Technologies du Web Sémantique

Claudine Métral
Gilles Falquet

Organisation

Mercredi 10h15-11h50 cours, 12h20-14h séminaire

Moodle

- https://moodle.unige.ch/course/view.php?id=9007

Evaluation

- 3 projets
- test écrit

Introduction to the Semantic Web

G. Falquet

Semantic Web Technologies

Main ideas (2001)

A web

readable/understandable by software agents

- pages on the web would be meaningful to programs

encompassing not just documents but every kind of data one could imagine

- interconnecting data (stored in different servers)

A use case

[...] His sister, Lucy, was on the line from the doctor's office: "Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments."

[...]At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved the information about Mom's prescribed treatment within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services. It then began trying to find a match between available appointment times (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules.

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

Required technologies

knowledge representation

formally represent the information/knowledge content of a web site

data representation

data representation framework for semi-structured data

interconnection

global/shared object identification technique (for cross-server links)

shared 'vocabularies' and concept description

reasoning/computing services

logical inferences; computation (spatial, temporal, ...); decision making; ...

decentralized web services

Machine readable web ⇒ Knowledge representation

a typical web page

Human understanding of text (and layout)

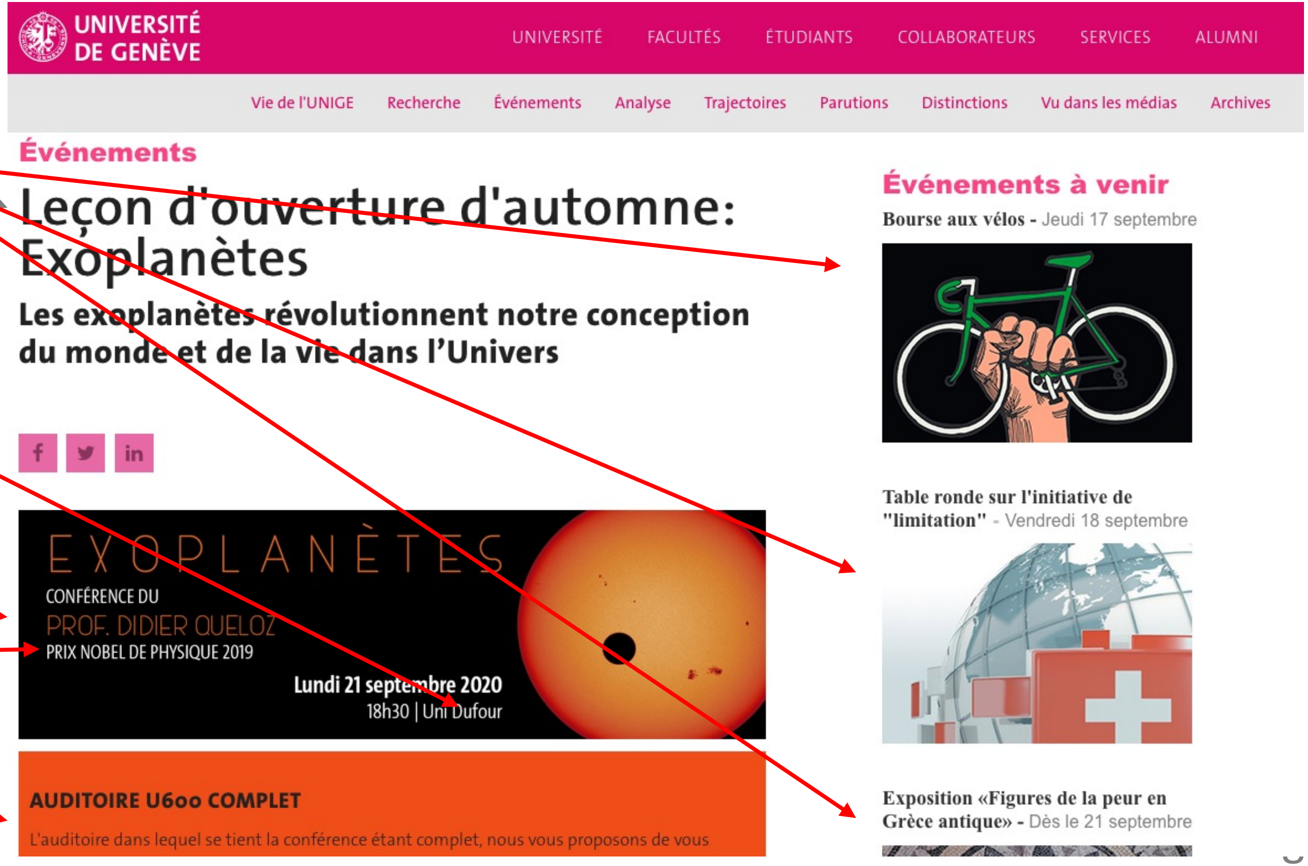
event descriptions

event place

presenter

presenter's attribute

special announcement



Machine understanding

text

text text text text text

text

text text text text

text text text text

text

text text

text

text text

text

towards a solution

- The first significant change required for this to happen is that data on the web [...] must be available in a machine-readable form with **defined semantics**. [...] general purpose languages could be defined in which assertions could be made, within which axiomatic concepts could be defined from time to time in human readable documents.
- In this case, the power of the language to combine concepts originating from different areas could lead to a very much more powerful system on which one could base **machine reasoning systems**.
- Knowledge Representation (KR) languages are something which, while interesting academically, have not had a wide impact on applications of computer. But then, the same was true of hypertext before the Web gave it global scope.

https://www.w3.org/People/Berners-Lee/1996/ppf.html

