

Introduction

Petr Křemen

petr.kremen@fel.cvut.cz

October 5, 2017



Outline

- 1 About Knowledge Management
- 2 Overview of Ontologies
- 3 Overview of Data Integration
- 4 Introduction to Semantic Web
 - Semantic Web Adopters
 - Semantic Web Principles
- 5 Linked Data
- 6 Linked Data



- 1 About Knowledge Management
- 2 Overview of Ontologies
- 3 Overview of Data Integration
- 4 Introduction to Semantic Web
 - Semantic Web Adopters
 - Semantic Web Principles
- 5 Linked Data
- 6 Linked Data

About Knowledge Management



About Knowledge

Knowledge is all around. But what is the difference among different types of knowledge ? How about their *machine reusability (R)*/*interpretability (I)*/*expressive power (E)*?

Book	R—	I+++	E+++
Java program	R	I—	E—
R/Matlab Script	R	I—	E—
Relational Database	R+	I	E
Prolog Program	R++	I	E+
SKOS Vocabulary	R++	I++	E
5* Linked Data	R+++	I++	E++



Is Knowledge Management Worth ?

What is the trend of **Runway Incursion** incidents at an airline operator ?



Airline Operator



Unauthorized entering the runway

Incorrect entering (without clearance) active runway



Civil Aviation Authority



Is Knowledge Management Worth ?

DID YOU KNOW



Just months before 9/11, the World Trade Center's lease was privatized and sold to Larry Silverstein.

Silverstein took out an insurance plan that 'fortuitously' covered terrorism.

After 9/11, Silverstein took the insurance company to court, claiming he should be paid double because there were 2 attacks.

Silverstein won, and was awarded \$4,550,000,000.

source:<https://www.metabunk.org/larry-silversteins-9-11-insurance.t2375>

What is an event ? How many events occurred at 9/11 – One or Two ?

Knowledge Management

9/11 ... matter of billions of USD

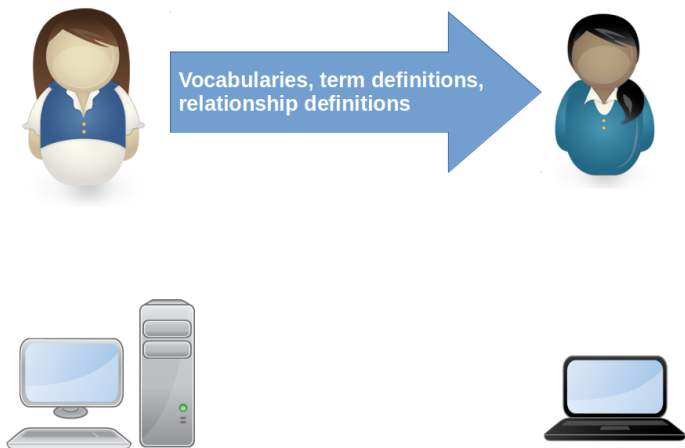


- 1 About Knowledge Management
- 2 Overview of Ontologies
- 3 Overview of Data Integration
- 4 Introduction to Semantic Web
 - Semantic Web Adopters
 - Semantic Web Principles
- 5 Linked Data
- 6 Linked Data

Overview of Ontologies



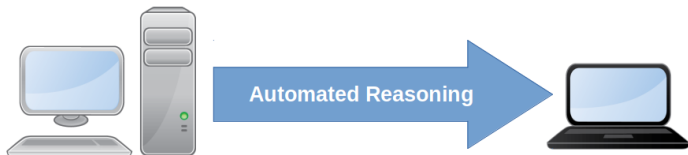
First, People Need to Understand Each Other



