

DBpedia e Wikidata

DBpedia

- “crowd-sourced community effort to extract structured content from the information created in various Wikimedia projects”
 - CMS MediaWiki (<https://www.mediawiki.org/wiki/MediaWiki>)
 - Licenza: Creative Commons Attribution-ShareAlike 3.0
- 4.58 milioni di entità, di cui 4.22 milioni classificati in una ontologia
 - 1,445,000 persone
 - 735,000 luoghi
 - 411,000 opere
- *localized editions* in 125 lingue
- Insieme, queste versioni descrivono 38.3 milioni di entità
- Contiene 3 miliardi di triple RDF

DBpedia nei LOD

- 25.2 milioni di link immagini
- 29.8 milioni di link to a pagine web;
- 80.9 milioni di link a categorie di Wikipedia,
- 41.2 milioni di link a YAGO.
- DBpedia collegata a altri datasets da circa 50 milioni di link RDF.

Dbpedia, come si presenta:

- «The resulting DBpedia knowledge base currently describes over 2.6 million entities.»
- «For each of these entities, DBpedia defines a globally unique identifier that can be dereferenced over the Web into a rich RDF description of the entity including:»
 - human-readable definitions in 30 languages,
 - relationships to other resources
 - classifications in four concept hierarchies
 - various facts
 - data-level links to other Web data sources describing the entity.

AUER, Sören, et al. Dbpedia: A nucleus for a web of open data.

In: *The semantic web*. Springer, Berlin, Heidelberg, 2007. p. 722-735.

Pagina di Dbpedia

http://dbpedia.org/page/Édouard_Manet



Browse using ▾

Formats ▾

Faceted Browser

Sparql Endpoint

About: Édouard Manet

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

Édouard Manet (* 23. Januar 1832 in Paris; † 30. April 1883 ebenda) war ein französischer Maler. Er gilt als einer der Wegbereiter der modernen Malerei.

Property	Value
dbo:abstract	<ul style="list-style-type: none">Manet non volle mai essere identificato col gruppo degli impressionisti, né partecipò mai alle loro esposizioni. Ciò perché, lungo tutta la sua carriera, egli preferì ottenere l'ammissione al Salon mediante un riconoscimento ufficiale dello Stato, e non «attraverso sotterfugi», come lui stesso affermò. Per questo si batté in difesa del principio della libertà espressiva dell'artista con opere che suscitavano scandalo presso i suoi contemporanei, come Colazione sull'erba e Olympia. A partire dal 1869 si dedicò alla pittura en plein air ("all'aperto") e le sue uscite ai giardini delle Tuileries, sul retro del Louvre, divennero quasi degli appuntamenti mondani. Restò in attività fino al 1883, l'anno della sua morte. Manet raggiunse una grandissima fama e ancora oggi è considerato il più grande interprete della pittura pre-impressionista. ^(it)
dbo:birthDate	<ul style="list-style-type: none">1832-01-23 (xsd:date)1832-1-23
dbo:birthName	<ul style="list-style-type: none">Édouard Manet (en)
dbo:birthPlace	<ul style="list-style-type: none">dbr:Paris
dbo:deathDate	<ul style="list-style-type: none">1883-04-30 (xsd:date)1883-4-30
dbo:deathPlace	<ul style="list-style-type: none">dbr:Paris

Pagina di DBpedia

http://dbpedia.org/page/Édouard_Manet

@prefix dbo: <<http://dbpedia.org/ontology/>> . @prefix dbr: <<http://dbpedia.org/resource/>> .

dbr:Edouard_manet **dbo:wikiPageRedirects** <http://dbpedia.org/resource/\u00C9douard_Manet> .

dbr:Edward_Manet **dbo:wikiPageRedirects** <http://dbpedia.org/resource/\u00C9douard_Manet> .

<http://dbpedia.org/resource/A_Bar_at_the_Folies-Berg\u00E8re> **dbo:author**
<http://dbpedia.org/resource/\u00C9douard_Manet> .

dbr:The_Fifer **dbo:author** <http://dbpedia.org/resource/\u00C9douard_Manet> .

