



PUBBLICARE LINKED DATA

Metodologie e strumenti

Maurizio Atzori

Dipartimento di Matematica e Informatica
Università di Cagliari

atzori@unica.it

SUMMER SCHOOL LDA
LIBRARIES IN THE DIGITAL AGE

29 Agosto - 1 Settembre 2016

Hello!



Maurizio Atzori



Ricercatore e Professore Aggregato presso
Dipartimento di Matematica e Informatica

Università di Cagliari

<http://atzori.webofcode.org/>

email: atzori@unica.it



Hello!

- Didattica: corso di *Programmazione 2*, Laurea triennale in Informatica
- Ricerca in area Graph Databases,

in collab. con



- Semantic Web
- Metodi di interrogazione di basi di dati semantiche
 - QA3
 - SWiPE



Google
Faculty Research Awards



Obiettivi del Corso

Imparare ad utilizzare i linked open data (LOD)

- Come si creano
- Come si pubblicano
- Come si interrogano



Obiettivi del corso: strumenti

Imparare ad utilizzare i linked open data (LOD)

- Come si creano
 - OpenRefine
- Come si pubblicano
 - Jena/Fuseki, Pubby/Jetty
- Come si interrogano
 - SPARQL, BEStQ/Swipe



Programma Day #1

comprendere i linked data e creare triple

- Intro a Linked Data
 - Web of Documents vs Web of (Linked) Data
 - grafi
 - RDF data (triple)
 - serializzazioni (formati)
 - Esempio DBpedia
 - data cleaning/reconciliation
- Esercitazione
 - come utilizzare OpenRefine
 - trasformare dati tabellari (bookstore.xls) in triple RDF
 - data cleaning/reconciliation



Materiale

- **Linked Data: Evolving the Web into a Global Data Space**

Autori: Tom Heath , Talis Christian Bizer ,
Freie Universität Berlin

<http://linkeddatabook.com/editions/1.0/>

- Materiale online:

<http://atzori.webofcode.org/lda2016>

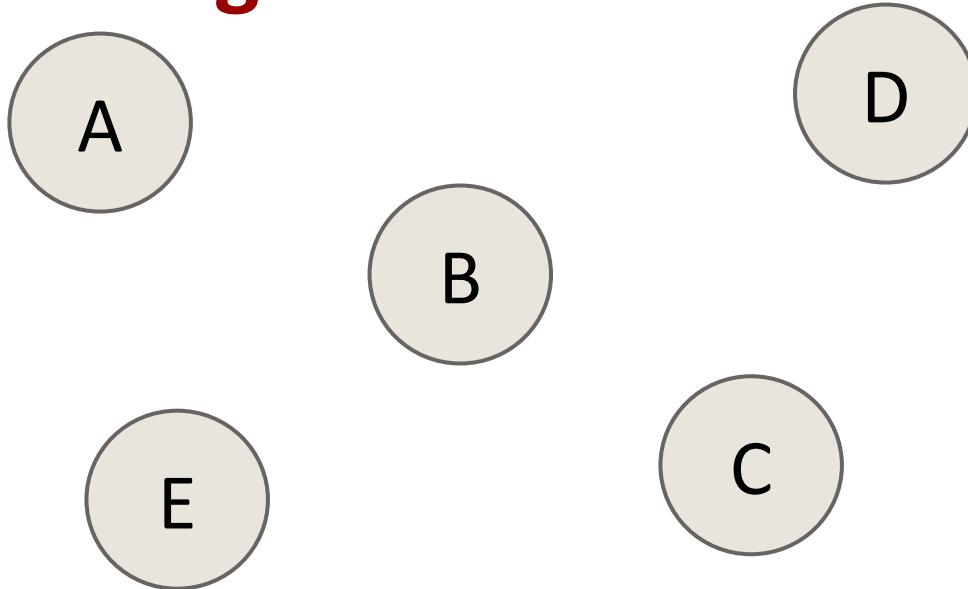


Intro sui grafi

- **Definizione:** un grafo e' un insieme finito non vuoto N ed un insieme di un copie A contenente coppie di elementi in N .
 - gli elementi in N vengono detti *nodi*
 - gli elementi in A vengono detti *archi*
- Esempio di grafo:
 - $N = \{ A, B, C, D \}$
 - $E = \{ (A,B) \ (A,D) \ (C,B) \ (C,D) \ (B,D) \ }$



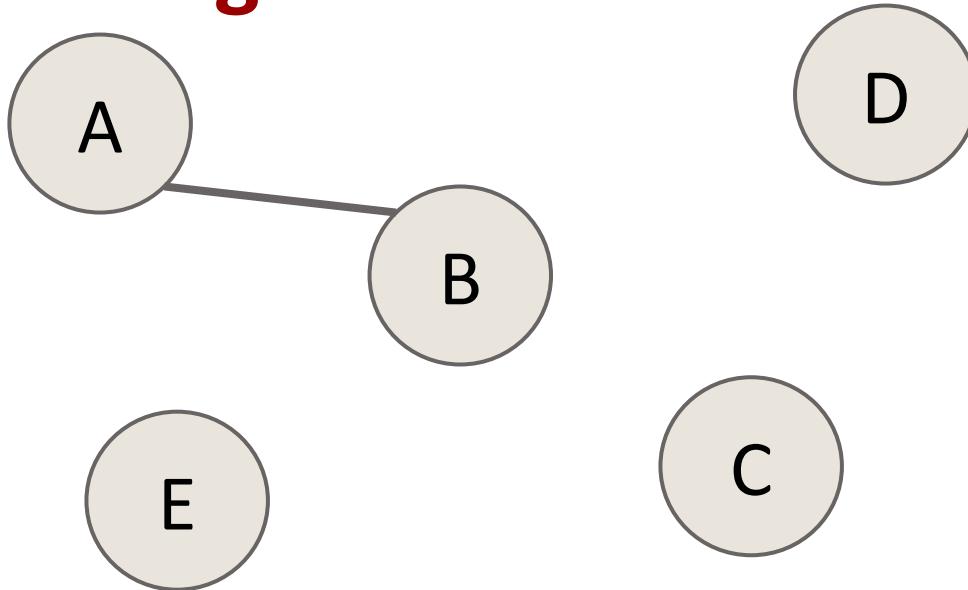
Intro sui grafi



- Esempio di grafo:
 - $N = \{ A, B, C, D, E \}$
 - $E = \{ (A,B) \quad (A,D) \quad (C,B) \quad (C,D) \quad (B,D) \quad (B, E) \}$



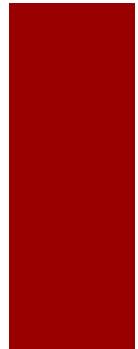
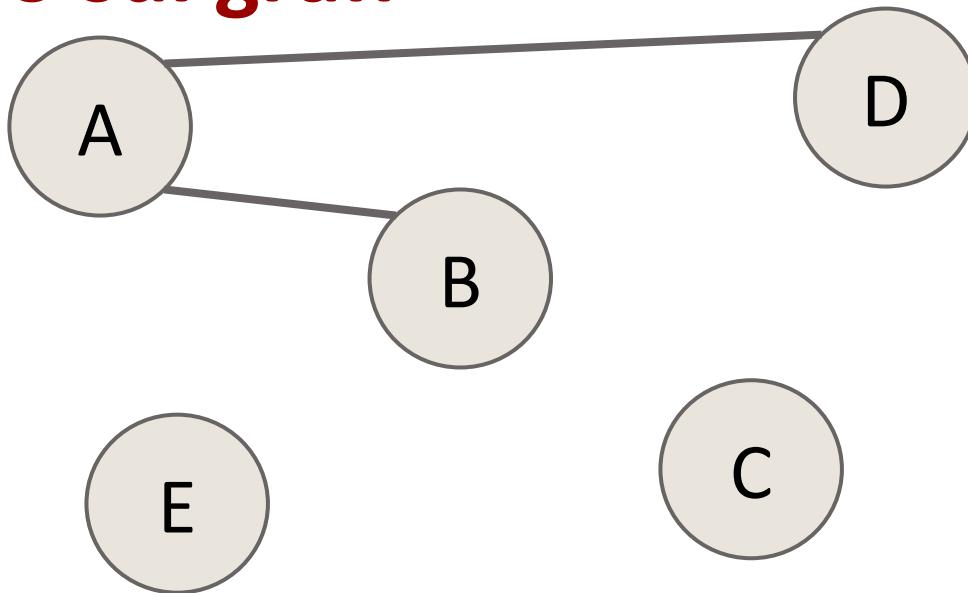
Intro sui grafi



- Esempio di grafo:
 - $N = \{ A, B, C, D, E \}$
 - $E = \{ (A,B) \quad (A,D) \quad (C,B) \quad (C,D) \quad (B,D) \quad (B, E) \}$



