Lecture 3

Knowledge Graphs: Vocabularies & Ontologies

COMP 474/6741, Winter 2024

René Witte
Department of Computer Science
and Software Engineering
Concordia University

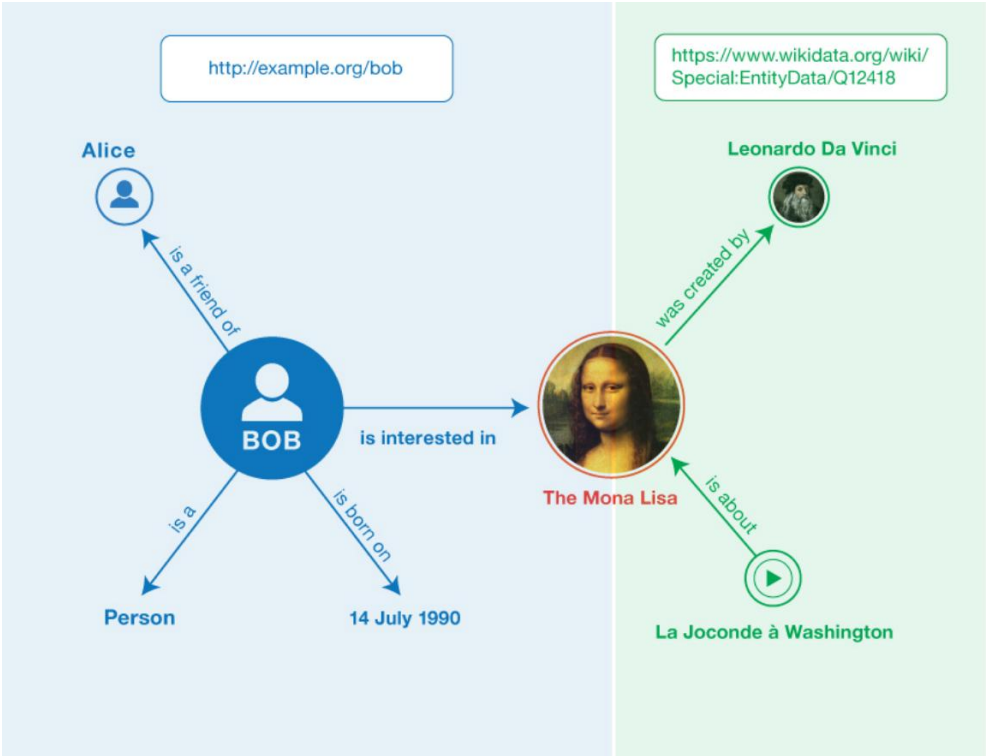
Outline

René Witte

Slides Credit

- Includes slides from Jay Pujara & Sameer Singh, *Mining Knowledge Graphs from Text*, https://kgtutorial.github.io/
- Includes slides by Ivan Herman, W3C [Her]

Outline



René Witte
Concordia University

Why knowledge graphs?

- Humans:
  - Combat information overload
  - Explore via intuitive structure
  - Tool for supporting knowledge-driven tasks
- AIs:
  - Key ingredient for many AI tasks
  - Bridge from data to human semantics
  - Use decades of work on graph analysis

RDF triples (cont.)

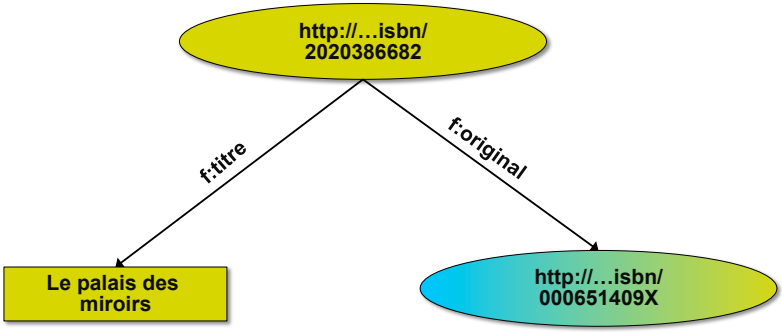
- An RDF Triple (s,p,o) is such that:
  - "s", "p" are URI-s, ie, resources on the Web; "o" is a URI or a literal
    - "s", "p", and "o" stand for "subject", "property", and "object"
  - here is the complete triple:

```
(<http://…isbn…6682>, <http://…/original>, <http://…isbn…409X>)
```

- *RDF* is a general model for such triples (with machine readable formats like RDF/XML, Turtle, N3, RDFa, Json, …)

René Witte
Concordia University

A simple RDF example (in Turtle)




```
<http://…/isbn/2020386682>
    f:titre "Le palais des mirroirs"@fr ;
    f:original <http://…/isbn/000651409X> .
```

René Witte
Concordia University

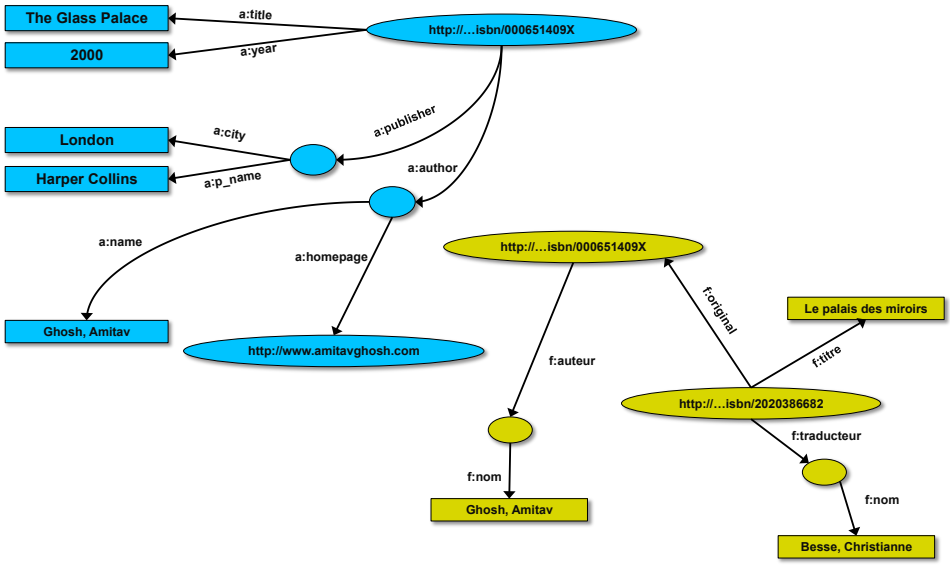
Anatomy of a URI

- A URI *(Uniform Resource Identifier)* uniquely identifies a resource (e.g., person, book, class of things) on the Web.
- A URI is not always a URL *(Uniform Resource Locator)*
  - URNs *(Uniform Resource Names)* are URIs that name resources without specifying how to retrieve them, e.g., `urn:isbn:0451450523`.
  - It's possible URLs are not available, e.g.,
    `http://www.concordia.ca/comp474/lecture03/slide5`
    is a valid URL, but nothing can be (currently?) retrieved from this address
- URIs have a generic syntax:

- Queries are also possible, following the format:
  scheme:[//authority]path[?query][#fragment]
- Namespaces are used to shorten URIs and prevent name clashes, e.g., `ex:me`
- IRIs *(Internationalized Resource Identifiers)* are URIs with Unicode characters

René Witte

3rd: start merging your data



René Witte

From the first week's lecture. . .