<http://en.wikipedia.org/wiki/\u00C9douard_Manet> **foaf:primaryTopic** <http://dbpedia.org/resource/\u00C9douard_Manet> .

<http://dbpedia.org/resource/Paul_C\u00E9zanne> **dbo:influencedBy** <http://dbpedia.org/resource/\u00C9douard_Manet> .

@prefix rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>> .

<http://dbpedia.org/resource/\u00C9douard_Manet> **rdf:type** yago:WikicatFrenchRealistPainters ,
yago:WikicatRealistPainters ,
yago:WikicatPeopleFromParis .

<http://dbpedia.org/resource/\u00C9douard_Manet> **rdf:type** owl:Thing ,
yago:CausalAgent100007347 ,
dbo:Artist .

@prefix schema: <<http://schema.org/>> .

<http://dbpedia.org/resource/\u00C9douard_Manet> **rdf:type** schema:Person .

http://dbpedia.org/page/Édouard_Manet (2)

```
http://dbpedia.org/resource/\u00C9douard_Manet> rdfs:label      "\u00C9douard Manet"@es ,
                                                                "\u00C9douard Manet"@de ,
                                                                "\u00C9douard Manet"@it ,
                                                                "\u00C9douard Manet"@en ,
                                                                "\u00C9douard Manet"@fr .

<http://dbpedia.org/resource/\u00C9douard_Manet> owl:differentFrom dbr:Claude_Monet ;
                                                    owl:sameAs <http://fr.dbpedia.org/resource/\u00C9douard_Manet> ,
                                                    <http://de.dbpedia.org/resource/\u00C9douard_Manet> .
                                                    <http://it.dbpedia.org/resource/\u00C9douard_Manet> ,
                                                    <http://viaf.org/viaf/97379936> ,
                                                    <http://data.nytimes.com/88965468228098832063> .

@prefix dct: <http://purl.org/dc/terms/> .

<http://dbpedia.org/resource/\u00C9douard_Manet>    dct:subject    dbc:French_Impressionist_painters ,
                                                    dct:description    "French painter"@en ;
                                                    foaf:givenName      "\u00C9douard"@en ;
                                                    foaf:gender         "male"@en ;
                                                    foaf:isPrimaryTopicOf <http://en.wikipedia.org/wiki/\u00C9douard_Manet> ;
                                                    foaf:surname        "Manet"@en .

<http://dbpedia.org/resource/\u00C9douard_Manet> dbp:caption "portrait by Nadar, 1874"^^rdf:langString ;

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://dbpedia.org/resource/\u00C9douard_Manet>    dbo:birthDate    "1832-01-23"^^xsd:date ,
                                                    dbo:deathDate        "1883-4-30"^^xsd:date ,
                                                                "1883-04-30"^^xsd:date ;
```

... E MOLTO ALTRO!

Classi dell'ontologia di DBpedia:

Agent

Agent ([Show in class hierarchy](#))

Label (it): agente

Label (el): πράκτορας

Label (ga): gníomhaire

Label (fr): agent

Label (de): Agent

Label (gl): axente

Label (es): agente

Label (da): agent

Label (ja): エージェント

Label (en): agent

Label (ko): 에이전트

Label (nl): agent

Comment (en): Analogous to a foaf:Agent, an agent is an entity that acts. This is intended to be the super class of Person and Organisation.

age (edit)	age	Agent	<i>xsd:integer</i>
------------------------------	-----	-----------------------	--------------------

<http://dbpedia.org/ontology/Agent>

Costruzione di DBpedia

- Mapping delle *infobox* verso un'unica ontologia di 320 classi e 1650 proprietà.
- Problemi di normalizzazione tra le edizioni nazionali.
 - Mapping classi: Assegnare una classe dell'ontologia di DBpedia alle entità che sono descritte dall'infobox di un certo tipo
 - Mapping proprietà: Infobox diverse per lo stesso tipo di entità: nomi di proprietà per la stessa proprietà
- Confrontare le pagine di Wikipedia in Inglese e Italiano della stesse città (es. Roma)