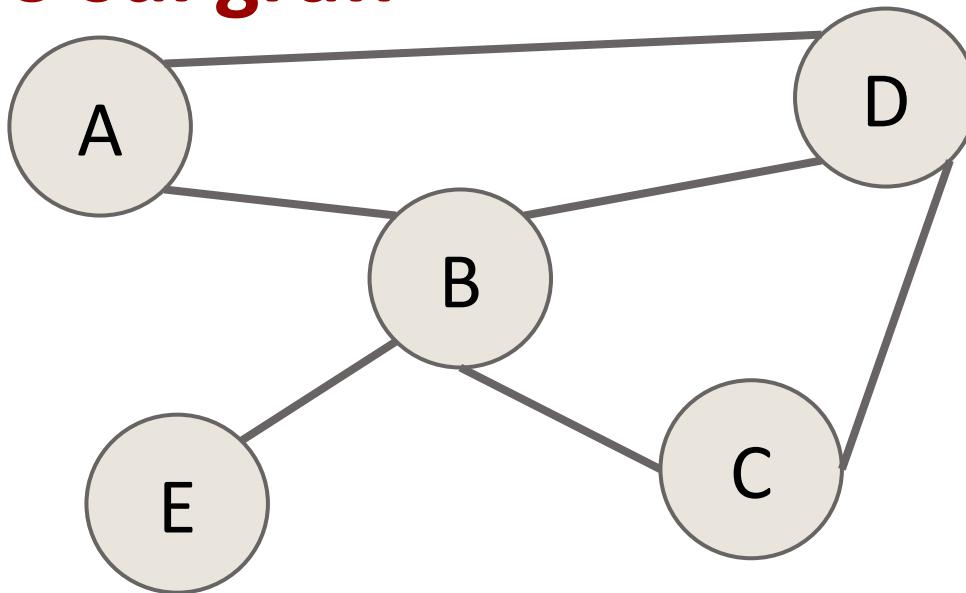
Intro sui grafi



- Esempio di grafo:
 - $N = \{ A, B, C, D, E \}$
 - $E = \{ (A,B) \quad (A,D) \quad (C,B) \quad (C,D) \quad (B,D) \quad (B, E) \}$



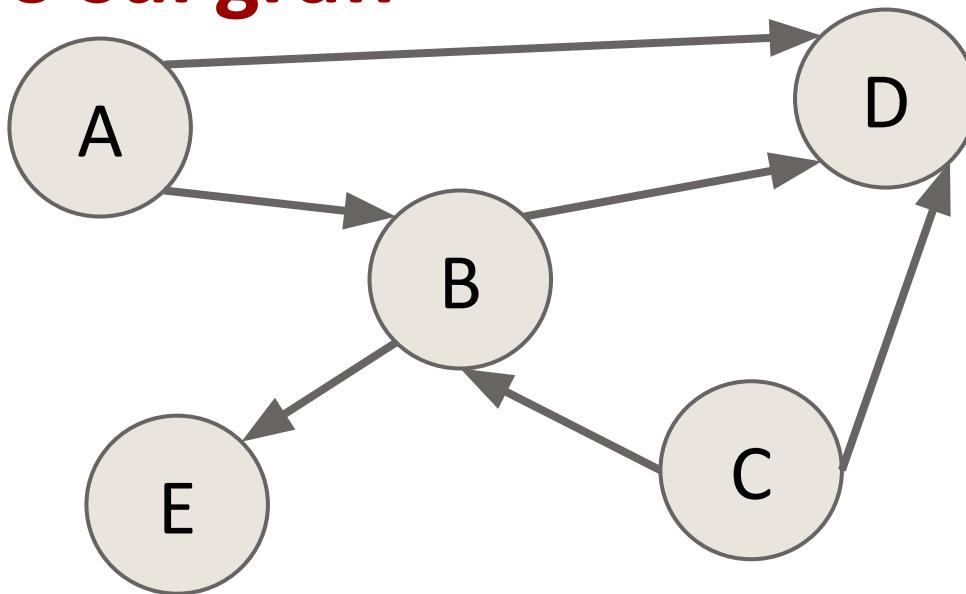
Intro sui grafi



- Esempio di grafo:
 - $N = \{ A, B, C, D, E \}$
 - $E = \{ (A,B) \quad (A,D) \quad (C,B) \quad (C,D) \quad (B,D) \quad (B, E) \}$



Intro sui grafi

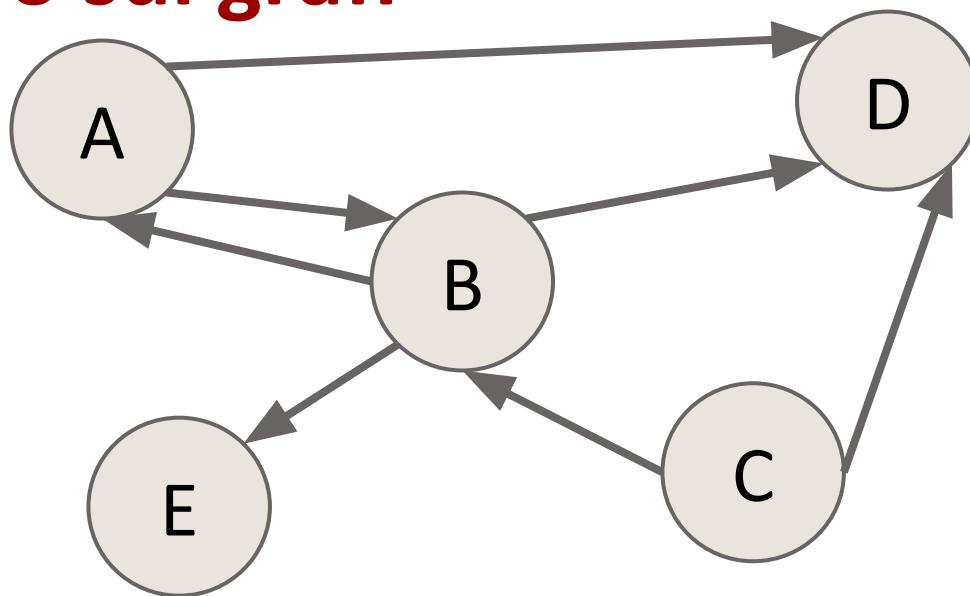


grafo direzionato

- Esempio di grafo:
 - $N = \{ A, B, C, D, E \}$
 - $E = \{ (A,B) \ (A,D) \ (C,B) \ (C,D) \ (B,D) \ (B, E) \}$



Intro sui grafi

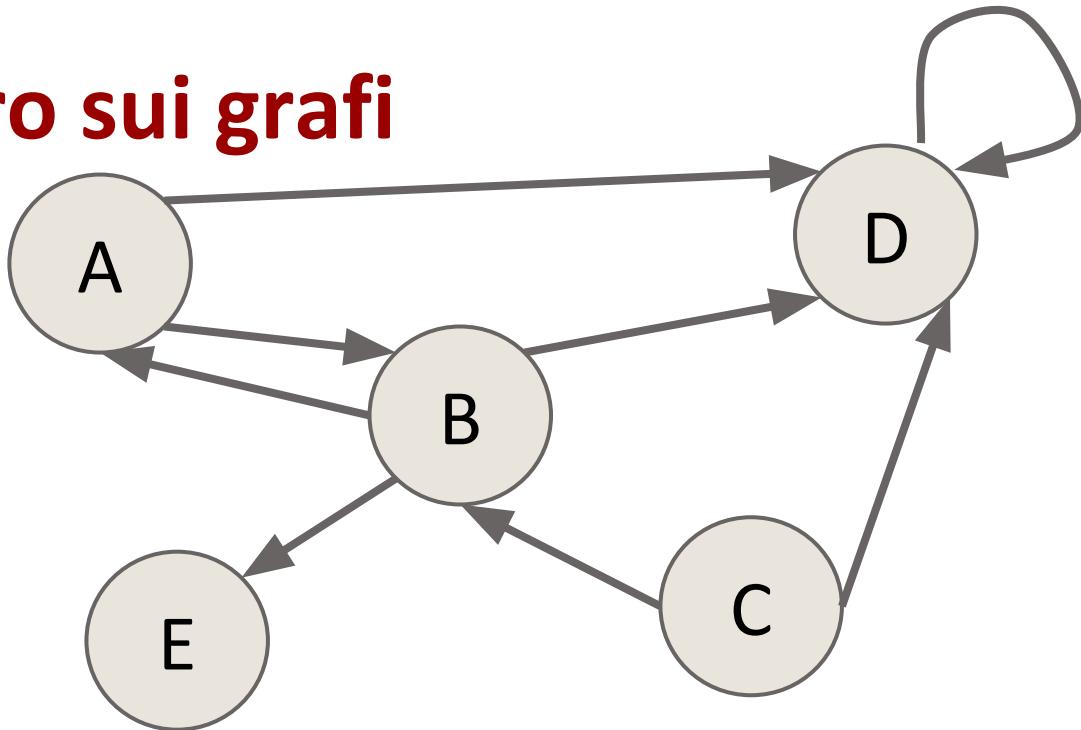


grafo direzionato

- Esempio di grafo:
 - $N = \{ A, B, C, D, E \}$
 - $E = \{ (A,B) \quad (B,A) \quad (A,D) \quad (C,B) \quad (C,D) \quad (B,D) \quad (B, E) \}$



