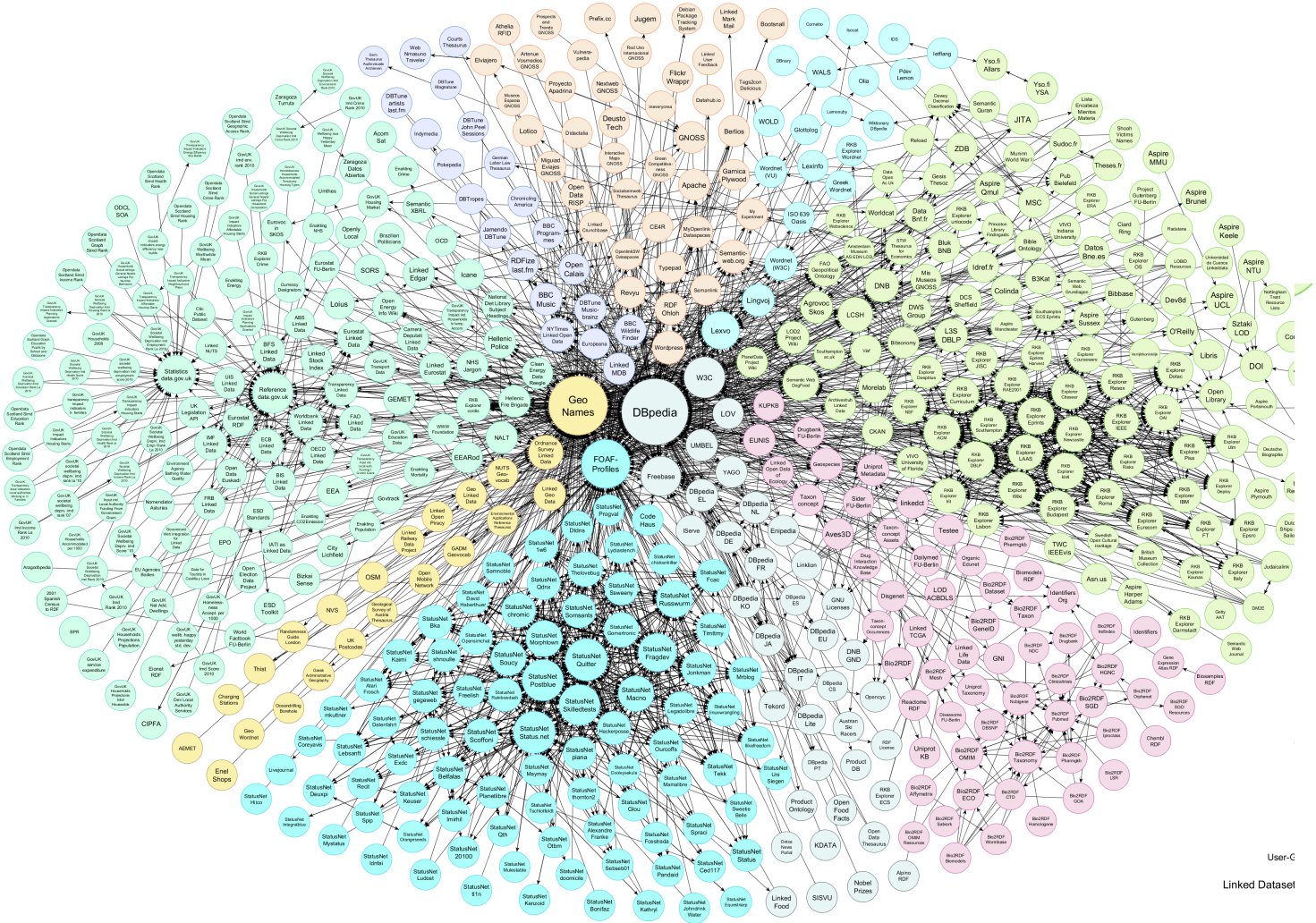
QwwwQ:
Querying Wikipedia without writing queries

Massimiliano Battan and Marco Ronchetti
DISI, Università degli Studi di Trento
38123 Povo di Trento
Italy

Linked Open Data
more and more public admins join!

The LOD cloud



- Linguistic
- Publications
- Life Sciences
- Cross-Domain
- Social Networking
- Geographic
- Government
- Media
- User-Generated Content

taken from http://data.dws.informatik.uni-mannheim.de/

Sometimes, too much data = no data

Christian Wolmar, railway historian:

"The huge corruption rate in the 19th century in USA railway administration was facilitated by an excessive level of accountability, where expense documentation for every single spike were collected".

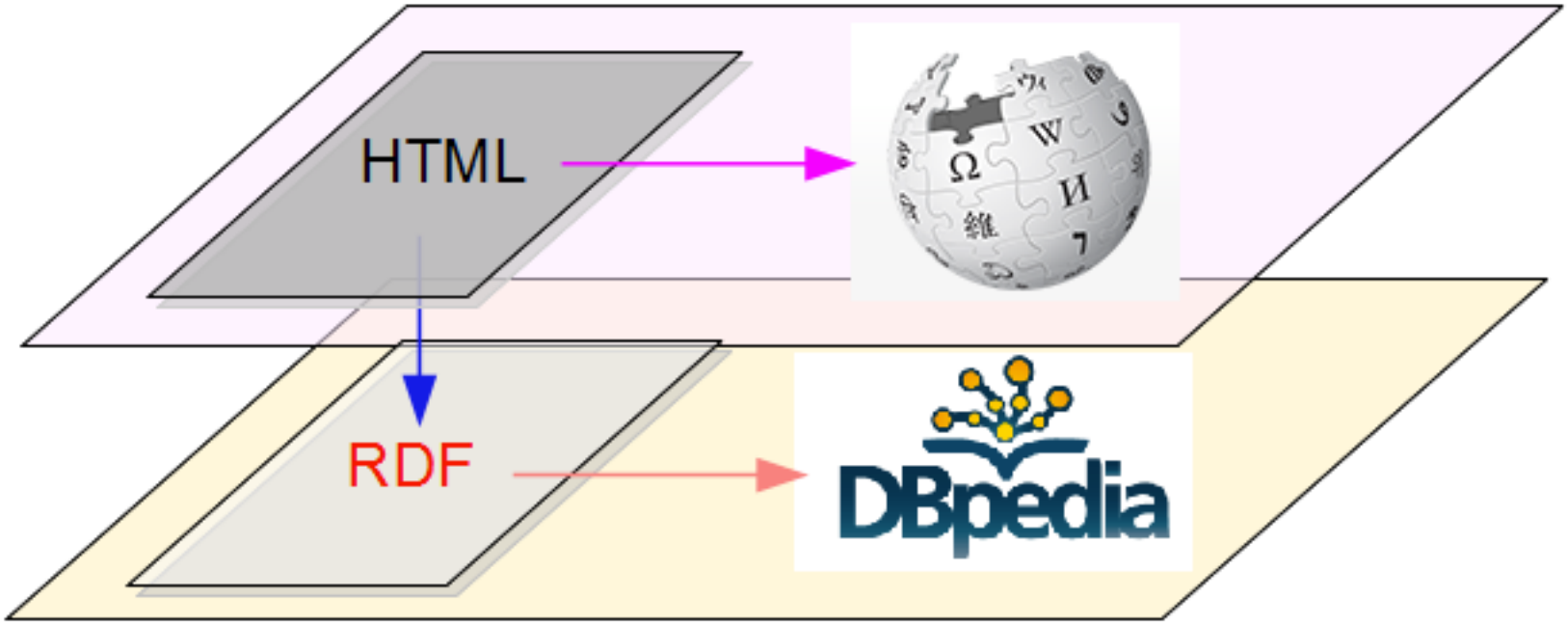
Data is only the first step…

- Data -> Information -> Knowledge
  - Data Mining, Search Engines, Recommendation…

Make sense of the data,
extract meaningful information
in simple and accessible way

user + data

DBPedia



S. Auer, and J. Lehmann, "What have Innsbruck and Leipzig in common? Extracting semantics from wiki content." 2007

S. Auer et al. "Dbpedia: A nucleus for a web of open data." 2007

A wikipedia page…

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
Deutsch
Français
Italiano
Magyar
Edit links

FS Class E.444

From Wikipedia, the free encyclopedia

The FS **E.444** is a class of Italian railways electric locomotives. They were introduced in the course of the 1960 until 1975. Starting from 1995, all E.444s were upgraded as E.444R.