Università degli Studi di Torino	
	
	
Il Rettorato dell'Ateneo torinese	
Ubicazione	
Stato	 Italia
Città	Torino
Altre sedi	Aosta, Alba, Asti, Biella, Collegno, Cuneo, Grugliasco, Ivrea, Moretta, Orbassano, Savigliano ^[1] ^[2] ^[3] , Venaria Reale ^[senza fonte]
Dati generali	
Nome latino	Universitas Augustae Taurinorum
Fondazione	1404
Tipo	Statale
Rettore	Stefano Geuna
Studenti	72 410 (2017) ^[4]
Dipendenti	1 700 amministrativi circa
Sport	CUS Torino
Mappa di localizzazione	

Esempio di query su DBpedia

PREFIX dbp: <http://dbpedia.org/ontology/>

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

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

SELECT ?person ?caption

WHERE { ?person rdf:type dbp:Person ; dbpprop:name ?name ;
dbpprop:caption ?caption FILTER (regex(?name, 'Manet'))
}

person	caption
http://dbpedia.org/resource/Édouard_Manet	"portrait by Nadar, 1874"^^<http://www.w3.org/1999/02/22-rdf-syntax-ns#langString>

Come hanno fatto?



Infobox

- «The type of wiki contents that is most valuable for the DBpedia extraction are Wikipedia infoboxes» *MA...*
 - Comunità diverse usano templates diversi per descrivere lo stesso tipo di cose
 - I diversi template usano nomi diversi per lo stesso attributo (es. birthplace and placeofbirth)
- Procedura
 - The URI dell'articolo di Wikipedia (in Dbpedia) diventa il soggetto;
 - Il predicato è creato concatenando il namespace <http://dbpedia.org/property/> con il nome dell'attributo nell'infobox
 - L'oggetto è creato dal valore dell'attributo

Dbpedia: extractors

- Gli Extractors trasformano un certo tipo di markup del wiki in triple
- 11 extractors per tipi specifici di informazione:
 - Labels
 - Abstracts
 - Interlanguage links
 - Images
 - Redirects
 - Disambiguation
 - External links
 - Homepages
 - Page Links
 - Categories
 - Geo-coordinates

Mapping

- I template di Wikipedia sono stati mappati su un'ontologia. Creata manualmente a partire dai 350 infobox template più comuni dell'edizione inglese di Wikipedia
- Il mappings delle proprietà si basa su regole che descrivono come gestire i valori dei campi negli infobox definendo i datatypes per il range
- Questo mapping è stato parzialmente compiuto tramite crowdsourcing
- Classificazione delle entità
 - Wikipedia Categories
 - YAGO
 - UMBEL (The Upper Mapping and Binding Exchange Layer)
 - Wikipedia Ontology

DBpedia ontology

- 320 classi
- 1650 proprietà.
- Gerarchia di classi di 5 livelli massimo (per scopi di navigazione e visualizzazione)
- Versioni locali di Dbpedia per 111 linguaggi
 - Caratterizzate dalla sigla del linguaggio come sotto-dominio:
<http://it.dbpedia.org>

