CC6202-1 LA WEB DE DATOS PRIMAVERA 2016

Lecture 1: Introduction

Aidan Hogan aidhog@gmail.com

THE WEB

The Web is now 27 years old



Evolution of the Web



The Future of the Web?

ssitues and the states of the second second

What will the Web be like in 27 more years?

CC6202-1 LA WEB DE DATOS PRIMAVERA 2016

Lecture 1: Introduction

Aidan Hogan aidhog@gmail.com

Hint:

THE "SEMANTIC WEB"

The "Semantic Web"

... what is the "Semantic Web"?

Semantic Web?



semantic web

Google Search

I'm Feeling Lucky

Ŷ

Semantic Web?



"The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for - Berners-Lee e Semantic Web" ricun. 284(5):34–43.

The Semantic Web is Tim Berners-Lee's vision of the future Web?



The Semantic Web is ...

... intelligent machines making sense of the Web?



The Semantic Web is Linking together Open Data?



Image by Cyganiak & Jentzsch ; <u>http://lod-cloud.net</u>

The Semantic Web is naming everything?



"Now! ... That should clear up a few things around here!"

The Semantic Web is ...

... breaking down the walls between sites?



The Semantic Web is Web 3.0?



Image by Radar Networks; Nova Spivak; http://memebox.com/futureblogger/show/824

The Semantic Web is an active area of research in Comp. Sci.?





Not pictured:



The Semantic Web is something that must die?



Tom Coates; See talk: <u>https://vimeo.com/21829271</u>

The Semantic Web is a Buzzword?







The Semantic Web is...



...cf. "What is the Web?"

HF

WHAT'S WRONG WITH THE CURRENT WEB?

The current Web is fantastic!



... but can the Web be better?

The current Web is document-centric







The current Web is document-centric



```
<html>
<body>
<h1>Ireland</h1>
<div class="flag"><img src="flag.jpg" /></div>
Ireland is a European country.
Its capital is <a href="http://ex.org/Dublin">Dublin</a>.
...
</body>
</html>
```

(Most of it) Makes sense to humans



... assuming they speak the language



Even worse for machines



Even worse for machines



What machines on the Web can do



What machines on the Web can do



This (with some "tricks") works really well



About 107,000 results (0.36 seconds)

Google



Beethoven - Symphony No. 5 in C Minor (1) - YouTube www.youtube.com/watch?v=W2qW6fOtAMY •

| sometimes wi | nen i mi | | |
|-----------------------------|----------------------------------|-------------------|---|
| sometimes w | nen i'm alone i us e | e comic sans | |
| sometimes wi | nen i'm alone i god | ogle myself | |
| sometimes wi | hen i'm <mark>alone i cry</mark> | • | |
| sometimes wi | hen i'm all alone | | |
| sometimes when i'm dreaming | | | |
| sometimes wi | hen i'm sad i like t | o cut myself | 1 |
| sometimes wi | nen i'm dreaming | lyrics | |
| sometimes w | nen i'm alone | | |
| sometimes wi | nen i'm driving on | the road at night | |
| sometimes wi | nen i'm alone i wo | nder | |
| | | | |
| | Google Search | I'm Feeling Lucky | |

Can even get "direct answers" now



THE WEB IS GREAT WHAT'S THE PROBLEM ...

Horses were great too ...



At its core, Google is still just doing ... (... but really *really* well) http://ex.org/h ħ ħ M, ≈ ೮ Υ. ତ ∂⁰ E <mark>୪</mark>. Google 0 ireland Ŷ **Google Search** I'm Feeling Lucky http://ex.org/Ireland ?

Google does not change the fact that ...


PROBLEM WITH CURRENT WEB: LITERATURE VETERANS

Doing a report for university ...



Wants to find *all*:

- <u>Nobel Prize winners in Literature</u>
- Who fought in at least one war
- The year they won the prize
- And the year the war(s) started



... how would you solve this?



Lots of Wikipedia Tabs ...



The Laureate Algorithm (27 years on)



results := Ø

```
for all nobel-lit-winner in wiki-list
    year-prize := year of nobel-lit-winner
    wars := search "war", "conflict", "battle" in nobel-lit-winner
    for all war in wars
        year-war := starting year of war
        add nobel-lit-winner, year-prize, war, year-war to results
    end
end
return results
```

The Laureate Algorithm (54 years on?)



results := Ø

for all nobel-lit-winner-and-wars in <u>magical-sem-web-results()</u> add nobel-lit-winner-and-wars to results

end

return results



PROBLEM WITH CURRENT WEB: FINDING A DESSERT

Going to dinner party ...