The locomotives are nicknamed *Tartaruga* (turtle).

History [edit]

E.444 standard [edit]

The E.444 locomotive was designed in the 1960s as the first Italian electric locomotive capable to reach 200 km/h (120 mph) (in that period first high-speed trains like the Japanese Shinkansen and the French TGV were appearing). Italian railways could boast fast trains like the ETR 200, but they were getting old and the Pendolino project was just moving its first phaces.

The first 4 prototypes, built at Savigliano, made their debut in 1967-1968: their power output 3,000 kW (4,000 hp) was respectable for the time, but they proved unable to held the fast international services required for the new locomotive, and the bogies were limited to 180 km/h (110 mph). They proved anyway that Italian industry could produce locomotive capable of more than 200 km/h (120 mph)[3] was registered on November 8, 1967 in the maiden trip Rome-Milan.

The series production saw the introduction of the more powerful T750 motors, which boosted the power to 4,200 kW (5,600 hp), while the boogies were upgraded for 200 km/h (120 mph). The frontal part was improved and made more aerodynamic. A characteristic livery with two blue stripes on a pale grey background was adopted. The first 50 units proved successful, and the Ferrovie dello Stato (FS) ordered 60 more to be built starting from 1972. 16 units were adapted for feeding at 1,5 kV DC. During 1974 two locomotives (units 056 and 057) were provided with a "shunt chopper" system (later the modifications applied

FS class E.444

An E.444 in Bari Centrale station in 1995.

Specifications	
Power type	Electric
Build date	1965-1975
UIC classification	Bo-Bo
Wheel diameter	1.250 m (49.21 in)
Wheelbase	9.000 m (29 ft 6.3 in) between bogies 3.020 m (9 ft 10.9 in) between axles on each bogie
Length	16.840 m (55 ft 3.0 in)
Width	3.020 m (9 ft 10.9 in)
Height	4.300 m (14 ft 1.3 in)
Locomotive weight	83 t (82 long tons; 91 short tons)
Electric system(s)	3,000 V DC Catenary
Current collection method	Pantograph
Traction motors	DC series
Transmission	41/77 gear ratio
Top speed	200 km/h (120 mph)
Power output	4,272 kW (5,729 hp)
Tractive effort	201 kN (45,000 lb_f)
Safety systems	RSC4; SCMT
Career	
Railroad(s)	FS Trenitalia

DBPedia

From dbpedia.org/about:

"DBpedia allows you to ask sophisticated queries against Wikipedia, and to link the different data sets on the Web to Wikipedia data. We hope that this work will make it easier for the huge amount of information in Wikipedia to be used in some new interesting ways.

Furthermore, it might inspire new mechanisms for navigating, linking, and improving the encyclopedia itself."

Querying DBPedia

Using RDF, you can formulate powerful queries

Which European countries have a capital with more than 3 million people in which flows a river longer than 300 km?

The problem:

How can a generic user express a query on DBPedia?

S(he) needs to know the schema…

S(he) needs to know a query language…

Proposed solutions

OpenLink Virtuoso
RelFinder
Lodlive
gFacet
DBPedia Mobile
Feceted Wikipedia Search (FWS)

DBpedia Blog | Get Involved | Get Help

About / News
Applications
Use Cases
Datasets
Online Access
DBpedia Live
Downloads
Interlinking
Development

DBpedia Applications

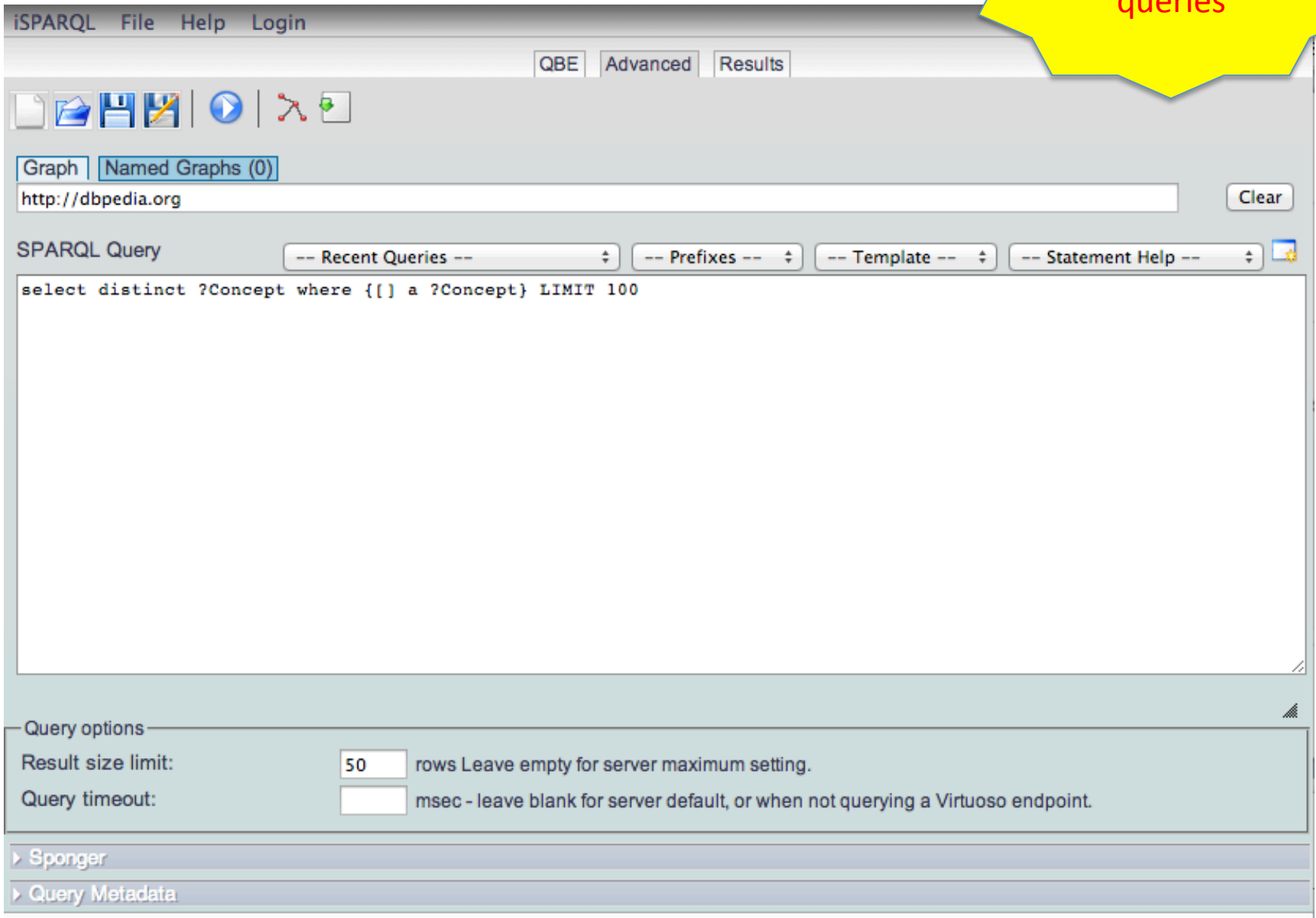
This page lists a number of applications (in no particular order) to get you started using DBpedia:

Contents

http://wiki.dbpedia.org/Applications

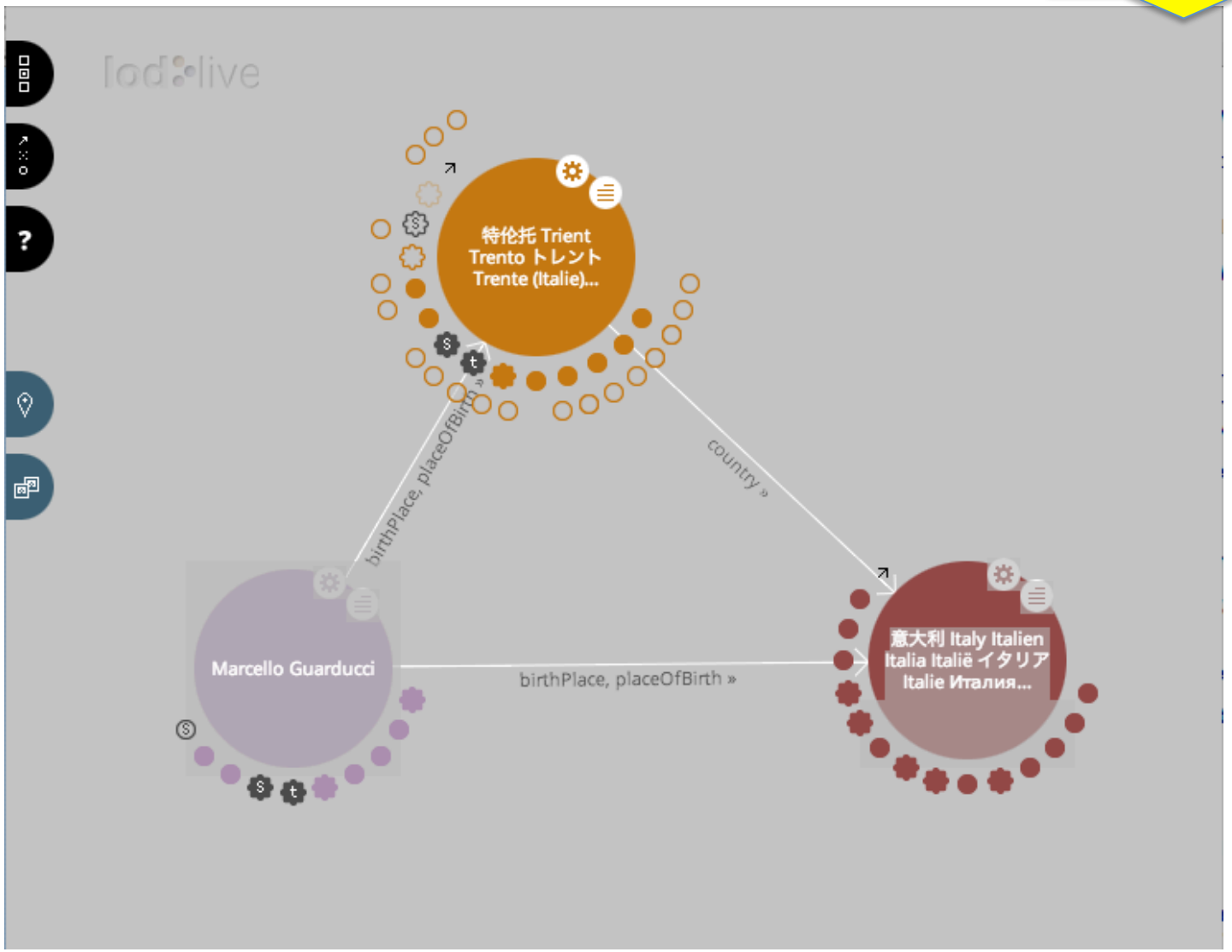
isparql

running queries



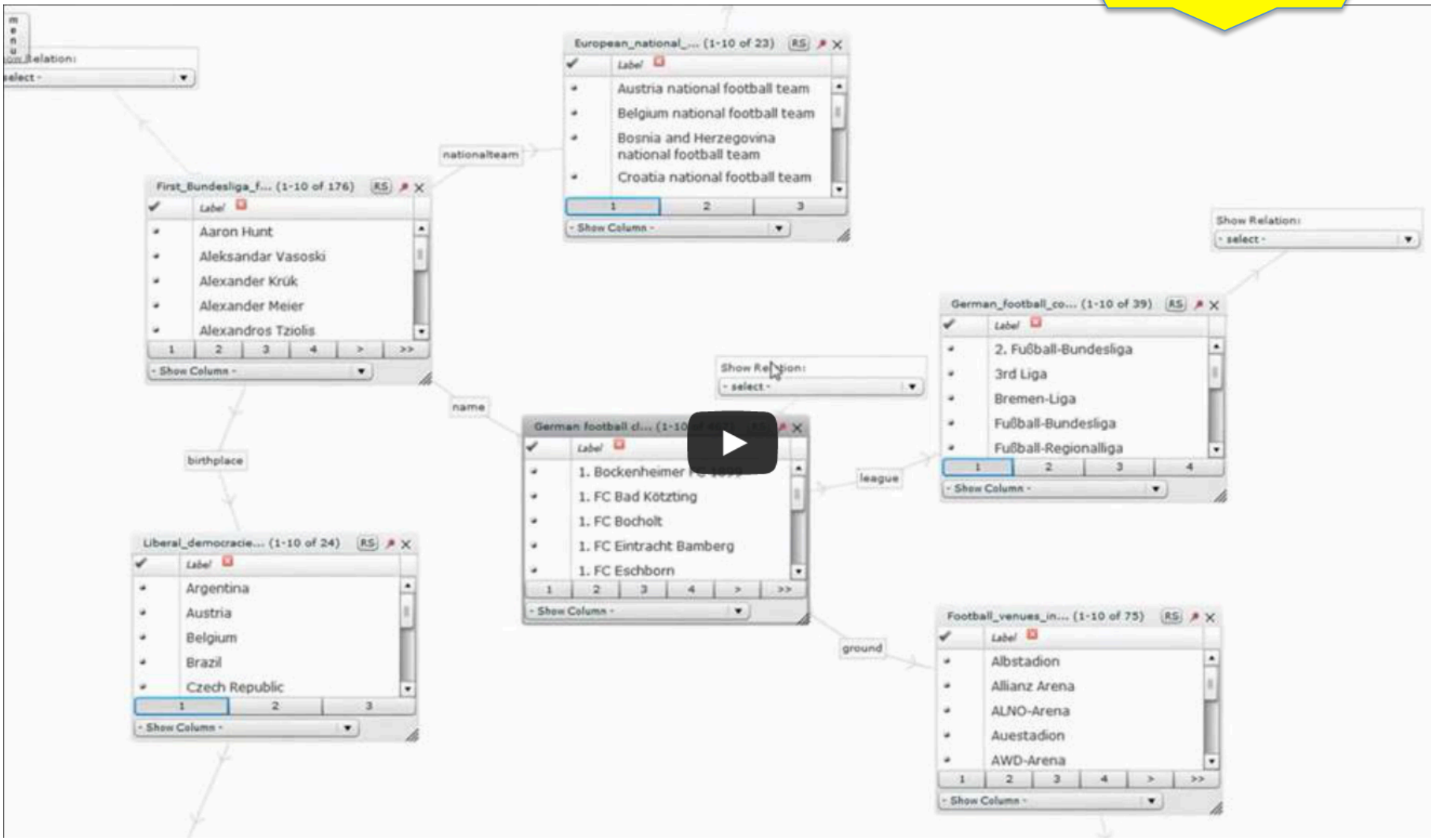
lodlive

linked data browsers



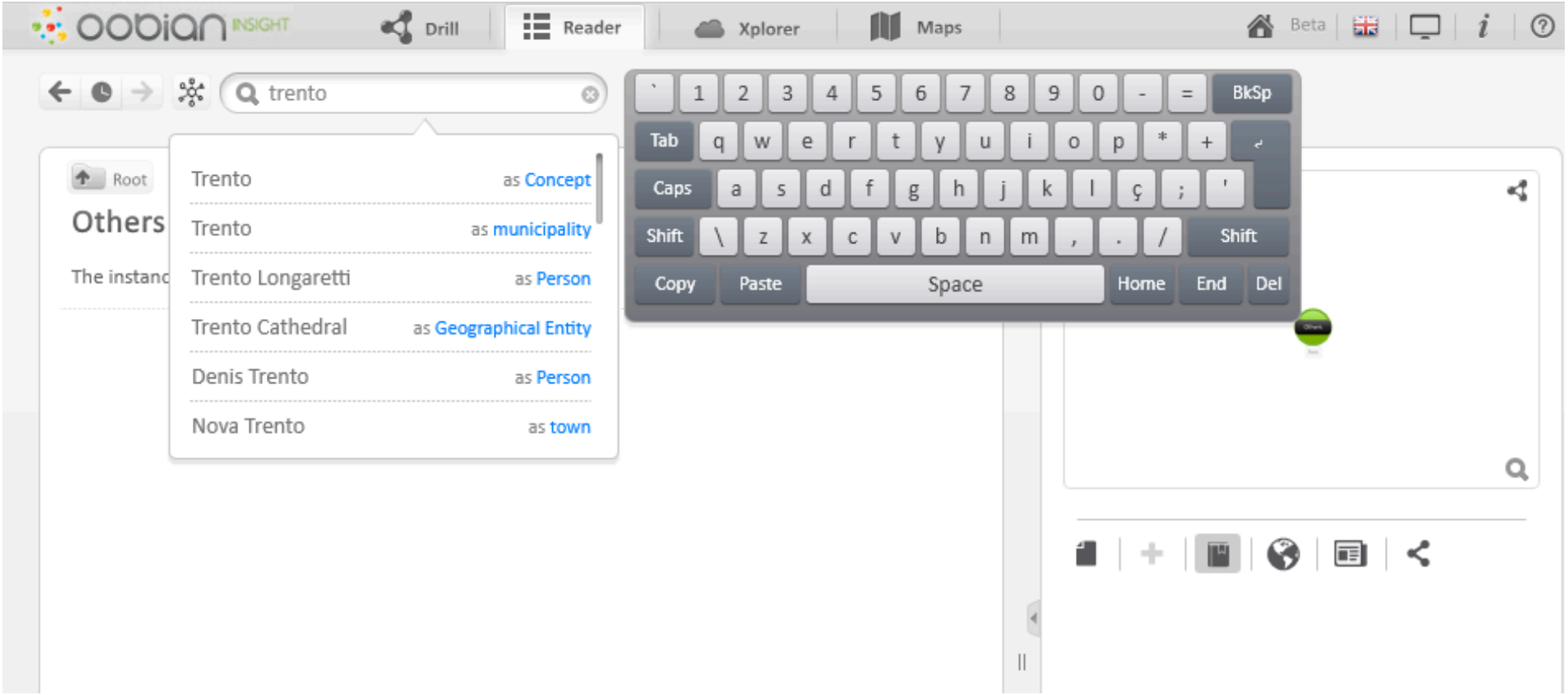
gfacet

exploring data and relations

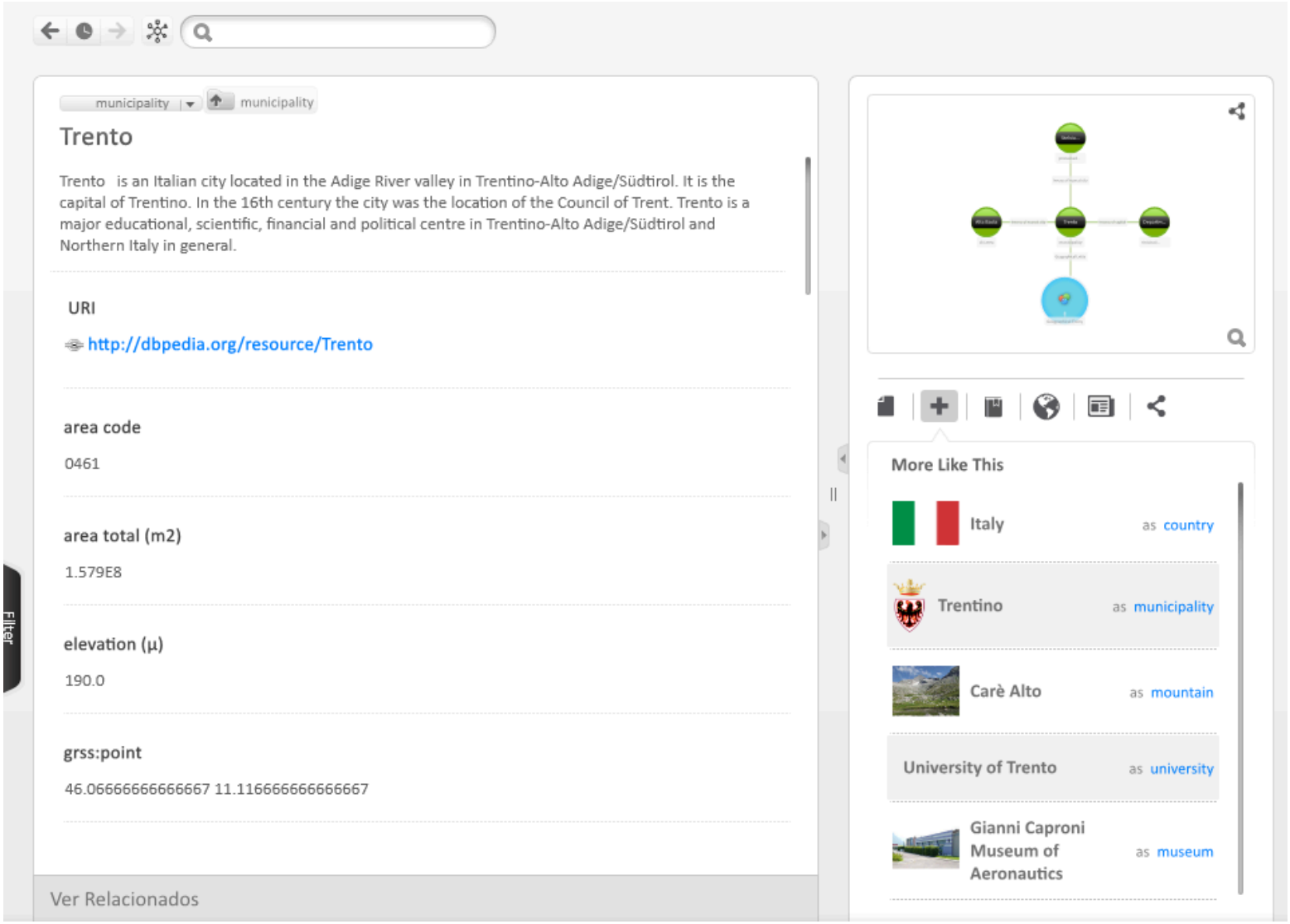


oobian

focusing on instances



oobian


municipality
municipality
Trento
Trento is an Italian city located in the Adige River valley in Trentino-Alto Adige/Südtirol. It is the capital of Trentino. In the 16th century the city was the location of the Council of Trent. Trento is a major educational, scientific, financial and political centre in Trentino-Alto Adige/Südtirol and Northern Italy in general.
URI
http://dbpedia.org/resource/Trento
area code
0461
area total (m2)
1.579E8
elevation (μ)
190.0
grss:point
46.06666666666667 11.116666666666667
Ver Relacionados
Filter
More Like This
Italy
as country
Trentino
as municipality
Carè Alto
as mountain
University of Trento
as university
Gianni Caproni Museum of Aeronautics
as museum

