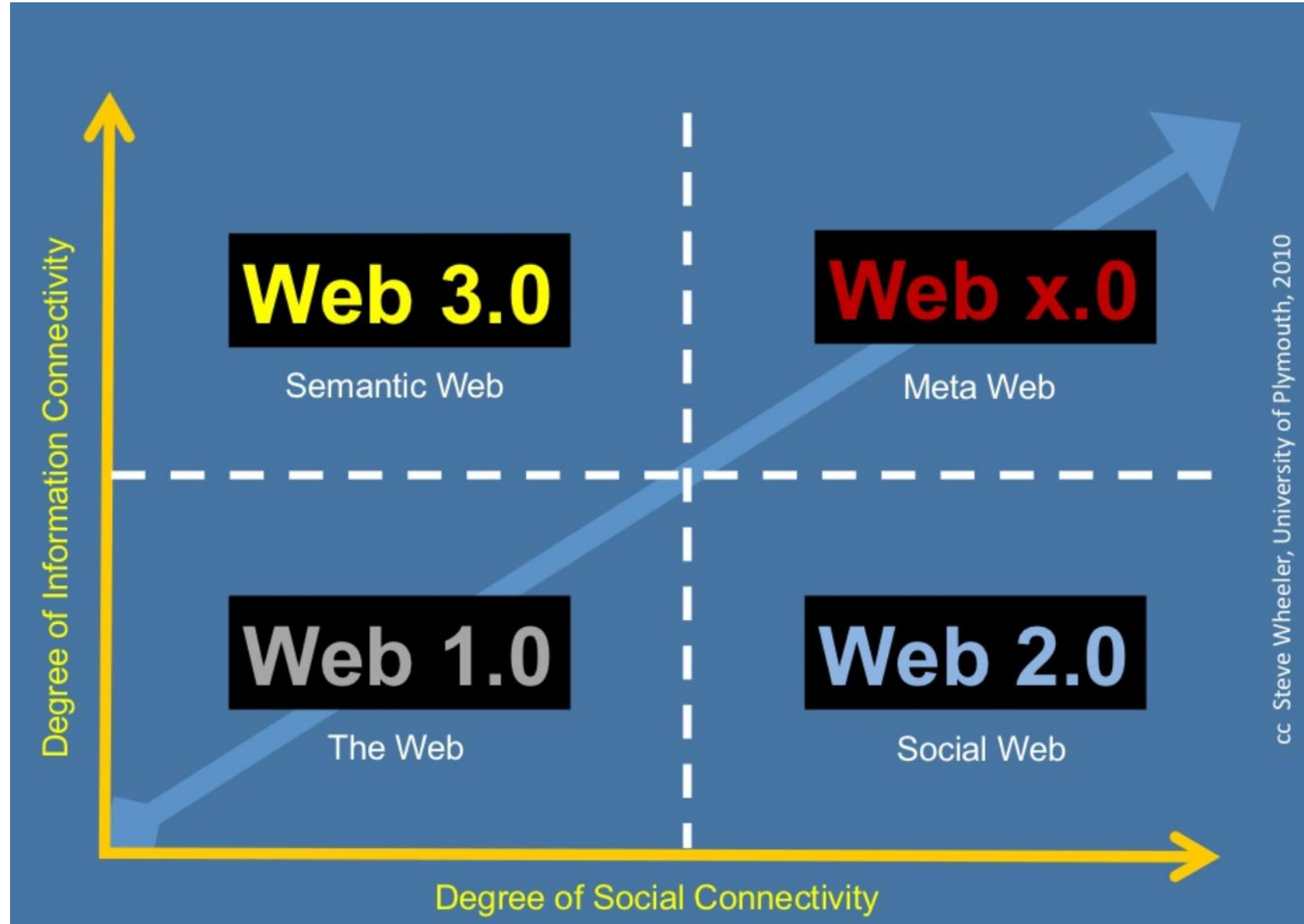


# Sémantický web

Ing. Michal Radecký, Ph.D.