Example: A representation in first order logic

```
Event(c1)
Conference(c1)
title(c1, "Exoplanètes …")
speaker(c1, dq)
…
Person(dq)
name(dq, "Didier Queloz")
…
Event(bav)
title(bav, "Bourse aux vélos")
```

Requires some background knowledge

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Print/export
Download as PDF
Printable version

In other projects
Wikimedia Commons

Not logged in Talk Contributions Creat

Article Talk Read Edit View history Search Wikipedia

Academic conference

From Wikipedia, the free encyclopedia

"Symposium (academic)" redirects here. For other uses of symposium, see Symposium (disambiguation).

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed.
Find sources: "Academic conference" – news · newspapers · books · scholar · JSTOR *(May 2009) (Learn how and when to remove this template message)*

An **academic conference** or **scientific conference** (also **symposium**, **workshop**, or **meeting**) is an event for researchers (not necessarily academics) to present and discuss their work. Together with academic or scientific journals and Eprint archives such as arXiv, conferences provide an important channel for exchange of information between researchers.

Conference on medicam
assisted rehabilitation in

Contents [hide]

Overview [edit]

Conferences usually encompass various presentations. They tend to be short and concise, with a time span of about 10 to 30 minutes; presentations are usually followed by a discussion. The work may be bundled in written form as academic papers and published as the

Formalized in FOL

```
∀x Conference(x) →
    Event(x) ∧
    ∃t String(t) ∧ title(x, t) ∧
    ∃s Person(s) ∧ speaker(x, s)

∀x Person(x) →
    HumanBeing(x) ∧
    ∃p Place(p) ∧ birthplace(x, p) ∧
    ∀e employer(x, e) → (Organization(e) ∨ Person(p))
```

or in Description logics

```
Conference
    ⊑ Event
    ⊑ title some String
    ⊑ speaker min 1 Person

Person
    ⊑ HumanBeing
    ⊑ birthPlace some Place
    ⊑ employer only (Organization or Person)
```

Data representation

"A Web ... encompassing not just documents but every kind of data one could imagine"

$\rightarrow$ Data models for

- structured data
- texts (no structure)
- **semi-structured data**

Data representation ⇒ Semi-structured Data

“Roughly speaking and emphasize its main aspects semi-structured data is data that is neither raw data nor very strictly typed as in conventional database systems” (Abiteboul 1997)

Exemples

- Web pages about restaurants
- BibTeX files
- ...

Serge Abiteboul, “Querying Semi-structured data,” in *International Conference on Data Base Theory (ICDT)*, pp. 1 – 18, Delphi, Greece, 1997. http://dbpubs.stanford.edu:8090/pub/1996-19.