Intro sui grafi



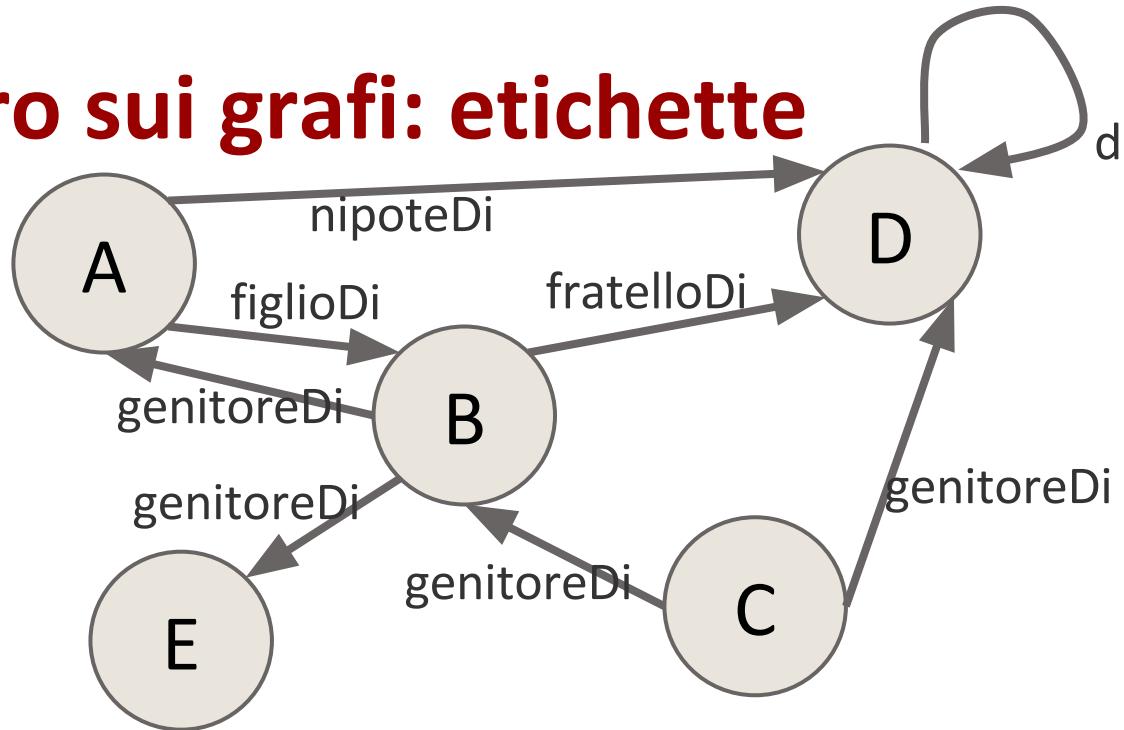
grafo direzionato

- Esempio di grafo:
 - $N = \{ A, B, C, D, E \}$
 - $E = \{ (A,B) \ (B,A) \ (A,D) \ (C,B) \ (C,D) \ (B,D) \ (B,E) \ (D,D) \ }$

Intro sui grafi: etichette



datoreDiLavoroDi



grafo direzionato

- Esempio di grafo:
 - $N = \{ A, B, C, D, E \}$
 - $E = \{ (A,B) \ (B,A) \ (A,D) \ (C,B) \ (C,D) \ (B,D) \ (B,E) \ (D,D) \ }$

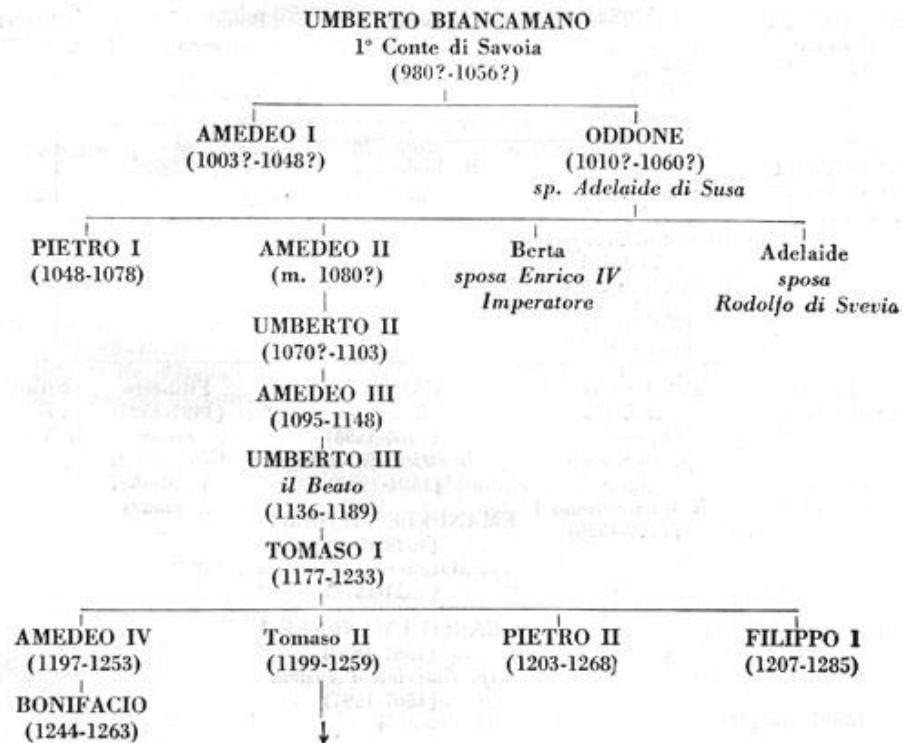


Grafo... cui prodest?

- I grafi sono una struttura di base della matematica discreta
- Hanno praticamente infinite applicazioni
 - l'albero genealogico è un grafo

un albero e' un grafo:

- con un solo arco entrante per nodo
- eccetto uno (la radice)





Grafo... cui prodest?

- I grafi sono una struttura di base della matematica discreta
- Hanno praticamente infinite applicazioni
 - l'albero genealogico è un grafo
 - chimica, biologia...