# Web x.0



# Sémantický web

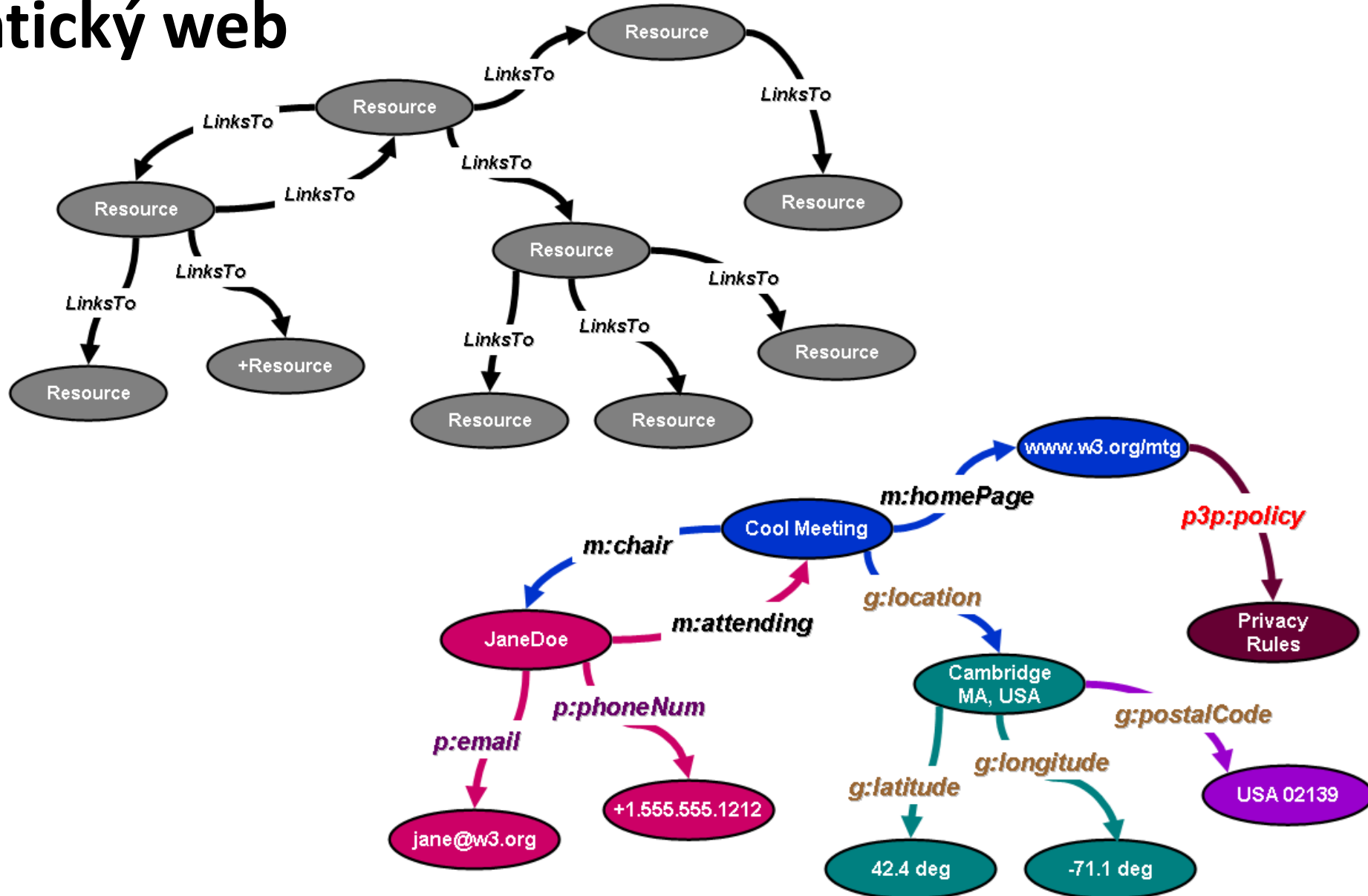
„I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A ‘Semantic Web’, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The ‘intelligent agents’ people have touted for ages will finally materialize.“

*Tim Berners-Lee, 2001*

# Proč sémantický web ...

- Kombinace a použití dat je při dnešním stavu webu poměrně snadné pro člověka
  - odvozování faktů z částečných informací
  - vytváření asociací mezi informacemi
  - využívání sensorických vjemů
- Pro počítač (software) se jedná pouze a shluk dat ve formě webových stránek, a to bez hlubší možnosti práce s těmito daty
  - částečné informace jsou nepoužitelné
  - problematické zachycení smyslu z informací
  - obtížná kombinace informací z různých zdrojů

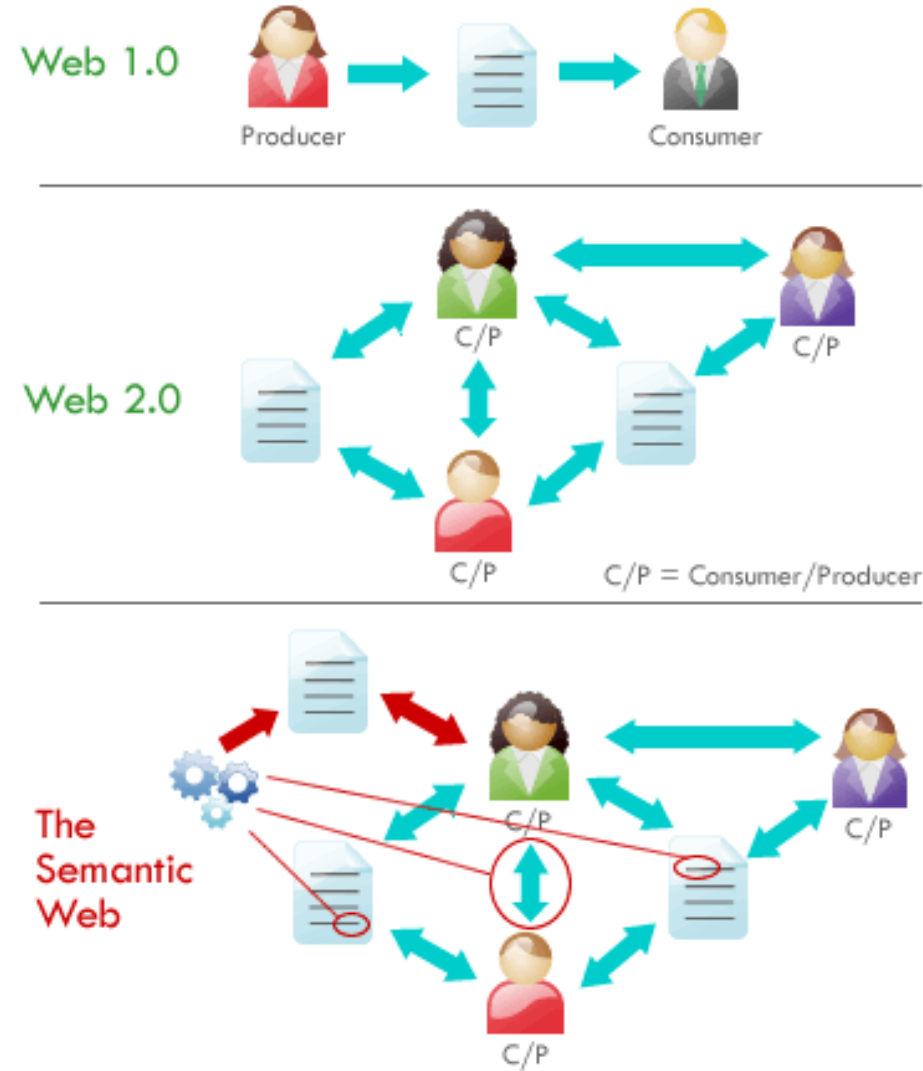
# Sémantický web



# Co je to sémantický web ...

- Sémantika je nauka o **významu jednotlivých slov**, morfémů a jiných znaků, případně též jejich vztahu ke skutečnosti, kterou označují. Slovo vzniklo z řeckého „séma“ znamená význam (např. semafor = nosič významu).
- Web 3.0 = sémantika + Web 2.0
- Web 3.0 = sémantický web
- Web 3.0 = „linked data“