Second, People Need to Explain Things to Computers

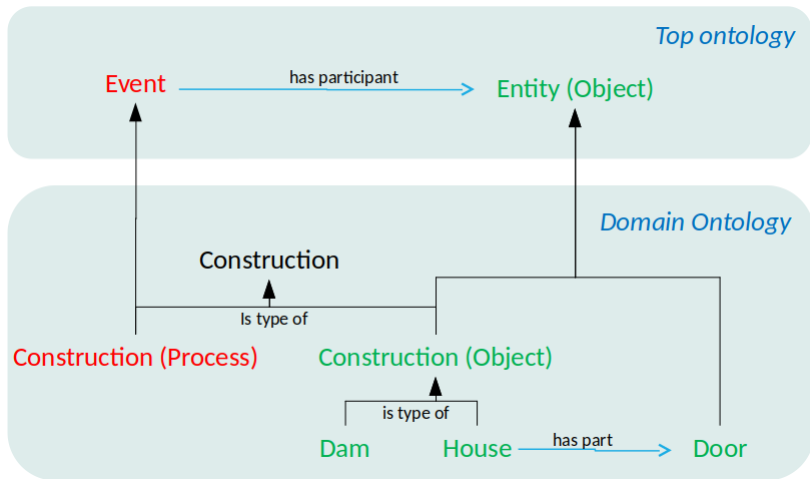


Third, Computers Can Understand One Another



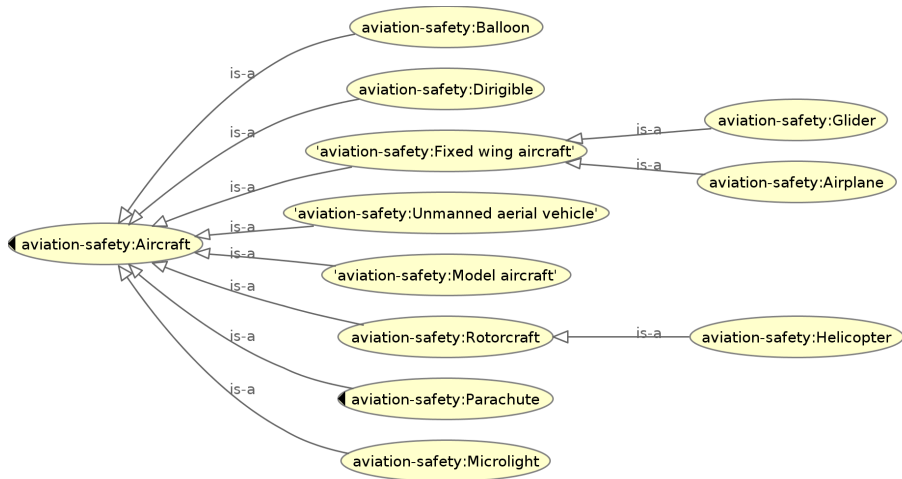
Solution = Ontology

Explicit Conceptualization of Shared Meaning



Example Ontology Hierarchy

Each helicopter is also an aircraft.



Ontologies \neq Taxonomies

Taxonomies = just a single type of relationship.

Construction	→ broad meaning (object, construction site, process)
Dam	
House	→ broad meaning (dwelling, construction)
Door	→ specific meaning (not type of house, but its part)

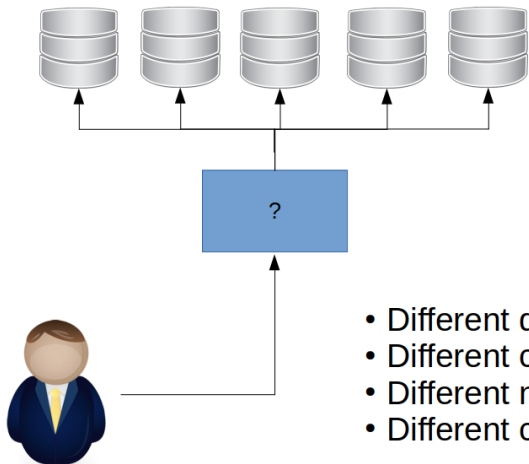


- 1 About Knowledge Management
- 2 Overview of Ontologies
- 3 Overview of Data Integration**
- 4 Introduction to Semantic Web
 - Semantic Web Adopters
 - Semantic Web Principles
- 5 Linked Data
- 6 Linked Data

