COMP318

Ontologies and Semantic Web

Introduction and RDF part 1

Dr Valentina Tamma

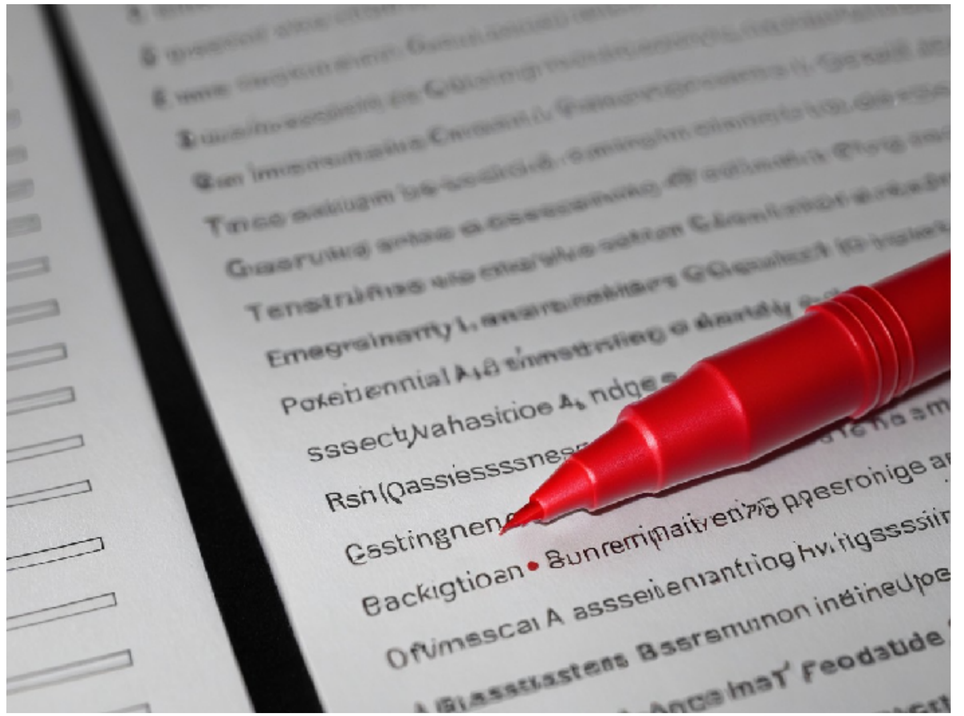
V.Tamma@liverpool.ac.uk

Organisation of the course

- Flipped classroom:
  - New videos released weekly
  - Quizzes and other activities at the end of each topic
  - Weekly face to face exercises and discussion
- Practicals:
  - Independent programming tasks in the lab
- Assessment
  - 70% exam
  - 10% assignment
  - 10% class test
  - 10% completion of 3 engagement tasks (in week 2, week 4 and week 10)

Engagement tasks

- They are checkpoints to assess your understanding of the subject, and encourage you to keep the pace:
  - You get marks primarily for trying;
  - A mix of practical exercises and MCQ questions;
  - Every engagement task is awarded:
    - 3 marks, if your overall score is ≧ 70;
    - 2 marks, if your overall score is between 50 and 70;
    - 1 mark, if your overall score is 50 and;
    - 0 marks if you don’t attempt;
  - At the end of the module, an additional mark is added to the total if you attempted all 3 engagement tasks;
  - Contribute 10% of the final module grade.