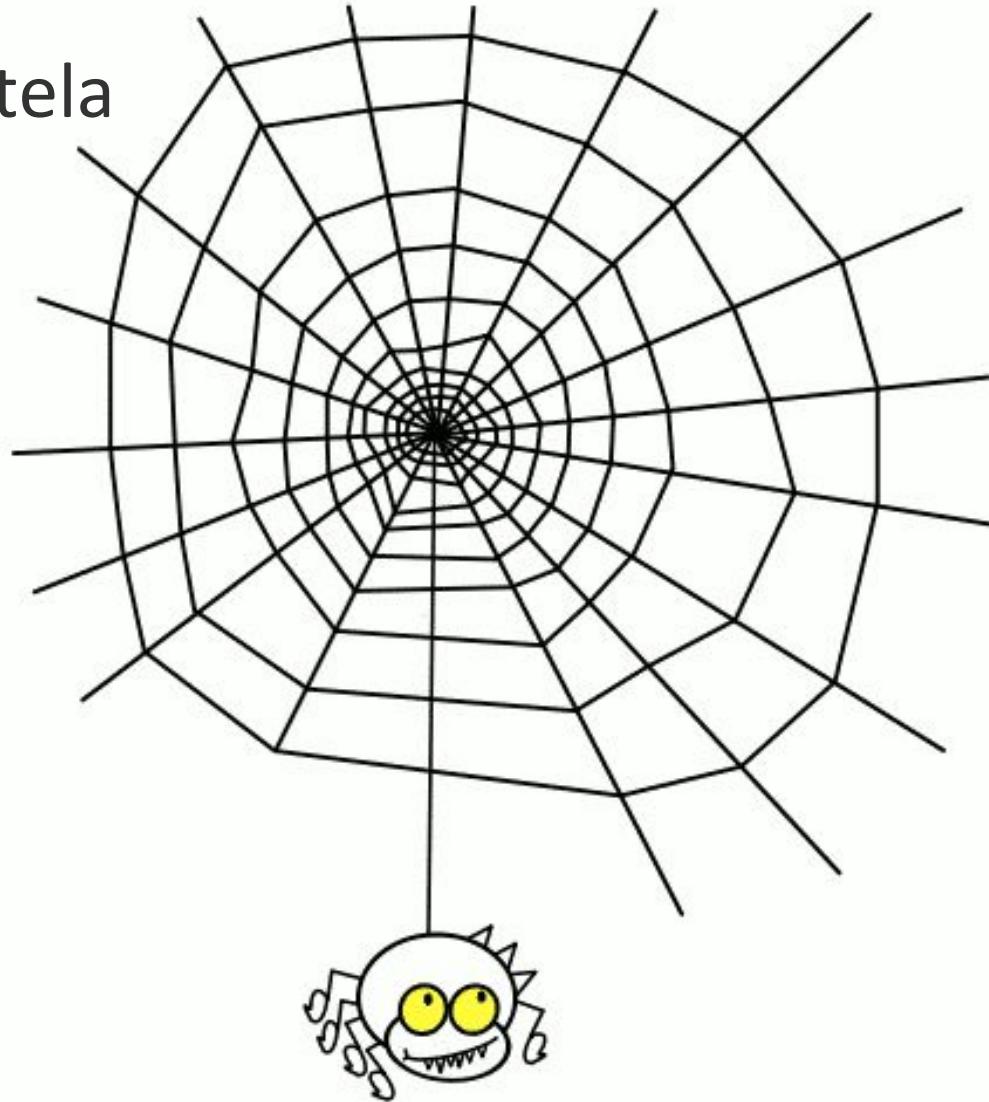
WEB



WEB



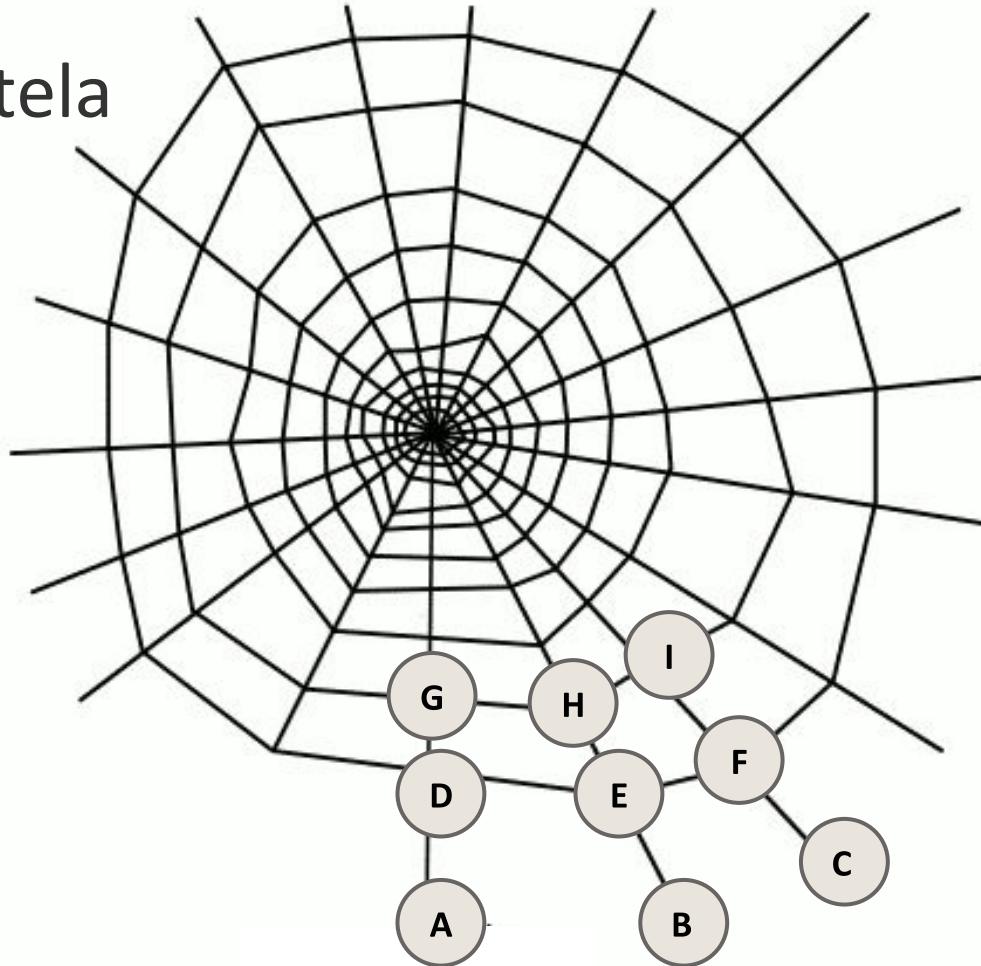
- Web = Ragnatela



WEB



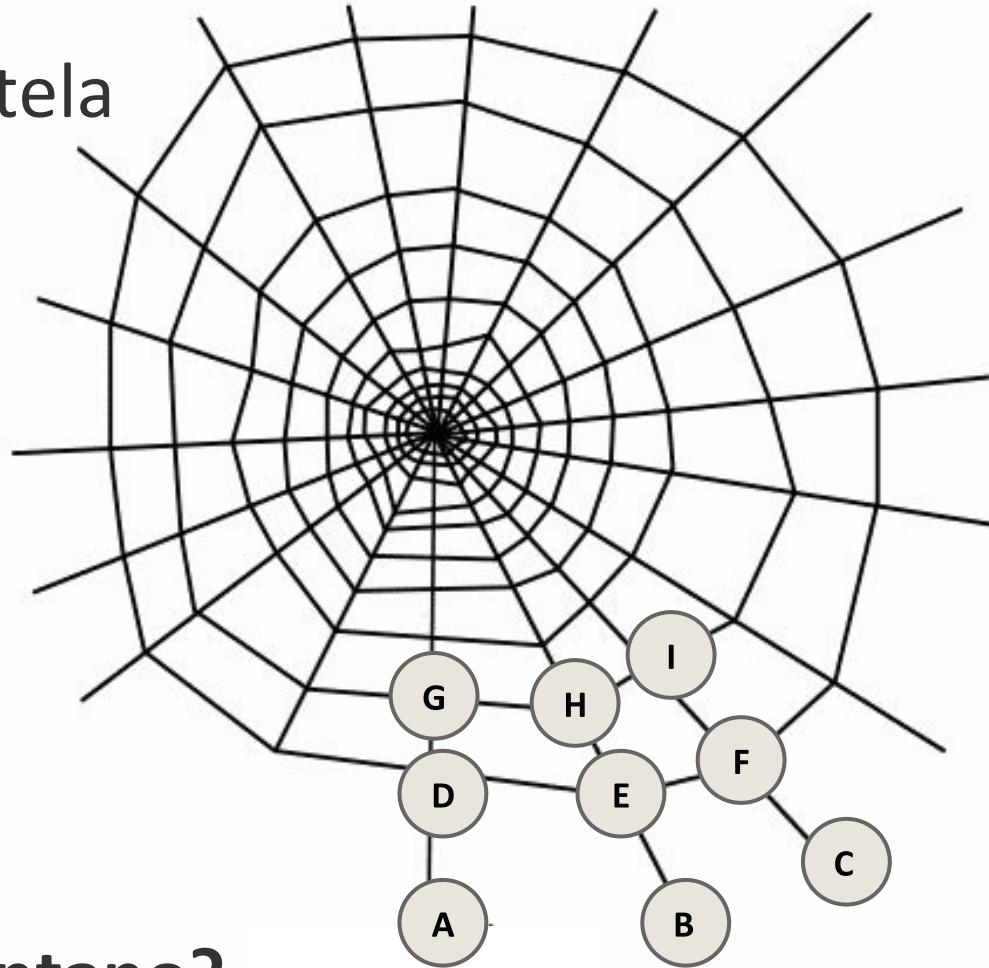
- Web = Ragnatela





WEB

- Web = Ragnatela
- Nodi e Archi



cosa rappresentano?



WEB

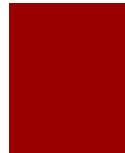
- World Wide Web: un grafo
 - i nodi sono le pagine di internet (URL)
 - un grafo enorme **>15 miliardi** di nodi
 - chiamato *webgraph*
 - gli archi sono determinati dagli hyperlink (i collegamenti, o link)





Un nodo del WWW

- Una pagina a caso



Voce Discussione

Leggi Modifica ▾

Ricerca



Stampo pilota

Da Wikipedia, l'enciclopedia libera.

Lo **stampo pilota** è una tipologia di **stampo** per realizzare oggetti in **plastica**, realizzato con materiali più facilmente e velocemente lavorabili (es. [alluminio](#)) anziché l'[acciaio](#).

Rispetto ad uno stampo tradizionale in acciaio, lo stampo pilota è uno stampo di "prova" più economico e di rapida realizzazione, con il solo scopo di stampare pochi pezzi (non avendo un sistema di raffreddamento si surriscalda rapidamente) e verificare la fattibilità di un medesimo stampo definitivo in acciaio.



Un nodo del WWW

http://it.wikipedia.org/wiki/Stampo_pilota

- Una pagina a caso

Voce [Discussione](#) Leggi [Modifica](#) ▾ Ricerca

Stampo pilota

Da Wikipedia, l'enciclopedia libera.

Lo **stampo pilota** è una tipologia di **stampo** per realizzare oggetti in **plastica**, realizzato con materiali più facilmente e velocemente lavorabili (es. **alluminio**) anziché l'**acciaio**.

Rispetto ad uno stampo tradizionale in acciaio, lo stampo pilota è uno stampo di "prova" più economico e di rapida realizzazione, con il solo scopo di stampare pochi pezzi (non avendo un sistema di raffreddamento si surriscalda rapidamente) e verificare la fattibilità di un medesimo stampo definitivo in acciaio.

- qual e' la stella uscente di questo nodo?

- (http://it.wikipedia.org/wiki/Stampo_pilota,
<http://it.wikipedia.org/wiki/Stampo>)
- (http://it.wikipedia.org/wiki/Stampo_pilota,
<http://it.wikipedia.org/wiki/Plastica>) ...



Un nodo del WWW

- Una pagina a caso

http://it.wikipedia.org/wiki/Stampo_pilota

- qual e' la stella entrante?

Voce [Discussione](#) Leggi [Modifica](#) ▾ Ricerca

Stampo pilota

Da Wikipedia, l'enciclopedia libera.

Lo **stampo pilota** è una tipologia di **stampo** per realizzare oggetti in **plastica**, realizzato con materiali più facilmente e velocemente lavorabili (es. **alluminio**) anziché l'**acciaio**.

Rispetto ad uno stampo tradizionale in acciaio, lo stampo pilota è uno stampo di "prova" più economico e di rapida realizzazione, con il solo scopo di stampare pochi pezzi (non avendo un sistema di raffreddamento si surriscalda rapidamente) e verificare la fattibilità di un medesimo stampo definitivo in acciaio.