# Sémantický web



# Pojmy sémantického webu

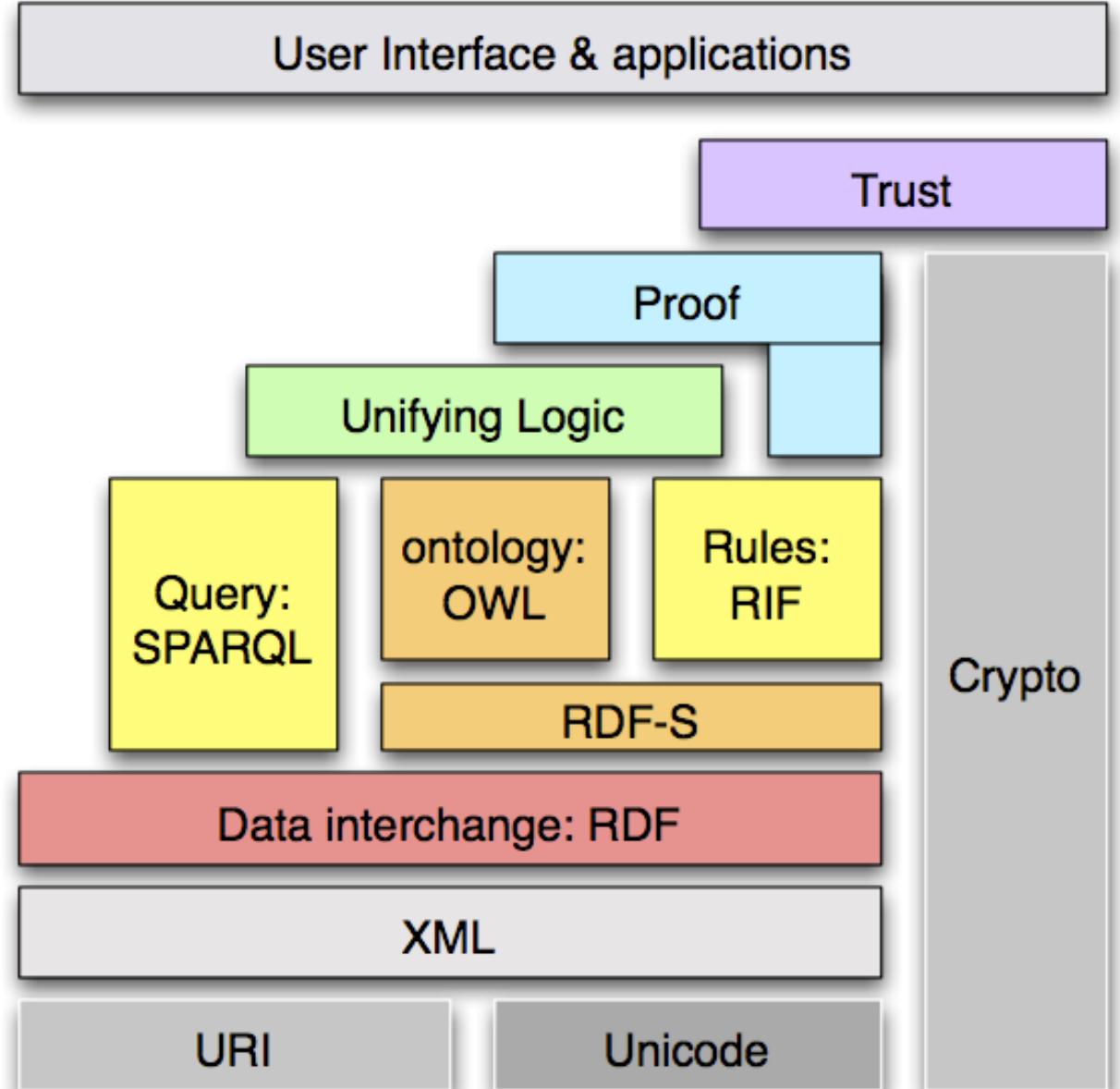
- Metadata
  - strukturovaná data o datech, kdy z jednoho pohledu se může jednat o data z jiného o metadata
- Ontologie
  - Ontologie je výslovný (explicitní) popis určité problematiky. Je to formální a deklarativní reprezentace, která obsahuje glosář (definici pojmů) a tezaurus (definici vztahů mezi jednotlivými pojmy). Ontologie je slovníkem, který slouží k uchovávání a předávání znalosti týkající se určité problematiky.
  - Datový model ontologie obecně obsahuje čtyři typy prvků: jedince, třídy, atributy a vazby
  - Nejvyšší forma metadat



# Sémantický web

- Rozšíření stávajícího webu (webových stránek) o definici významu (pomocí metadat)
- Platforma pro snadné hledání, sdílení a integraci informací a dat v internetu s využitím sémantiky
- Web zpracovatelný = srozumitelný (automatizovaně, strojově) počítači a umožňující komunikaci mezi nimi
- Aplikace databázových principů (datový model)
- Prostředky
  - RDF, URI, XML
  - ontologie
  - softwaroví agenti

