF(S) schema vocabulary is itself provided in the form of an RDF vocabulary
 - resources in the RDF Schema vocabulary have URIrefs with the prefix
 http://www.w3.org/2000/01/rdf-schema# (conventional abreviation rdfs:)



What is an Ontology ?

"ontology is the philosophical study of the nature of being, becoming, existence, or reality, as well as the basic categories of being and their relations. Traditionally listed as a part of the major branch of philosophy known as metaphysics, ontology deals with questions concerning what entities exist or can be said to exist, and how such entities can be grouped, related within a hierarchy, and subdivided according to similarities and differences." http://en.wikipedia.org/wiki/Ontology



Parmenides (c. 515 BCE - c. 460 BCE)

 "In computer science ..., an ontology is a formal framework for representing knowledge. This framework names and defines the types, properties, and interrelationships of the entities in a domain of discourse. The entities are conceptualizations (limited abstractions) of phenomena."

http://en.wikipedia.org/wiki/Ontology_%28information_science%29

An ontology is an explicit specification of a conceptualization. [...] A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose.

Thomas R. Gruber, *Towards Principles for the Design of Ontologies Used for Knowledge Sharing* in Formal Ontology in Conceptual Analysis and Knowledge Representation, Kluwer Academic Publishers, 1993 <u>http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.6025&rep=rep1&type=pdf</u> <u>http://tomgruber.org/writing/ontology-definition-2007.htm</u>



Thomas R. Gruber (1959 -)



Web Ontology Language (OWL)

- OWL vocabulary : a set of primitives (described in RDF that extends RDFS vocabulary

 - owl:Class, owl:ObjectProperty, owl:DataProperty, owl:sameAs, owl:unionOf, owl:InteresctionOf...
- Far more expressive than RDFS
 - Classes can be describe by union, intersection, complement, properties restrictions.
 - notions of classes or properties equivalence, resources equality,
 - notions of inverse, symmetric, transitive ... properties
 - properties cardinality...
- Formal specification (based on Description Logics)
 → support for automated reasoning
- a W3C standard
 - OWL 1 : W3C recommendation 10 Feb. 2004
 - OWL 2 : W3C recommendation 11 Dec. 2012 <u>http://www.w3.org/TR/owl2-overview/</u>





http://www.w3.org/TR/owl-features/

Simple Knowledge Organization System (SKOS)

- A lightweight OWL ontology designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary.
 - "l'objectif d'un thésaurus est de constituer des vocabulaires normalisés et d'organiser la liste des termes de ces vocabulaire sans forcément les définir, dans le but notamment d'indexer un corpus de ressources documentaires et de faciliter les recherches dans ce corpus" Le web Sémantique, Fabien Gandon, Catherine Faron-Zucker, Olivier Corby, ed. Dunod 2012
- Hierarchical relationships
 - Generic term (BT: broader term) , specific term (NT: narrower term).
 - Partitives relationships (whole-part relations) , instantiation relationships (to give examples)
- Associative relationships
 - RT: related term.
- Equivalent terms
 - ...
- W3C Standard (recommendation) : SKOS Simple Knowledge Organization System Reference (August 2009) <u>http://www.w3.org/TR/2009/REC-skos-reference-20090818/</u>



Vocabulary vs. Ontology

• There is no clear division between what is referred to as "vocabularies" and "ontologies".

The trend is to use the word "ontology" for more complex, and possibly quite formal collection of terms, whereas "vocabulary" is used when such strict formalism is not necessarily used or only in a very loose sense. Vocabularies are the basic building blocks for inference techniques on the Semantic Web. <u>http://www.w3.org/standards/semanticweb/ontology</u>





Merci de votre attention