Overview of Data Integration



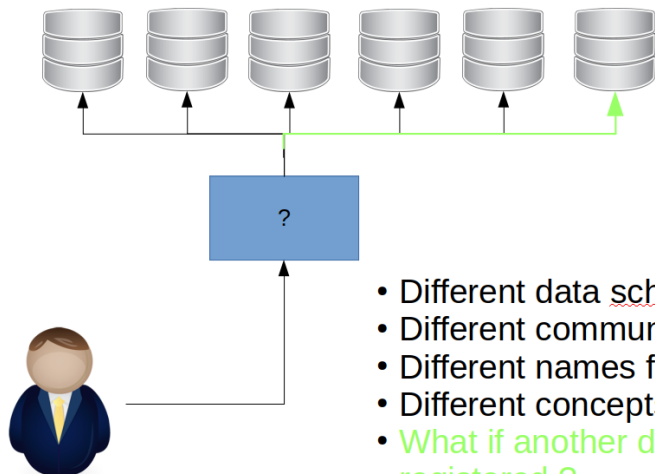
Data Integration Scenario



- Different data schemas
- Different communication speeds
- Different names for a concept
- Different concepts for one term



Data Integration Scenario



- Different data schemas
- Different communication speeds
- Different names for a concept
- Different concepts for one term
- What if another data source gets registered ?



Ontologies for Data Integration

Ontologies help to share data meaning.

Modeling and Inference for different data schemas, different data quality

OWL sameAs for different naming of the same thing

IRI identification for different namings of the same thing

Open World Assumption to react on new data source emergence



- 1 About Knowledge Management
- 2 Overview of Ontologies
- 3 Overview of Data Integration
- 4 Introduction to Semantic Web**
 - Semantic Web Adopters
 - Semantic Web Principles
- 5 Linked Data
- 6 Linked Data

Introduction to Semantic Web



Current Web vs. Semantic Web

- SoA – semistructured HTML or XML data. There is vast amount of search engines like Google, Yahoo, MSN, etc. Many of them are invaluable, but as the engines use just keywords and/or some natural language preprocessing methods, the search results contain lots of irrelevant results that need to be processed manually.



Current Web vs. Semantic Web

- SoA – semistructured HTML or XML data. There is vast amount of search engines like Google, Yahoo, MSN, etc. Many of them are invaluable, but as the engines use just keywords and/or some natural language preprocessing methods, the search results contain lots of irrelevant results that need to be processed manually.
- How to make web search more efficient ?



Current Web vs. Semantic Web

- SoA – semistructured HTML or XML data. There is vast amount of search engines like Google, Yahoo, MSN, etc. Many of them are invaluable, but as the engines use just keywords and/or some natural language preprocessing methods, the search results contain lots of irrelevant results that need to be processed manually.
- How to make web search more efficient ?
 - more expressive power for web designers to capture complexities – SW languages (RDF(S), OWL),



Current Web vs. Semantic Web

- SoA – semistructured HTML or XML data. There is vast amount of search engines like Google, Yahoo, MSN, etc. Many of them are invaluable, but as the engines use just keywords and/or some natural language preprocessing methods, the search results contain lots of irrelevant results that need to be processed manually.
- How to make web search more efficient ?
 - more expressive power for web designers to capture complexities – SW languages (RDF(S), OWL),
 - more efficient search engines to handle SW languages – new inference techniques for these languages,



Current Web vs. Semantic Web

- SoA – semistructured HTML or XML data. There is vast amount of search engines like Google, Yahoo, MSN, etc. Many of them are invaluable, but as the engines use just keywords and/or some natural language preprocessing methods, the search results contain lots of irrelevant results that need to be processed manually.
- How to make web search more efficient ?
 - more expressive power for web designers to capture complexities – SW languages (RDF(S), OWL),
 - more efficient search engines to handle SW languages – new inference techniques for these languages,
 - better search engines interfaces – more expressive query languages