Looking for a recipe:

- Must be a <u>dessert</u>
- Must be <u>citrus-free</u>
- Must be gluten-free
- Ingredients available in <u>local</u> supermarket(s)
- <u>Cheap</u> (student)
- Must be <u>delicious</u>





Using the Current Web ...



The Dessert Algorithm (27 years on)

```
candidates := \mathbf{0}
for all recipe-site in google-results
   for all dessert-recipe in recipe-site
      if dessert-recipe type looks-delicious
         suitable := true
         for all ingredient in dessert-recipe
            if searchNutrition(ingredient) type <u>wheat</u> or <u>lemon</u> or <u>lime</u> ...
               suitable := false
            else if searchShops(ingredient) type <u>null</u> or <u>expensive</u>
               suitable := false
            end
         end
         if suitable add dessert-recipe to candidates end
      end
   end
end
return candidates
```



The Dessert Algorithm (54 years on?)



candidates := Ø

for all safe-cheap-local-dessert-recipe in <u>magical-sem-web-results()</u> if safe-cheap-local-dessert-recipe type looks-delicious add safe-cheap-local-dessert-recipe to candidates end end

end

return candidates



The Dessert Algorithm (54 years on?)



for all safe-cheap-local-dessert-recipe in magical-sem-web-results()



BEING MORE EXPLICIT ON STEP 2

Let's look at this recipe ...







Searle's Chinese Room: B214



... what should the output be?

Оитрит: "ð х"

Multiple symbols, one meaning ...



One symbol, multiple meanings ...





Multiple ways to say the same thing ...

. . .

. . .

. . .

. . .

Dublin's population is one million. Dublin has a population of one million. Dublin's population is 1,000,000. Dublin has 1,000,000 inhabitants. One million people live in Dublin. [Dublin] Its population is one million. La población de Dublín es un millón.





Multiple meanings for the same saying ...

Sherlock saw the man using binoculars.









BRIDGING THE SEMANTIC GAP ON THE WEB

Searle's Chinese Room





One symbol, one meaning ...



One symbol, <u>one</u> meaning ...





One (simple) way to say one thing ...

Dublin's population is one million. Dublin has a population of one million. Dublin's population is 1,000,000. Dublin has 1,000,000 inhabitants. One million people live in Dublin. [Dublin] Its population is one million. La población de Dublín es un millón. (Dublin,population,1000000)







INPUT: " $(\hbar, \sigma^{O}, x), (x, \sigma, y)$ "



... what might the output be?

OUTPUT: $\{(x \mapsto \mathcal{Z}, y \mapsto 100000)\}$

INPUT: "(Ireland, capital, x), (x, population, y)"



OUTPUT: {($x \mapsto \text{Dublin}, y \mapsto 100000$)}

```
INPUT: "(x, \odot, y)"
```



... what might the output be?

OUTPUT: $\{(x \mapsto h, y \mapsto \uparrow)\}$?

INPUT: "(*x*,partOf,*y*)"



 $(x \mapsto \mathsf{Dublin}, y \mapsto \mathsf{Ireland})$?

OUTPUT: {($x \mapsto$ Ireland, $y \mapsto$ Europe)}

INPUT: " (x, \odot, y) "



OUTPUT: $\{(x \mapsto h, y \mapsto \uparrow), (x \mapsto J, y \mapsto h)\}$

INPUT: "(*x*,partOf,*y*)"



RULES: $(a, capital, b) \rightarrow (b, partOf, a)$

OUTPUT: {($x \mapsto$ Ireland, $y \mapsto$ Europe), ($x \mapsto$ Dublin, $y \mapsto$ Ireland)}

INPUT: " (x, \odot, y) "



OUTPUT: $\{(x \mapsto h, y \mapsto \uparrow), (x \mapsto F, y \mapsto h), (x \mapsto F, y \mapsto \uparrow)\}$

INPUT: "(*x*,partOf,*y*)"


(1) Data, (2) Query, (3) Rules/Ontologies

INPUT: "(*x*,partOf,*y*)"



THE SEMANTIC WEB

The Semantic Web is now about 16 years old



... so where is it then?

Hidden within the Web ... let's have a look



GOOGLE'S RICH SNIPPETS

Fancy-looking search results ...

Publishers add structured data!!

```
itemtype="http://schema.org/NutritionInformation">

<span class="nutrition_label">kcal</span>
<span class="nutrition_value" itemprop="calories">480</span>
[...]
```

Publishers get more clicks on their results
 Google gets data to make fancy results

schema.org (Bing, Google, Yahoo!, Yandex)

schema.rdfs.org

JSON

Home Learn Tools Mappings FAQ About

What is Schema.RDFS.org?