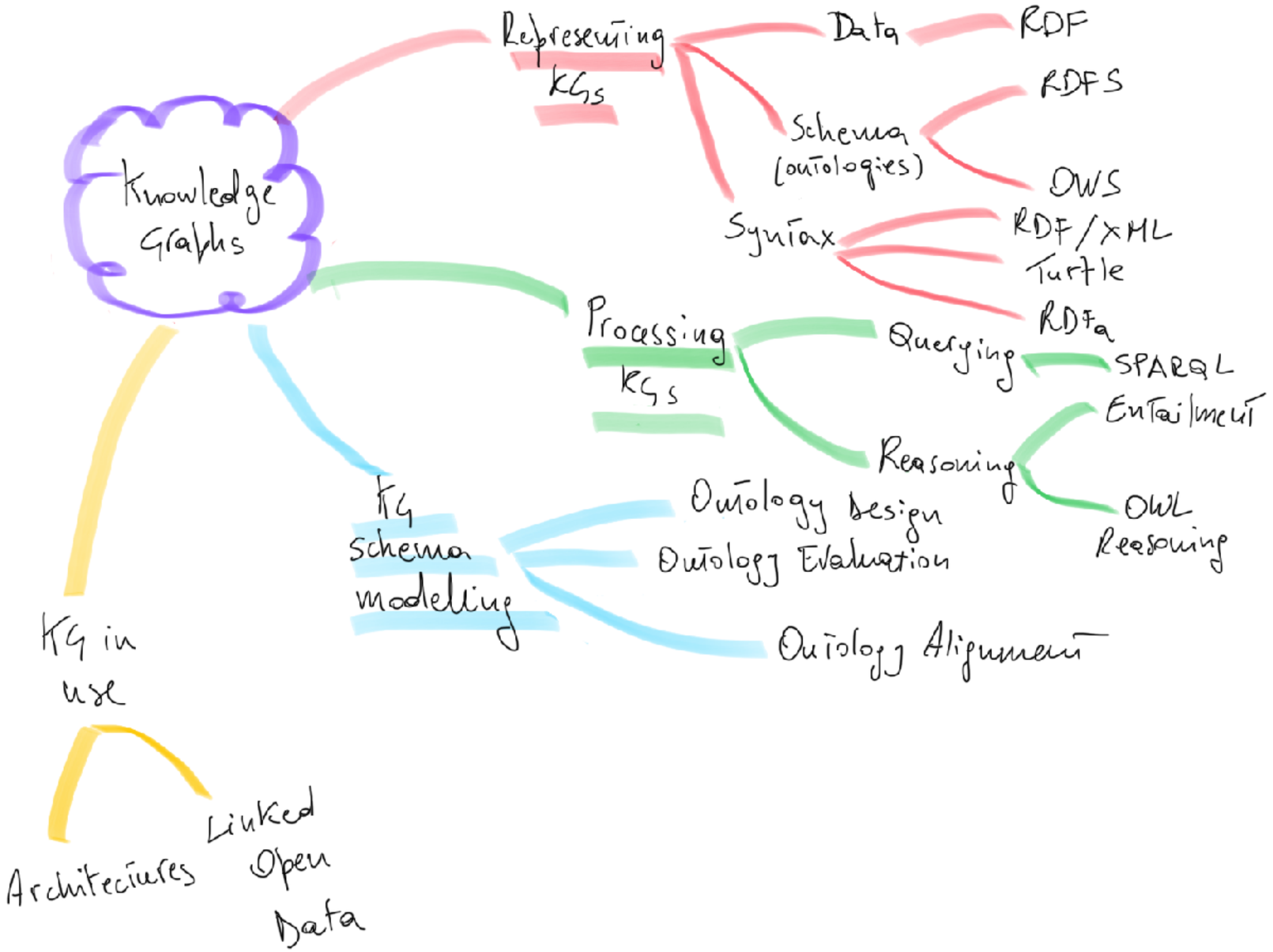


What to study

- Lecture slides and videos:
  - Slides are available on Canvas;
- Text books:
  - There isn't a single book, but rather a collection of book chapters. All of these are available from the library:
    - Antoniou, Groth, Hoekstra, van Harmelen: A Semantic Web Primer. MIT Press, 2012
    - Uschold: Demistifying OWL for the enterprise. Morgan & Claypool 2018.
    - Heath, Bizer: Linked Data - Evolving the Web into a Global Data Space. Morgan & Claypool 2011.
    - Euzenat, Shvaiko: Ontology matching. Springer, 2012
    - Gomez-Perez, Fernandez-Lopez, Corcho: Ontological Engineering. Springer-Verlag, 2003
- Useful readings:
  - Papers of interest and web resources will be indicated during the lectures, and are available on the course web pages. They will allow you to get a better understanding of the topic, and ultimately better marks!

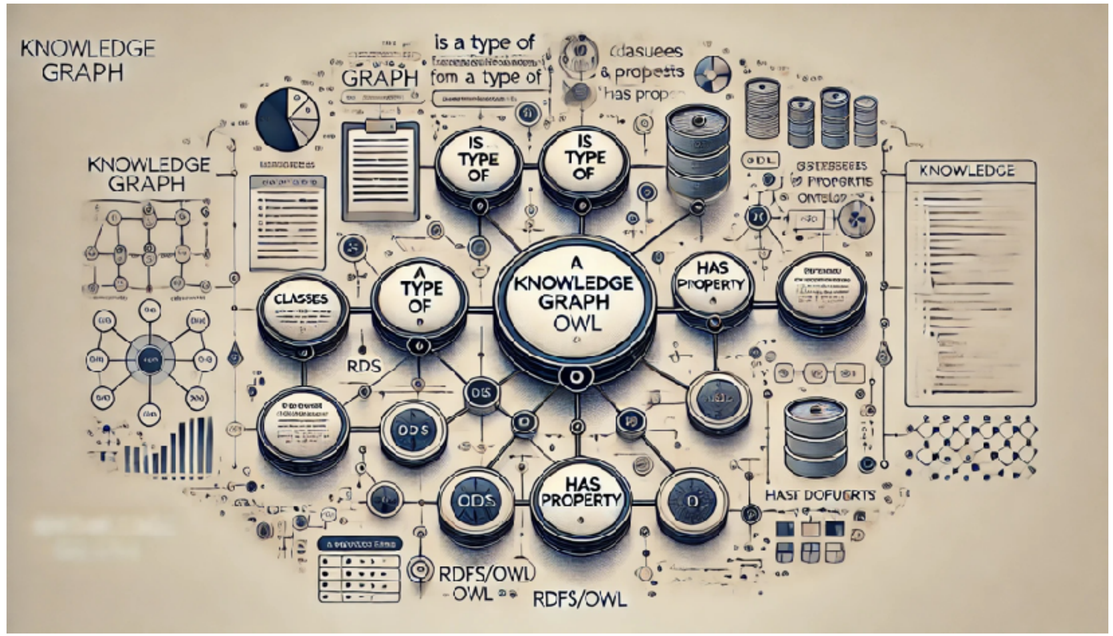

E WORLD WADE WEB
INTERN
Content

Course content



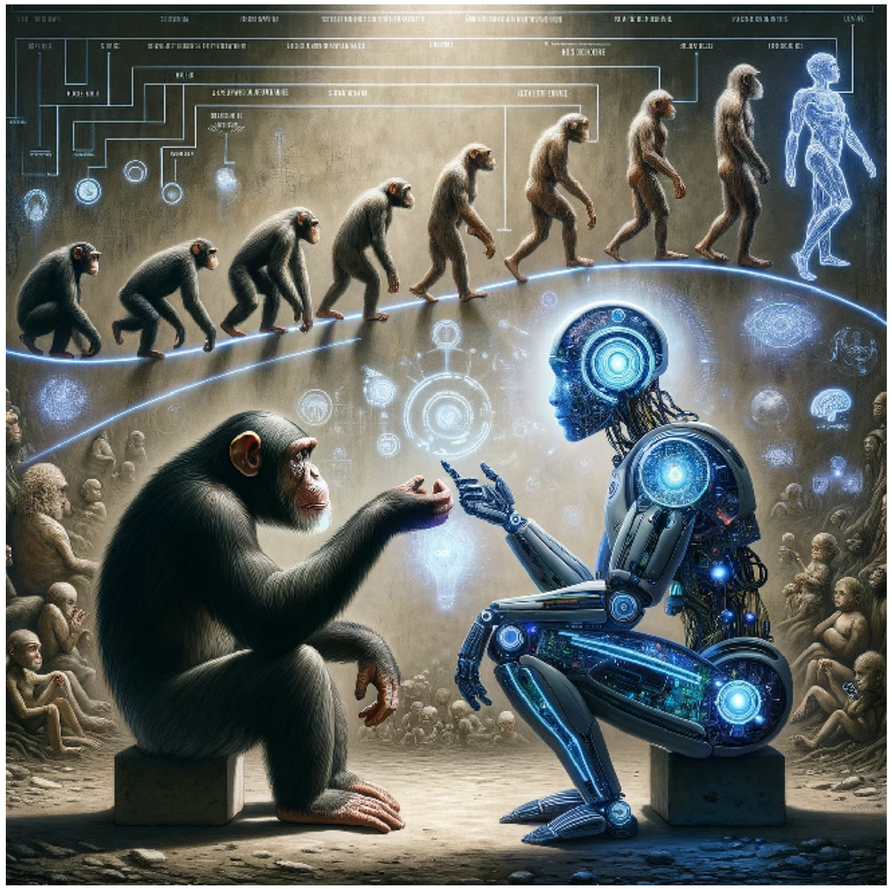
Focus on knowledge graphs

- Knowledge graphs provide an intuitive and yet powerful way to combine:
  - Complex data, often heterogeneous;
  - Background knowledge about a topic;
  - Common-sense knowledge, about the world.
- Similar to the way human recall concepts or entities, and reason about things;
- Foundation of knowledge-based AI
  - Now in mainstream devices, e.g mobile phones



The evolution of the web

- Initially the WWW was:
  - Mainly text document centric;
  - Expressed to be understood by humans;
  - Different languages are a barrier.
- More document formats were slowly introduced.
- Searching functionalities were limited:
  - Yahoo folder hierarchy.