# Sémantický web

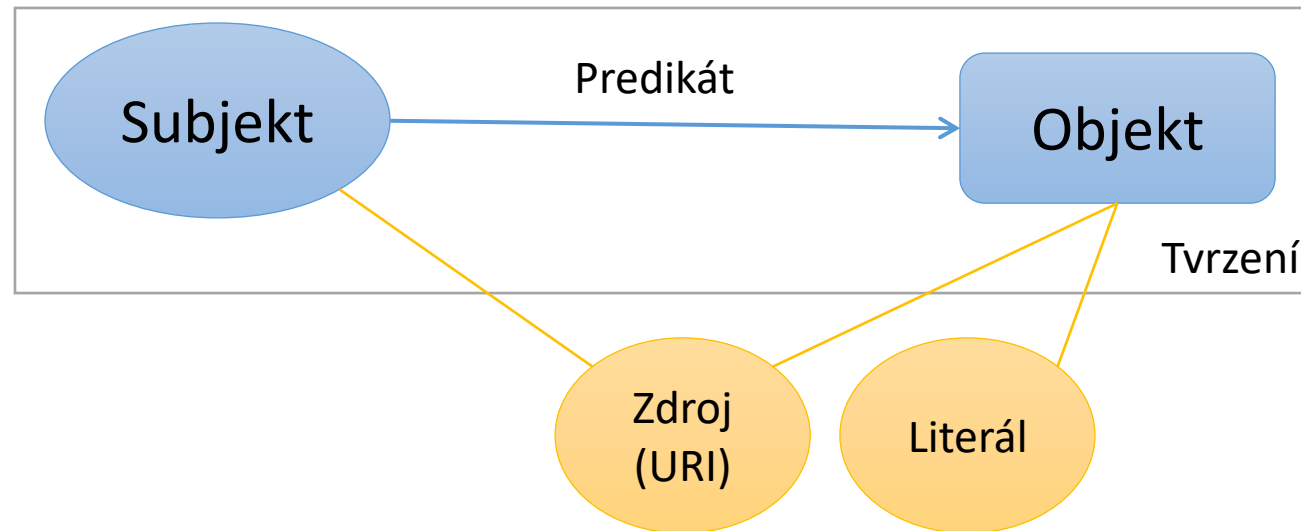


# URI (Uniform Resource Identifier)

- Standard pro **jednoznačnou** identifikaci a lokalizaci zdrojů (nejen v prostředí WWW)
- V sémantickém webu je základem pro identifikaci zdrojů v rámci webu i mimo něj (objekty reálného světa, např. RČ, login)
- URI neurčuje cestu ke zdroji, ale určuje přímo daný zdroj
- Zohledňuje jmenné prostory
  - Dublin Core (metadata dokumentů, dc:creator, dc:title, dc:subject, atd., název podle města, kde se uskutečnila první konference – Dublin, Ohio USA)

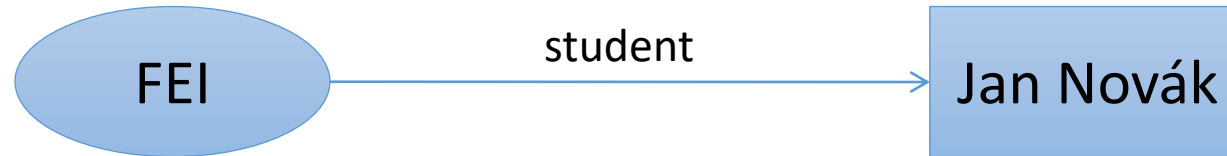
# RDF (Resource Description Framework)

- Doporučení W3C z roku 1999
- Slouží k reprezentaci struktury webových metadat pomocí tvrzení
- RDF Graf propojující tvrzení pomocí zdrojů (obecný orientovaný multigraf)
- Základ struktury tvoří: **subjekt**, **predikát** a **objekt**



# RDF (Resource Description Framework)

„Jan Novák je studentem FEI“



(FEI, student, Jan Novák)

student(FEI, Jan Novák)

*N-Triple:* <FEI> <student> "Jan Novák"

*RDF/XML:*

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://www.description.org/schema/">
```

```
  <rdf:Description about="FEI">
```

```
    <s:Student>Jan Novák</s:Student>
```

```
  </rdf:Description>
```

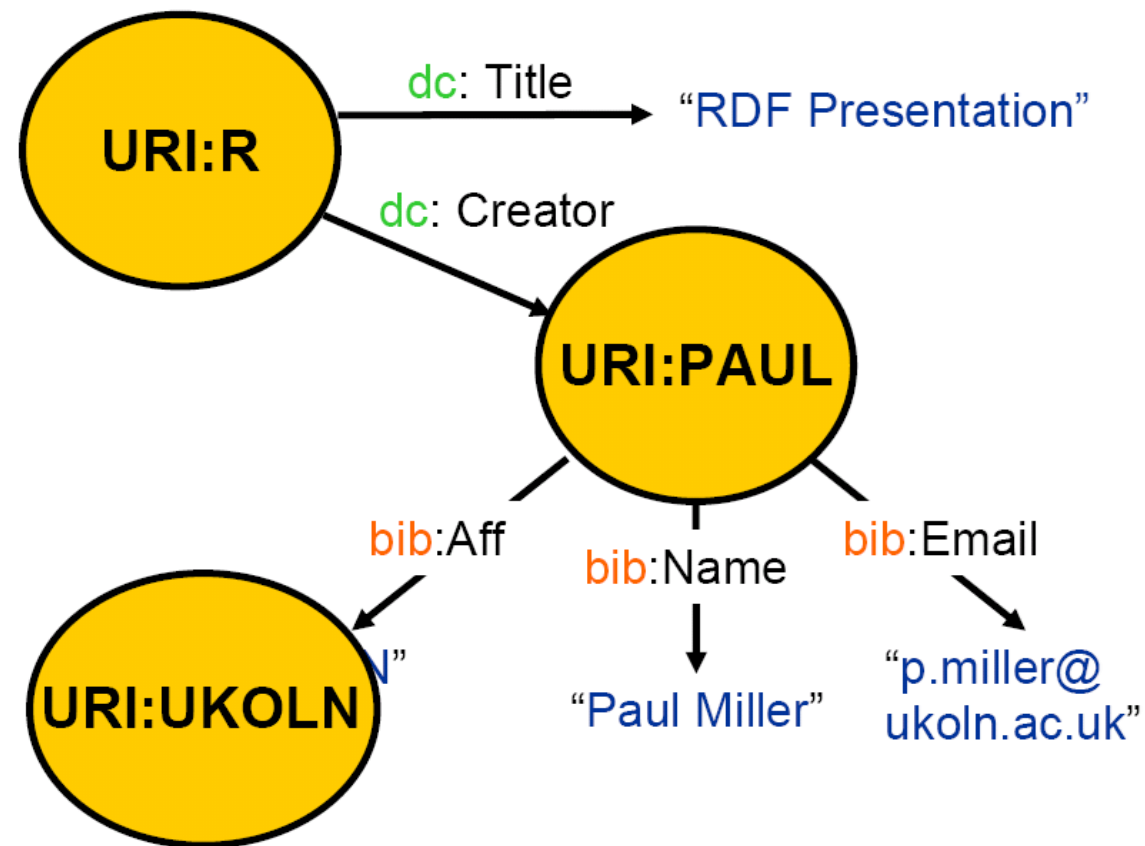
```
</rdf:RDF>
```

# RDF Graf/Model