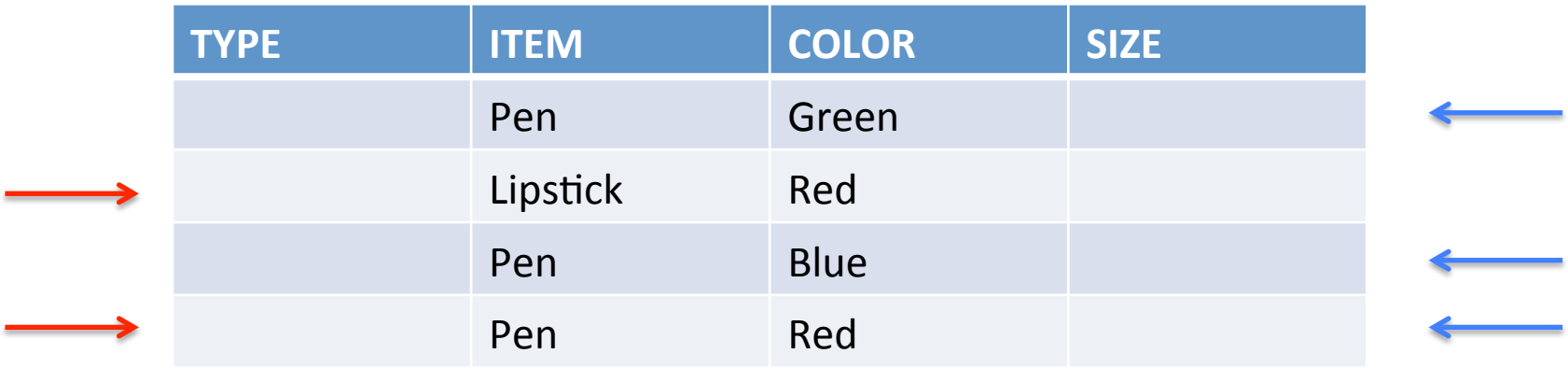
What do we really need?

Get closer to the final user!

Our proposal: QwwwQ

Get me something like this!

- Query by example
- Zloof 1975, relational data model



TYPE	ITEM	COLOR	SIZE
	Pen	Green	
	Lipstick	Red	
	Pen	Blue	
	Pen	Red	

ITEM=P.PEN COLOR=RED -> {Lipstick, Pen}

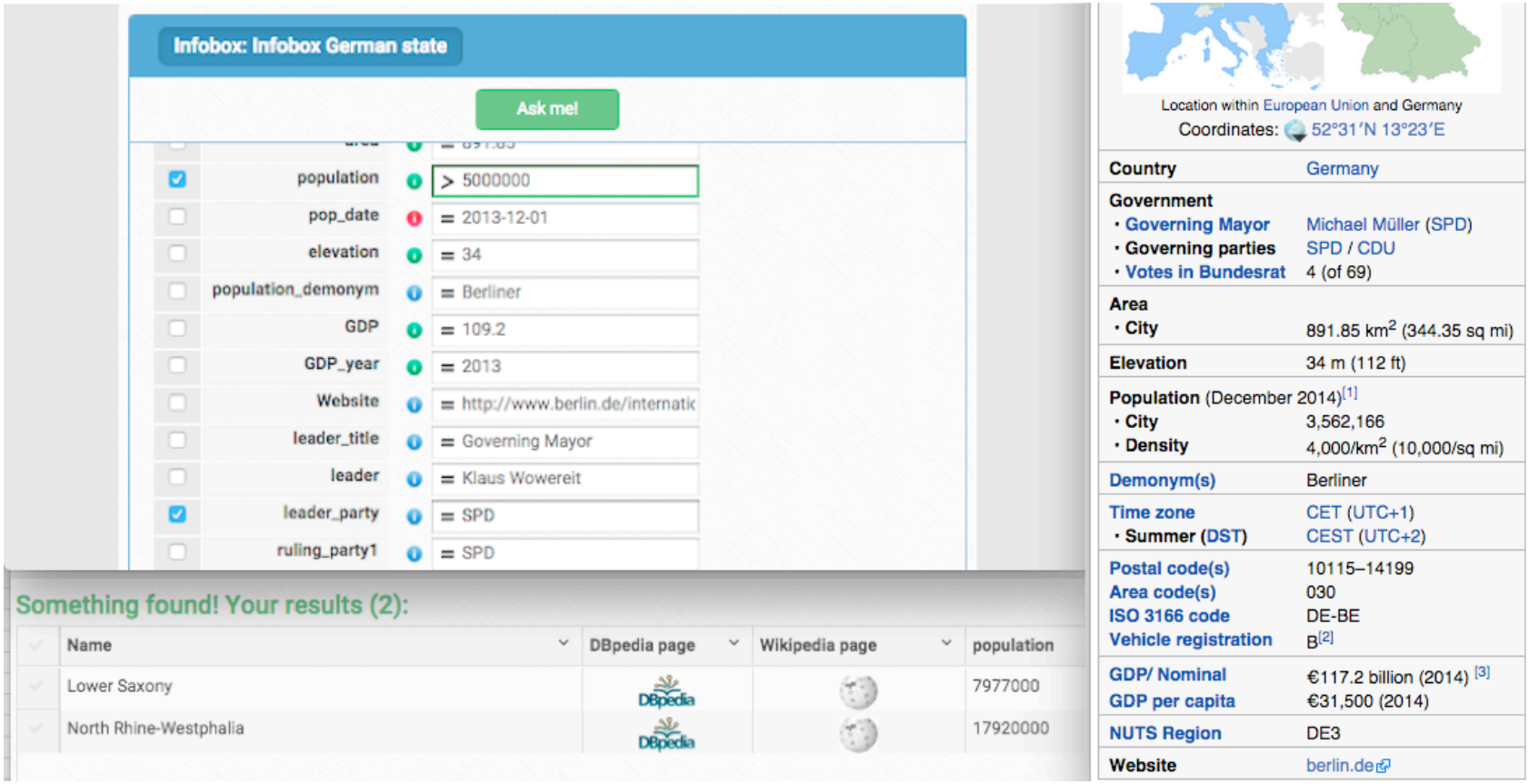
ITEM=PEN COLOR=P.RED -> {Green, Red, Blue }

Get me something like this!