Example: A BibTeX file

@article{**miller1995wordnet**,
  Author = {Miller, George A},
  Journal = {Communications of the ACM},
  Number = {11},
  Pages = {39--41},
  Publisher = {ACM},
  Title = {WordNet: a lexical database for English},
  Volume = {38},
  Year = {1995}}

@techreport{**masolo2003wonderweb**,
  Author = {Masolo, Claudio and Borgo, Stefano and Gangemi, Aldo and Guarino, Nicola and Oltramari, Alessandro},
  Institution = {LOA-ISTC-CNR},
  Title = {The WonderWeb library of foundational ontologies and the DOLCE ontology. WonderWeb (EU IST project 2001-33052) deliverable D18},
  Year = {2003}}

@inproceedings{**niles2001towards**,
  Author = {Niles, Ian and Pease, Adam},
  Booktitle = {Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001},
  Organization = {ACM},
  Pages = {2--9},
  Title = {Towards a standard upper ontology},
  Year = {2001}}

Main aspects

Irregular structure

heterogeneous, incomplete elements

Implicit structure

structure in textual parts => parsing

Partial structure

unstructured parts: plain text, images, external data

Indicative structure vs. constraining structure

schema adds information

Main aspects

A-posteriori schema/data guide

created to structure existing data (from the data)