```

<RDF xmlns = "http://www.w3.org/TR/WD-rdf-syntax#"
  xmlns:dc = "http://purl.org/dc/elements/1.0/"
  xmlns:bib = "http://www.bib.org/persons#">
  <Description about = "URI:R">
    <dc>Title> RDF Presentation </dc>Title>
    <dc:Creator>
      <Description>
        <bib>Name> Paul Miller </bib>Name>
        <bib>Email> p.miller@ukoln.ac.uk </bib>Email>
        <bib:Aff resource = "http://www.ukoln.ac.uk" />
      </Description>
    </dc:Creator>
  </Description>
</RDF>

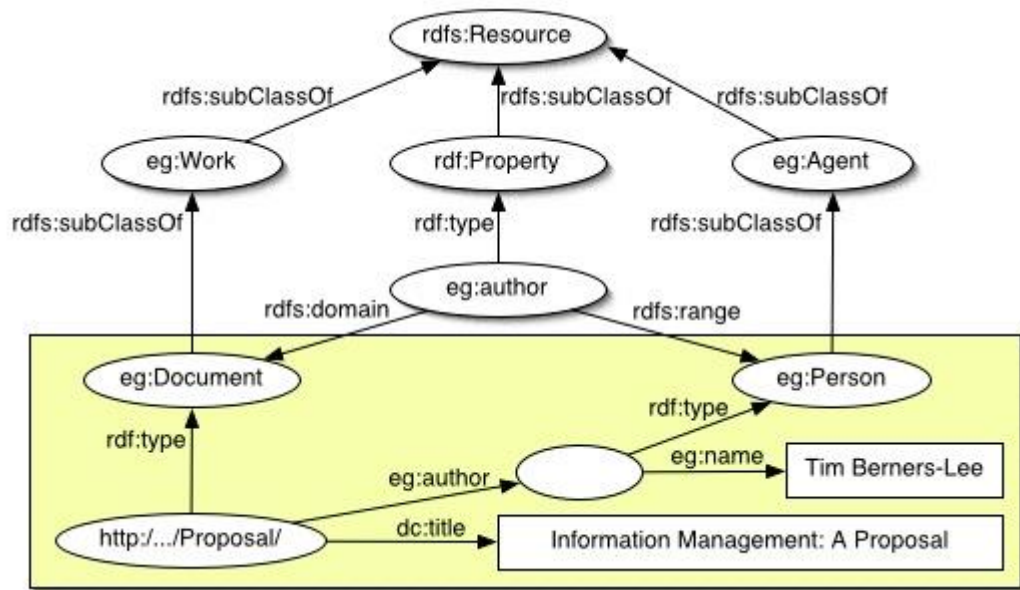
```



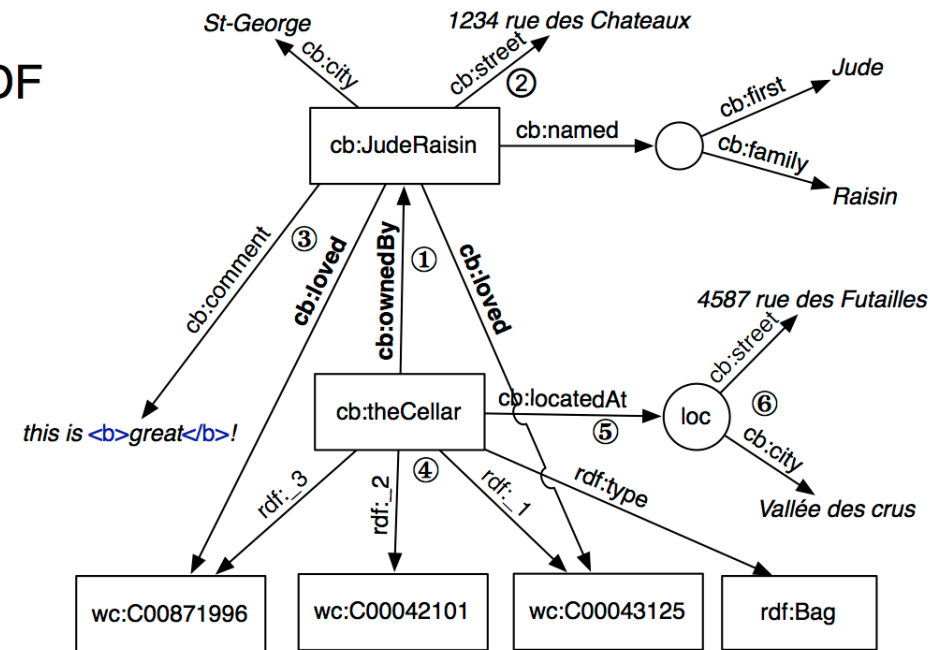
# RDFS (RDF Schema)

- Typový modelovací jazyk pro popis tříd zdrojů, jejich vlastností a vazeb
  - definiční obor a obor hodnot vlastností
  - vztah třídy a podtřídy, vlastnosti a podvlastnosti
- Poskytuje framework pro jednoduché možnosti odvozování
- Založený na objektovém přístupu (třídy, ISA hierarchie, vícenásobné dědění) a množinové sémantice
- Nedisponuje nástroji pro preciznější specifikace podmínek, datové typy, apod.

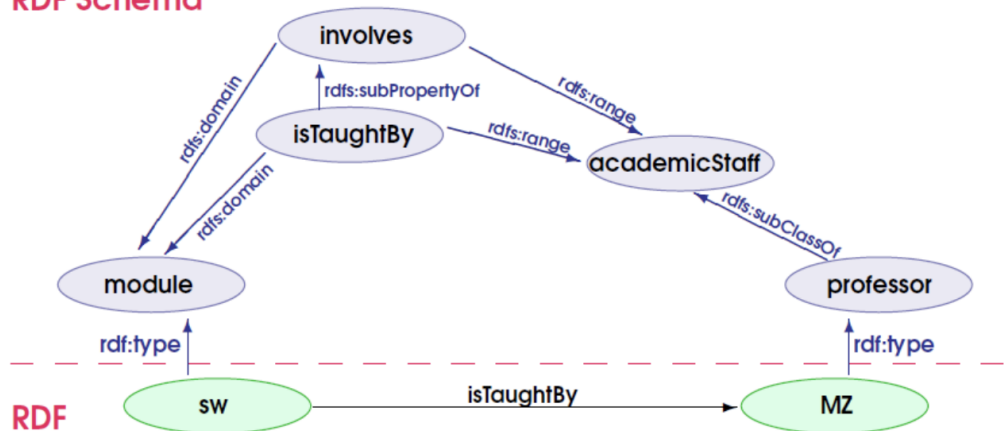
# RDFS (RDF Schema)



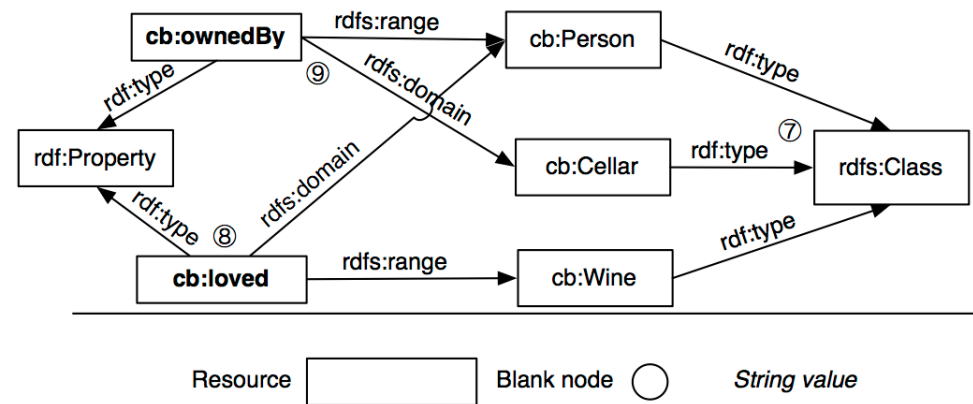
## RDF



## RDF Schema



## RDF-S



Resource  Blank node  String value



# DAML+OIL

- DAML (DARPA Agent Mark-up Language)
  - sémantický jazyk pro RDF s vyšší vyjadřovací silou než RDFS
- OIL (Ontology Inference Layer)
  - jazyk založený na deskripční logice pro popis ontologií

intersectionOf unionOf complementOf	Booleovský výraz nad třídami	subClassOf sameClassAs disjointWith	vztah obecnější a speciálnější třídy ekvivalence tříd prázdný průnik tříd
oneOf	Třída jako množina primitivních hodnot	subPropertyOf	vztah obecnější a speciálnější vlastnosti