Web e Internet

- Il Web è l'insieme dei documenti identificabili con una URL (<http://...>)
 - ciascuno è un ipertesto, ovvero un testo in cui possono essere presenti link ad altri documenti
 - le pagine (documenti) sono nodi, i link gli archi
- Internet è una rete di computer
 - i nodi sono i computer (inclusi router, ecc)
 - i computer sono collegati, formando una rete (grafo)

Web of Documents and Web of Data



- Il **Web of Documents** è l'insieme dei *documenti* identificabili con una URL (<http://...>)
 - ciascuno e' un ipertesto, ovvero un testo in cui possono essere presenti link ad altri documenti
 - le pagine (documenti) sono nodi, i link gli archi
- Il **Web of Data (o Linked Data)** è l'insieme dei *dati* identificabili con una URL (<http://...>)
 - ciascuno e' un pezzo di informazione, ovvero un testo in cui possono essere presenti link ad altri documenti
 - le entità sono nodi, le proprietà gli archi

Tim Berners-Lee, inventore di Web e Semantic Web: “Linked Data is the Semantic Web done right”

Semantic Web

- It is about giving semantics to information in the Web
 - In the Web, URLs correspond to resources (*mostly html pages*)
 - This forms a big network among pages (*through links*)
- In the Semantic Web a URL indicates a concept.
 - a URL may represent an **entity** or a **property**

Semantic Web: an apple

(examples of Entities)

- an apple
 - <<http://dbpedia.org/resource/Apple>>
 - <<http://www.wikidata.org/entity/Q89>>
 - <<http://www.dbpedialite.org/things/18978754>>
- another kind of "Apple"
 - <http://dbpedia.org/resource/Apple_Inc.>
 - <<http://www.wikidata.org/entity/Q312>>
 - <<http://www.dbpedialite.org/things/856>>
- no room for ambiguity
- we can define prefixes to shorten urls:
 - dbpedia:Apple
 - wd:Q89



Semantic Web: not only entities (*examples of Properties, aka Attributes*)

- How much proteinic is something?
 - it is relevant to "apple" (Q89), but also other foods
 - <<http://dbpedia.org/property/protein>>
 - is it relevant to the other "apple" (Q312)?
- Properties can be themselves entities.
 - What is the range, unit measure, the english word for attribute <<http://dbpedia.org/property/protein>> ?

Resource Description Framework (RDF)

- Entities and properties are linked together through RDF "data format"
- Multigraph-like structure made of triples
 - subject, predicate, object

not a graph (not even multigraph) in the CS sense (edge labels are also nodes in RDF)

dbpedia:Alan_Turing dbpedia-owl:owner dbpedia:Apple.
~~dbpedia:Alan_Turing dbpedia-owl:owner dbpedia:Apple_Inc.~~
dbpedia:Alan_Turing dbpedia-owl:field dbpedia:Computer_science.



An example of RDF dataset: DBpedia

<http://dbpedia.org/resource/Apple> (Dereferencing a URL)

About: Apple

An Entity of Type : [species](#), from Named Graph : <http://dbpedia.org>, within Data Space : dbpedia.org

| | |
|-----------------------|---|
| dbpprop:note | ▪ http://ndb.nal.usda.gov/ndb/search/list?ql lookup=090 |
| dbpprop:opt1n | ▪ dbpedia:Fluoride |
| dbpprop:opt1v | ▪ 3.300000 (xsd:double) |
| dbpprop:ordo | ▪ dbpedia:Rosales |
| dbpprop:pantothenicMg | ▪ 0.061000 (xsd:double) |
| dbpprop:phosphorusMg | ▪ 11 (xsd:integer) |
| dbpprop:potassiumMg | ▪ 107 (xsd:integer) |
| dbpprop:protein | ▪ 0.26 |
| dbpprop:q | ▪ Apples |
| dbpprop:regnum | ▪ Plantae |
| dbpprop:riboflavinMg | ▪ 0.026000 (xsd:double) |
| dbpprop:s | ▪ 1911 (xsd:integer) |
| dbpprop:sign | ▪ dbpedia:Plato |
| dbpprop:sodiumMg | ▪ 1 (xsd:integer) |

How is DBpedia content generated?

Featured content
Current events
Random article

Robbie Williams

- mainly by exploiting "structured" data in Wikipedia page infoboxes

to perform both as a member of Take That and as a solo artist.

Williams has sold more than 77 million albums and singles worldwide.^[5] He is the best-selling British solo artist in the United Kingdom and the best selling non-Latino artist in Latin America. Six of his albums are among the top 100 biggest-selling albums in the United Kingdom, and in 2006 he entered the *Guinness Book of World Records* for selling 1.6 million tickets of his *Close Encounters Tour* in a single day.^[6] He has also been honoured with seventeen BRIT Awards—more than any other artist—and eight ECHO Awards. In 2004, he was inducted into the UK Music Hall of Fame after being voted as the "Greatest Artist of the 1990s." Williams is married to actress [Ayda Field](#). He has a net worth of £130 million (2014).^[7]

Williams has currently (based on BPI certifications) sold a minimum of 18.9 million albums and 5.4 million singles in the UK alone as a solo artist.^[8] Williams also topped the 2000–2010 UK airplay chart, racking up almost 50% more plays than the Sugababes at number 2. In 2014 he was awarded the freedom of Stoke on Trent, as well as having a tourist trail created and streets named in his honour.

Contents [\[hide\]](#)

1 Early life

2 Take That

2.1 1990–1995: First stint

Robbie Williams



Williams filming the [music video](#) for "Candy" at Old Spitalfields Market, in August 2012

Background information

| | |
|--------------|--|
| Birth name | Robert Peter Williams |
| Born | 13 February 1974 (age 40) Stoke-on-Trent, Staffordshire, England |
| Genres | Pop rock, soft rock, dance, Britpop (early solo work) |
| Occupations | Singer-songwriter, musician, record producer, actor |
| Instruments | Vocals, guitar, bass guitar, synthesiser, drums, piano, violin, harmonica, marimba |
| Years active | 1990–present |
| Labels | Island, Universal (2011–present) Virgin (2009–2010) |

Infobox

| Background information | |
|------------------------|--|
| Birth name | Robert Peter Williams |
| Born | 13 February 1974 (age 38) Stoke-on-Trent, Staffordshire, England |
| Genres | Pop rock, soft rock, dance |
| Occupations | Singer-songwriter, musician, vocal coach, record producer, actor |
| Instruments | Vocals, guitar, bass guitar, synthesiser, drums, piano, violin, harmonica, marimba |
| Years active | 1990–present |
| Labels | Universal (2011–present) Virgin (2009–2010) Chrysalis (1996–2008) |
| Associated acts | Take That |
| Website | robbiewilliams.com |

Markup Source

```
 {{Infobox musical artist
|name      = Robbie Williams
|image     = Robbie Williams at Sunderland
           2011a crop.jpg
|caption   = Williams performing as part of
           the [[Progress Live]] tour in
           May 2011
|Img_size  =
|background =
|birth_name =
|Born      =
|instrument =
|genre     =
|occupation = Singer-songwriter, musician, vocal
               coach, record producer, actor
|years_active = 1990–present
|label     = [[Universal Music Group|Universal]]
|associated_acts = [[Take That]]
|website   = {{url|robbiewilliams.com}}
}}
```

Infobox

| Background information | |
|------------------------|--|
| Birth name | Robert Peter Williams |
| Born | 13 February 1974 (age 38) Stoke-on-Trent, Staffordshire, England |
| Genres | Pop rock, soft rock, dance |
| Occupations | Singer-songwriter, musician, vocal coach, record producer, actor |
| Instruments | Vocals, guitar, bass guitar, synthesiser, drums, piano, violin, harmonica, marimba |
| Years active | 1990–present |
| Labels | Universal (2011–present) Virgin (2009–2010) Chrysalis (1996–2008) |
| Associated acts | Take That |
| Website | robbiewilliams.com |

Markup Source

```
 {{Infobox musical artist
|name          = Robbie Williams
|image         = Robbie Williams at Sunderland
               2011a crop.jpg
|caption       = Williams performing as part of
               the [[Progress Live]] tour in
               May 2011
|Img_size      =
|background    =
|birth_name    = Robert Peter Williams
|Born          = {{Birth date and age|1974|2|13|
               df=yes}}<br />[[Stoke-on-Trent]],
               [[Staffordshire]], [[England]]
|Instrument   = Vocals, guitar, bass guitar,
               synthesiser, drums, piano, violin,
               harmonica, marimba
|genre         = [[Pop rock]], [[soft rock]],
               [[Dance music|dance]]
|occupation   = Singer-songwriter, musician, vocal
               coach, record producer, actor
|years_active = 1990–present
|label         = [[Universal Music Group|Universal]]
|associated_acts = [[Take That]]
|website       = {{url|robbiewilliams.com}}
}}
```

DBpedia

A Nucleus for a Web of Open Data (ISWC 2007,
Semantic Web Journal 2014)

- Univ of Leipzig, Univ of Mannheim, OpenLink SW
- 583 million “facts” in terms of RDF triples (en14)
 - 3 billions is the union of the 125 localized versions
- 4.58 million entities (en14)
- some triples imported from other datasets
- Introduces an ontology
 - e.g. :Actor :subClassOf :Person

WikiData: Cagliari (city)

| | | | |
|---------------|---|----------|-------------------|
| population | <p>149,883</p> <p>point in time 9 October 2011</p> <p>determination method census</p> <p>criterion used questionnaire</p> <p>legal population of Italy</p> <p>1 reference</p> | [edit] | Editable by Users |
| Time changing | <p>164,249</p> <p>point in time 21 October 2001</p> <p>determination method census</p> <p>criterion used questionnaire</p> <p>legal population of Italy</p> <p>1 reference</p> | [edit] | |
| | <p>149,038</p> <p>point in time 30 September 2013</p> <p>determination method demographic balance</p> <p>1 reference</p> | [edit] | Provenance |

SPARQL Protocol and RDF Query Language

Give me the proteins that an apple has.