René Witte
Concordia University

Common Issue

- Data in Information Silos
  Documents, databases, spreadsheets, emails, . . .
- Disconnected, missing knowledge

Knowledge Integration

- Connect silo-ed knowledge
- Leverage existing, external Knowledge Bases
- Freely available, many domains
- Continuously updated

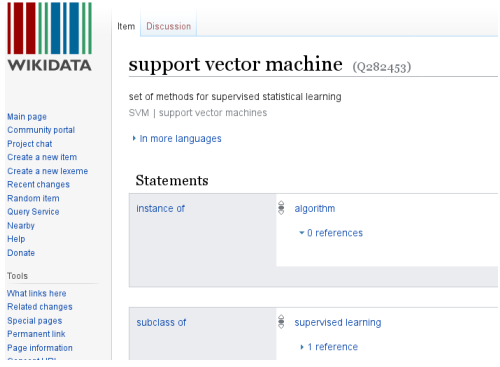
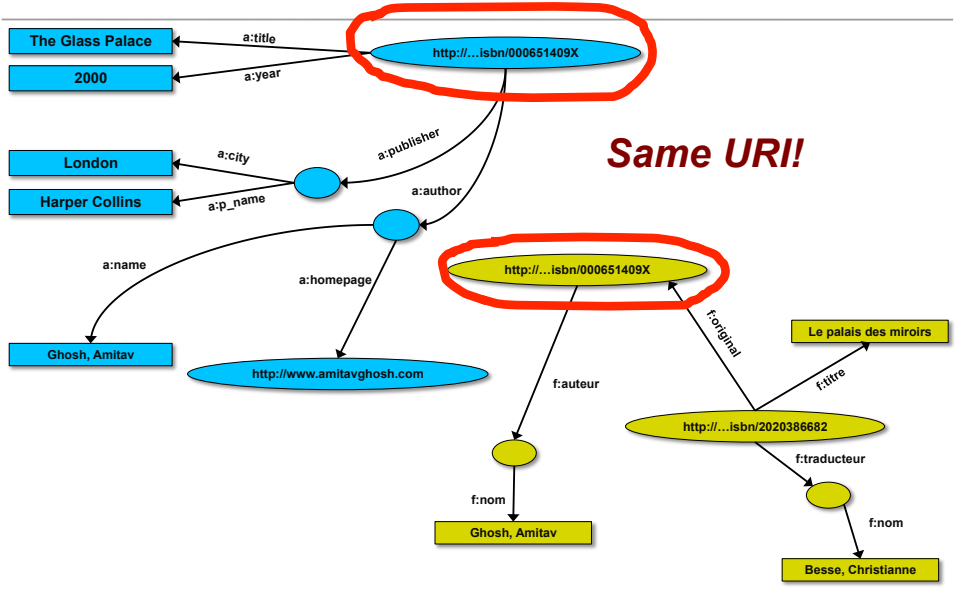


Photo credit "nepatterson", Flickr

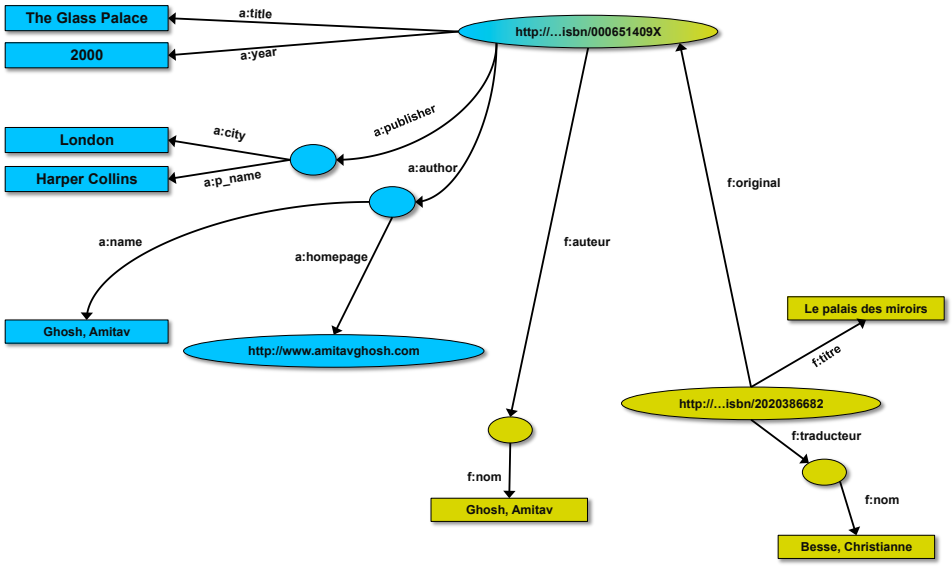
René Witte
Concordia University

3rd: start merging your data (cont)



René Witte
Concordia University

3rd: start merging your data



René Witte

Semantics

René Witte

What's an "author"?

- author
- auteur
- Autor
- book author
- writer
- editor
- ghostwriter
- co-author
- blogger
- ...

How can we define their meaning? And relations?

Everything Search

The Library of Congress > Linked Data Service > **LC Subject Headings (LCSH)**

Authors and publishers

Here are entered works on the relations between author and publisher.

URI(s)
- http://id.loc.gov/authorities/subjects/sh85010023
- info:lc/authorities/sh85010023
- http://id.loc.gov/authorities/sh85010023#concept

Instance Of
- MADS/RDF Topic
- MADS/RDF Authority
- SKOS Concept

Scheme Membership(s)
- Library of Congress Subject Headings

Collection Membership(s)
- LCSH Collection - Authorized Headings
- LCSH Collection - General Collection
- LCSH Collection - May Subdivide Geographically

Variants
- Author and publisher
- Authors and publishers--Law and legislation
- Publishers and authors
- Publishing contracts

Broader Terms
- Authorship
- Contracts

Narrower Terms
- Queries (Authorship)

Related Terms
- Book proposals

Subject Of Works
243 resources filter Page 1 of 5

Alden, Chevy. How to get published, guaranteed
Alden, Chevy. How to get published-guaranteed
Allen, Marilyn. complete idiot's guide to book proposals & query letters
Allfeld, Philipp, 1852- gesetze betreffend das urheberrecht an werken der literatur und der tonkunst und über das verlagsrecht
Allfeld, Philipp, 1852- verlagsrecht
Allison, Alida. Grad student's guide to getting published
Amir, Nina, How to blog a book
Amir, Nina, author training manual
Amir, Nina. How to blog a book
Anderson, Rick, 1965- Scholarly communication
Anderson, Rick, 1965- Scholarly communication
Appelbaum, Judith. How to get happily published