Esempio

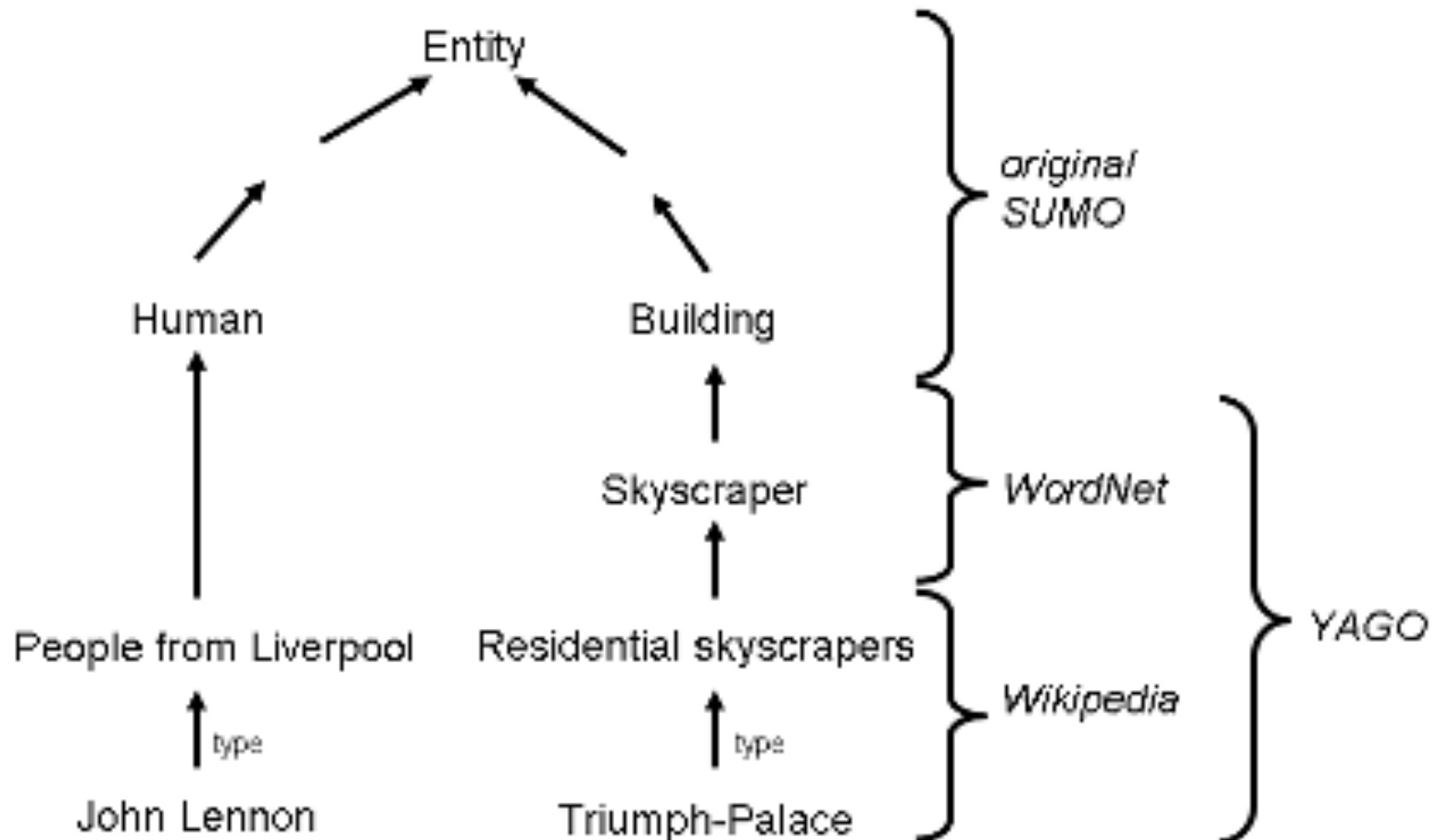
```
{{TemplateMapping
| mapToClass = Actor
| mappings =
  {{ PropertyMapping | templateProperty = name | ontologyProperty = foaf:name }}
  {{ PropertyMapping | templateProperty = birth_place | ontologyProperty = birthPlace }}
}}
```

- <http://mappings.dbpedia.org>

WordNet in Yago2

- YAGO2 costruita automaticamente da Wikipedia, GeoNames e WordNet.
- Contiene 447 milioni di fatti su circa 9.8 milioni di entità
- Utilizzo di risorse linguistiche per estrarre e mappare le informazioni sull'ontologia

YAGOSUMO: architettura e sorgenti di conoscenza



Procedura di mapping

- Per ogni categoria della pagina, determina la testa nel nome della categoria (parsing). Se è plurale, propone la categoria come classe e l'entità come membro della classe.
 - Es. “Poeta canadese”: “poeta” è la testa.
- La classe viene collegata a WordNet scegliendo il synset più frequente della testa
 - Questa strategia si è rivelata piuttosto accurata
- Un insieme di pattern definiti a mano mappano le categorie e gli attributi della infobox su un insieme di template
 - Es. Leonard Cohen: l'attributo born=Montreal, che è mappato sul fatto “wasBornIn(LeonardCohen, Montreal)”