- endpoint: <http://dbpedia.org/sparql>
- query (proteins quantity of an apple):

```
PREFIX dbpedia: <http://dbpedia.org/resource/>
PREFIX dbpprop: <http://dbpedia.org/property/>
```

```
SELECT ?proteins WHERE {
    dbpedia:Apple dbpprop:protein ?proteins.
}
```

- graph pattern matching
 - variables can appear in the triple in any point

SPARQL: People who were born in Berlin before 1900

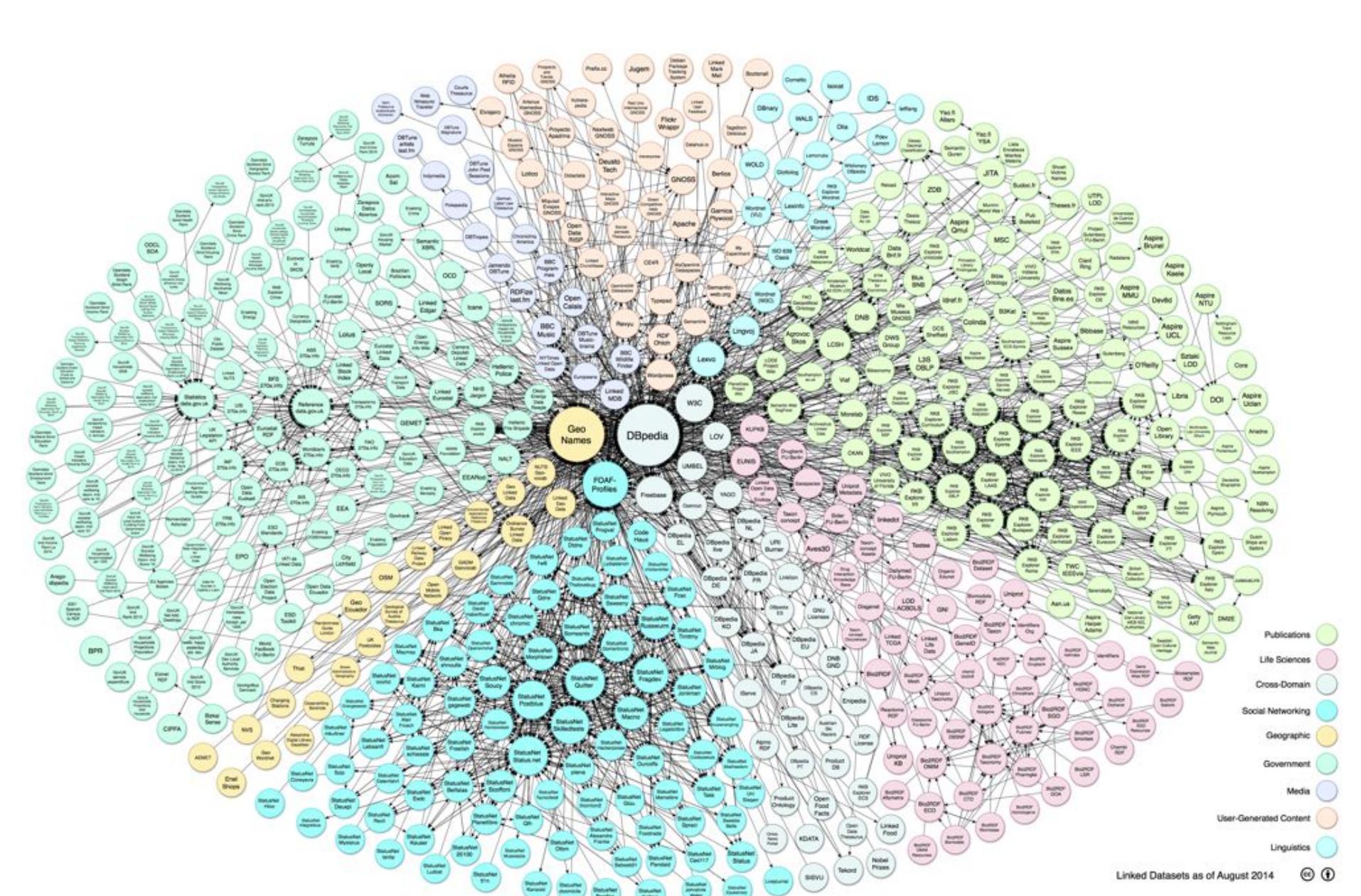
```
PREFIX dbo: <http://dbpedia.org/ontology/>

SELECT ?name ?birth ?death ?person WHERE {
    ?person dbo:birthPlace :Berlin .
    ?person dbo:birthDate ?birth .
    ?person foaf:name ?name .
    ?person dbo:deathDate ?death .
    FILTER (?birth < "1900-01-01"^^xsd:date) .
}
ORDER BY ?name
```

SPARQL 1.1 also supports features such as aggregates, having, group by, built-in and user-defined functions, path properties, unions, intersections, difference, etc.

Moltissimi dati disponibili come Linked Data

- Do we have data in a structured form?
 - DBpedia (information from Wikipedia)
 - Wikidata (Google-funded crowdsourced structured data)
 - Musicbrainz/DBtune (encyclopedia of music)
 - SIDER (<http://sideeffects.embl.de>)
 - Diseasome
(<http://wifo5-03.informatik.uni-mannheim.de/diseasome/>)
 - Drugbank (<http://www.drugbank.ca>)
 - plenty of other sources
 - **Yes we do have structured data!**

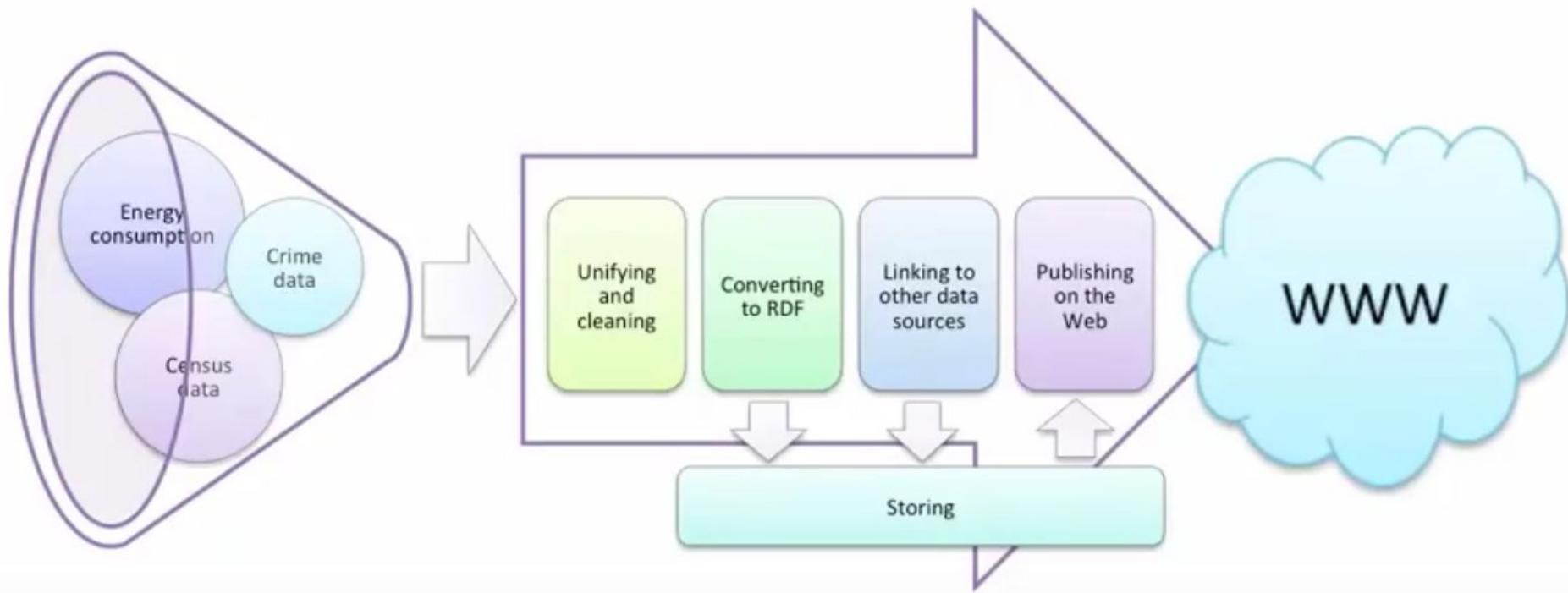
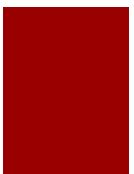