The evolution of the web

- The current Web is made of text, pictures, music, movies ….
  - Still great for People!
    - We can make sense of what we see, and select, combine and integrate information from these various sources
  - Useless for Computer Automation
    - Computers can’t easily do “intelligent” tasks, which requires understanding this information

What made the web scale

- The success of the web is due to:
  - ability to search for content
    - search engines
  - ability to exchange and communicate data
    - standardised formats and protocols
      - HTML, XML

- The web is both a ***technological*** artefact and a ***social*** environment

User based Interpretation & Aggregation of results

- The problem is that some entity (person or organisation) has knowledge of something that they want to share:
  - E.g. knowledge about their business: opening hours, availability, locations, business category
- The entities producing the data need to advertise it to possible consumers using many channels: the entity's web page itself, but also search engines, personal assistants, mash-ups, review sites, maps, etc.
- But the data is too idiosyncratic, or too detailed, or just too complex to simply advertise it by writing its description.

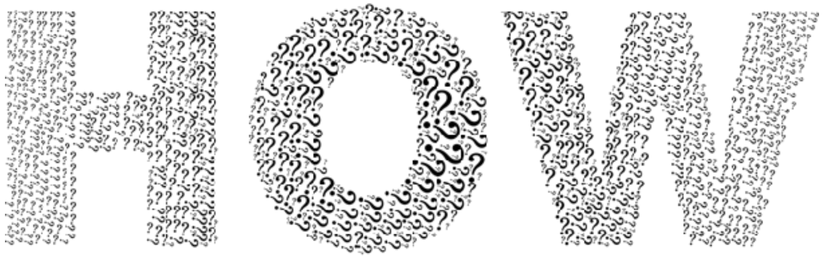
How do you…?

- Suppose you consult a web page for Sherwood forest. You want to find a list of all hotels that have branches in the vicinity
  - Your favourite chain, TavernLodge, does not appear, but some hotels near Sherwood forest are listed on the TavernLodge website
  - You want to the hotel that is nearest to the forest. You have the address of the forest, and the addresses of your hotel locations, and you can access several mapping services on the Web.
  - One of them shows the forest, and some hotels nearby, but they don't show all the TavernLodge locations.
  - Solution: **you** copy and paste the addresses from the TavernLodge page to the map, and you do the same for the forest.

- You have an appointment with your hairdresser, who just moved premises, and you want to check the new address.
  - You think you remember where the new shop is, somewhere on Parr Street, and the name of the hairdresser, but not the name of the salon.
    - You look for hairdressers on Parr Street, but none of them sounds familiar. When you finally find their web page, the hairdresser is listed as "beauty salon".
  - How do you know that a hairdresser is a type of beauty salon?

How do you…?

- Suppose you consult a web page for Sherwood forest. You want to find a list of all hotels that have branches in the vicinity
  - Your favourite chain, TavernLodge, does not appear, but some hotels near Sherwood forest are listed on the TavernLodge website
  - You want to the hotel that is nearest to the park. You have the address of the park, and the addresses of your hotel locations, and you can access several mapping services on the Web.
  - One of them shows the park, and some hotels nearby, but they don't show all the TavernLodge locations.
  - Solution: you copy and paste the addresses from the TavernLodge page to the map, and you do the same for the park.



- You have an appointment with your hairdresser, who just moved premises, and you want to check the new address.
  - You think you remember where the new shop is, somewhere on Parr Street, and the name of the hairdresser, but not the name of the salon.
  - You look for hairdressers on Parr Street, but none of them sounds familiar. When you finally find their web page, the hairdresser is listed as “beauty salon”.
  - How does an app know that a hairdresser **is a type of** beauty salon?

What if?

- But what if the Web was made of text, pictures, music, movies etc that were also readable and processable by machines?
  - **From** a presentation of information for a specific audience, requiring interpretation from a human being,
  - **To** an exchange of data between machines.
    - Instead of relying on human intuition just in the interpretation of the data, we have data providers make it easier to consume the data. The desire to share data, facilitates its consumption.

- Each data source is no longer a single point that is communicating one thing to a group of people but it is part of an interconnected network of data, consumed by different stakeholders.

- What would be the impact on tasks like search, query answering and knowledge aggregation?

