

# **Semantic Web, SW Services, Grid, Cloud**

Martin Kuba, ÚVT MU

# Semantic Web

- idea introduced by Tim Berners Lee (inventor of WWW) in 2001
- *“The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation”*
- web instead of platform for distributed presentations would be platform for distributed knowledge

# Semantic Continuum

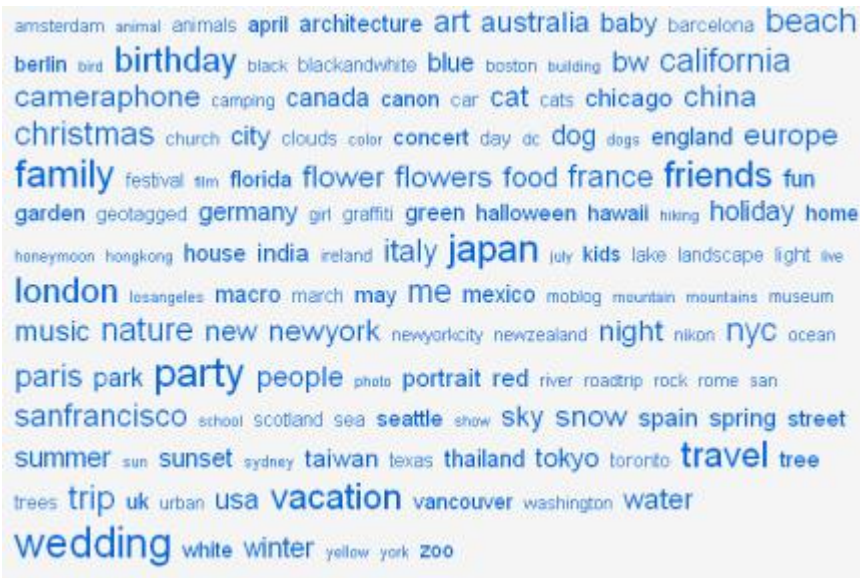
- semantics = meaning
- semantic in SW means *machine-processable*
- semantic continuum (Uschold 2003)
  - a. implicit semantics in the minds of humans
  - b. explicit informal semantics (text description in natural language, e.g. HTML specification)
  - c. formal semantics for humans (in formal language processed by humans)
  - d. formal semantics for machine processing
- goal is to create robotic decision-making devices
- metadata - data about data

# Expressing Semantics

- folksonomies
- microdata
- RDF triples and RDF Schema vocabularies
- OWL-DL ontologies for automated reasoning

# Folksonomies

- keyword metadata as **tags**
- e.g. an image of a dog may be tagged with tags *dog*, *collie* or *pet*
- (+) low entry barrier, no user training
- (-) no synonym control, flat structure
- tag clouds



amsterdam animal animals april architecture art australia baby barcelona beach  
berlin bird birthday black blackandwhite blue boston building bw california  
cameraphone camping canada canon car cat cats chicago china  
christmas church city clouds color concert day dc dog dogs england europe  
family festival film florida flower flowers food france friends fun  
garden geotagged germany girl graffiti green halloween hawaii hiking holiday home  
honeymoon hongkong house india ireland italy japan july kids lake landscape light live  
london losangeles macro march may me mexico moblog mountain mountains museum  
music nature new newyork newyorkcity newzealand night nikon nyc ocean  
paris park party people photo portrait red river roadtrip rock rome san  
sanfrancisco school scotland sea seattle show sky snow spain spring street  
summer sun sunset sydney taiwan texas thailand tokyo toronto travel tree  
trees trip uk urban usa vacation vancouver washington water  
wedding white winter yellow york zoo

# Microdata

- competing Microdata, Microformats, RDFa
- nesting semantics within existing content on web pages
- RDFa only inside XML, not in HTML5
- Microdata provides JavaScript API
- Microdata use namespace-qualified vocabularies predefined at data-vocabulary.org or schema.org
- supported by Google search engine
- opposite vision than in 2000:
  - XML with CSS or XSLT - semantic markup with presentational metadata
  - HTML5 with Microdata - presentational markup with semantic metadata

# Comparison of Microdata and others

Microformat:

```
<div class="vcard">
<span class="fn">Bob Smith</span>
<span class="title">engineer</span> at <span class="org">ACME Corp</span>.
<span class="adr">
<span class="locality">Albuquerque</span>,
<span class="region">NM</span>
</span>
</div>
```

Microdata:

```
<div itemscope itemType="http://data-vocabulary.org/Person">
<span itemprop="name">Bob Smith</span>
<span itemprop="title">engineer</span> at <span itemprop="affiliation">ACME Corp</span>.
<span itemprop="address" itemscope itemType="http://data-vocabulary.org/Address">
<span itemprop="locality">Albuquerque</span>, <span itemprop="region">NM</span>
</span>
</div>
```