Suggest Alternative Terminology

René Witte
Concordia University

Library of Congress

René Witte

RDF Triples (SKOS vocabulary, introduced later)

```
<http://id.loc.gov/authorities/subjects/sh85010023>
    <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
    <http://www.w3.org/2004/02/skos/core#Concept> .

<http://id.loc.gov/authorities/subjects/sh85010023>
    <http://www.w3.org/2004/02/skos/core#prefLabel>
    "Authors and publishers"@en .

<http://id.loc.gov/authorities/subjects/sh85010023>
    <http://www.w3.org/2004/02/skos/core#broader>
    <http://id.loc.gov/authorities/subjects/sh85031620> .

<http://id.loc.gov/authorities/subjects/sh85010023>
    <http://www.w3.org/2004/02/skos/core#narrower>
    <http://id.loc.gov/authorities/subjects/sh85109817> .
...
```

http://id.loc.gov/authorities/subjects/sh85010023.html

Photo credit "kxlly", Flickr

René Witte

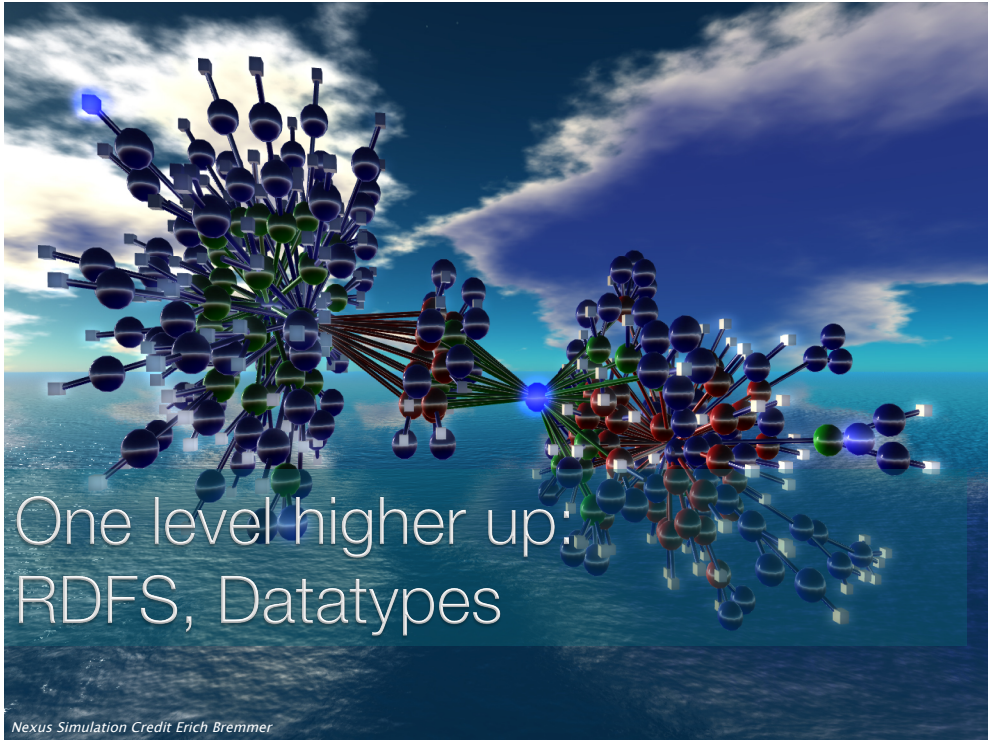
TBL at TED on “The year open data went worldwide” (2010)

Tim Berners-Lee: The year open data went worldwide

https://www.youtube.com/watch?v=3YcZ3Zqk0a8

René Witte
Concordia University

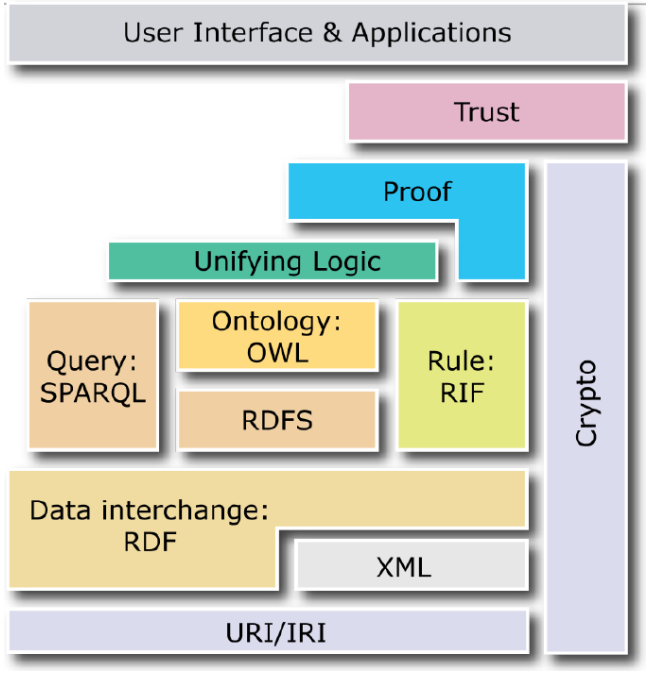
Outline



Nexus Simulation Credit Erich Bremmer

René Witte
Concordia University

The W3C “Layer Cake”



René Witte
Concordia University

RDF Schema (RDFS)

René Witte
Concordia University

W3C Recommendation

- "RDF Vocabulary Description Language 1.0: RDF Schema" (RDFS 1.0)
- Current version (2014): "RDF Schema 1.1"

Used together with RDF

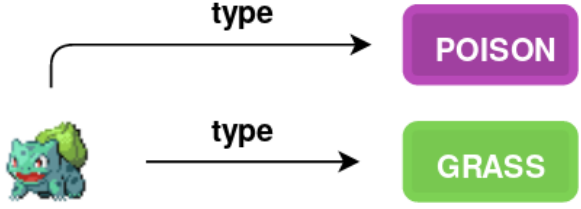
- RDF provides "a way to make statements about resources" (IRIs)
- RDFS provides *semantics* about what the IRIs stand for (Schemas aka Vocabularies aka Ontologies aka. . .)

Classes and Instances

René Witte
Concordia University

Classes

- Resources may be divided into groups called classes
- The members of a class are known as instances of the class
- An instance can be member of more than one class



What is a knowledge graph – Pokémon edition: https://pieterheyvaert.com/blog/2019/12/27/kg-pkmn/

Defining Classes

We define that an URI in a triple is a class using . . . a triple!
(sounds weird the first time you hear it, but you get used to it)

Classes and Instances (contd.)

René Witte
Concordia University

Defining Classes

To define that **C** (a resource) is an RDFS `class`, write:

C *rdf:type rdfs:Class*

with `rdfs` defined as http://www.w3.org/2000/01/rdf-schema#

Example

```
ex:Novel rdf:type rdfs:Class
```

Turtle

In Turtle, `rdf:type` can be abbreviated as `a`

→ **Worksheet #2: Task 2**

Classes and Instances (contd.)

René Witte
Concordia University

Instances

To define that **I** (a resource) is an instance of **C** (a class), write:

I `rdf:type` ***C***

(or `a` instead of `rdf:type` in Turtle.)