As we read it

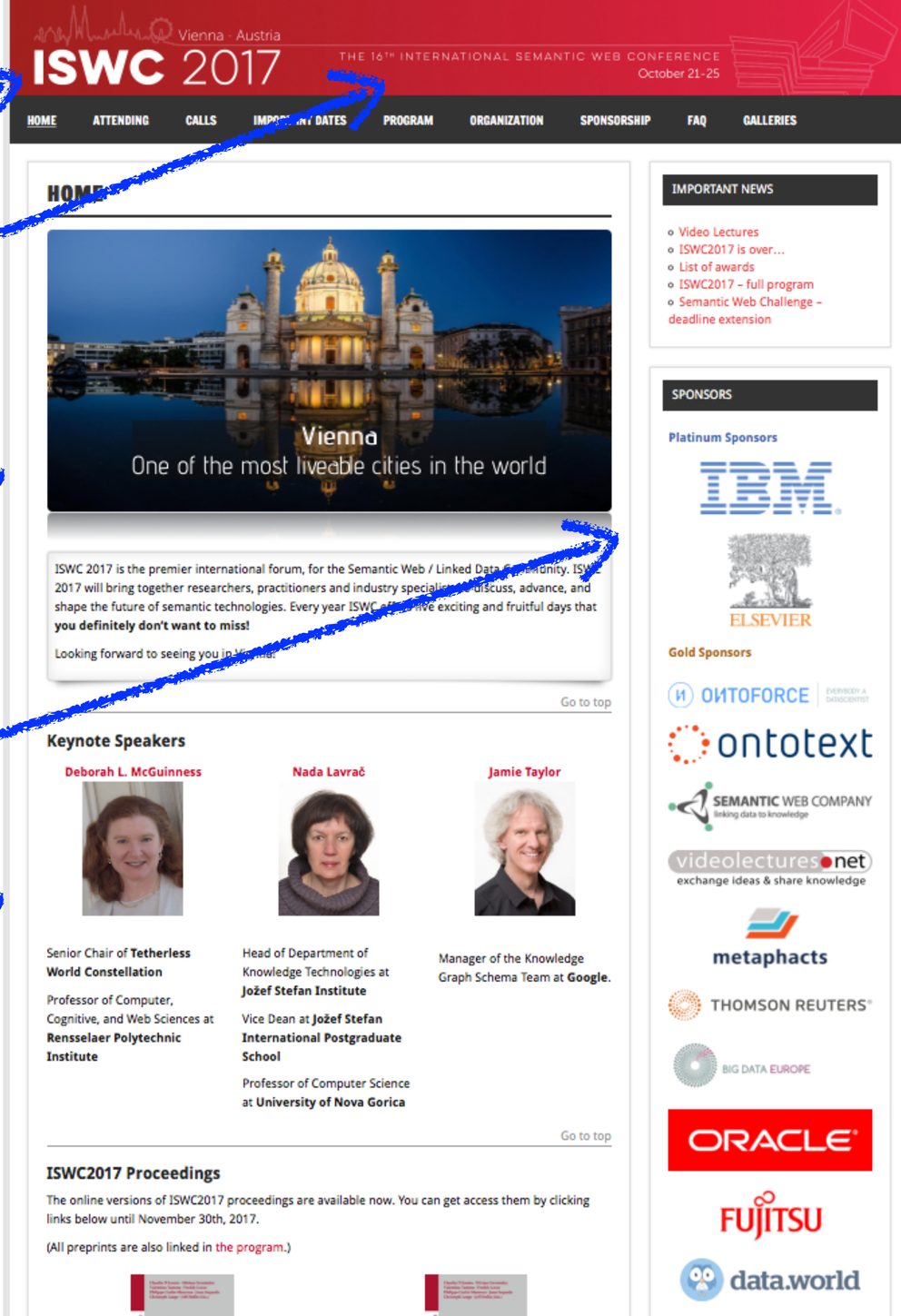
What is the page about

When does it take place

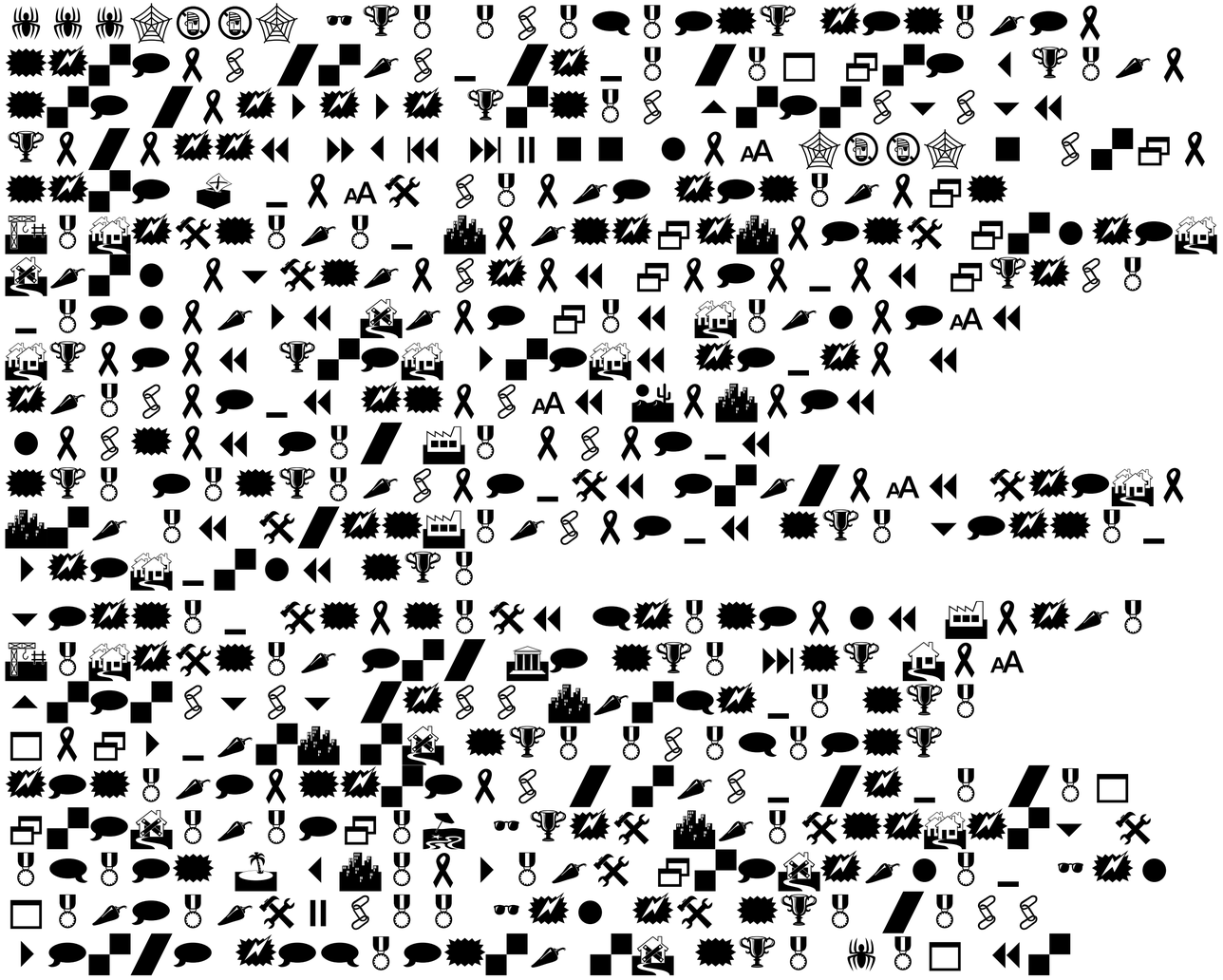
Where does it take place

Companies sponsoring

Keynote speakers



As a computer reads it



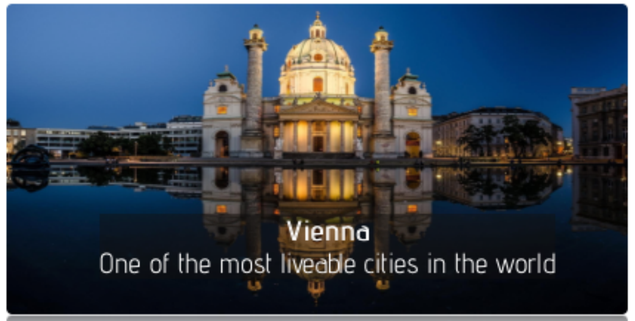
Vienna · Austria

ISWC 2017

THE 16TH INTERNATIONAL SEMANTIC WEB CONFERENCE
October 21-25

HOME | ATTENDING | CALLS | IMPORTANT DATES | PROGRAM | ORGANIZATION | SPONSORSHIP | FAQ | GALLERIES

HOME



ISWC 2017 is the premier international forum, for the Semantic Web / Linked Data Community. ISWC 2017 will bring together researchers, practitioners and industry specialists to discuss, advance, and shape the future of semantic technologies. Every year ISWC offers five exciting and fruitful days that **you definitely don't want to miss!**

Looking forward to seeing you in Vienna!

Go to top

Keynote Speakers

Deborah L. McGuinness

Senior Chair of **Tetherless World Constellation**

Professor of Computer, Cognitive, and Web Sciences at **Rensselaer Polytechnic Institute**

Nada Lavrač

Head of Department of Knowledge Technologies at **Jožef Stefan Institute**

Vice Dean at **Jožef Stefan International Postgraduate School**

Professor of Computer Science at **University of Nova Gorica**

Jamie Taylor

Manager of the Knowledge Graph Schema Team at **Google**.