Regole

- La creazione di Yago2 si avvale di alcuni set di regole per migliorare quanto estratto
 - Ci sono regole specifiche per migliorare la conoscenza linguistica
- Implication rules dicono che se nella base di conoscenza compare un certo fatto, deve esserne aggiunto un altro
 - Se A property1 B \rightarrow A property2 B
- Factual rules: collega le categorie di Wikipedia ai **synset** di WordNet
 - "capital" hasPreferredMeaning wordnet_capital_108518505
- Replacement rules se un dato segue una certo pattern, rimpiazzalo con un altro:
 - "USA" \rightarrow "United States of America"
- Extraction rules se una parte del testo sorgente segue un certo pattern, vengono aggiunti determinati fatti.
 - Convertono le feature delle info boxes in proprietà (per es. "births" \rightarrow wasBornOnDate)

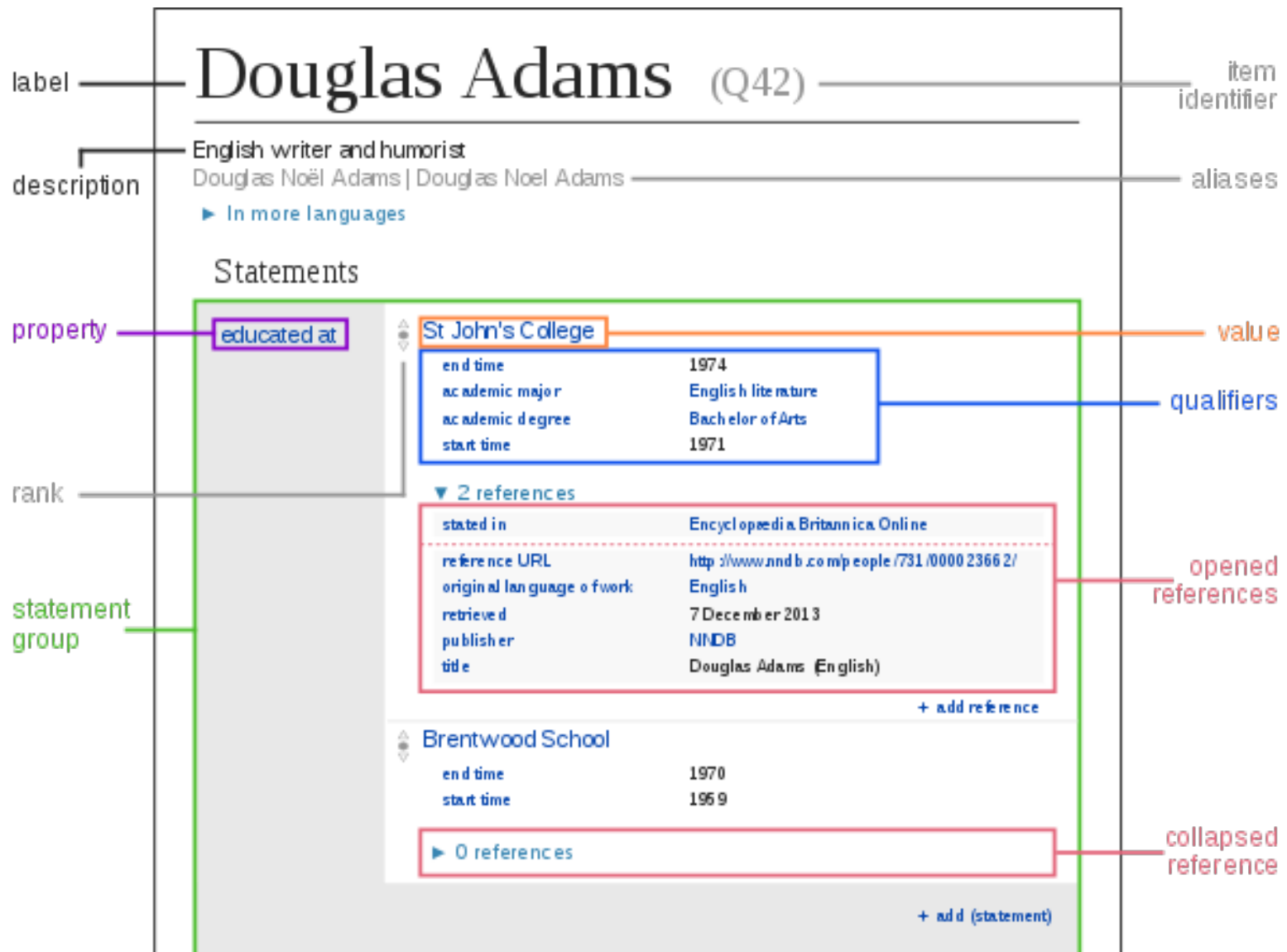
Wikidata

Introduzione

Cos'è Wikidata

- Un database **collaborativo** basato sui contenuti di Wikipedia e Wikimedia
- Licenza Creative Commons
 - CC0 1.0 Universal (CC0 1.0) Public Domain Dedication
- *Secondary database*: le informazioni sono etichettate con la **sorgente** da cui provengono
- Dati **strutturati** per supportare l'accesso da parte di persone e macchine
- La descrizione delle entità viene effettuata tramite un'**interfaccia** utente predefinita, in modo tale che i dati inseriti siano necessariamente strutturati in modo conforme al modello dati di Wikimedia

Datamodel di Wikimedia



<https://www.wikidata.org/wiki/Wikidata:Introduction>

Inserimento dei dati

No label defined (Q17339402)



No description defined

 edit

▼ In more languages [Configure](#)

Language	Label	Description	Also known as
English	No label defined	No description defined	
Italian	No label defined	No description defined	
French	No label defined	No description defined	
Sardinian	No label defined	No description defined	

Statements

highest point	 Mount Everest	 edit
	▼ 0 references	+ add reference
		+ add value

+ add statement

Items