Large schema

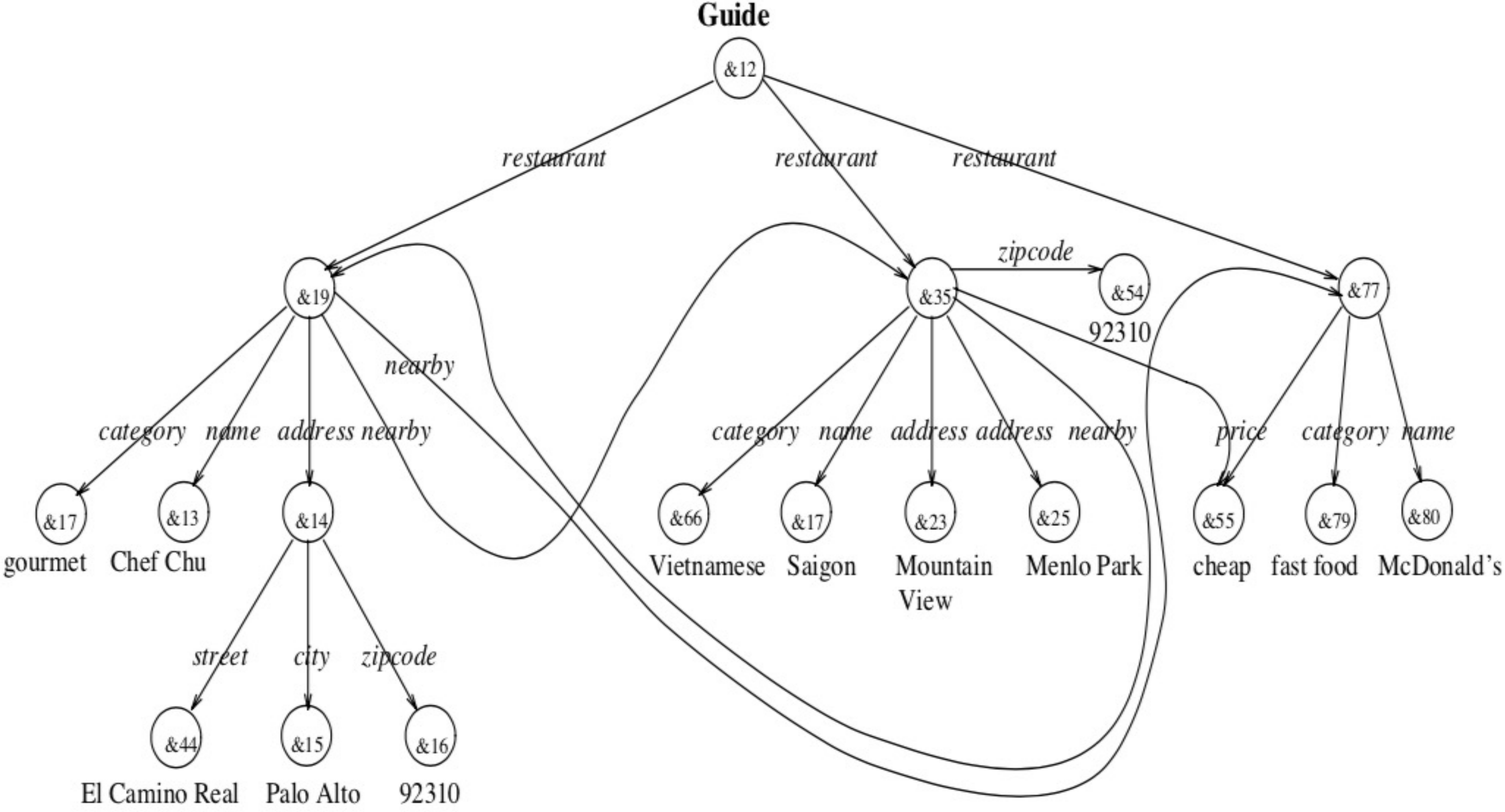
e.g., Wikidata has 3'000'000 classes

Schema ignored

in discovery/navigation queries the schema must be ignored

Rapidly evolving schema

e.g. in scientific databases (new techniques/knowledge)



Guide
&12
restaurant
restaurant
restaurant
&19
&35
zipcode
&54
92310
&77
nearby
category
name
address
nearby
category
name
address
address
nearby
price
category
name
&17
gourmet
&13
Chef Chu
&14
&66
Vietnamese
&17
Saigon
&23
Mountain View
&25
Menlo Park
&55
cheap
&79
fast food
&80
McDonald's
street
city
zipcode
&44
El Camino Real
&15
Palo Alto
&16
92310

Hierarchical Solutions

- XML
  - + XML Schema, XSL transformations, XML APIs

- JSON
  - + JSON APIs, JSON Schemas

in XML

```xml
<restaurant id="r19">
   <category>gourmet</category>
   <name>Chef Chu</name>
   <address>
      <street>El Camino Real</street>
      <city>Palo Alto</city>
      <zip>92310</zip>
   </address>
   <nearby ref="r35"/><nearby ref="r77"/>
</restaurant>
<restaurant id="r35">
   ...
</restaurant>
<restaurant id="r44">
   ...
</restaurant>
<restaurant id="r77">
   ...
</restaurant>
```

XML

- hierarchical model (+ references)
  - only one type of relation between elements: E1 sub-element of E2
- either no schema or a constraining schema
  - but XML schemas may have variants (optional elements)
- ...

in JSON

```
[{"type": "restaurant", "id": "r19"
  "category": "gourmet", "name": "Chef Chu",
  "address":
     {"street": "El Camino Real", "city": "Palo Alto", "zip": 92310},
  "nearby": ["r35", "r77"]
  },
  {"type": "restaurant", "id": "r35",
  ...},
  {"type": "restaurant", "id": "r44",
  ...},
  {"type": "restaurant", "id": "r77",
  ...}
]
```

Data Interconnection

Problems with databases

- no standard way to refer to entities across databases
- different databases may use different names for the same entity
- the same name may have different meanings in different databases
- ...

Interconnection

Problem: different databases use different identifiers for the same entity

Part	Origin	Color	Supplier
Motor	DE	White	
Windows	FR	--	
Wheels	USA	Orange	IBM
...			

in Database 1

Company	Headquarters	...
IBM corp.	Unites States	
Telefónica	Spain	
Orange	France	
...		

in Database 2

The Linked-data Solution

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

A resource is the main information building block

Anything that can be named is a resource.

Information resources entities that convey information and can be completely represented in binary code:

- documents, images, video, software ...

Non-information resources cannot be represented as bits:

- people, phenomena, concepts, ideas ...

Web resources are conceptual relations uniquely identified by HTTP URLs

- An HTTP URL points to at most one resource.
- If it is an information resource, HTTP allows clients to retrieve a representation of it.
    - The **concept** pointed to by an URL shouldn't change.
    - The **value** and **representations** retrieved when looking up an URL might change over time.

Using HTTP URIs ensures that anybody can look up the resource

An HTTP URI of a resource can be dereferenced:
use an HTTP client to retrieve a representation.

- Information resources result in a representation.
- Non-information resources result in a 303 redirect.

- Relies on the double role of an HTTP URI as identifier and locator.

- Principle: If you don't know something, look it up. Follow your nose.