Go to top

ISWC2017 Proceedings

The online versions of ISWC2017 proceedings are available now. You can get access them by clicking links below until November 30th, 2017.

(All preprints are also linked in the program.)

IMPORTANT NEWS

- Video Lectures
- ISWC2017 is over…
- List of awards
- ISWC2017 – full program
- Semantic Web Challenge – deadline extension

SPONSORS

Platinum Sponsors

IBM

ELSEVIER

Gold Sponsors

ONTOFORCE

ontotext

SEMANTIC WEB COMPANY
linking data to knowledge

videolectures.net
exchange ideas & share knowledge

metaphacts

THOMSON REUTERS

BIG DATA EUROPE

ORACLE

FUJITSU

data.world

```
<!DOCTYPE html><!-- HTML 5 -->
<html lang="en-US">

<head>
<meta charset="UTF-8" />
<meta name="viewport" content="width=device-width, initial-scale=1">
<link rel="profile" href="http://gmpg.org/xfn/11" />
<link rel="pingback" href="https://iswc2017.ai.wu.ac.at/xmlrpc.php" />

<title>ISWC2017 The 16th International Semantic Web Conference &#8211; ISWC2017</title>
<link rel='dns-prefetch' href='//iswc2017.ai.wu.ac.at' />
<link rel='dns-prefetch' href='//maps.google.com' />
<link rel='dns-prefetch' href='//fonts.googleapis.com' />
<link rel='dns-prefetch' href='//s.w.org' />
<link rel="alternate" type="application/rss+xml" title="ISWC2017 The 16th International Semantic Web Conference &raquo; Feed" href="https:
<link rel="alternate" type="application/rss+xml" title="ISWC2017 The 16th International Semantic Web Conference &raquo; Comments Feed" hre
        <script type="text/javascript">
            window._wpemojiSettings = {"baseUrl":"https:\/\/s.w.org\/images\/core\/emoji\/2\/72x72\/","ext":".png","svgUrl":"https:\/\/s.w
            !function(a,b,c){function d(a){var c,d,e,f,g,h=b.createElement("canvas"),i=h.getContext&&h.getContext("2d"),j=String.fromCharC
        </script>
        <style type="text/css">
img.wp-smiley,
img.emoji {
    display: inline !important;
    border: none !important;
    box-shadow: none !important;
    height: 1em !important;
    width: 1em !important;
    margin: 0 .07em !important;
    vertical-align: -0.1em !important;
    background: none !important;
    padding: 0 !important;
}
</style>
<link rel='stylesheet' id='dropshadowboxes_css-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/plugins/drop-shadow-boxes/css/dropshado
<link rel='stylesheet' id='ce_responsive-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/plugins/simple-embed-code/css/video-container
<link rel='stylesheet' id='faqwd_front_end_style-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/plugins/faq-wd/css/front_end_style.cs
<link rel='stylesheet' id='faqwd_front_end_default_style-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/plugins/faq-wd/css/default.cs
<link rel='stylesheet' id='dynamicnewslite-stylesheet-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/themes/dynamic-news-lite/style.c
<link rel='stylesheet' id='genericons-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/themes/dynamic-news-lite/css/genericons/generico
<link rel='stylesheet' id='dynamicnewslite-default-fonts-css'  href='//fonts.googleapis.com/css?family=Droid+Sans%3A400%2C700%7CFrancois+O
<link rel='stylesheet' id='wps_gfw-widget-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/plugins/easy-customizable-text-widget/css/fr
<link rel='stylesheet' id='wpgmp-frontend-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/plugins/wp-google-map-plugin//assets/css/fro
<link rel='stylesheet' id='ms-main-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/plugins/masterslider/public/assets/css/masterslider
<link rel='stylesheet' id='ms-custom-css'  href='https://iswc2017.ai.wu.ac.at/wp-content/uploads/masterslider/custom.css?ver=7.5' type='te
<link rel='stylesheet' id='wp-add-custom-css-css'  href='https://iswc2017.semanticweb.org?display_custom_css=css&#038;ver=4.6.10' type='te
<!-- This site uses the Google Analytics by MonsterInsights plugin v5.5.3 - Universal enabled - https://www.monsterinsights.com/ -->
<script type="text/javascript">
    (function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){
        (i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new Date();a=s.createElement(o),
        m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(a,m)
    })(window,document,'script','//www.google-analytics.com/analytics.js','__gaTracker');

    __gaTracker('create', 'UA-85705863-1', 'auto');
    __gaTracker('set', 'forceSSL', true);
    __gaTracker('send','pageview');

</script>
<!-- / Google Analytics by MonsterInsights -->
<script type='text/javascript' src='https://iswc2017.ai.wu.ac.at/wp-includes/js/jquery/jquery.js?ver=1.12.4'></script>
<script type='text/javascript' src='https://iswc2017.ai.wu.ac.at/wp-includes/js/jquery/jquery-migrate.min.js?ver=1.4.1'></script>
<!--[if lt IE 9]>
<script type='text/javascript' src='https://iswc2017.ai.wu.ac.at/wp-content/themes/dynamic-news-lite/js/html5shiv.min.js?ver=3.7.3'></scri
<![endif]-->
<script type='text/javascript'>
/* <![CDATA[ */
```

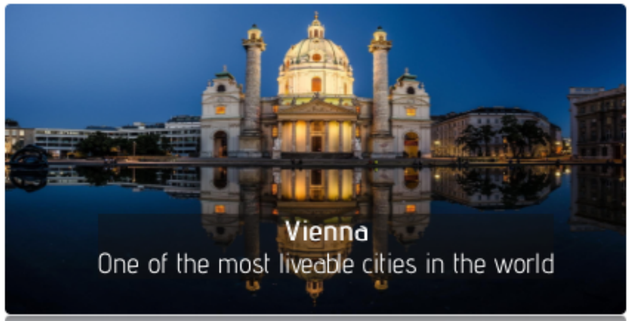
Vienna · Austria

ISWC 2017

THE 16TH INTERNATIONAL SEMANTIC WEB CONFERENCE
October 21-25

HOME | ATTENDING | CALLS | IMPORTANT DATES | PROGRAM | ORGANIZATION | SPONSORSHIP | FAQ | GALLERIES

HOME



ISWC 2017 is the premier international forum, for the Semantic Web / Linked Data Community. ISWC 2017 will bring together researchers, practitioners and industry specialists to discuss, advance, and shape the future of semantic technologies. Every year ISWC offers five exciting and fruitful days that **you definitely don't want to miss!**

Looking forward to seeing you in Vienna!