toClass hasClass	Třída splňující univerzální resp. existenční omezení na slot (tj. na třídy, které jsou jeho hodnotami)	samePropertyAs inverseOf	(slotu) ekvivalence vlastností
hasValue	Třída splňující omezení na konkrétní hodnotu slotu (kombinace hasClass a oneOf)	transitiveProperty	vztah vzájemně inverzních vlastností
minCardinalityQ maxCardinalityQ CardinalityQ	Třída splňující omezení na kardinalitu slotu	uniqueProperty unambiguousProperty	tranzitivita vlastnosti vlastnost je funkcí inverzní vlastnost k dané vlastnosti je funkcí
		sameIndividualAs differentIndividualFrom	totožnost individuí odlišnost individuí

# DAML+OIL

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE rdf:RDF [
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <!ENTITY daml "http://www.daml.org/2001/03/daml+oil#">
  <!ENTITY old "http://www.daml.org/xyz/20000911.daml#">
  <!ENTITY xsd "http://www.w3.org/2000/10/XMLSchema#">
]>
<rdf:RDF xmlns:rdfs="&rdfs;"
         xmlns:rdf="&rdf;"
         xmlns:daml="&daml;">

  <daml:Ontology rdf:about="">
    <daml:versionInfo>xyz</daml:versionInfo>
    <rdfs:comment>Ontology to describe organizations and individuals
participating in a government R&D program.</rdfs:comment>
  </daml:Ontology>

  <rdfs:Class rdf:ID="Agency">
    <rdfs:subClassOf rdf:resource="#Organization"/>
    <daml:sameClassAs rdf:resource="&old;Agency"/>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#name"/>
        <daml:toClass
rdf:resource="http://www.w3.org/2000/10/XMLSchema#string"/>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#partOf"/>
        <daml:toClass rdf:resource="#Agency"/>
      </daml:Restriction>
    </rdfs:subClassOf>
  </rdfs:Class>
```