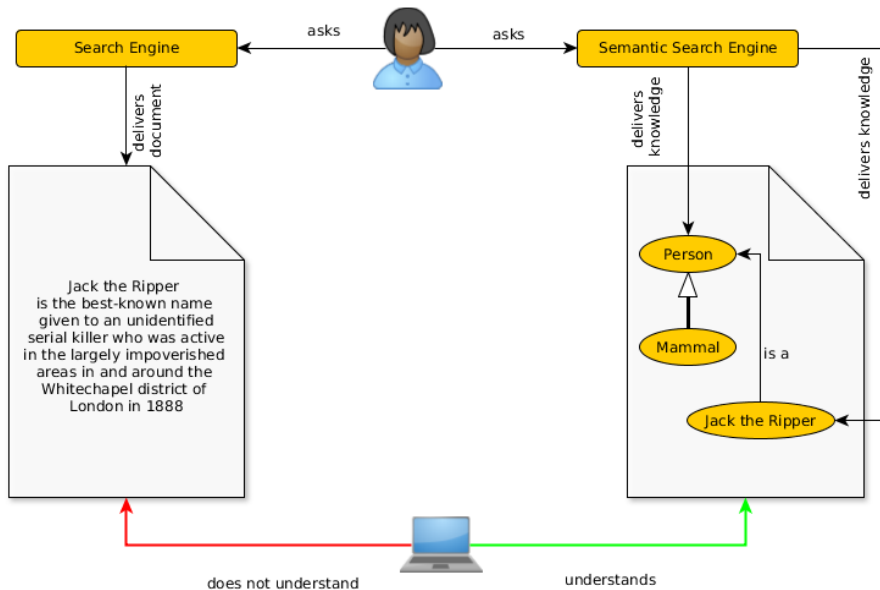


Current Web vs. Semantic Web

- SoA – semistructured HTML or XML data. There is vast amount of search engines like Google, Yahoo, MSN, etc. Many of them are invaluable, but as the engines use just keywords and/or some natural language preprocessing methods, the search results contain lots of irrelevant results that need to be processed manually.
- How to make web search more efficient ?
 - more expressive power for web designers to capture complexities – SW languages (RDF(S), OWL),
 - more efficient search engines to handle SW languages – new inference techniques for these languages,
 - better search engines interfaces – more expressive query languages
- **the amount of (unstructured) data is steadily growing**



Semantic search



Ontologies and Semantic Web

ontology has many definitions, but let's consider it **a formal representation of a complex domain knowledge that is shared with others to ensure intelligent system interoperability,**

semantic web *is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.* (cit. Semantic Web. Tim Berners-Lee, James Hendler and Ora Lassila, Scientific American, 2001)



Idea of Semantic Web

- W3C web page - <http://www.w3.org/2001/sw>



Idea of Semantic Web

- W3C web page - <http://www.w3.org/2001/sw>
- The data format will be either RDF(S) or OWL,



Idea of Semantic Web

- W3C web page - <http://www.w3.org/2001/sw>
- The data format will be either RDF(S) or OWL,
- Reasoners for RDF(S) can be used for partial derivation in OWL,



Idea of Semantic Web

- W3C web page - <http://www.w3.org/2001/sw>
- The data format will be either RDF(S) or OWL,
- Reasoners for RDF(S) can be used for partial derivation in OWL,
- Reasoners for OWL can be used for derivation in RDF(S)



Semantic Web Adopters

- 1 About Knowledge Management
- 2 Overview of Ontologies
- 3 Overview of Data Integration
- 4 Introduction to Semantic Web**
 - Semantic Web Adopters**
 - Semantic Web Principles
- 5 Linked Data
- 6 Linked Data