In early June 2011, the three big search engines Bing, Google and Yahoo! introduced Schema.org, a collection of terms that webmasters can use to markup their pages to improve the display of search results. This site is a complementary effort by people from the Linked Data community to support Schema.org deployment and usage with a special focus on Linked Data:

- We provide markup examples and tutorials about publishing & consuming data with Schema.org terms.
- We maintain mappings from Web Data vocabularies such as the DBpedia ontology to Schema.org terms.
- · We list tools and libraries that are able to consume or produce Schema.org-based data.
- We automatically scrape the Schema.org terms on a daily basis and generate the following formats:

Note that the official OWL version of the terms is directly maintained at Schema.org and independent from the above presented formats. Various tools that use or produce Schema.org terms are already available or in preparation. Stay tuned!

GOOGLE'S KNOWLEDGE GRAPH

Google's info-box

Sully Prudhomme

Poet

René François Armand Prudhomme was a French poet and essayist. He was the first ever winner of the Nobel Prize in Literature in 1901 Wikipedia

Born: March 16, 1839, Paris, France

Died: September 6, 1907, Châtenay-Malabry, France

Books: Les vaines tendresses

Awards: Nobel Prize in Literature

People also search for

Theodor

Leconte de Lisle

Frédéric Mommsen Mistral

Paul Verlaine

View 10+ more

Gabriel Fauré

Google's Semantic Search

| G wife of william golding | g- C X | Aidan 🗖 🗖 🗙 | | | | | |
|--------------------------------------|--|------------------|--|--|--|--|--|
| $\leftarrow \rightarrow C^{\bullet}$ | s://www.google.cl/?gfe_rd=cr&ei=AfHMV9-YCMqnxgT356mQDw&gws_rd=ssl#q=wife+of+willia | m+g 🔍 🗣 🏠 🔥 🔳 | | | | | |
| Aplicaciones 🗶 Lingu | Jee 🕒 SGICM 🎦 Apache Any23: Anythi | Otros marcadores | | | | | |
| Google | Google wife of william golding | | | | | | |
| | All Images Videos News More - Search tools | 1 | | | | | |
| | About 388,000 results (0.50 seconds) | | | | | | |
| | William Golding / Spouse | | | | | | |
| | Ann Brookfield m. 1939–1993 | | | | | | |
| | Feedback | | | | | | |
| | People also ask | | | | | | |
| 4 | Who is William Golding? | | | | | | |

Using Semantic Web knowledge-bases

FACEBOOK OPEN GRAPH PROTOCOL

On Facebook ...

Facebook want to know what pages linked in comments are about

The latest trailer for "Game of Thrones" is out, and contains so many fantastic details that we don't even know where to begin. Take a look: http://imdb.to/1KXLPFJ

İ

"Game of Thrones": The Wheel "I'm not going to stop the wheel. I'm going to break the wheel."

Like · Comment · Share · Send · 🖒 6,469 🖓 395 🖒 1,381

Elsewhere on the Web ...

Facebook wants to know what you like on other sites ... is it a movie, a recipe ...

Facebook Open Graph Protocol

DATA-FYING WIKIPEDIA: WIKIDATA

What is Wikidata?

Problem 1: Different language versions manually edited by users

Alexis Sánchez

| Full name | Alexis Alejandro Sánchez Sánchez ^[1] | | | |
|----------------|--|--|--|--|
| Date of birth | 19 December 1988 (age 26) ^{[1][2]} | | | |
| Place of birth | Tocopilla, Chile ^{[3][2]} | | | |

| | National team [‡] | | |
|-------|----------------------------|----|------|
| 2007 | Chile U20 | 12 | (2) |
| 2006– | Chile | 82 | (26) |

Nombre completo Alexis Alejandro Sánchez Sánchez

| Carrera internacional | | | | | | | |
|-----------------------|---------|--|--|--|--|--|--|
| Selección | Chile | | | | | | |
| Part. (goles) | 82 (26) | | | | | | |
| Debut | 2006 | | | | | | |

Alexis Sánchez im Dezember 2011

Spielerinformationen

Voller Name Alexis Alejandro Sánchez SánchezGeburtstag 19. Dezember 1988Geburtsort Tocopilla, Chile

| Nationalmannschaft ² | | | | | |
|---------------------------------|------------|---------|--|--|--|
| 2007 | Chile U-20 | | | | |
| 2006- | Chile | 76 (25) | | | |

Problem 2: Complex lists of things manually edited by users

Not logged in Talk Contributions Create account Log in
 Article Talk
 Read Edit View history Search Q
 Chile national football team
 From Wikipedia, the free encyclopedia

Most capped players [edit]

As of September 1, 2016 Players in **bold** are still active, at least at club level.