RDFa:

```
<div xmlns:v="http://rdf.data-vocabulary.org/#" typeof="v:Person">
<span property="v:name">Bob Smith</span>
<span property="v:title">engineer</span> at <span property="v:affiliation">ACME Corp</span>.
<span rel="v:address">
<span typeof="v:Address">
<span property="v:locality">Albuquerque</span>, <span property="v:region">NM</span>
</span>
</span>
</div>
```

# RDF - Resource Description Framework

- statements about web resources
- triples *subject-predicate-object*
- subject and predicate are URIs
- object can be a URI or a data value
- **reification** - an RDF statement is assigned a URI and treated as a resource
- producers and consumers of RDF statements must agree on the semantics of the resource identifiers, conveyed by some controlled vocabulary



# RDF Schema

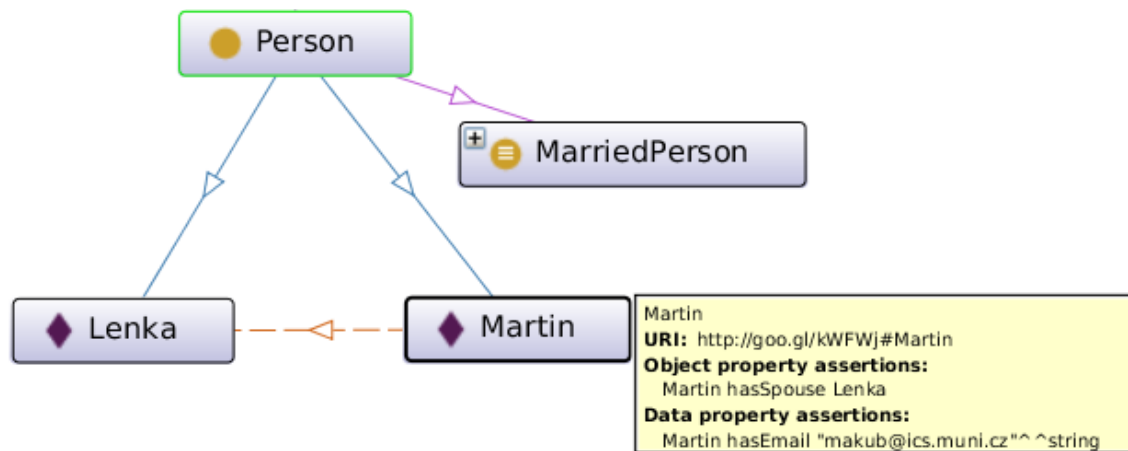
- tool for defining controlled vocabularies
- defines
  - classes of things
  - properties (binary predicates)
  - subsumption relationships (subclasses, subproperties)
  - `rdf:type` - resource is an instance of a class
- **SPARQL** (SPARQL Protocol and RDF Query Language) is an SQL-like language for querying RDF graphs
- *entailment rules* allow to entail e.g. that when a resource is in a particular class, then it is also in all its superclasses

# OWL

- Web Ontology Language defined by W3C
- **ontology** is a term from artificial intelligence
- ontology is “*an explicit (written) formal conceptualization*”, used for capturing knowledge about some domain of interest
- OWL 1 released in 2004, OWL 2 in 2009
- two different (incompatible) semantics
  - RDF based - OWL Full
  - DL (Description Logics) based - OWL DL

# OWL DL

- Description Logics is a family of logics
- decidable fragment of First Order Predicate Logic (FOL) plus decidable extensions
- **reasoners** - software able to entail complete inferrable knowledge in finite time
- OWL DL uses:
  - classes
  - individuals
  - properties (binary relations)
    - object properties (between two objects)
    - data properties (between object and data literal)
- can use SWRL (Sem. Web. Rule Language)



Prefix( :=<<http://goo.gl/kFWj#>> )

Prefix( xsd:=<<http://www.w3.org/2001/XMLSchema#>> )

Ontology( <<http://goo.gl/kFWj#>> )

Declaration( Class( :Person ) )

Declaration( Class( :MarriedPerson ) )

Declaration( NamedIndividual( :Martin ) )

Declaration( NamedIndividual( :Lenka ) )

Declaration( ObjectProperty( :hasSpouse ) )

Declaration( DataProperty( :hasEmail ) )

SymmetricObjectProperty( :hasSpouse )

FunctionalObjectProperty( :hasSpouse )

ClassAssertion( :Person :Lenka )

ClassAssertion( :Person :Martin )

DifferentIndividuals( :Martin :Lenka )

ObjectPropertyAssertion( :hasSpouse :Martin :Lenka )

DataPropertyAssertion( :hasEmail :Martin "makub@ics.muni.cz"^^xsd:string )

SubClassOf( :MarriedPerson :Person )

EquivalentClasses( :MarriedPerson ObjectSomeValuesFrom( :hasSpouse :Person ) )

# OWL DL Tools

- ontology editor with GUI - Protege 4.1
  - <http://protege.stanford.edu/>
- reasoners
  - Pellet
  - HermiT
  - FACT++
  - Stardog
- Java API for OWL - OWL API 3
  - <http://owlapi.sourceforge.net/>