Example

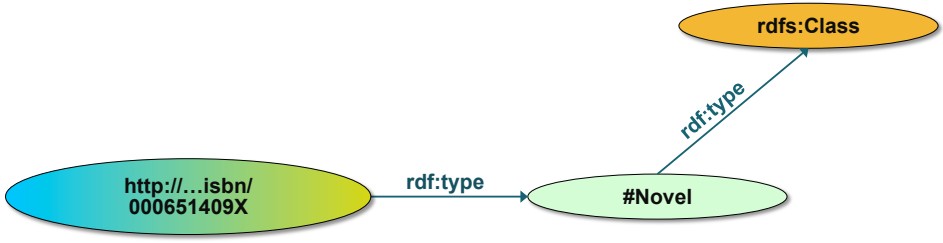
```
<http://...isbn/000651409X> rdf:type ex:Novel
```

Note

This is just another triple, so we can read both *data* and *schema* at run-time!

→ **Worksheet #2: Task 3**

Classes, resources in RDF(S)



- RDFS defines the meaning of these terms
  - (these are all special URI-s, we just use the namespace abbreviation)

René Witte
Concordia University

Label & Comment

René Witte
Concordia University

Human-Readable Content

By convention, always provide:

`rdfs:label` a human-readable label

`rdfs:comment` a short (one paragraph) description

using language tags for multiple languages.

Examples (dbpedia:The_Glass_Palace)

```
<http://dbpedia.org/resource/The_Glass_Palace>
  rdfs:label     "The Glass Palace"@en ,
                 "Le Palais des miroirs"@fr ;

  rdfs:comment   "The Glass Palace is a 2000 historical novel..."@en ,
                 "Le Palais des miroirs est un roman..."@fr ;
```

→ **Worksheet #2: Task 4**

Subclasses

Defining a subclass

To define that **C1** (a class) is a subclass of **C2** (a class), write:

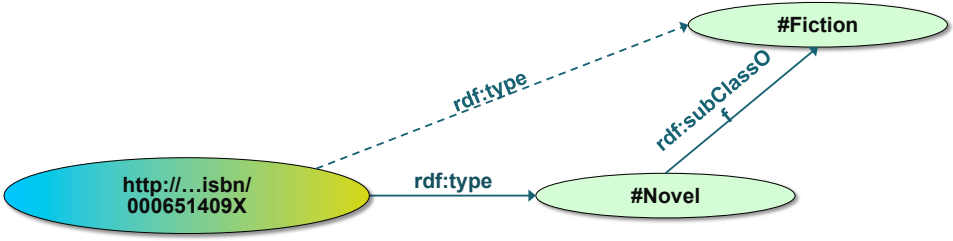
C1 `rdfs:subClassOf` ***C2***

Semantics

This states that all the instances of C1 are also instances of C2.
The `rdfs:subClassOf` property is transitive.

→ **Worksheet #2: Task 5**

Inferred properties




```
(<http://…/isbn/000651409X> rdf:type #Fiction)
```

- is not in the original RDF data…
- …but can be inferred from the RDFS rules
- RDFS environments return that triple, too

René Witte
Concordia University

Inference: let us be formal…

- The RDF Semantics document has a list of (33) entailment rules:
  - "if such and such triples are in the graph, add this and this"
  - do that recursively until the graph does not change
- The relevant rule for our example:

```
If:
  uuu rdfs:subClassOf xxx .
  vvv rdf:type uuu .
Then add:
  vvv rdf:type xxx .
```

→ **Worksheet #2: Task 6**

René Witte
Concordia University

Properties

René Witte

Example

`<studies at> <type> <Property>`

Defining a Property

To define that **P** (a resource) is a property, write:

P `rdf:type rdf:Property`

Properties are used to define relations between subject resources and object resources.

→ **Worksheet #2: Task 7**

Domain and Range

René Witte
Concordia University

Domain

To define a class **C** as the domain of a property **P**, write:

P `rdfs:domain` ***C***

This states that resources denoted by the subjects of triples whose predicate is **P** are instances of the class **C**.

Range

To define a class **C** as the range of a property **P**, write:

P `rdfs:range` ***C***

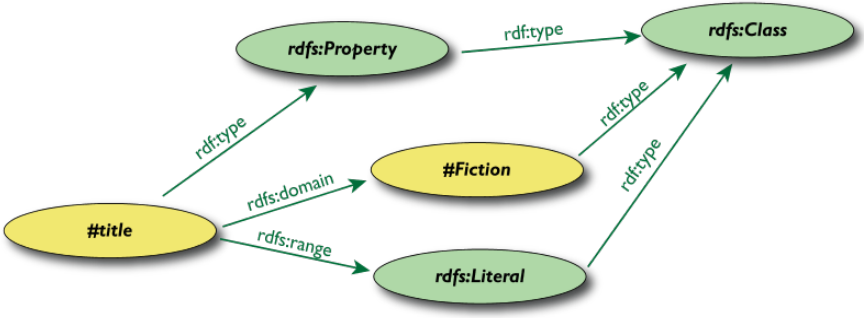
This states that the resources denoted by the objects of triples whose predicate is **P** are instances of the class **C**.

Note

- Properties are also resources (named with URIs)
- So we define properties of properties using. . . RDF properties!
- Again, you'll get used to it. . .

→ Worksheet #2: Task 8

Property specification example



René Witte
Concordia University

Property specification serialized

- In RDF/XML:

```
<rdf:Property rdf:ID="title">
  <rdfs:domain rdf:resource="#Fiction"/>
  <rdfs:range rdf:resource="http://...#Literal"/>
</rdf:Property>
```

- In Turtle:

```
:title
  rdf:type     rdf:Property;
  rdfs:domain  :Fiction;
  rdfs:range   rdfs:Literal.
```

René Witte

Subproperties

René Witte

Defining a Subproperty

To define that **P1** (a property) is a subproperty of **P2** (a property), write:

P1 `rdfs:subPropertyOf` ***P2***

With a subproperty, we can state that all resources related by one property are also related by another.

Example

Like inheritance for classes, we can have inheritance for properties:

`<is father of>` `<subPropertyOf>` `<is parent of>`

RDFS Utility Vocabulary

René Witte
Concordia University

Some "helper" constructs

rdfs:seeAlso a property that links a resource to another for more information (can be in any format)

rdfs:isDefinedBy a property typically used to refer to a vocabulary (RDF Schema) defining the subject IRI

There are also some datastructures (bag, list etc.) – read more before using!