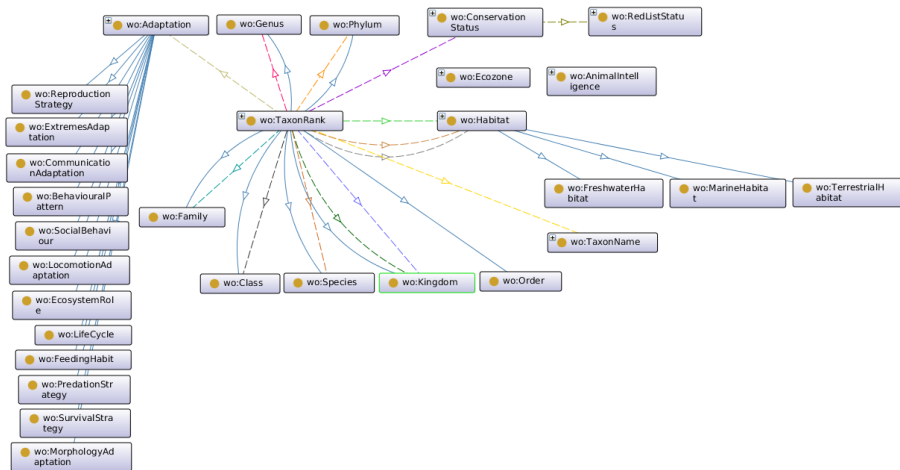
Who is Using Semantic Web Technologies

Let's name a few:

- Google – *Knowledge Graph* (although they do not name it Semantic web – <http://semanticweb.com/google-just-hi-jacked-the-semantic-web-vocabulary-b29092>)
- Microsoft – Satori, <http://research.microsoft.com/en-us/projects/trinity/query.aspx>
- Facebook – Open Graph Protocol <http://ogp.me/>
- BBC – various datasets in RDF – <http://www.bbc.co.uk/developer/technology/apis.html>
- Ordnance Survey – geographic datasets in RDF – <http://data.ordnancesurvey.co.uk>

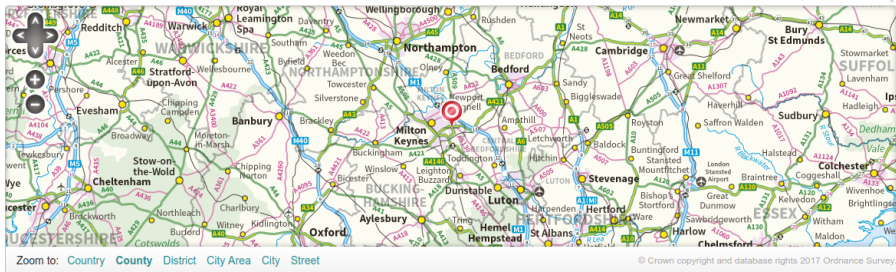



BBC Wildlife Ontology



Ordnance Survey Linked Data

Kents Hill, Monkston and Brinklow

Map powered by OS OpenSpace 

Kents Hill, Monkston and Brinklow is a Parish in Milton Keynes.

Objects related to "Kents Hill, Monkston and Brinklow"

Extent	41649-49
In European Region	South East
Within	Milton Keynes
In District	Milton Keynes
Touches	Walton Broughton Old Woughton Milton Keynes Wavendon

Core facts about "Kents Hill, Monkston and Brinklow"

Type	Parish
Label	Kents Hill, Monkston and Brinklow
Pref Label	Kents Hill, Monkston and Brinklow
Alt Label	Kents Hill, Monkston and Brinklow CP
Northing	238013.803835
Easting	489602.596729
Lat	52.0333028515
Long	-0.695254366017
Area Code	CPC

Semantic Web Principles

- 1 About Knowledge Management
- 2 Overview of Ontologies
- 3 Overview of Data Integration
- 4 Introduction to Semantic Web**
 - Semantic Web Adopters
 - Semantic Web Principles**
- 5 Linked Data
- 6 Linked Data



Unique Data Identification – URIs

Semantic web speaks about resources.

URI is a unique identifier for addressing web resources in the form

`<scheme name> : <hier. part> [? <query>] [# <fragment>]`

. HTTP scheme is used typically.