# OWL (Ontology Web Language)

- Nástupce DAML+OIL, v současnosti základ pro tvorbu ontologií sémantického webu
- Další možnosti specifikace tříd, vlastností, individuí, axiomů, odvozování, atd., využívá hierarchii
- OWL Full
  - absolutně „svobodná“ verze OWL na úrovni samotného RDF bez výpočetních záruk
- OWL DL
  - maximální výraznost při co nejmenší výpočetní náročnosti a bezespornost, dodržování pravidel
- OWL Lite
  - hierarchie tříd a jednoduchá omezení, nižší formální složitost

# OWL (Ontology Web Language)

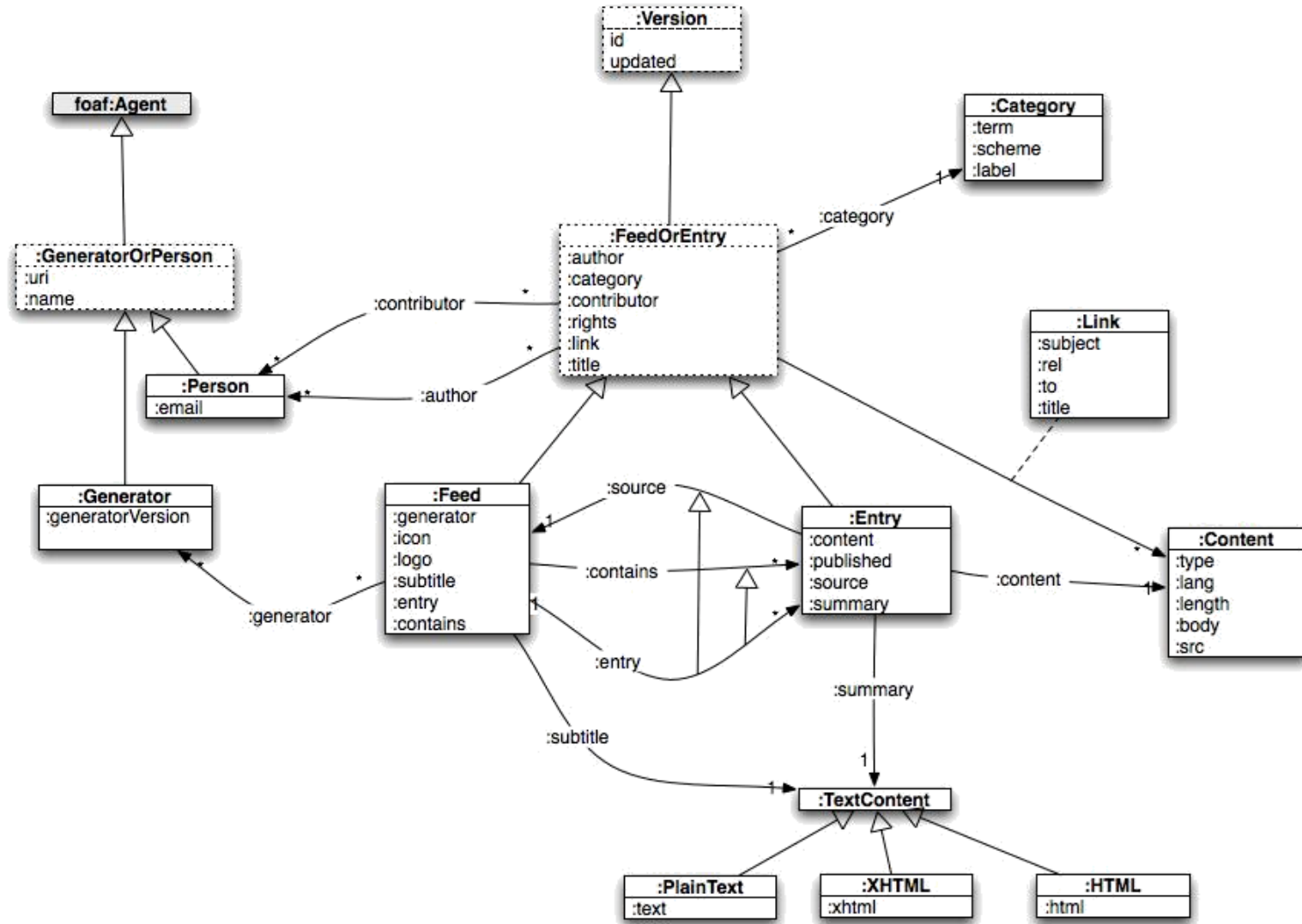
- Forma zápisu
  - abstraktní zápis
  - RDFS
  - grafický zápis

## Enumerations

```
EnumeratedClass (weekdays Monday  
Tuesday  
Wednesday  
Thursday  
Friday  
Saturday  
Sunday)
```

```
<!DOCTYPE owl [  
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >  
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >  
>  
<rdf:RDF  
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"  
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"  
  xmlns:owl="http://www.w3.org/2002/07/owl#"  
>  
<owl:Ontology rdf:about="xml:base"/>  
  
<owl:Class rdf:ID="weekdays">  
  <owl:oneOf rdf:parseType="Collection">  
    <owl:Thing rdf:about="#Monday"/>  
    <owl:Thing rdf:about="#Tuesday"/>  
    <owl:Thing rdf:about="#Wednesday"/>  
    <owl:Thing rdf:about="#Thursday"/>  
    <owl:Thing rdf:about="#Friday"/>  
    <owl:Thing rdf:about="#Saturday"/>  
    <owl:Thing rdf:about="#Sunday"/>  
  </owl:oneOf>  
</owl:Class>  
  
</rdf:RDF>
```

# OWL – ontologie Atom formátu (UML notace)



# Ontologie

- Typy ontologií
  - terminologické – pokročilejší tezaury
  - informační – rozvinutí databázových konceptuálních schémat
  - znalostní – znalosti v rámci umělé inteligence
- Ontologické slovníky
  - terminologické ontologie pro klasifikaci obsahu
  - Dublin Core, lékařské ontologie, knihovnické ontologie, atd.