Go to top

Keynote Speakers

Deborah L. McGuinness

Senior Chair of **Tetherless World Constellation**

Professor of Computer, Cognitive, and Web Sciences at **Rensselaer Polytechnic Institute**

Nada Lavrač

Head of Department of Knowledge Technologies at **Jožef Stefan Institute**

Vice Dean at **Jožef Stefan International Postgraduate School**

Professor of Computer Science at **University of Nova Gorica**

Jamie Taylor

Manager of the Knowledge Graph Schema Team at **Google**.

Go to top

ISWC2017 Proceedings

The online versions of ISWC2017 proceedings are available now. You can get access them by clicking links below until November 30th, 2017.

(All preprints are also linked in the program.)

IMPORTANT NEWS

- Video Lectures
- ISWC2017 is over…
- List of awards
- ISWC2017 – full program
- Semantic Web Challenge – deadline extension

SPONSORS

Platinum Sponsors

IBM

ELSEVIER

Gold Sponsors

ONTOFORCE

ontotext

SEMANTIC WEB COMPANY
linking data to knowledge

videolectures.net
exchange ideas & share knowledge

metaphacts

THOMSON REUTERS

BIG DATA EUROPE

ORACLE

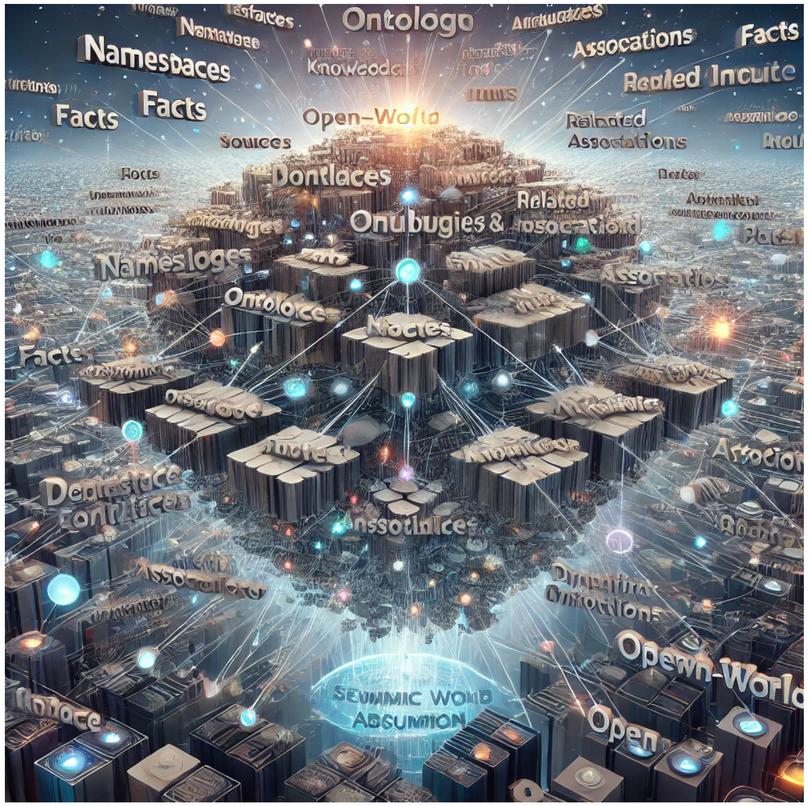
FUJITSU

data.world

Semantic Web

- Evolution of the current web towards a distributed web of data:
  - At the level of data rather than presentation
  - From information to knowledge based
- Represent knowledge on the Web in a form that is more easily machine-processable:
  - Improve retrieval;
  - Use intelligent techniques to take advantage of these representations:
  - Delegate autonomous software components (agents)

But what is the Semantic Web?



- It NOT just another centralised Knowledge Base framework !!!
  - The Semantic Web is a web of (partially connected) vocabularies (ontologies) and knowledge bases

- Facts scattered in many locations…
  - Breaking the “file” paradigm
    - Related instances & ontology fragments can be asserted anywhere
    - Scope & association is managed through namespaces
    - References to ontologies and instances managed through URI
  - Dynamic knowledge source that is typically:
    - Inconsistent / chaotic
    - Incomplete (open-world)
    - Composed of Several Ontologies
    - Generated by novices and machines, as well as experts

The principles of the Semantic Web

- ***AAA: Anyone can say anything about anything***
  - Any individual can assert data about some entity, ideally in a way that can be combined with data from other sources.
    - Foundation for the design of RDF
  - Issues about maintenance, trustability, veracity
- ***Linked Open data cloud and knowledge graphs***
  - Much of the information in the Semantic Web was originally in HTML, **sometimes in the form of tables, spreadsheets, or databases**
  - Now there are several large, public data sources available in RDF, including an RDF version of Wikipedia called dbpedia, a large number of government datasets and enterprise data stores
- ***Information is distributed, networked, disparate and maybe inconsistent***
  - In general, we cannot assume at any time that we have all the information in the network, or even that we know every thing that has been asserted about one single topic (***Open world assumption***)
  - But this is fine, this is an artefact of the nature of knowledge and information

Distributing data on the Web

- Data are most typically represented in tabular form
  - each row represents a data item we are describing,
  - each column represents some property (or attribute) of these items.
  - the cells in the table represent specific values for those properties for the data item

- Let us consider an example, the works from the Beatles band members
  - What are the strategies to distribute this data on the Web

Works from the Beatles band members

ID	Title	Attribution	Type	Year
1	Here comes the sun	George Harrison	Song	1969
2	In his own write	John Lennon	Book	1964
3	Here Today	Paul McCartney	Song	1982
4	A Spaniard oin the works	John Lennon	Book	1965
5	A hards day night	Beatles	Movie	1964
6	Blackbird	John Lennon, Paul McCartney	Song	1968
7	Don’t pass me by	Richard Starkey	Song	1968
8	Dig it	Lennon, McCartney, Harrison, Starkey	Song	1970

Distributing by Rows

1	Here comes the sun	George Harrison	Song	1969

4	A Spaniard oin the work			5
6	Blackbird			8
8	Dig it			0

Needs to reference the schema: Which column is which?