Top goalscorers [edit]

As of September 1, 2016 Players in **bold** are still active, at least at club level.

| # | Name | International Career | Caps | Goa | ıls# | Name | International Career | Goals | Caps |
|----|-------------------|-------------------------|------|-----|------|-------------------|-------------------------|-------|------|
| 1. | Claudio Bravo | 2004 — | 106 | 0 | 1. | Marcelo Salas | 1994–2007 | 37 | 70 |
| 2. | Alexis Sánchez | 2006 – | 102 | 34 | 1 | lván Zamorano | 1987–2001 | 34 | 69 |
| 3. | Gary Medel | 2007 – | 96 | 7 | 2. | Alexis | | | |
| 4. | Gonzalo Jara | 2006 – | 95 | 3 | | Sánchez (list) | 2006 – | 34 | 102 |
| | | | | | | Educanda | | | |

Alexis scores a goal ...

different languages, lists, ...

Solution: Wikidata

• Collaboratively edit structured data in one place, with multi-lingual labels

Wikidata Activity

- 42k contributors 5k contributors with 5+ edits in Jan 2014
- Well over 100M edits so far up to 500k per day

Wikidata facts about Abraham Lincoln

Abraham Lincoln (Q91)

American politician, 16th President of the United States in office from 1861 to 1865 Abe Lincoln | Lincoln | Honest Abe

In more languages

Query Service ...

Query Service ...

| and the second second | | | Aidan 🗖 🗖 🗙 | | | | | |
|---|---|----------------------------------|-----------------------|--|--|--|--|--|
| III Wikidata Query Service × | sectors and the local division of the local | the state of a local data | and it is not a | | | | | |
| ← → C 🔒 https://query.wikidata.c | org/#PREFIX%20wd%3A%20%3Ch | http%3A%2F%2Fwww.wikidata.org%2F | entity%2F%3Q 🗟 🟠 🚺 🛢 | | | | | |
| 🛗 Aplicaciones 🚺 Linguee 🗅 SGICM 🕒 Apache Any23: Anythi | | | | | | | | |
| 10 :War Wul:PSo0 :WarStart . # Get date the war started 17 BIND(YEAR(?warStart) as ?warYear) # Get year from date 18 ?laureate rdfs:label ?laureateName . # Get name of laureate 19 FILTER(lang(?warName)="en" # Filter for English 20 && lang(?laureateName)="en") # names only 21 } ORDER BY ?awardYear # Order by year 22 Press (CTPL-SPACEIto activate auto completion Deta undeted is a completion | | | | | | | | |
| Run Clear 12 Results in 1589 ms | | Display | - Ł Download- ⊗ Link- | | | | | |
| laureateName | awardYear 👙 | warName | warYear 👙 | | | | | |
| Carl Spitteler | 1919 | World War I | 1914 | | | | | |
| Winston Churchill | 1953 | World War I | 1914 | | | | | |
| Ernest Hemingway | 1954 | World War I | 1914 | | | | | |
| Ernest Hemingway | 1954 | World War II | 1939 | | | | | |
| Jean-Paul Sartre | 1964 | Algerian War | 1954 | | | | | |
| Jean-Paul Sartre | 1964 | World War II | 1939 | | | | | |
| Heinrich Böll | 1972 | World War II | 1939 | | | | | |
| Eugenio Montale | 1975 | World War I | 1914 | | | | | |
| William Golding | 1983 | World War II | 1939 | | | | | |
| Claude Simon | 1985 | Spanish Civil War | 1936 | | | | | |
| Camilo José Cela | 1989 | Spanish Civil War | 1936 | | | | | |
| Günter Grass | 1999 | World War II | 1939 | | | | | |

(Hidden slide)

- PREFIX wd: <http://www.wikidata.org/entity/>
- PREFIX wdt: <http://www.wikidata.org/prop/direct/>
- PREFIX p: <http://www.wikidata.org/prop/>
- PREFIX ps: <http://www.wikidata.org/prop/statement/>
- PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
- PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
- SELECT DISTINCT ?laureateName ?awardYear ?warName ?warYear
- WHERE {
- ?laureate p:P166 ?award . # Winner of some prize
- ?award ps:P166 wd:Q37922 . # Prize is Nobel Pr. in Lit.
- ?award pq:P585 ?awardDate . # Get the date of the award
- BIND(YEAR(?awardDate) as ?awardYear) # Get the year of the award
- ?laureate wdt:P607 ?war . # Find war(s) laureate was in
- ?war rdfs:label ?warName . # Get name(s) of war(s)
- ?war wdt:P580 ?warStart .