# SPARQL

- Simple Protocol and RDF Query Language
- Dotazovací jazyk nad RDF zdroji
- Vrací graf (část RDF informací)

```
PREFIX abc: <http://mynamespace.com/exampleOntologie#>
SELECT ?capital ?country
WHERE {
  ?x abc:cityname ?capital.
  ?y abc:countryname ?country.
  ?x abc:isCapitalOf ?y.
  ?y abc:isInContinent abc:africa.
}
```

# Linked Data

- Rozrůstající se řada významových slovníků v podobě ontologií resp. RDF Schémat
- **Dublin Core** - Ontologie pro CMS a publikaci dat, <http://dublincore.org>
- **FOAF** - Friend of a Friend
  - Projekt podporující tvorbu strojově zpracovatelných stránek s osobními profily a dalšími osobnostními vazbami (Dublin Core, apod.), sémantický systém pro propojování sociálních sítí
  - <http://www.foaf-project.org>
- **DBPedia** – sémantické aktivity kolem wikipédie, <http://www.dbpedia.org>
- **Wikidata** – databáze dokumentů, <http://www.wikidata.org>
- **Geonames** – databáze geografických informací, <http://www.geonames.org>
- **SKOS** - popis znalostní a organizační struktury
- **SCIOS** - webové výsledky komunit (blogy, diskusní fóra, atd.)
- **Triplydb** – databáze různých schémat a slovníků, <https://triplydb.com>





# Aplikace sémantické webu

- Znalostní databáze – zachycení podstaty znalostí, umožnění vyhledávání a zajištění konzistence
- Pojmové vyhledávání informací – vyhledávání založené na využívání všech přístupů sémantického webu (odvozování, automatizace doplňování pojmů, softwaroví agenti, atd.)
- Zpracování přirozeného jazyka
- Elektronické obchodování
- Sdílení a integrace informací – RSS, Atom

# Další nástroje sémantického webu

- RDFa (RDF in Attributes)
  - Mechanismus pro vkládání RDF do HTML resp. XHTML
  - 100% závislý na slovnících vkládaných pomocí namespace
  - schválená specifikace pro XHTML 1.1
  - podporováno Google a Yahoo
  - atributy: about, rel, rev, href, src, resource, property, content, , datatype, typeof

# RDFa

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
  "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  version="XHTML+RDFa 1.0" xml:lang="en">
  <head>
    <title>John's Home Page</title>
    <base href="http://example.org/john-d/" />
    <meta property="dc:creator" content="Jonathan Doe" />
  </head>
  <body>
    <h1>John's Home Page</h1>
    <p>My name is <span property="foaf:nick">John D</span> and I like
      <a href="http://www.neubauten.org/" rel="foaf:interest"
        xml:lang="de">Einstürzende Neubauten</a>.
    </p>
    <p>
      My <span rel="foaf:interest" resource="urn:ISBN:0752820907">favorite
      book</span> is the inspiring <span about="urn:ISBN:0752820907"><cite
      property="dc:title">Weaving the Web</cite> by
      <span property="dc:creator">Tim Berners-Lee</span></span>
    </p>
  </body>
</html>
```

# Další nástroje sémantického webu

- Microformats
  - nepřímé využití technologií sémantického webu
  - hCard, hCalendar

```
<p class="vcard">  
  During my visit to <span class="locality">Atlanta</span>,  
  <abbr title="Georgia" class="region">GA</abbr>,  
  I plan to check out <span class="fn org">Centennial Olympic Park</span>.  
</p>
```

# Další nástroje sémantického webu

- Strukturovaná data

<https://developers.google.com/search/docs/guides/intro-structured-data>

- Microdata – HTML5

```
1. <section itemscope itemtype="http://data-vocabulary.org/Person">
2.   <h1 itemprop="name">Jan Sládek</h1>
3.   <p></p>
4.   <p><a itemprop="url" href="http://honzasladek.com">osobní
      stránky</a></p>
5. </section>
```

- Microdata – JSON-LD (JSON for Linking Data)

<https://json-ld.org/>

- Facebook OpenGraph

<https://developers.facebook.com/docs/sharing/overview>

# Nástroje

- <https://search.google.com/structured-data/testing-tool>
- <https://developers.facebook.com/tools/debug/>
- [www.schema.org](http://www.schema.org)

# Problémy sémantického webu

- Složitost deskriptivních jazyků a technologií specifikace obsahu
- RDF vs. XML vs. JSON, ...
- Jak rozsáhlé a automatizované mají být možnosti odvozování nad ontologiemi
- Jak vytvářet ontologie, do jaké hloubky, zajištění důvěry, apod.
- Rozsáhlé, heterogenní a nekonzistentní zdroje a současné technologie
  
- Jak technicky a efektivně ukládat a spravovat data (NoSQL, RDF database)
- Jak řídit a udržovat celý ekosystém



# Další zdroje

- [https://www.ted.com/talks/tim\\_berners\\_lee\\_the\\_next\\_web/transcript#t-38928](https://www.ted.com/talks/tim_berners_lee_the_next_web/transcript#t-38928)
- <https://www.youtube.com/watch?v=vioCbTo3C-4>
- <https://www.youtube.com/watch?v=kMI6qEg12RM>