User starts from a wikipedia page, and invokes a Chrome plug-in

- Infobox provides the schema
- User selects the variables of interest
- Add filtering and based on wikipedia categories, add ordering on parameters

Query on Berlin



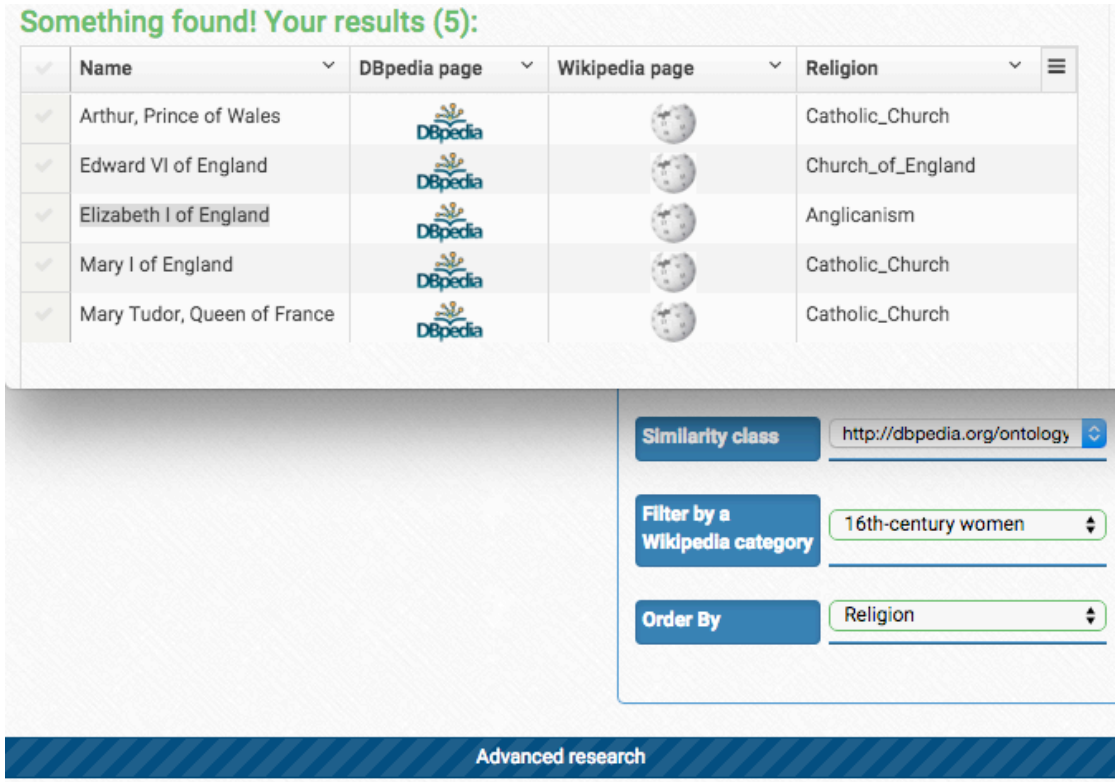
Output: DBPedia page, Wikipedia page

Get me something like this!

User starts from a wikipedia page, and invokes a Chrome plug-in

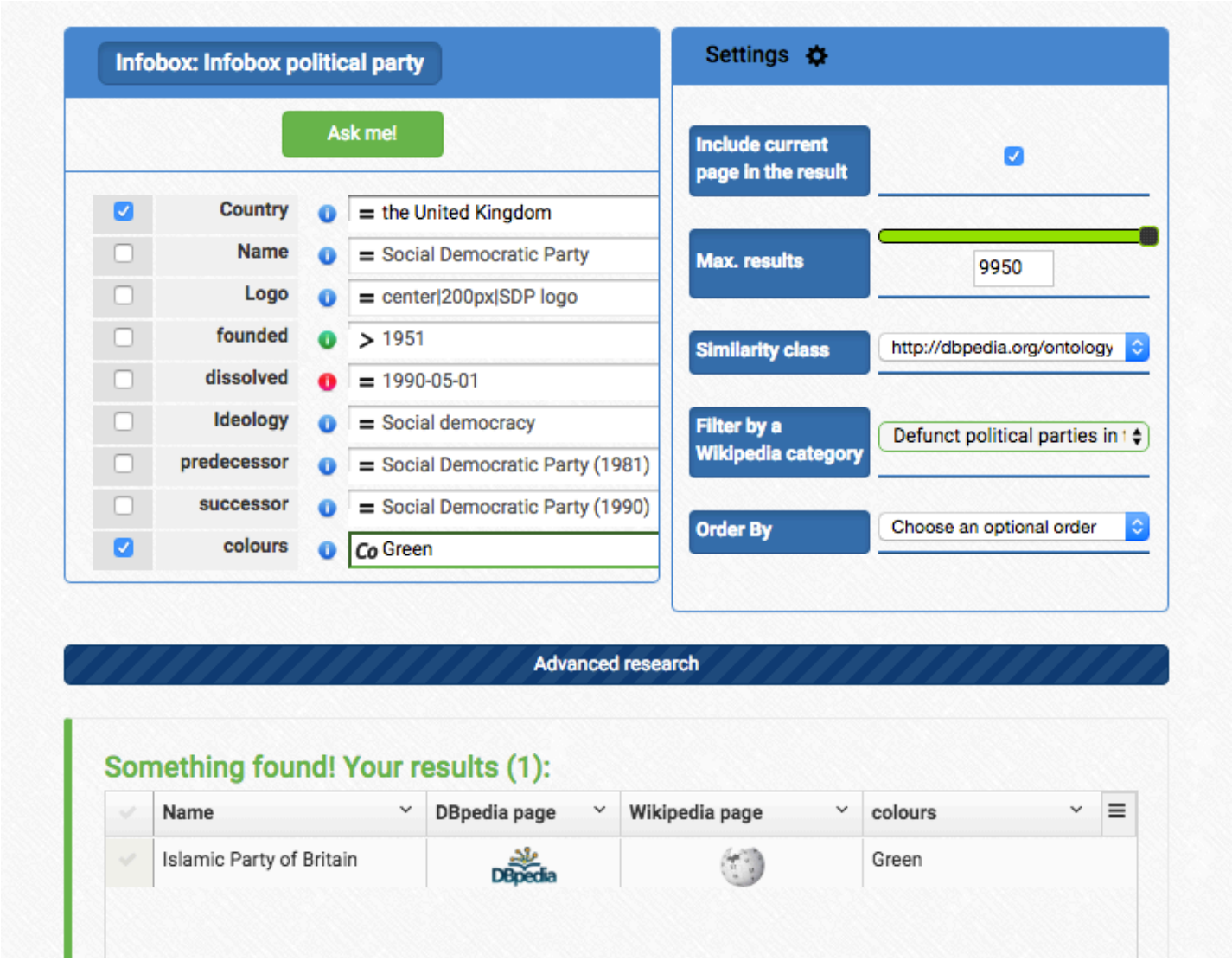
- Infobox provides the schema
- User selects the variables of interest
- Add filtering and based on wikipedia categories, add ordering on parameters

Queen Elizabeth I, House=House of Tudor, Religion=anything



exportable to csv, pdf

How many UK defunct political parties have green among their colors?