7	Don’t pass me by	Richard Starkey	Song	1968
3	Here Today	Paul McCartney	Song	1982

Distributing by Columns

1	Here comes the sun
2	In his own write
3	Here Today
4	A Spaniard oin the works
5	A hards day night
6	Blackbird
7	Don’t pass me by
8	Dig it

George Harrison
John Lennon
Paul McCartney
John Lennon
Beatles
John Lennon, Paul McCartney
Richard Starkey
Lennon, McCartney, Harrison, Starkey

Song	1969
Song	1968
Song	1970

Needs to reference entities: what are the things that we’re talking about

V.Tamma

Distributing by Cell

Needs to reference both schema and entities

	Type
Row 7	Song

	Title
Row 2	In his own write

	Type
Row 5	Movie

	Year
Row 2	1964

	Attribution
Row 8	Lennon, McCartney, Harrison, Starkey

XML: data exchange mechanism

- Exchange languages components
  - Syntax: how to write the data
  - Data model: how to structure or organise the data
  - Semantics: how to interpret (meaning) this data

```
<lecturer name John Smith>
<university> University of Liverpool</university>
<course>SemWeb Technologies</course>
<url>http://www.csc.liv.ac.uk/~jsmth</url>
</lecturer>
```

- Semantics clarifies the meaning, specifies assumptions and allows deductions

```
<course name SemWeb Technologies>
<university> University of Liverpool</university>
<lecturer>John Smith</lecturer>
<url>http://www.csc.liv.ac.uk/~Comp356</url>
</course>
```

XML is not the answer

- Meaning of XML documents is intuitively clear
  - “semantic” markup tags are domain terms
  - no unique way to express the same information
- But computers do not have intuition
  - Tag names per se do not provide semantics
  - The semantics are encoded outside the XML specification
- XML makes no commitment on:
  - Domain specific ontological vocabulary
  - Ontological modelling primitives
    - requires pre-arranged agreement on 1. & 2.
- Feasible for closed collaboration
  - agents in a small & stable community
  - pages on a small & stable intranet

RDF: Resource Description Framework

- Semantic Web: beyond machine readable to machine understandable.
  - Resource Description Framework (RDF) is the W3C language for describing metadata on the Web.

- **Graph data model that formally describes the semantics (meaning of information)**
  - Models Meta-Data about resources on the Web using subject-predicate-object triples
  - Triples define the relationship or predicate between two entities (the subject and object)

A richer data model

- Semantic Web: beyond machine readable to machine understandable.
  - data model:
    - data model that can be used by multiple applications
      - not only for describing documents
      - for people to describe application-specific information
    - data model that is domain independent
      - any application can use it to describe information
    - **data model that can be distributed in a seamless way**
  - semantics:
    - mechanism to interpret the data model
    - describes the interpretations of the data items wrt the domain
  - syntax:
    - standardised exchange mechanism

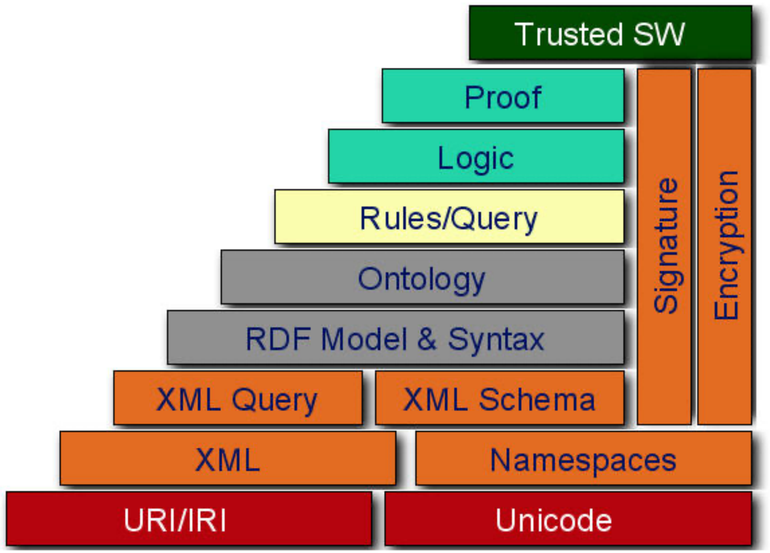
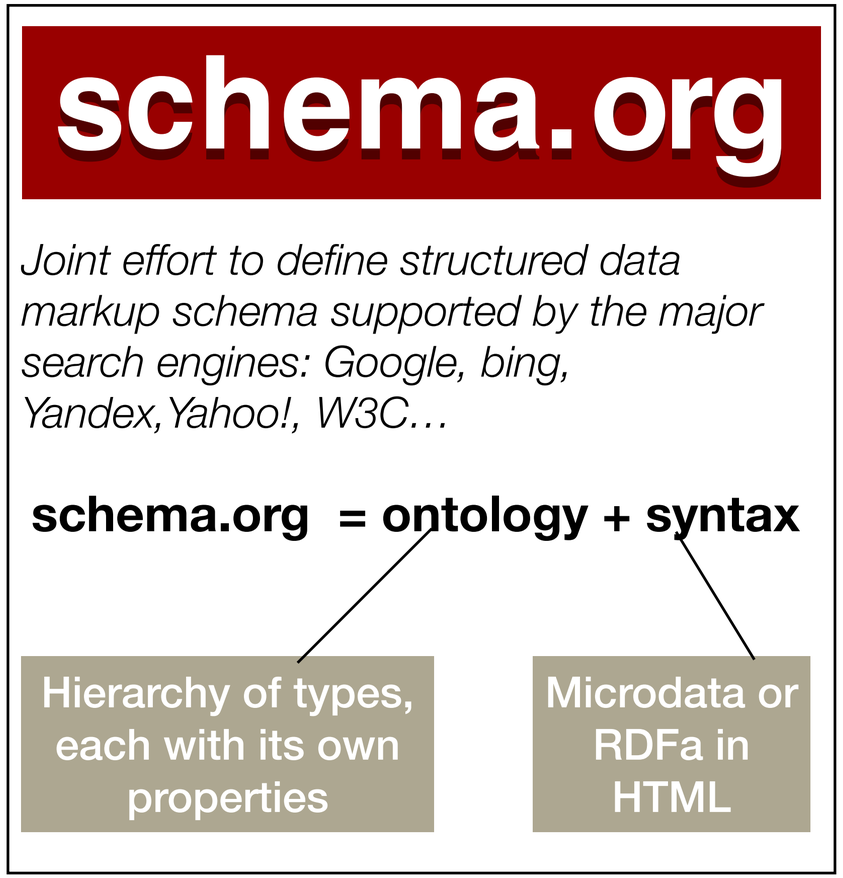


image: Wikimedia Commons

A richer data model

- Semantic Web: **beyond machine readable** to **machine understandable.**
  - data model:
    - **data model that can be used by multiple applications**
      - not only for describing documents
      - for people to describe application-specific information
    - **data model that is domain independent**
      - any application can use it to describe information
  - semantics:
    - mechanism to interpret the data model
    - describes the interpretations of the data items wrt the domain
  - syntax:
    - standardised exchange mechanism



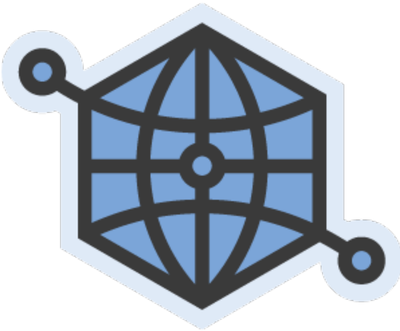
Schema.org, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=51534848

V.Tamma

The role of models

- Models are abstractions that help people understand (a fragment of) their world and assemble their knowledge
  - They hide some details whilst spotlighting others

schema.org



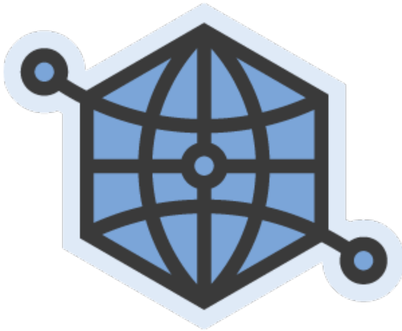
The role of models

- Models help people communicate:
  - by describing a situation in detail so other people can understand it

- Models explain and make predictions:
  - they relate primitive phenomena to one another and to more complex phenomena,
  - provide explanations and predictions about the world.

schema.org

- Models mediate among multiple viewpoints
  - no two people agree completely on what they want to know about a phenomenon;
  - **models represent their commonalities while allowing them to explore their differences.**



RDF Building Blocks

Statements

Statements are subject-predicate-object triples.