- Il database contiene **items**, ognuno dei quali ha almeno una **label**, una **descrizione** e degli **alias**
 - Identificati da una **Q** seguito da un numero
- Gli **statement** sull'item hanno lo scopo di descriverlo e consistono of una **proprietà** e un **valore**.
 - Le proprietà sono identificate da una **P** seguita da un numero, es. *educated at* (P69).
- Gli item rappresentano concetti astratti o concreti

Item	Property	Value
Q42	P69	Q691283
<i>Douglas Adams</i>	<i>educated at</i>	<i>St John's College</i>

Accesso ai dati

- Via SPARQL endpoint
- Via strumento di ragionamento (Reasonator, sperimentale)
- Tramite un insieme di tool

(<https://www.wikidata.org/wiki/Wikidata:Tools>)

Struttura degli item

- Identificativo Q###
- Insieme di label linguistiche (solo 1 per ogni lingua)
 - Convenzione: maiuscole per i nomi propri, il resto minuscolo
- Descrizioni: accompagnano le label (1:1)
- Si possono aggiungere alias (possibili più alias)
- «Properties limit and define what kind of data values can be added to an item. For example, the “discovery date” property requires a date value and won’t accept just anything to be entered in like “blue” or “France”. These constraints also protect Wikidata from vandalism.»

Esempio: Ocean's Eleven

Ocean's Eleven (Q205447)

2001 film by Steven Soderbergh

"Ocean's Eleven"



 [edit](#)

[▼ In more languages](#) [Configure](#)

Language	Label	Description	Also known as
English	Ocean's Eleven	2001 film by Steven Soderbergh	"Ocean's Eleven"
Italian	Ocean's Eleven - Fate il vostro gioco	film del 2001 diretto da Steven Soderbergh	
French	Ocean's Eleven	film sorti en 2001	L'Inconnu de Las Vegas Ocean's 11
Sardinian	No label defined	No description defined	

[All entered languages](#)

Statements

instance of	 film	 edit
-------------	--	--

<https://www.wikidata.org/wiki/Q205447>

Statement

- Sono attaccati agli item
- Divisi per tipo di proprietà, a cui è associato un valore
 - Il dato che descrive un item per una data proprietà si dice valore.