<http://lod-cloud.net/versions/2014-08-30/lod-cloud.svg>

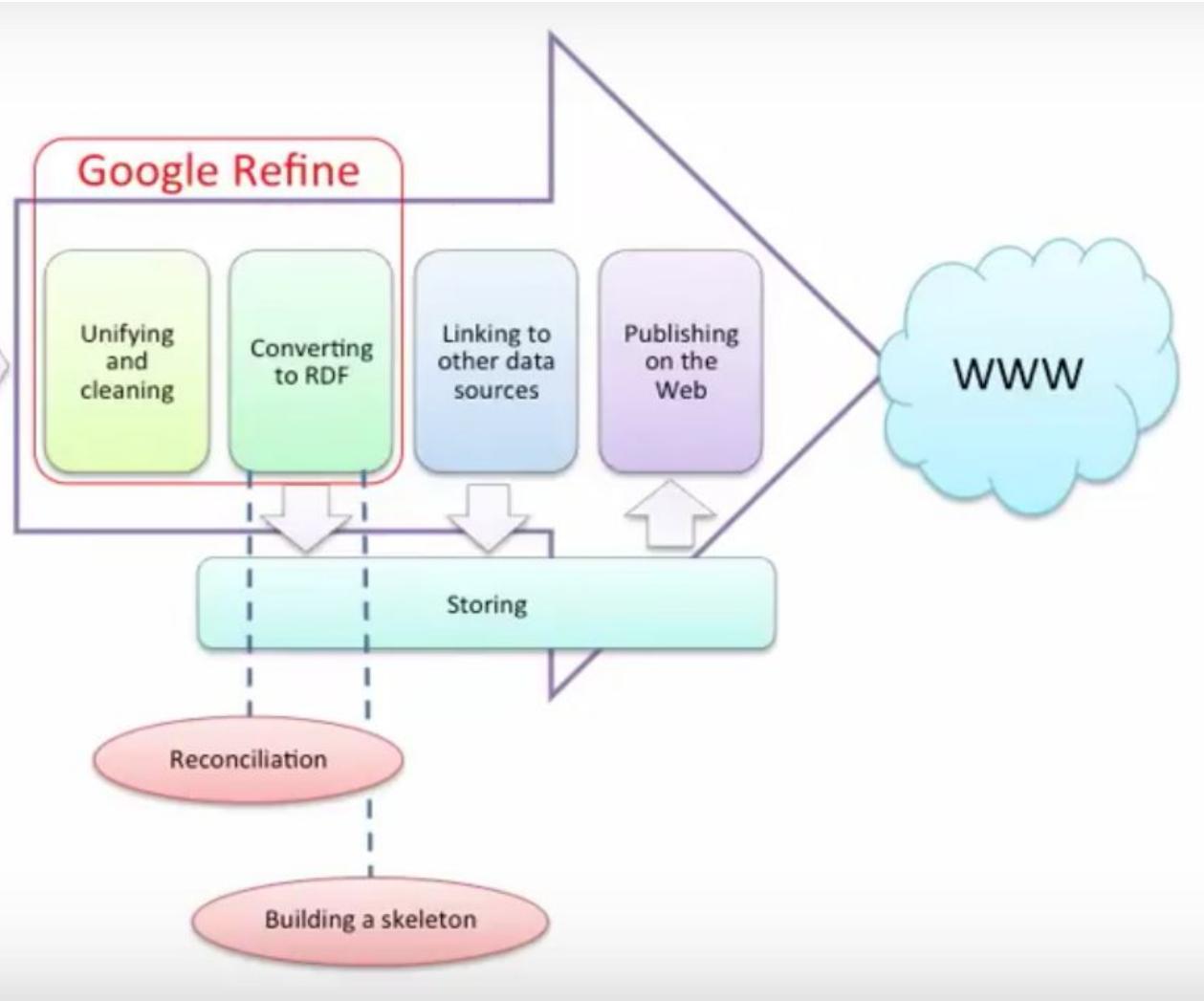
Linked Datasets as of August 2014



Il processo di pubblicazione dei propri dati come *Linked Open Data*



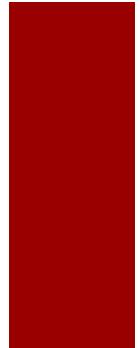
OpenRefine (formerly Google Refine)





Question Time

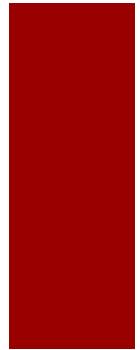
- Vantaggi/Svantaggi di Excel vs Linked Data





OpenRefine

- Consente di elaborare dati tabulari (come quelli di excel)
 - ripulirli
 - unirli ad altri dati
 - salvare in RDF





Esercitazione #1

- Assicurarsi che Java JDK8 sia installato
- Avviare OpenRefine (OR) o **LODRefine**
 - posizionarsi nella directory contenente OR (preinstallati\lodrefine)
 - cliccare su: *lodrefine.exe*
 - se non si avvia il browser, avviare quest'ultimo e digitare:
localhost:3333
- Esercizio
 - importare file bookstore.xls
 - conversione colonne numeriche (edit cell transform)
 - filtrare solo libri con autore di nome "laura" (text filter)
 - data cleaning / reconciliation con DBpedia (via sparql)
 - creazione RDF skeleton
 - id URI
 - creazione triple RDF (formato turtle: bookstore.ttl)



PUBBLICARE LINKED DATA

Metodologie e strumenti

Maurizio Atzori

Dipartimento di Matematica e Informatica
Università di Cagliari

atzori@unica.it

SUMMER SCHOOL LDA
LIBRARIES IN THE DIGITAL AGE

29 Agosto - 1 Settembre 2016



Programma Day #2

pubblicare triple nel web (linked open data)

- Intro a Web
 - Browser e siti (Client & & Server)
 - HTML
 - Content negotiation
 - RDF e microformati/RDFa
- Strumenti software
 - Intro Triplestores: Jena/Fuseki, OpenLink Virtuoso, Sesame
 - SPARQL Endpoint
 - Pubby
- Esercitazione con Fuseki e Pubby



Web (Client/Server)

- Al web si accede tramite browser
 - scaricano file html/js/css
 - client di visualizzazione di file html/js/css
- I siti che forniscono i documenti web
 - sono computer (server) che ad ogni richiesta del browser rilasciano i file html, immagini, ecc

HTML - HyperText Markup

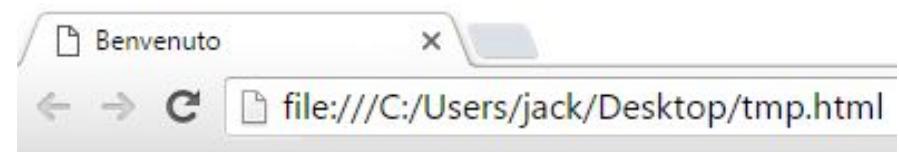
Language



Linguaggio di markup (a marcatori) per ipertesti

- file di testo comprensibili dai browser
 - determinano il contenuto della pagina e la sua struttura (titolo, pie' di pagina, colonne)

```
<html>
  <head>
    <title>Benvenuto</title>
  </head>
  <body>
    <h1>Ciao!</h1>
    <p>Il mio primo
       paragrafo</p>
  </body>
</html>
```



Ciao!

Il mio primo paragrafo



HTTP - HyperText Transfer Protocol

- E' il "protocollo" che browser e server usano per dialogare
 - GET /page (<http://example.com/page>)
 - Headers
- prevede la negoziazione del formato del contenuto (content negotiation)
 - browser "standard"
 - Accept: text/html,application/xhtml+xml
 - browser "semantici"
 - MIME type: application/rdf+xml



RDF e serializzazioni

RDF è un modello a triple senza schema

Diverse "serializzazioni" (formati di salvataggio delle triple):

- Turtle (ttl)
- rdf/xml
- N-Triple
- ...



Serializzazioni

XML/RDF

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:au="http://description.org/schema/">
  <rdf:Description
    about="http://www.book.it/Rosso_di_sera_bel_tempo_si_spera/">
    <au:author>Mario_Rossi</au:author>
  </rdf:Description>
</rdf:RDF>
```

N-Triple

```
<http://www.book.it/Rosso\_di\_sera\_bel\_tempo\_si\_spera/>
<http://description.org/schema/author>
"Mario_Rossi".
```



Content Negotiation

Risorsa: <http://dbpedia.org/resource/Cagliari>

Il browser html restituisce un redirect a:

<http://dbpedia.org/page/Cagliari>

Se si usa un browser semantico viene restituito cio'
che e' possibile osservare a questo indirizzo (RDF):

<http://dbpedia.org/data/Cagliari>



Content Negotiation

Risorsa: <http://dbpedia.org/resource/Cagliari>

Il browser html restituisce un redirect a:

<http://dbpedia.org/page/Cagliari>

Se si usa un browser semantico viene restituito cio'
che e' possibile osservare a questo indirizzo (RDF):

<http://dbpedia.org/data/Cagliari>



<http://dbpedia.org/resource/Cagliari>

HTML

```
<?xml version="1.0" encoding="UTF-8" ?>

<!DOCTYPE html PUBLIC "-//W3C//DTD
XHTML+RDFa 1.0//EN"
"http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">

<html
    xmlns="http://www.w3.org/1999/xhtml"
    xmlns:dbpprop="http://dbpedia.org/property/"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    " version="XHTML+RDFa 1.0"
    xml:lang="en">

<head
    profile="http://www.w3.org/1999/xhtml/vocab">

    <title>About: Cagliari</title>
```

XML/RDF

```
<rdf:RDF

    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

    xmlns:dbo="http://dbpedia.org/ontology/"

    xmlns:dbp="http://dbpedia.org/property/"

    . . .