Dereferencing a URI should lead to useful information about that resource

"Useful" means the information is available using standard technologies.

(RDF and SPARQL)

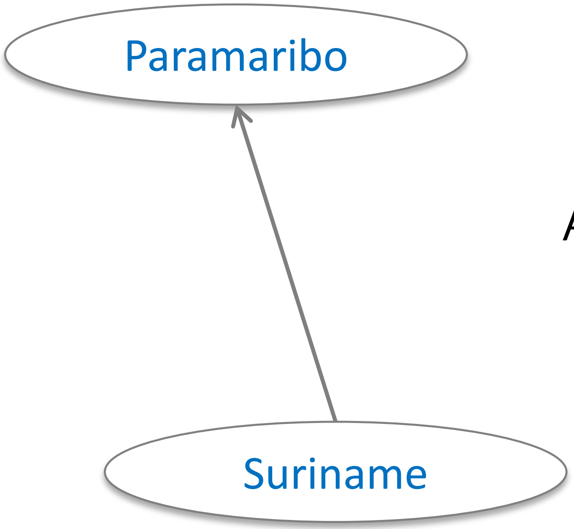
"Useful" also means the information provides explanations and/or context for the resource

Define the resource in terms of concepts the client already knows or can look up.

By including links to other resources, we create a Web of Data

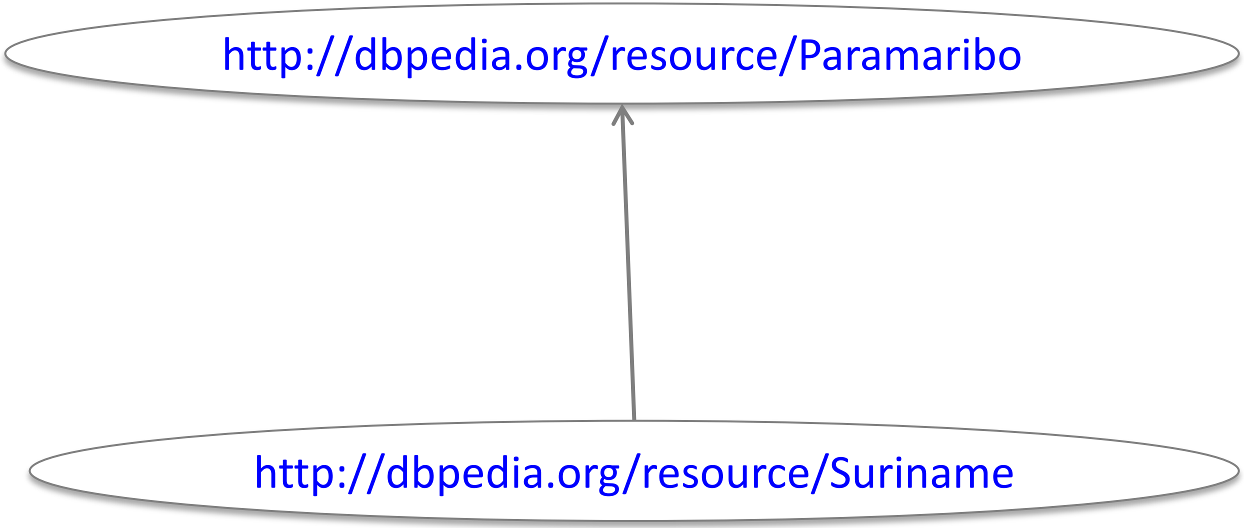
- Links connect a resource to known concepts.
  - Alberto is a researcher at U. of Toronto
- Links give meaning to data.
  - These temperatures are measured in degrees Celsius.
- Links allow exploration of related data.
  - Find more by the same author.