RDF Schema Constructs: Summary

Construct	Syntactic form	Description
Class (a class)	**C** `rdf:type rdfs:Class`	**C** (a resource) is an RDF class
Property (a class)	**P** `rdf:type rdf:Property`	**P** (a resource) is an RDF property
type (a property)	**I** `rdf:type` **C**	**I** (a resource) is an instance of **C** (a class)
subClassOf (a property)	**C1** `rdfs:subClassOf` **C2**	**C1** (a class) is a subclass of **C2** (a class)
subPropertyOf (a property)	**P1** `rdfs:subPropertyOf` **P2**	**P1** (a property) is a sub-property of **P2** (a property)
domain (a property)	**P** `rdfs:domain` **C**	domain of **P** (a property) is **C** (a class)
range (a property)	**P** `rdfs:range` **C**	range of **P** (a property) is **C** (a class)

René Witte

Outline

René Witte

Vocabularies

René Witte
Concordia University

Goal: Knowledge Integration

Two major principles:

1. Reuse of vocabularies
   E.g., always use FOAF to describe names, emails, etc., instead of making up your own schema
2. Make your data self-describing
   Embed metadata using RDF to ensure data can be understood and processed independently.

Adhering to these principles supports interoperability and semantic understanding across different systems.

Self-Describing Datasets

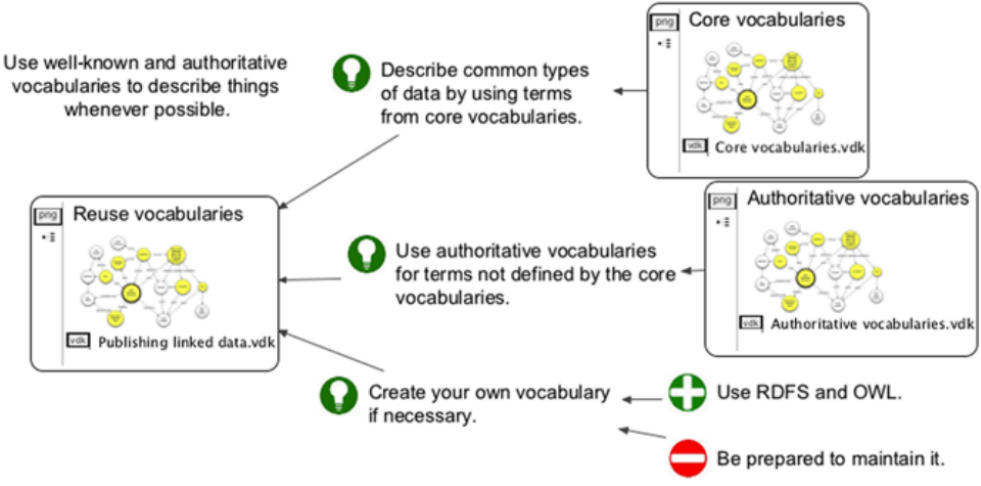
René Witte

Provide useful information about new terms

For example, if we create our own proprietary term, like SmallMediumEnterprise, we could describe it as [HB11]:

```
1  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
3  @prefix owl: <http://www.w3.org/2002/07/owl#> .
4  @prefix co: <http://biglynx.co.uk/vocab/sme#> .
5
6  <http://biglynx.co.uk/vocab/sme#SmallMediumEnterprise>
7    rdf:type rdfs:Class ;
8    rdfs:label "Small or Medium-sized Enterprise" ;
9    rdfs:subClassOf <http://dbpedia.org/ontology/Company> .
10   rdfs:subClassOf <http://umbel.org/umbel/sc/Business> ;
11   rdfs:subClassOf <http://sw.opencyc.org/concept/Mx4rvVjQNpwpEbGdrcN5Y29ycA> ;
12   rdfs:subClassOf <http://rdf.freebase.com/ns/m/0qb7t> .
```

Reuse vocabularies whenever possible



René Witte

Core Vocabularies

René Witte

FOAF (Friend-of-a-Friend) Vocabulary

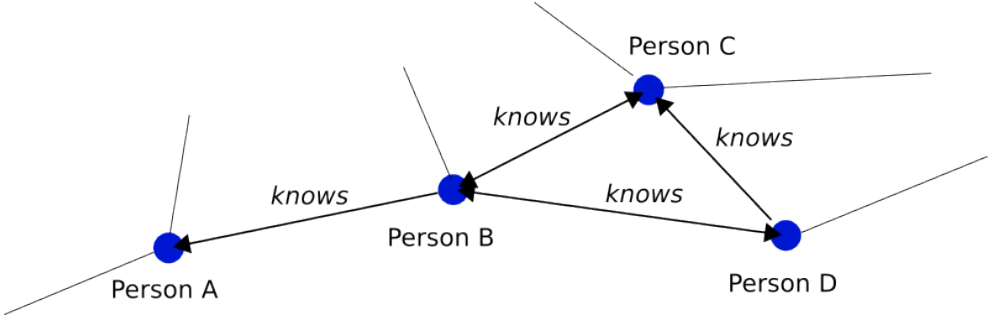
René Witte
Concordia University

FOAF

Model people and their connections in a social network.

```
@prefix foaf: <http://xmlns.com/foaf/0.1/>.

<http://example.org/joe> a foaf:Person ;
    foaf:name "Joe Doe" ;
    foaf:mbox <mailto:joe.doe@example.com> .
```



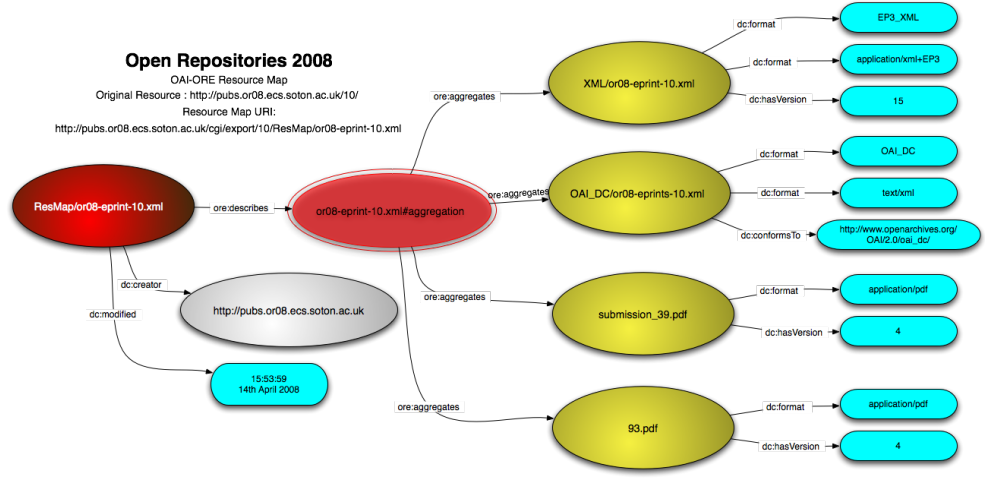
→ Worksheet #2: Task 9

Dublin Core

René Witte

Dublin Core® Metadata Initiative

Making it easier to find information.