    <dbp:populationTotal
        rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">154543</dbp:populationTotal>
```



Includere dati strutturati nell'HTML

Microformats A way to use html pages as both a human readable document and machine readable data, without repetition (e.g. sticking CSV in the head element).

RDF A data model designed for the web. Schemaless, uses URLs to name types and relations.

RDFa A way to encode (write) RDF in html, following the style of microformats (i.e. minimising repetition). Works by adding a few attributes to html.

Microdata An alternative to RDFa, using different attributes and a different data model.

Dati strutturati nei motori di ricerca



Google supporta i formati JSON-LD, Microdata e RDFa con vocabolario schema.org

le ali della libertà film

Google Search Results for "le ali della libertà film":

Tutti Video Immagini Notizie Maps Altro ▾ Strumenti di ricerca

Circa 398.000 risultati (0,62 secondi)

Le ali della libertà - Wikipedia
https://it.wikipedia.org/wiki/Le_ali_della_libertà ▾
Le ali della libertà (The Shawshank Redemption) è un film del 1994 diretto da Frank Darabont, con Tim Robbins e Morgan Freeman, tratto dal racconto di ...
Trama · Personaggi del film · Differenze con il racconto ... · Distribuzione

Le ali della libertà - MYmovies
[www.mymovies.it › film › 1994](http://www.mymovies.it/film/1994) ▾
★★★★★ Valutazione: 4,4 - 92 recensioni
Un film di Frank Darabont con James Whitmore, Tim Robbins, Clancy Brown, Morgan Freeman. Da un racconto di Stephen King una storia commovente.



Eventi a Cagliari (data, link, luogo)

eventa cagliari



Tutti

Notizie

Maps

Immagini

Shopping

Altro ▾

Strumenti di ricerca

Circa 22.500 risultati (0,34 secondi)

Eventa: i migliori eventi a Cagliari oggi, domani e nel weekend.

<https://www.eventa.it/eventi/cagliari> ▾

Mostre ed **eventi a Cagliari** regalano occasioni di svago, mentre gli aperitivi nel largo Carlo Felice permettono di avere una vista sul mare mentre si sta in ...

ven 5 ago Mercatinando CON NOI Santa Susanna

ven 5 ago Sconti d' estate Farmacia Carta Dr. Stefano

ven 5 ago Paratissima Cagliari Salotto Dell'Arte



Recap

- L'RDF consiste nel descrivere informazioni sotto forma di triple
 - queste posso essere salvate in vari formati
 - possono essere include dentro l'HTML
- Data una URL, se acceduta questa puo' restituire:
 - un file html (web of documents)
 - un file xml+rdf (web of linked data)
 - un html + rdfa (entrambi i web!)



Triplestore

- I triplestore sono software DBMS che consentono di immagazzinare triple
 - confrontabili con Oracle DB, Access o Excel, ma focalizzati su grafi anziche' tabelle
 - accedibili tramite un endpoint (indirizzo web)
 - Esempi: Fuseki, Virtuoso, RDF4J (ex Sesame)
- Una volta immagazzinate, le triple possono essere interrogate
 - SPARQL (SPARQL Protocol and RDF Query Language)



Apache Jena/Fuseki

Apache è una fondazione che rilascia software libero e aperto

- Jena è una libreria per programmatori
- Fuseki è un triplestore basato su Jena
- Necessita di Java
- immagazzina grandissime quantità di triple e le gestisce



Esempi di endpoints online

- DBpedia:
 - <http://dbpedia.org/sparql>
 - <http://live.dbpedia.org/sparql>
- Camera e Senato della Repubblica:
 - <http://dati.senato.it/23>
 - <http://dati.camera.it/sparql>



Pubby

Pubby consente di pubblicare Linked Data prendendo i dati da uno SPARQL endpoint

- Realizzato da Richard Cyganiak, Chris Bizer (DBpedia)
- Necessita di Jetty e di Java
- Per ogni soggetto (della tripla) crea una pagina web html+rdfa contenente le triple che lo riguardano



<http://dbpedia.org/resource/Apple>



About: Apple

An Entity of Type : [species](#), from Named Graph : <http://dbpedia.org>, within Data Space : dbpedia.org

The apple is the pomaceous fruit of the apple tree, species *Malus domestica* in the rose family (Rosaceae). It is one of the most widely cultivated tree fruits, and the most widely known of the many members of genus *Malus* that are used by humans. Apples grow on small, deciduous trees. The tree originated in Central Asia, where its wild ancestor, *Malus sieversii*, is still found today.

| Property | Value |
|---|---|
| dbpedia-owl:abstract | <ul style="list-style-type: none">The apple is the pomaceous fruit of the apple tree, species <i>Malus domestica</i> in the rose family (Rosaceae). It is one of the most widely cultivated tree fruits, and the most widely known of the many members of genus <i>Malus</i> that are used by humans. Apples grow on small, deciduous trees. The tree originated in Central Asia, where its wild ancestor, <i>Malus sieversii</i>, is still found today. Apples have been grown for thousands of years in Asia and Europe, and were brought to North America by European colonists. There are more than 7,500 known cultivars of apples, resulting in a range of sizes, shapes, and colors. Apples are often eaten raw, but can also be found in many prepared foods (especially desserts and pies), and are used in many non-food products, such as cider, juice, and vinegar. Apples are a good source of fiber, vitamins C and K, and potassium. They contain various phytochemicals, including quercetin and ellagic acid, which have antioxidant properties. Apples are also a good source of dietary fiber, which can help to promote gut health and reduce the risk of certain diseases, such as heart disease and type 2 diabetes. Apples are a low-calorie food, making them a popular choice for weight loss and maintenance. They are also a good source of water, which can help to keep you hydrated. Apples are a versatile fruit that can be enjoyed in many ways, from fresh to cooked, and from raw to processed. They are a healthy and delicious addition to any diet. |
| dbpedia-owl:binomialAuthority | <ul style="list-style-type: none">dbpedia:Moritz_Balthasar_Borkhausen |
| dbpedia-owl:class | <ul style="list-style-type: none">dbpedia:Eudicots |
| dbpedia-owl:division | <ul style="list-style-type: none">dbpedia:Flowering_plant |
| dbpedia-owl:family | <ul style="list-style-type: none">dbpedia:Rosaceae |
| dbpedia-owl:genus | <ul style="list-style-type: none">dbpedia:Malus |
| dbpedia-owl:kingdom | <ul style="list-style-type: none">dbpedia:Plant |
| dbpedia-owl:order | <ul style="list-style-type: none">dbpedia:Rosalesdbpedia:Rosids |
| dbpedia-owl:synonym | <ul style="list-style-type: none"><i>Malus communis</i> (Desf.)<i>Malus pumila</i> (auct.)<i>Pyrus malus</i> L. |



Esercitazione #2

- Avviare Jena Fuseki
 - posizionarsi nella directory contenente Fuseki
(preinstallati\apache-jena-fuseki)
 - cliccare su: *fuseki.bat*
 - avviare il browser e digitare: *localhost:3030*
- Esercizio
 - caricare le triple del file turtle (ttl) su fuseki
 - visione contenuto su fuseki
- Avviare Pubby (tramite Jetty)
 - posizionarsi nella directory contenente Jetty e Pubby
(preinstallati\jetty-with-pubby)
 - cliccare su: *pubby.bat*
 - avviare il browser e digitare: *localhost:3333*
 - visionare il file di configurazione
preinstallati\jetty-with-pubby\webapps\root\WEB-INF\config.ttl



PUBBLICARE LINKED DATA

Metodologie e strumenti

Maurizio Atzori

Dipartimento di Matematica e Informatica
Università di Cagliari

atzori@unica.it

SUMMER SCHOOL LDA
LIBRARIES IN THE DIGITAL AGE

29 Agosto - 1 Settembre 2016



Programma Day #3

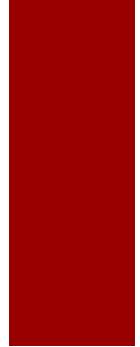
interrogare triple RDF usando SPARQL

- Intro a SPARQL (esempi)
 - intro linguaggi di interrogazione
 - Graph patterns
 - Filter e text search
 - Union
- BEStQ/Swipe
- Esercitazione con Fuseki e Virtuoso
 - interrogare i dati di bookstore su Fuseki
 - interrogare i dati di wikipedia usando DBpedia
 - interrogare i dati di Camera e Senato usando
 - endpoint <http://dati.senato.it/23>
 - endpoint <http://dati.camera.it/sparql>



SPARQL slides

- [sparql-tutorial.pdf](#)
 - same as "sparql by example.html"
- optional: [intro-rdf-sparql.pdf](#)
 - from page 22





Esercitazione #3

- Avviare Jena Fuseki
 - posizionarsi nella directory contenente Fuseki
(preinstallati\apache-jena-fuseki)
 - cliccare su: *fuseki.bat*
 - avviare il browser e digitare: *localhost:3030*
- Esercizio
 - usando Fuseki
 - chiedere la lista di tutti i libri
 - chiedere la lista di tutti i libri con prezzo maggiore di X
 - chiedere quanti libri con prezzo maggiore di X
 - sull'endpoint di DBpedia: <http://dbpedia.org/sparql>
 - trovare le coppie libro/autore
 - trovare i libri con autore di nome "laura"
 - serie TV di genere "crime"