Basic information unit: the link

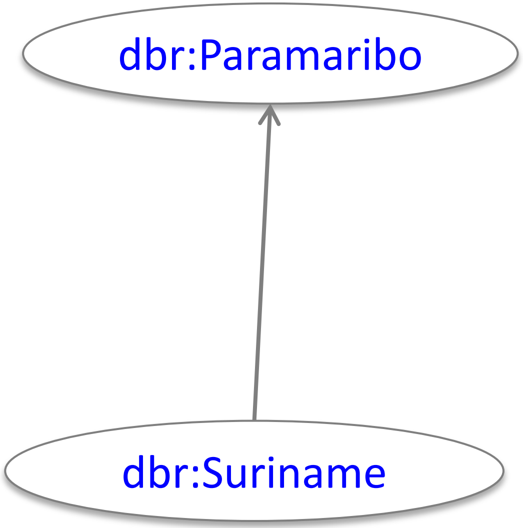


A link connects two resources

The resources are identified by URIs

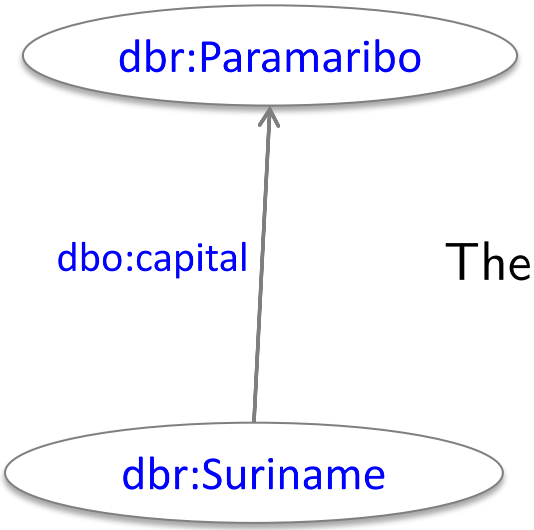


Using prefixes to abbreviate the URIs



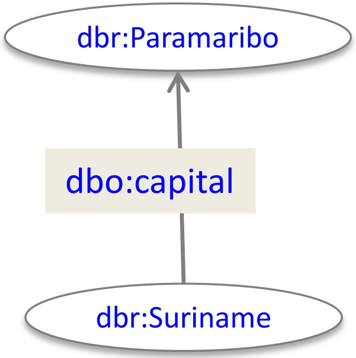
dbr = http://dbpedia.org/resource/

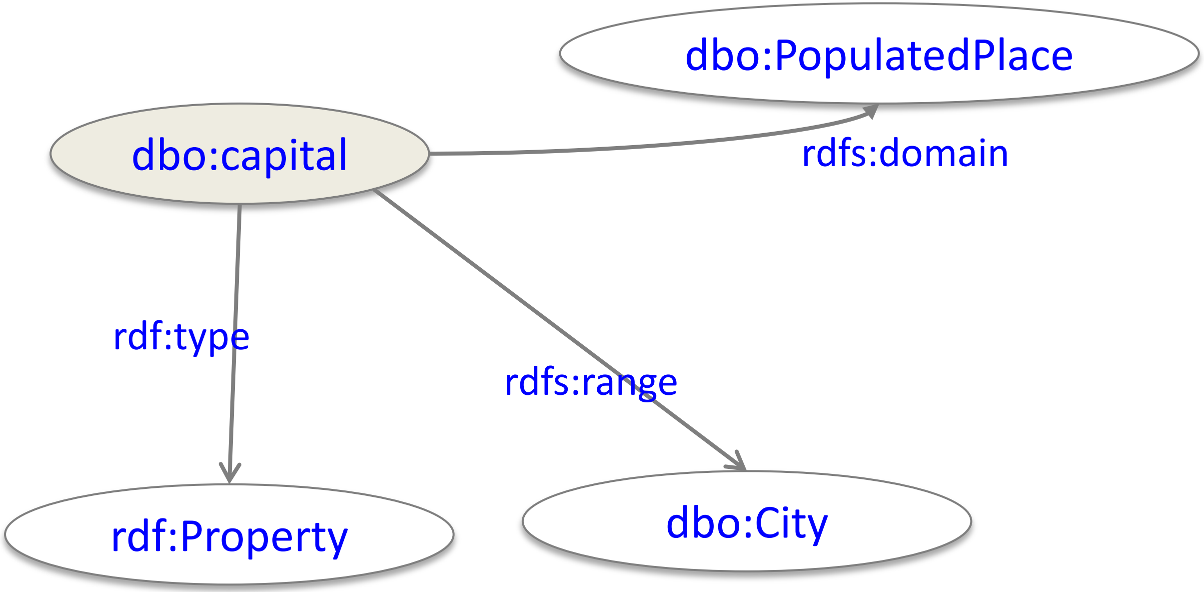
The links are typed (unlike Web links)