Implementation

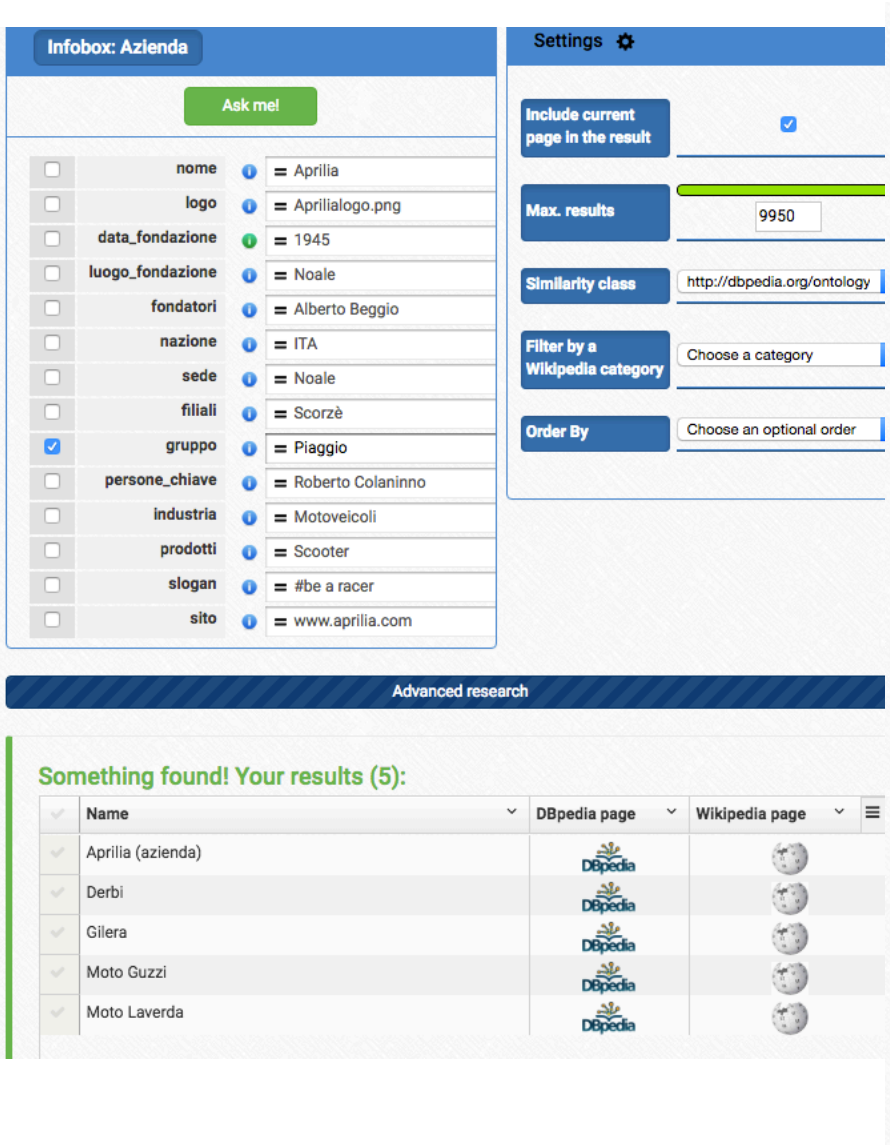
- core library:
  - wikipedia parser module (local copy of wp)
  - sparql query builder
- restful web services
  - callable via Ajax
- client code:
  - AngularJS
- browser embedding
  - Google Chrome extension

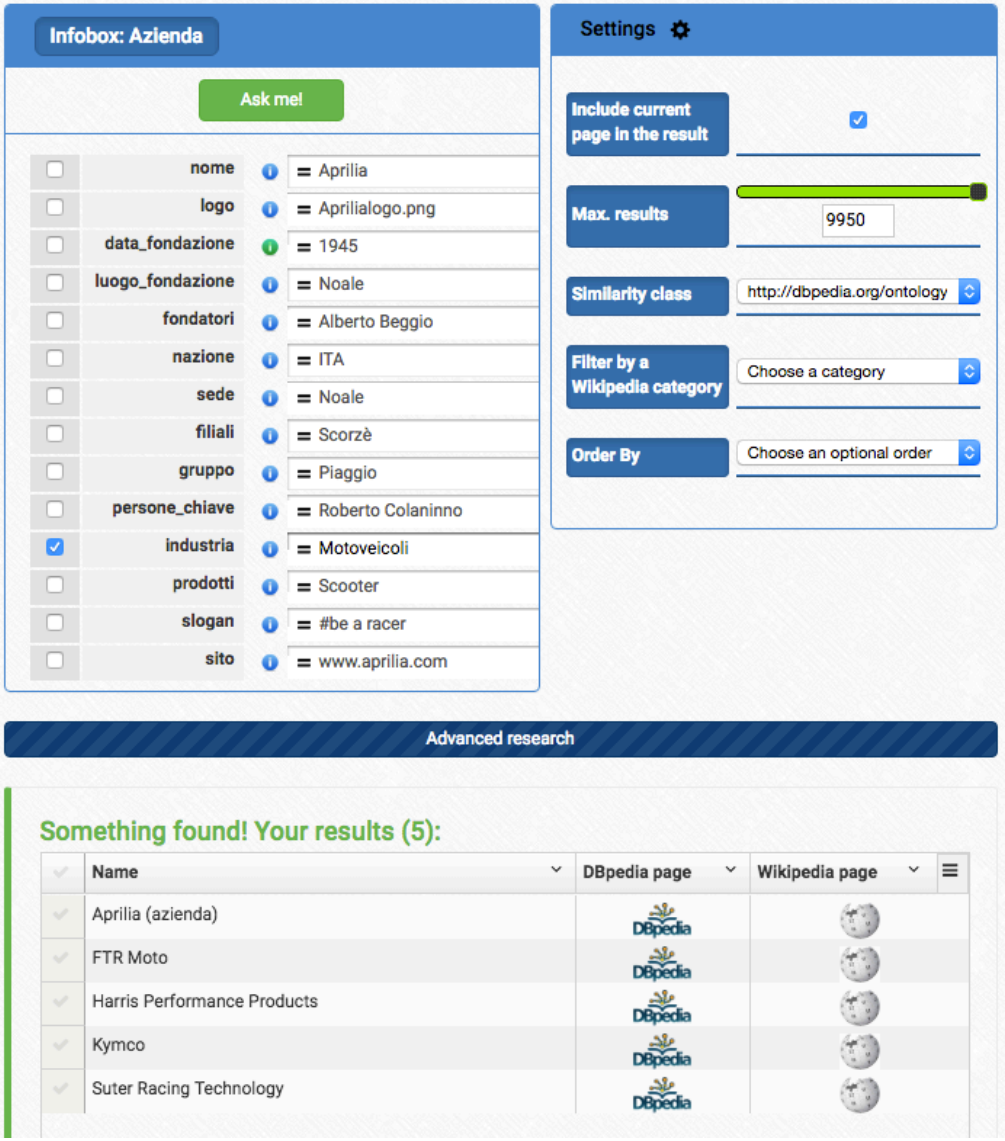
Problems

- mostly wikipedia-related:
  - Wikipedia in not complete, nor fully correct
  - presence of infoboxes
    - Berlin is a town or a a state?
  - Infoboxes have "lack of standard" problems:
    - terminology, units, or even comments
    - lists in infoboxes
  - Wikipedia categorization
    - incomplete: Berlin is not in the category of "University towns in Germany"
    - not even an acyclic graph!

people are good at collaborative writing,
not at collaborative creating categorizations

Get me something like Aprilia…





Inconsistencies..

Gilera & C. SpA

GILERA®

Stato	Italia
Tipo	Società per azioni
Fondazione	1909 a Milano in Corso XXII Marzo
Sede principale	Pontedera
Gruppo	Piaggio
Settore	Casa motociclistica
Prodotti	motocicli
Slogan	«Nata dal desiderio»
Sito web	www.it.gilera.com