URN a URI with *scheme name* equal to 'urn'; used e.g. in SWRL atom identification,

URL a URI that can be resolved to a content using the protocol (e.g. HTTP),

IRI generalization of URIs allowing non-ascii characters. IRI is the standard identifier for OWL.



Open World Assumption

The semantic web inference must take into account that we handle *incomplete knowledge*.

Description

Open world (OWA): Everything that cannot be proven is unknown,
Closed world (CWA): Everything that cannot be proven is false.

Statement : "John is a Man."

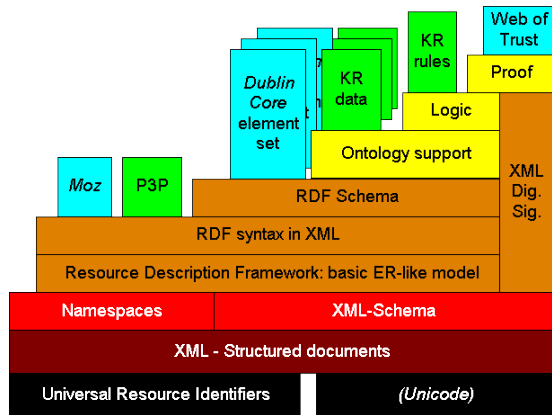
Query: "Is Jack a Man ?"

OWA Answer: "I don't know."

CWA Answer: "No."



Semantic Web Stack



Taken from <http://www.w3.org/2000/Talks/0906-xmlweb-tbl/slide9-0.html>, by Tim Berners Lee.



- 1 About Knowledge Management
- 2 Overview of Ontologies
- 3 Overview of Data Integration
- 4 Introduction to Semantic Web
 - Semantic Web Adopters
 - Semantic Web Principles
- 5 Linked Data**
- 6 Linked Data

Linked Data



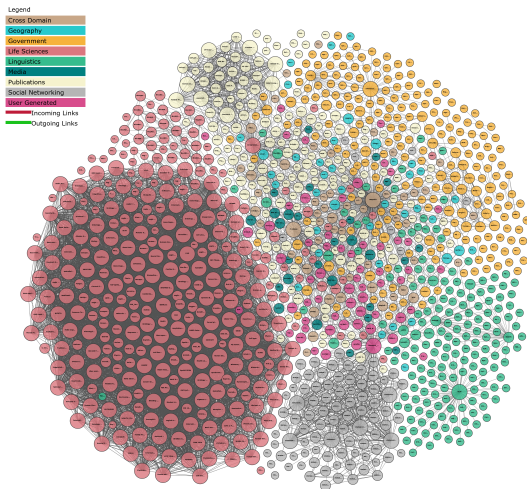
How to publish data related to other ?

Based on semantic web principles, Linked Data provide means to efficiently connect data created by different publishers.

- Web of Documents – WWW
 - webpage – readable by human
 - identifiers – IRI
 - transfer protocol – HTTP
 - unified language – HTML
- Web of Data – Linked Data
 - webpage – readable by machine
 - identifiers – IRI
 - transfer protocol – HTTP
 - unified language – RDF



Linked Open Data Cloud



“Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak.



- 1 About Knowledge Management
- 2 Overview of Ontologies
- 3 Overview of Data Integration
- 4 Introduction to Semantic Web
 - Semantic Web Adopters
 - Semantic Web Principles
- 5 Linked Data
- 6 Linked Data**

Linked Data



Selected Materials

- OSW pages – <https://cw.fel.cvut.cz/wiki/courses/osw>
- RDF Primer – <https://www.w3.org/TR/rdf11-primer/>
- SPARQL Query Language Spec – <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/>
- OWL Primer – <https://www.w3.org/TR/owl2-primer/>
- SKOS Primer – <https://www.w3.org/TR/skos-primer/>
- Description Logic Reasoning – P. Křemen, *Ontologie a Deskripční logiky*. In *Umělá inteligence VI.*, Academia, 2013.
- Linked Data – <http://linkeddata.org>
- Nice supplementary tutorial on RDF/OWL – <https://www.obitko.com/tutorials/ontologies-semantic-web/>