Get year the war started

- BIND(YEAR(?warStart) as ?warYear)
- ?laureate rdfs:label ?laureateName . # Get name of laureate
- FILTER(lang(?warName)="en" # Filter for English
- && lang(?laureateName)="en") # ... names only
- } ORDER BY ?awardYear # Order by year or award
- https://query.wikidata.org/

results := Ø

for all nobel-lit-winner-and-wars in <u>magical-sem-web-results()</u> add nobel-lit-winner-and-wars to results

end

return results

One day: automatically update articles

Not logged in Talk Contributions Create account Log in

 Article
 Talk
 Read
 Edit
 View history
 Search
 Q

 Chile national football team

 From Wikipedia, the free encyclopedia

Most capped players [edit]

As of September 1, 2016 Players in **bold** are still active, at least at club level.

Top goalscorers [edit]

As of September 1, 2016 Players in **bold** are still active, at least at club level.

| # | Name | International Career | Caps | Goa | ıls# | Name | International | Goals | Caps |
|----|-------------------|-------------------------|------|-----|---------|-------------------|---------------|-------|------|
| 1. | Claudio Bravo | 2004 – | 106 | 0 | 1. 龚 | Marco Salas | E. | 5. | 70 |
| 2. | Alexis Sánchez | 2006 – | 102 | 35 | | Iván Zamo | | Pr- | 69 |
| 3. | Gary Medel | 2007 – | 96 | 7 | 2. | Alexis | | | ł |
| 4. | Gonzalo Jara | 2006 – | 95 | 3 | | Sánchez (list) | 2006 – | 35 | 102 |
| | | | | | | Educardo | | | |

STANDARDISING OPEN DATA: LINKED OPEN DATA

How to publish Open Data?

5-Star Linking Open Data Scheme

The Linked Data Cloud

Geographical Data: Geonames

Linked Government Data: data.gov

Linked Government Data: data.gov.uk

Linked Government Data: datos.gob.cl

| datos. | gob.cl | | | ¿qué estás buscando? | Q |
|--------|--------------------------|-----------------------------|------|----------------------|------------------|
| INICIO | E CATĂLOGO | C VISUALIZACIONES | APPS | PIDE DATOS | ACERCA DEL SITIO |
| < A | PROXIMAME SANTIAGO, C | | | | |

CATÁLOGO

DATASETS PUBLICADOS: 1.232
Life Sciences



Life Sciences



Copyright | Privacy | Accessibility | Viewers & Players | Freedom of Information Act | USA.gov U.S. National Library of Medicine | U.S. National Institutes of Health | U.S. Department of Health & Human Services

Life Sciences



E-Commerce

GoodRelations The Web Vocabulary for E-Commerce



<<Prev Next>>

Main Quickstart Cookbook User's Guide Specification Tools Community

The most powerful Web vocabulary for e-commerce

A paradigm shift for e-commerce. Since 2008.

Only 5% of all potential visitors of your site will actually see your offers in their original beauty. 95% will never get beyond a reduced preview of your great products and services as provided by a Web search engine.

GoodRelations is the most powerful vocabulary for publishing all of the details of your products and services in a way friendly to search engines, mobile applications, and browser extensions. By adding a bit of extra code to your Web content, you make sure that potential customers realize all the great features and services and the benefits of doing business with you, because their computers can extract and present this information with ease.

News from Twitter

Video



Who uses GoodRelations?

Google Yahoo! BestBuy sears.com kmart.com ... and 10,000 more See here for additional references.

Contact

Univ.-Prof. Dr. Martin Hepp

License

The GoodRelations ontology is creative

Acknowledgments

Many organizations and individuals have supported

New York Times Meta-data



http://data.nytimes.com/schools/schools.html

BBC Music



BBC World Cup



BBC Olympics





THE WEB OF DATA



ABOUT THE COURSE ...

Topics Covered

- RDF (triple-based data model)
- RDFS/OWL (ontological languages)
- SPARQL (query language)
- Linked Data / Web of Data
- RDB2RDF (importing databases to Sem. Web.)
- Applications, research, discussion



Structure of the course

- Each week:
 - Class on Monday (learn concepts)
 - Lab on Wednesday (see concepts in practice)
- Marking structure:
 - 45% labs
 - 25% project / reading group (TBC)
 - 30% exam

Outcomes: Learn About the Semantic Web!

- An ongoing research topic here in the DCC
- Apply database, logic, AI, etc., to the Web!
- Mix of concepts and practical exercises
- See what the future Web (might) look(s) like!