The link type is also identified by a URI

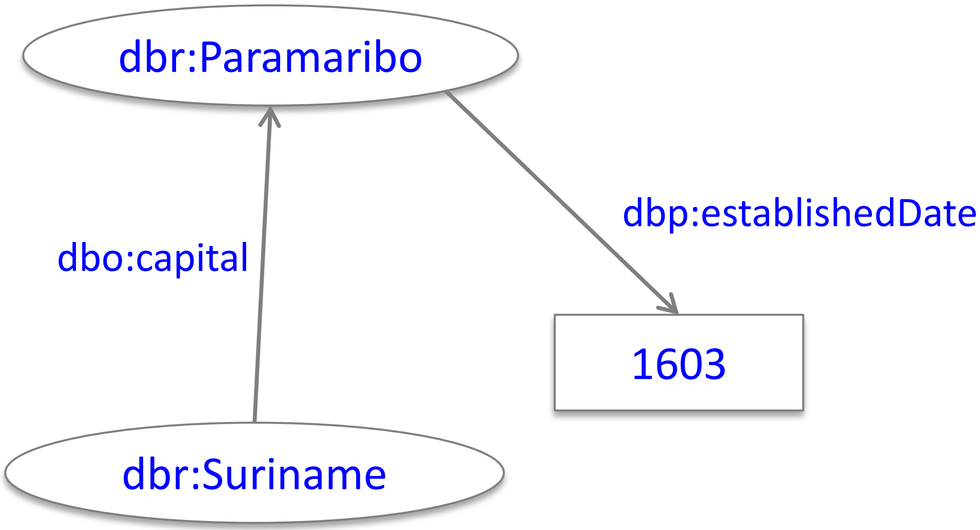
… so the link type can be described



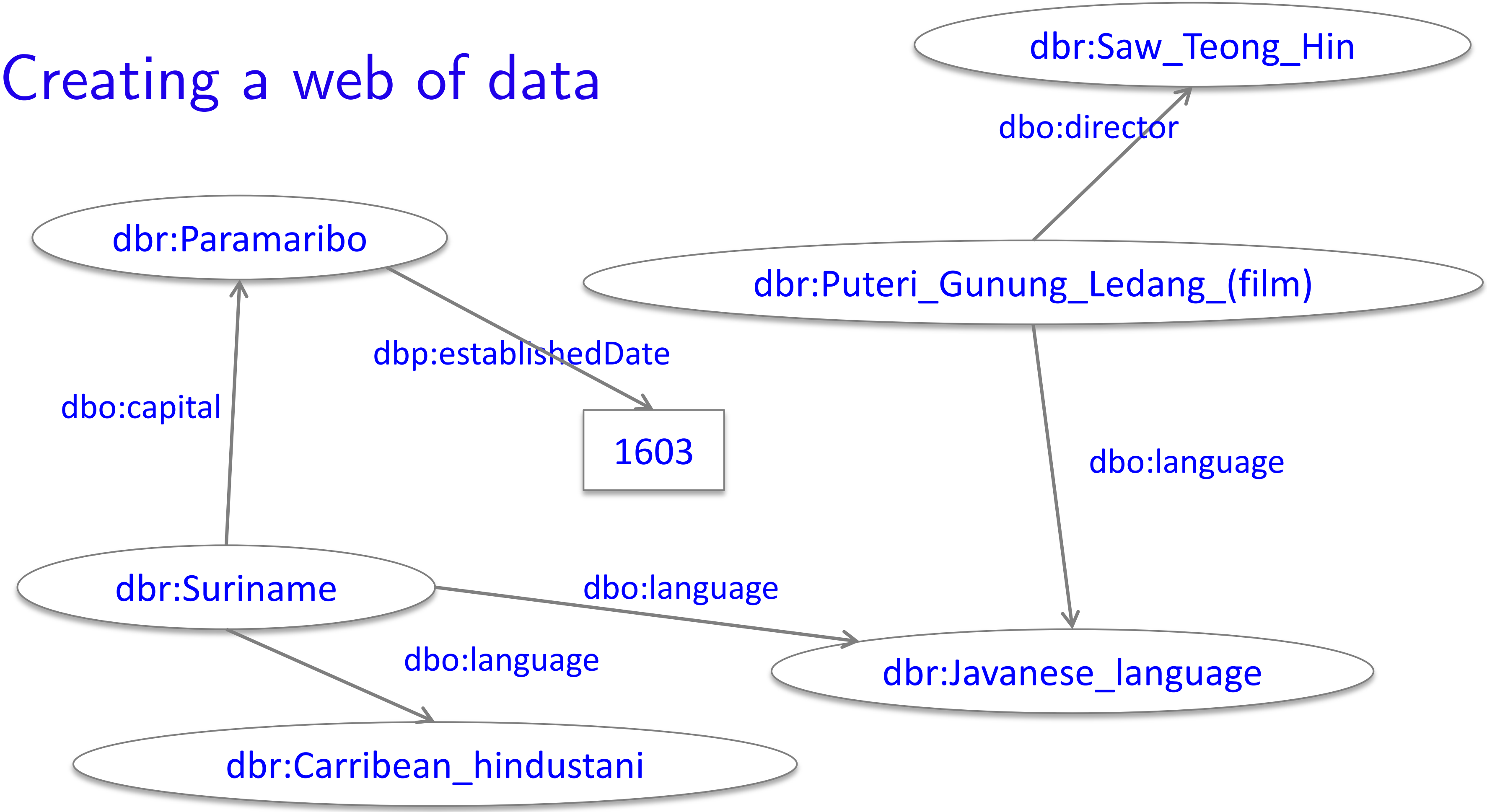


dbr = http://dbpedia.org/resource/
dbo = http://dbpedia.org/ontology/
rdf = http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs = http://www.w3.org/2000/01/rdf-schema#

Links can point to typed literal values



Creating a web of data



In a machine-readable form

```
@prefix dbr: <http://dbpedia.org/resource/>
@prefix dbo: <http://dbpedia.org/ontology/>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>


dbr:Suriname dbo:language dbr:Carribean_hindustani.
dbr:Suriname dbo:language  dbr:Javanese_language.
dbr:Suriname dbo:capital dbr:Parmaribo.

dbr:Parmaribo dbp:establishedDate 1603.

dbr:Puteri_Gunung_Ledang_(film) dbo:dirctor dbr:Saw_Teons_Hin.
dbr:Puteri_Gunung_Ledang_(film) dbo:language  dbr:Javanese_language.
```

Compared to relational databases

No constraining database schema

- DB: putting data in predefined boxes (tables, rows, columns)
- SW: linking data

Open world

- DB: what is not in the database is *true*, what is absent is *false*, (closed world)
- SW: what is described is *true*, what is absent is *unknown*
  - but we may have negative descriptions

Global vocabulary (identifiers)

- the resource and property names (URIs) are globally visible

Interconnection: semantic heterogeneity problem

Problem: the same term may have different meaning in different databases