Esempio: statements

title



Ocean's Eleven (English)

edit

▼ 0 references

+ add reference

+ add value

based on



Ocean's 11

edit

▼ 0 references

+ add reference

+ add value

series



Ocean's Trilogy

edit

followed by

Ocean's Twelve

series ordinal

1

follows

no value

▼ 0 references

+ add reference

+ add value

main subject



gambling

edit

Qualifiers

- I qualifiers permettono di specificare il significato di una coppia property-value
 - Unità di misura, precisazioni, metodo con cui è stato attribuito il valore ecc.
- Si possono aggiungere dal menu *edit* degli statament
 - «add qualifier»
 - Si attiva una selezione di possibili qualificatori

publication date

1816

✓ publish 🗑 remove

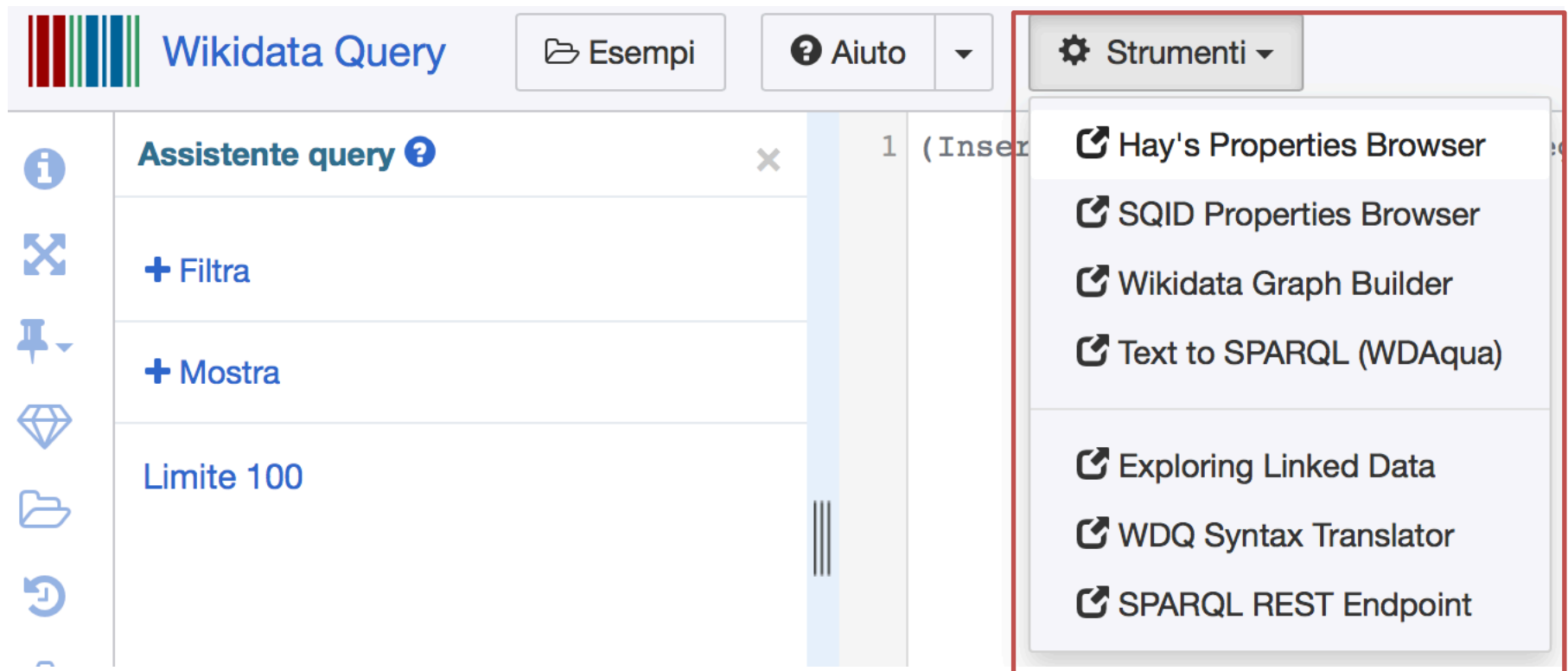
Property 🗑 remove

Property 🗑 remove

place of publication
geographical place of publication of the edition (use 1st edition when referring to works)