# Limits of OWL DL

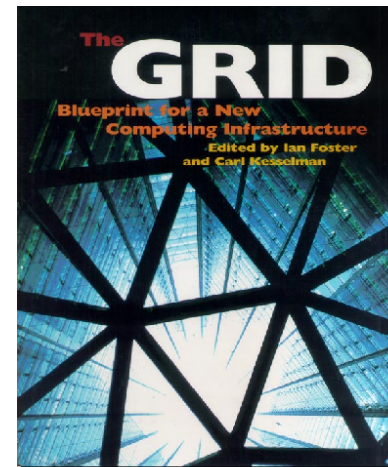
- based on FOL  $\forall x \exists y (P(x) \rightarrow Q(f(y)))$
- cannot express
  - fuzzy expressions - “It **often** rains in autumn.”
  - non-monotonicity - “Birds fly, penguin is a bird, but penguin does not fly.”
  - propositional attitudes - “Eve **thinks** that 2 is not a prime number.”
  - modal logic
    - possibility and necessity - “It is **possible** that it will rain today.”
    - epistemic modalities - “Eve **knows** that 2 is a prime number.”
    - temporal logic - “I am **always** hungry.”
    - deontic logic - “You **must** do this.”
- Transparent Intensional Logic (TIL)
  - can express anything that can be said
  - has no calculus or reasoning algorithms

# Semantic Web Services

- research efforts OWL-S, WSDL-S, WSMO
- semantics can enhance discovery
  - on the semantic continuum move it from b) to d)
  - e.g. search for "getHardDriveQuote" can find also "getQuoteForHardDrive" (synonym) and "getSCSIDriveQuote" (subsumed term)
- web service semantics
  - **Data semantics** - it defines meaning of the data, i.e. inputs and outputs of operations
  - **Functional semantics** - it defines meaning of the operations, i.e. how they transform input to output
  - **QoS semantics** - it provides meaning for quality aspects, like price, availability, level of trust etc. Service selection may be based on such characteristics.
  - **Execution semantics** - it provides details like preconditions and effects of service invocation, conversation patterns of service invocation etc

# Grid

- term introduced in 1998 by Carl Kesselman and Ian Foster in book "The Grid: Blueprint for a New Computing Infrastructure"
- analogy to electrical power grid
- *"A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities."*
- in 2001 in article "The Anatomy of the Grid" added Virtual Organizations





# What is The Grid ?

- coordinates resources that are not subject to centralized control
- using standard, open, general-purpose protocols and interfaces
- to deliver nontrivial qualities of service.

# Grid Usage

- high performance computing (HPC)
  - research of medical drugs
  - gravitation waves research
  - earthquake prediction
  - electronic chip engineering
  - ...
- large data
  - Large Hadron Collider in CERN
- expensive scientific instruments
  - large microscope in Japan
- remote cooperation
  - teleconferences, remote surgery, ...

# Grid Middleware

- not a single middleware
  - in U.S.A. Globus
  - in Europe gLite
  - in Germany UNICORE
- services
  - information services (Globus: MDS, gLite: BDII)
  - gridFTP - striped transfer, third party transfer
  - resource allocation (Globus: GRAM, gLite: WMS)
  - virtual organization membership (VOMS)
- Computing Element
  - grid gate, batch system, cluster of worker nodes
- Storage Element
  - disk servers, disk arrays, tape storage

# Grid Security

- based on X509 certificates and PKI (Public Key Infrastructure)
- list of selected grid CA (Certification Authorities) maintained by IGTF (International Grid Trust Federation) resp. EUGridPMA
- allows so-called **proxy certificates**
  - short-lived (24 hours, 1 week)
  - a certificate signed by a user certificate or proxy cert.
  - can be delegated to a running job
- VOMS (Virtual Organisation Membership Service) issues attribute certificates

# European Grid History

- in 2001-2003 project DataGrid
  - for processing massive data produced by Large Hadron Collider in CERN
- in 2004-2010 projects EGEE I, II, III
- in 2010-2014 EGI (European Grid Infrastructure) is built in project InSPIRE
- EGI in April 2013
  - 373800 CPU, 170 PB disk storage
  - 22078 users
- EGI consists of NGIs (National Grid Infrastructures)
- Czech NGI is MetaCentrum, operated by CESNET, collects 8580 CPU

# Cloud Computing

- use of computing resources (hardware and software) that are delivered as a service over a network
- in 1960 utility computing
- in 2006 Amazon released AWS (Amazon Web Service)
  - EC2 (Elastic Compute Cloud)
  - S3 (Simple Storage Service)
- virtual machines created on demand
- independence on user device

# Cloud Computing Middleware

- **middleware**
  - OpenNebula
  - Eucalyptus
  - Nimbus
- **service models**
  - Infrastructure as a service (IaaS) - only VM
  - Platform as a service (PaaS) - VM with OS
  - Software as a service (SaaS)
  - Network as a service (NaaS)