Ontology solution: shared concept descriptions

1. use common (standard) concept description languages (RDFS, OWL, ...)
2. make the descriptions available

   schema.org, Linked open vocabularies, BioPortal, ...

Reasoning

- Make logical inferences
  - find the logical consequences of facts and rules
  - test the consistency of a set of logical formulae

State of the SW

- the Semantic Web does not exist
  - not as imagined by TBL et al.
  - big companies have created their SW (e.g. Apple Siri, Google services, ...)

- Many SW technologies are operational
  - Resource description framework, Ontology languages
  - Querying and reasoning software
  - Semi-structured databases (RDF triple stores, graph databases, ...)
  - Knowledge graphs

Content of the course

- semi-structured data models and processing
- resource description and interconnection with RDF graphs
- querying and logical inference on graphs
- ontologies and logical reasoning for description logic and logic programming
- representing time and space
- interoperability
- knowledge graphs

Projects

1. RDF(S), linked data, sparql
   - data mapping (CSV, Relational, ...) to RDF
   - linking to dbpedia
   - query demonstrator (generate documents)
2. OWL (and SWRL)
   - OWLizing a db schema, with consistency checking
   - populating and showing instance level inferences (OWL-RL + SWRL)
3. Knowledge graph
   - Representing highly contextual data (historical facts in books)
   - inference or induction