They assert the properties of a resource the resource, a property, and a value

Objects can be resources or literals (atomic values - strings)

Resources are similar to entities in ER models

- –"something" we want to describe
  - E.g. *authors*, *books*, *publishers*, *places*, *people*, *hotels*

Every resource has a URI

- –a URL (Web address) or
- –some other kind of unique identifier: URNs

Advantages of using URIs:

- –a global, worldwide, unique naming scheme
- –reduces the homonym problem in distributed data

Properties

Properties are special types of resources

- –they describe semantic relations between resources
  - E.g. *written by, age, smaller than, etc*
- –they are also identified by a URI

Three Views of a Statement

- A statement can be:
  - A triple
    - The triple (x,P,y) corresponds to the logical formula P(x,y)
      - Binary predicate P relates object x to object y
    - RDF offers only binary predicates (properties)
    - db:location(BaronWay_Building, db:Amsterdam)
  - A piece of a graph
  - A XML code fragment
- An RDF document can be viewed as:
  - A set of triples
  - A graph (semantic net)
  - An XML document

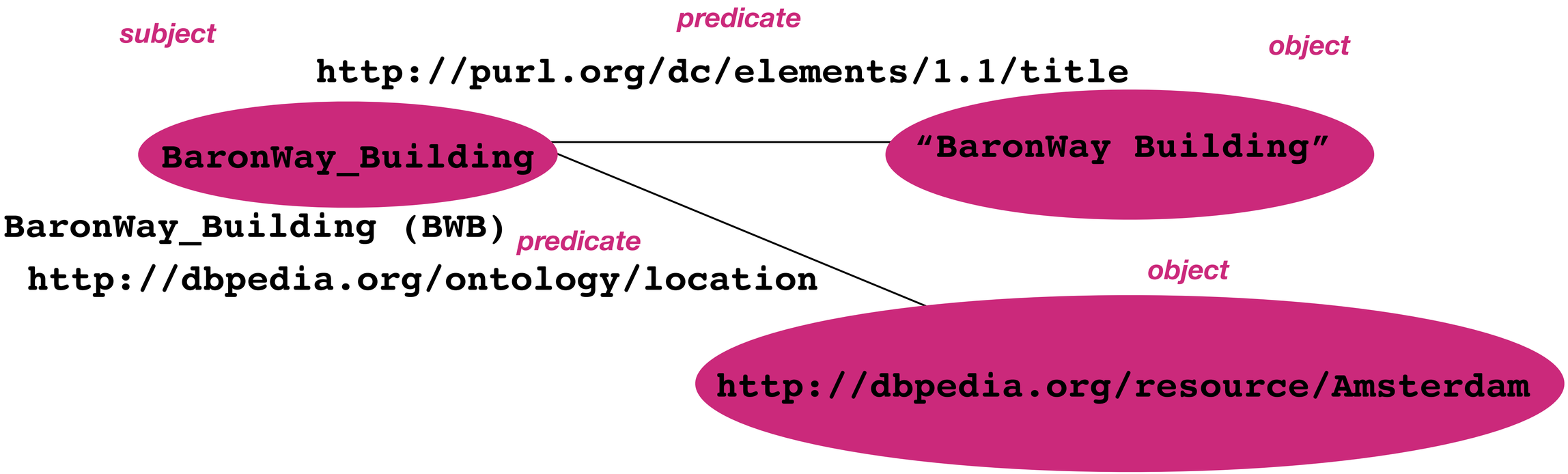
```
(BaronWay_Building,
  http://dbpedia.org/ontology/location,
  http://dbpedia.org/resource/Amsterdam)
```

Prefixes

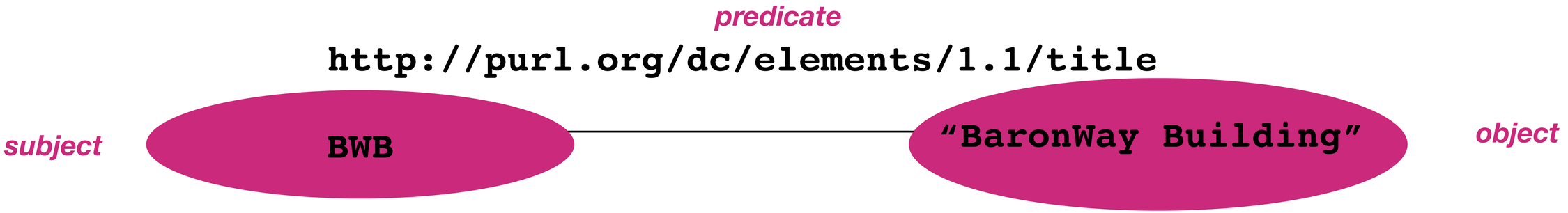
- Throughout the modules, we use real schema and vocabularies, that are widely used in Linked data or KG:
  - ***Dbpedia:***
    - ommunity effort to extract structured content from the information created in various Wikimedia projects (e.g Wikipedia).
    - This structured information is an open knowledge graph;
  - ***Foaf:***
    - descriptive vocabulary modellin persons, their activities and their relations to other people and objects.
    - ***FOAF*** is represented in RDF and OWL
  - ***DC:***
    - set of fifteen "core" elements (properties) for describing both digital resources (video, images, web pages, etc), as well as physical resources such as books or CDs, and objects like artworks.
    - formally standardised as ISO, ANSI and RFC standard,
    - DC Metadata Terms are represented using multiple RDF vocabularies,

RDF Graphs

- The subject/predicate/object triples found in an RDF document form a graph:



RDF Graph



- A directed graph with labeled nodes and arcs
  - **from** the resource (the **subject** of the statement)
  - **to** the value (the **object** of the statement)

- The value of a statement may be a resource
  - It may be linked to other resources

COMP318

Ontologies and Semantic Web

End of Introduction and RDF part 1

Dr Valentina Tamma

V.Tamma@liverpool.ac.uk