Interrogare Wikidata:

- La pagina di interrogazione mette a disposizione dell'utente un insieme di strumenti per facilitare l'esplorazione delle proprietà



The screenshot displays the Wikidata Query interface. At the top, there is a header bar with the Wikidata logo, the text 'Wikidata Query', and buttons for 'Esempi' (Examples), 'Aiuto' (Help), and 'Strumenti' (Tools). The 'Strumenti' button is highlighted with a red border, and its dropdown menu is open, showing a list of tools: Hay's Properties Browser, SQID Properties Browser, Wikidata Graph Builder, Text to SPARQL (WDAqua), Exploring Linked Data, WDQ Syntax Translator, and SPARQL REST Endpoint. On the left side, there is a sidebar with icons for information, filters, and other query-related functions. The main area shows a query editor with the text 'Assistente query' and a filter button '+ Filtra'.

Esplorare la terminologia: proprietà

<https://tools.wmflabs.org/prop-explorer/>

Wikidata Property Explorer

HomeAbout

🔄🌐 Engl

Search the tree...

Show properties in sub-branches ☒

▼ Wikidata property related to art

Wikidata property for authority control for artists

▶ Wikidata property related to architecture

Wikidata property related to fashion

▶ Wikidata property related to literature

▼ Wikidata property related to movies

▶ Wikidata property for identification in the film industry

Wikidata property related to music

Wikidata property related to photography

ID	Label	Description	Type
<input type="text"/>	<input type="text"/>	<input type="text"/>	All ▼
▶ P1265	AlloCiné film ID AlloCiné film ID AlloCiné movie ID AlloCiné movie ID	identifier for a film in the French AlloCiné database	ExternalId
▶ P4781	MYmovies director ID	identifier for a director in MYmovies.it	ExternalId
▶ P4834	Deutsche Synchronkartei series ID	identifier for a TV series in the synchronisation (dubbing) database Deutsche Synchronkartei	ExternalId
▶ P4782	Movieplayer character ID	identifier for a character in movieplayer.it	ExternalId
▶ P4785	MYmovies name ID	identifier for producer, cinematographer, screenwriter and company in MYmovies.it	ExternalId

Proprietà per tipo

Id proprietà

label

descrizione

range

Descrizione proprietà

<https://www.wikidata.org/wiki/Property:P3383>

subproperty of



image

proprietà più generale

subject item of this property



film poster

di cosa parla la proprietà

property constraint



type constraint

class

film

relation

instance of

vincoli (sul domain)

Esplorare la terminologia: classi

Class item ID

Q21162272

Name/description language code

en

Submit

[view root classes](#)

<https://bambots.brucemyers.com/WikidataClasses.php>

Name:

[means \(Q21162272\)](#)

Description:

means by which an item performs a function

Parent class:

root class

Direct subclasses:

7 ([SPARQL query](#))

Indirect subclasses:

628

Direct instances:

0 ([SPARQL query](#))

Indirect instances:

7 965

Direct subclasses

Name	Wikidata link	Direct subclasses	Indirect subclasses	Direct instances	Indirect instances
electrical means	Q38215788	0	0	0	0
electronic means	Q38215930	1	4	1	7
herbouwwaardemeter	Q2410290	0	0	0	0
means of production	Q260747	0	0	0	0
mechanism of action	Q3271540	28	253	0	21
mode of transport	Q334166	13	329	17	7 919

Anche: <https://www.npmjs.com/package/wikidata-taxonomy>
Tool da linea di comando per estrarre porzioni della gerarchia


Descrizione classi

<https://www.wikidata.org/wiki/Q334166>


descrizione

Language	Label	Description
English	mode of transport	way to perform transport/transportation

classe più generale

subclass of	 type of transport
-------------	---

relazioni part-of

part of	 transport infrastructure
---------	--

identificativi in altre sorgenti di conoscenza)

BNCF Thesaurus ID	 619
-------------------	---