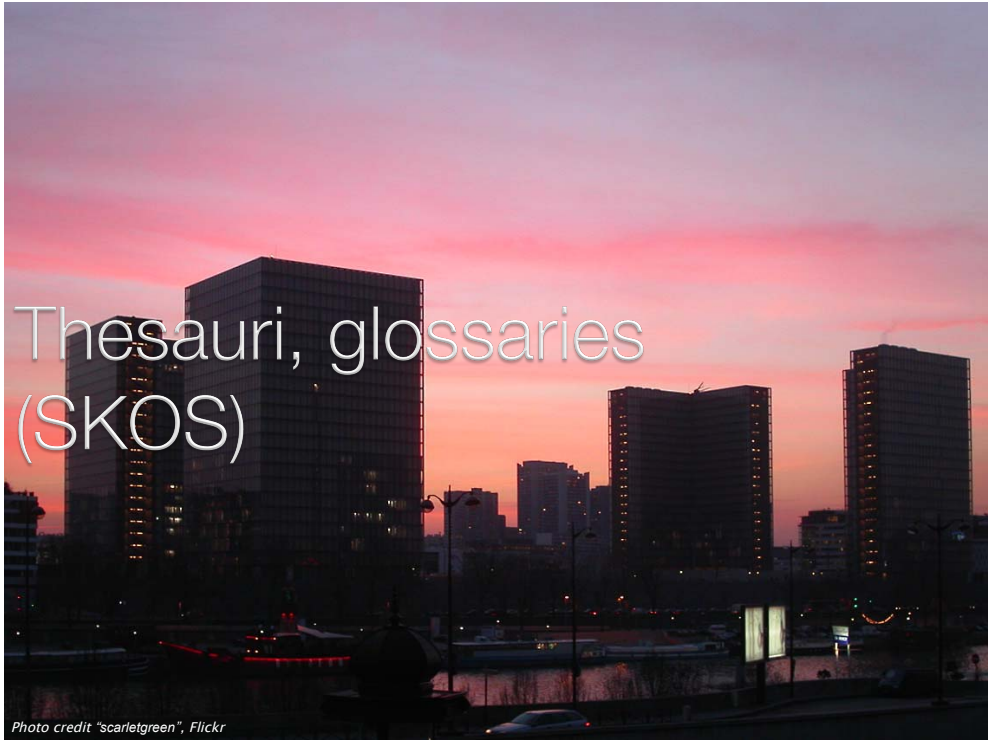


Photo credit "scarletgreen", Flickr

René Witte

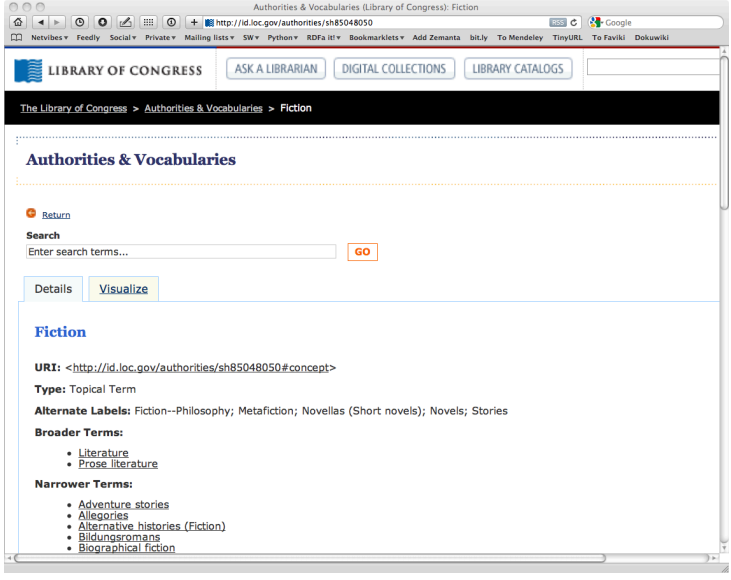
Concordia University

SKOS

- Represent and share classifications, glossaries, thesauri, etc
  - for example:
    - Dewey Decimal Classification, Art and Architecture Thesaurus, ACM classification of keywords and terms…
    - classification/formalization of Web 2.0 type tags
- Define classes and properties to add those structures to an RDF universe
  - allow for a quick port of this traditional data, combine it with other data

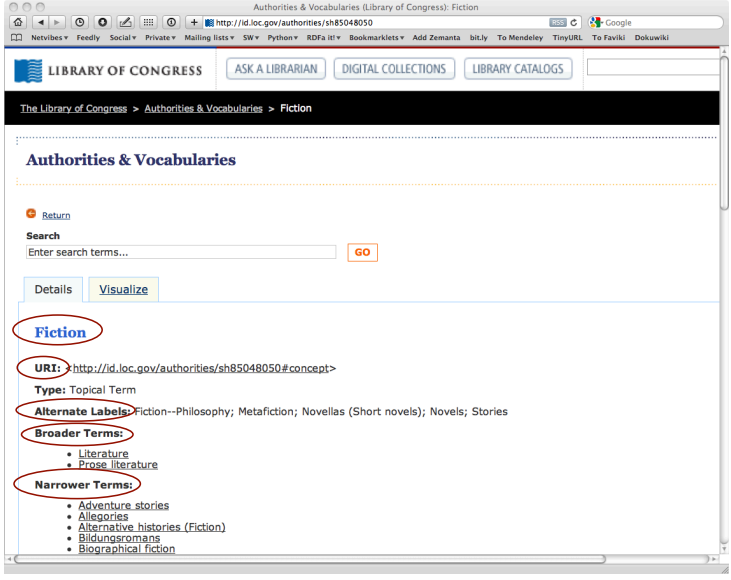
René Witte
Concordia University

Example: the term “Fiction”, as defined by the Library of Congress



René Witte
Concordia University

Example: the term “Fiction”, as defined by the Library of Congress



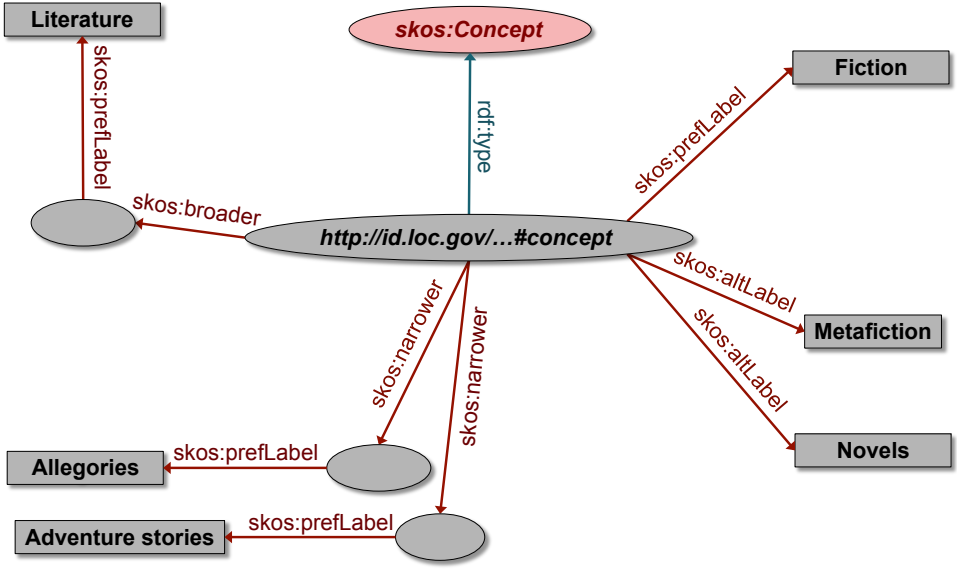
René Witte
Concordia University

Thesauri have identical structures…

- The structure of the LOC page is fairly typical
  - label, alternate label, narrower, broader, …
  - there is even an ISO standard for these
- SKOS provides a basic structure to create an RDF representation of these

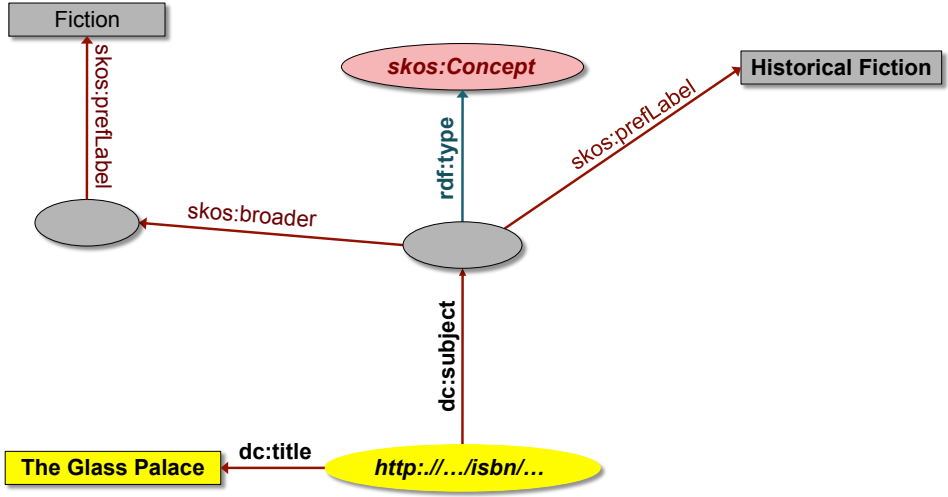
René Witte
Concordia University

LOC's "Fiction" in SKOS/RDF



René Witte
Concordia University

Usage of the LOC graph



René Witte
Concordia University

Same serialized

```
<http://…/isbn/000651409X>
    dc:title "The Glass Palace"@en;
    dc:subject <http://id.loc.gov/authorities/sh85061165#concept>;
    ...

<http://id.loc.gov/authorities/sh85061165#concept>
    a    skos:Concept;
    skos:prefLabel "Historical Fiction"@en;
    skos:broader <http://id.loc.gov/authorities/sh85048050#concept>;
    ...

<http://id.loc.gov/authorities/sh85048050#concept>
    a skos:Concept;
    skos:prefLabel "Fiction"@en;
    skos:narrower <http://id.loc.gov/authorities/sh85061165#concept>;
    ...
```

René Witte
Concordia University

SKOS terms overview

- Classes and Properties:
  - Basic description (Concept, ConceptScheme,…)
  - Labeling (prefLabel, altLabel,…)
  - Documentation (definition, historyNote,…)
  - Semantic relations (broader, narrower, related,…)
  - Collections (Collection, OrderedCollection,…)
  - Concept mappings (broadMatch, narrowMatch,…)

René Witte
Concordia University

Importance of SKOS

- SKOS provides a simple bridge between the “print world” and the (Semantic) Web
- Thesauri, glossaries, etc, from the library community can be made available
  - LOC is a good example
- SKOS can also be used to organize, eg, tags, annotate other vocabularies, …

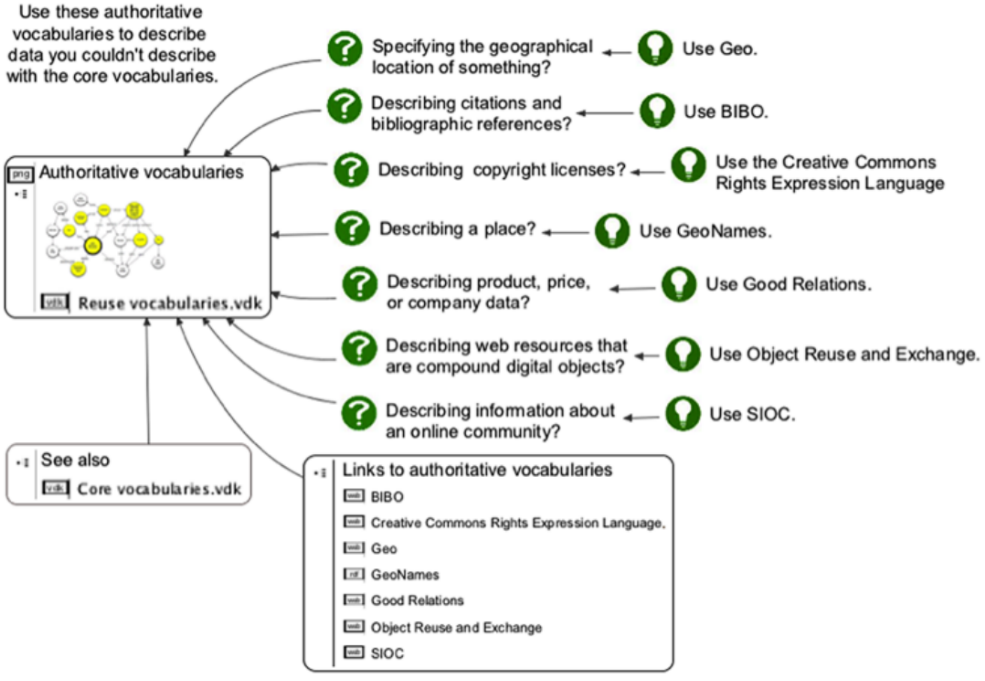
René Witte
Concordia University

Importance of SKOS

- Anybody in the World can refer to common concepts
  - they mean the same for everybody
- Applications may exploit the relationships among concepts
  - eg, SPARQL queries may be issued on the library data+LOC

René Witte
Concordia University

More authoritative vocabularies



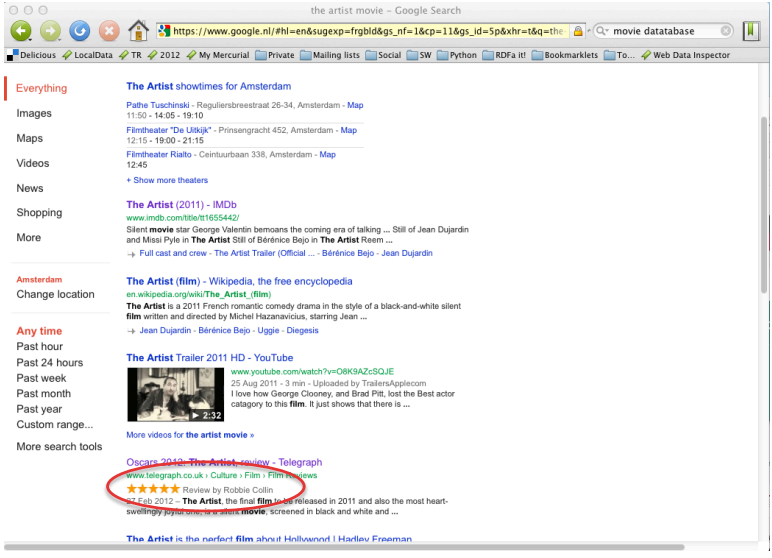
René Witte

Outline

René Witte

Concordia University

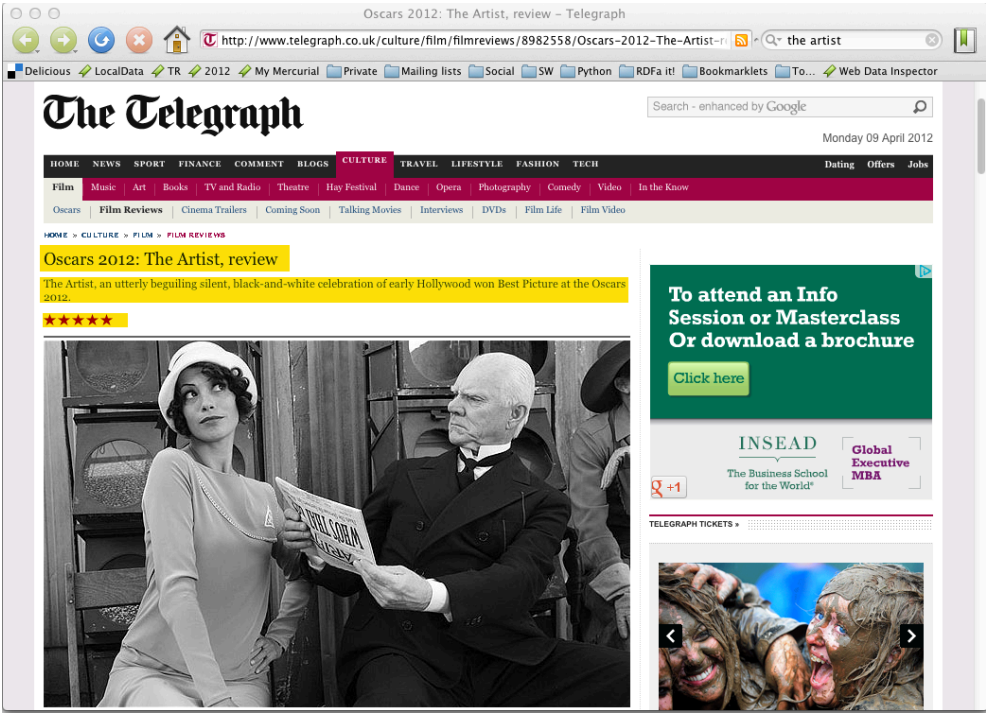
Typical usage of structured data



René Witte
Concordia University

René Witte

Concordia University



René Witte

Concordia University

René Witte

Concordia University

Source of http://www.telegraph.co.uk/culture/film/filmreviews/8982558/Oscars-2012-The-Artist-review.html

```
                        <li class="first"><a href="/">Home</a><span>&raquo;</span></li>
                        <li><a href="http://www.telegraph.co.uk/culture/">Culture</a><span>&raquo;</span></li>
                                <li><a href="http://www.telegraph.co.uk/culture/film/">Film</a><span>&raquo;</span></li>
                                <li class="styleSix"><a href="http://www.telegraph.co.uk/culture/film/filmreviews/">Film R
                <div class="cl"></div>
        </div>
</div>

<!-- googleon: all -->
<div id="tmglBody" >
        <div class="access"><a name="article"></a></div>

        <div class="twoThirdsThird2 gutterUnder">
                <div class="twoThirds gutter"  itemscope itemtype="http://schema.org/Review">
                        <div class="storyHead">

        <h1 itemprop="name">Oscars 2012: The Artist, review</h1>
                <h2 itemprop="description">
The Artist, an utterly beguiling silent, black-and-white celebration of early
  Hollywood won Best Picture at the Oscars 2012.

</h2>
        <div class="rating" itemprop="reviewRating" itemscope itemtype="http://schema.org/Rating">
                        <meta itemprop="worstRating" content = "0.5">
                        <meta itemprop="bestRating" content = "5">
                        <span itemprop="ratingValue" class="hidden">5</span>
                        <img src="/template/ver1-0/i/ratings/star_5_styleSix.png" alt="5 out of 5 stars" width="73" height="14" />
                </div>
        <div class="artIntro">
                                <div id="storyEmbSlide">
                        <div class="slideshow ssIntro">
                                <div class="nextPrevLayer">
                                        <div class="ssImg">
                                                <img src="http://i.telegraph.co.uk/multimedia/archive/0209
                                                <div class="artImageExtras" >

                                                        <div class="imgCaptionCredit">
                                                                <span class="caption">Bérénice Bejo as ris
                                                        </div>
                                                </div>
                                        </div>
                                </div>
                <div class="cl"></div>

        </div>
                        </div>
<div class="oneHalf gutter">
                        <div class="story">
                                <div class="cl"> </div>
                <!-- remove the whitespace added by escenic before end of </a> tag -->
```

René Witte
Concordia University

In a slightly more readable format…

```
<div itemscope itemtype="http://schema.org/Review">
    ...
  <h1 itemprop="name">Oscars 2012: The Artist, review</h1>
  <h2 itemprop="description">The Artist, an utterly beguiling…</h2>
  ...
  <span itemprop="ratingValue" class="hidden">5</span>
  ...
```

René Witte

Yielding…

```
[ rdf:type schema:Review ;
  schema:name "Oscars 2012: The Artist, review" ;
  schema:description "The Artist, an utterly beguiling…" ;
  schema:ratingValue "5" ;
  …
]
```

René Witte
Concordia University

Outline

René Witte

Concordia University

Reading Material

René Witte

Required

- [Yu14, Chapter 4] (RDFS)
- [Yu14, Chapter 7] (FOAF)

Supplemental

- [Wor14] (RDF Primer)
- [Yu14, Chapter 10] (Schema.org)
- [WZRH14, Chapters 2, 4] (RDF, FOAF)

References

[HB11] Tom Heath and Christian Bizer.
Linked Data: Evolving the Web into a Global Data Space.
Morgan & Claypool, 2011.
https://concordiauniversity.on.worldcat.org/oclc/704257552.

[Her] Ivan Herman.
Tutorial on Semantic Web Technologies.
http://www.w3.org/People/Ivan/CorePresentations/RDFTutorial/.

[Wor14] World Wide Web Consortium (W3C).
RDF 1.1 Primer.
http://www.w3.org/TR/rdf11-primer/, 24 June 2014.

[WZRH14] David Wood, Marsha Zaidman, Luke Ruth, and Michael Hausenblas.
Linked Data: Structured Data on the Web.
Manning, 2014.
https://concordiauniversity.on.worldcat.org/oclc/871683907.

[Yu14] Liyang Yu.
A Developer's Guide to the Semantic Web.
Springer-Verlag Berlin Heidelberg, 2nd edition, 2014.
https://concordiauniversity.on.worldcat.org/oclc/897466408.