Modifica dati su Wikidata · Manuale

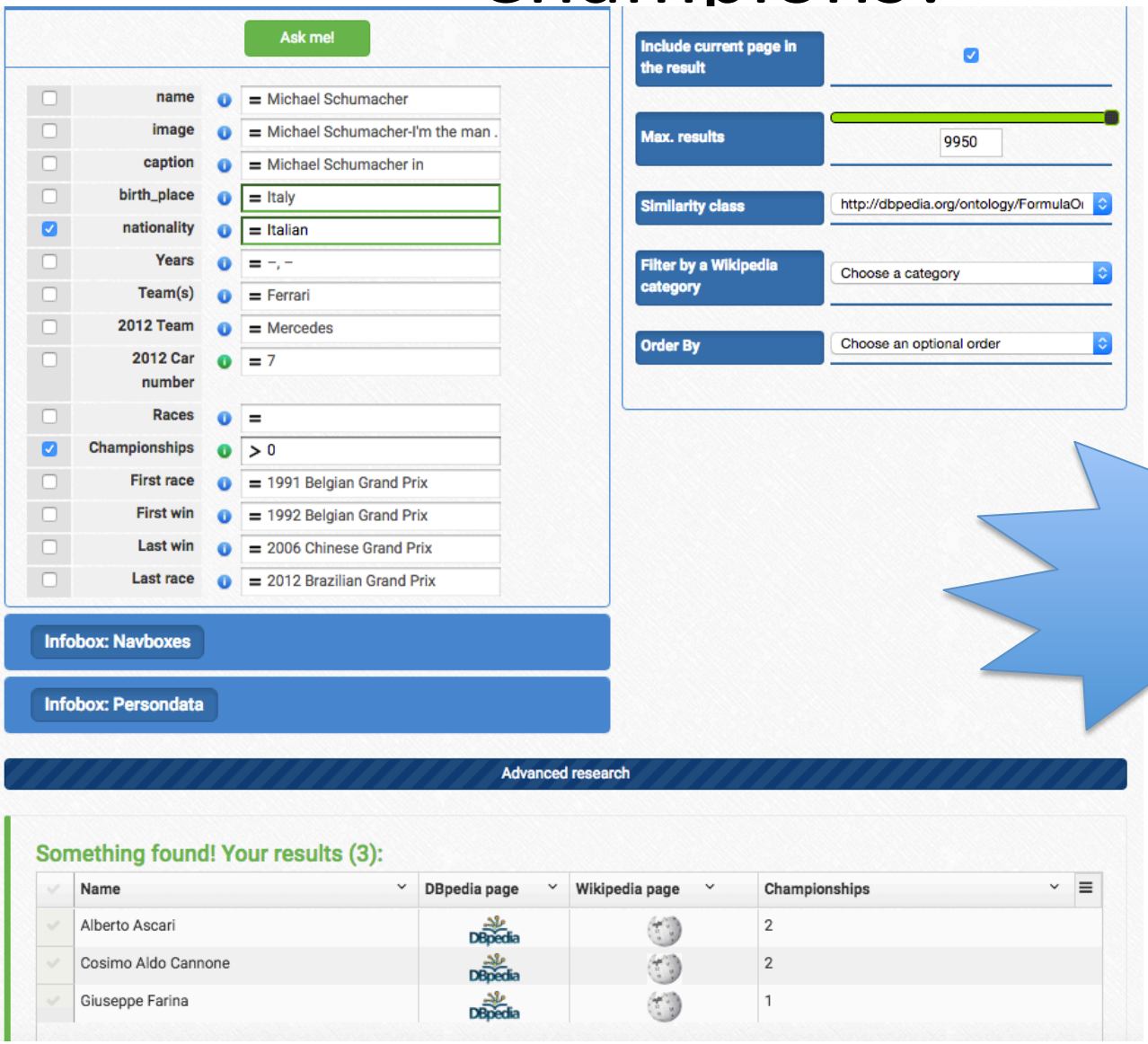
Aprilia

aprilia

Stato	Italia
Fondazione	1945 a Noale
Fondata da	Alberto Beggio
Sede principale	Noale
Gruppo	Piaggio
Filiali	Scorzè
Persone chiave	Roberto Colaninno (presidente)
Settore	Motoveicoli
Prodotti	Motociclette Scooter
Slogan	«#be a racer»
Sito web	www.aprilia.com/

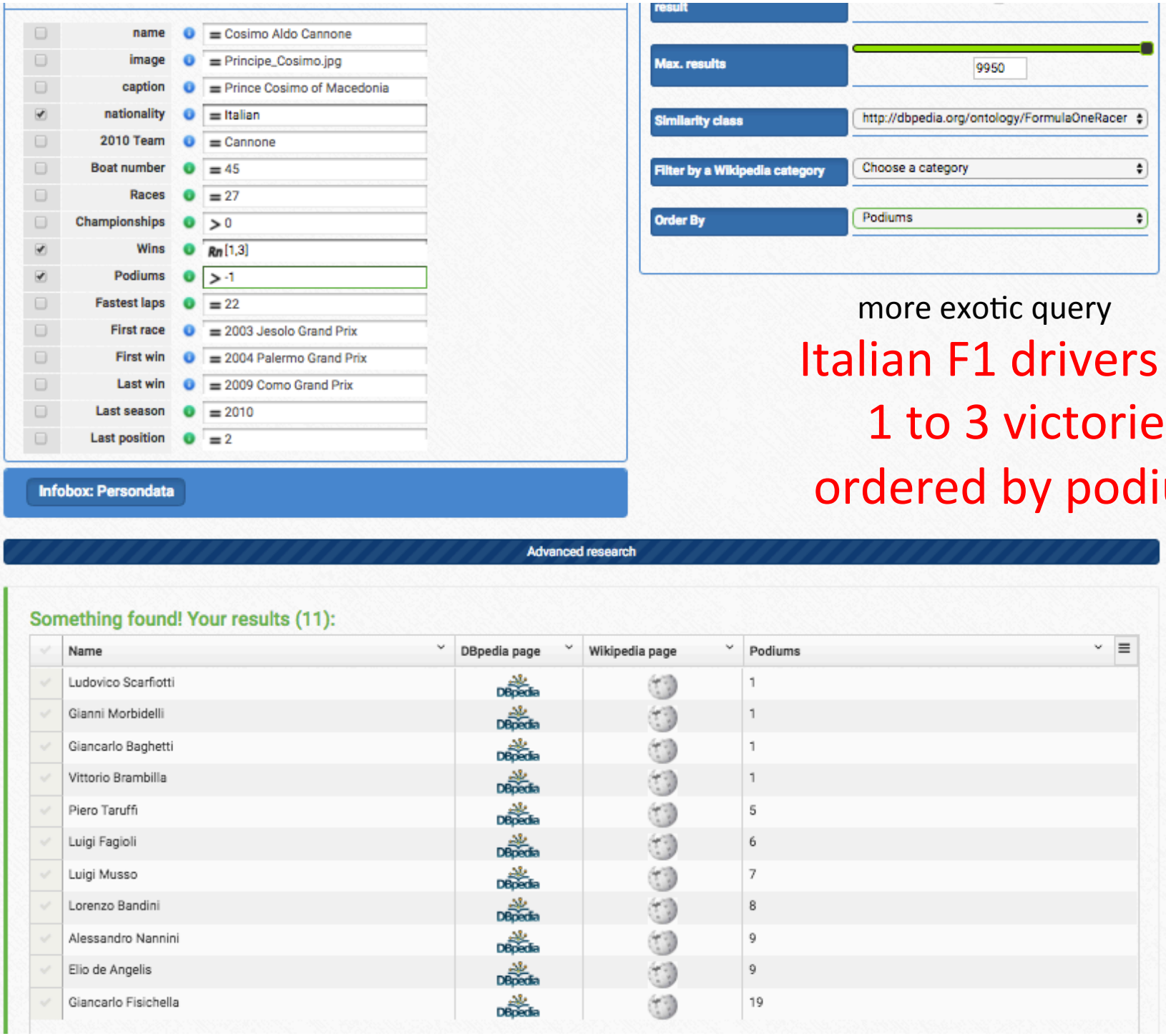
Modifica dati su Wikidata · Manuale

Which Italian F1 drivers became World Champions?



Answer present in wikipedia

Spotted a wikipedia error!



more exotic query

Italian F1 drivers with 1 to 3 victories, ordered by podiums

Problems are non only ours…

- Same problems affect DBPedia, and hence LOD…
- Some effort could be done to mitigate them
  - e.g., recognize unusable properties

Spacetime: a Two Dimensions Search and Visualisation Engine Based on Linked Data

Fabio Valsecchi and Marco Ronchetti
DISI, Università degli Studi di Trento
Povo di Trento, Italy

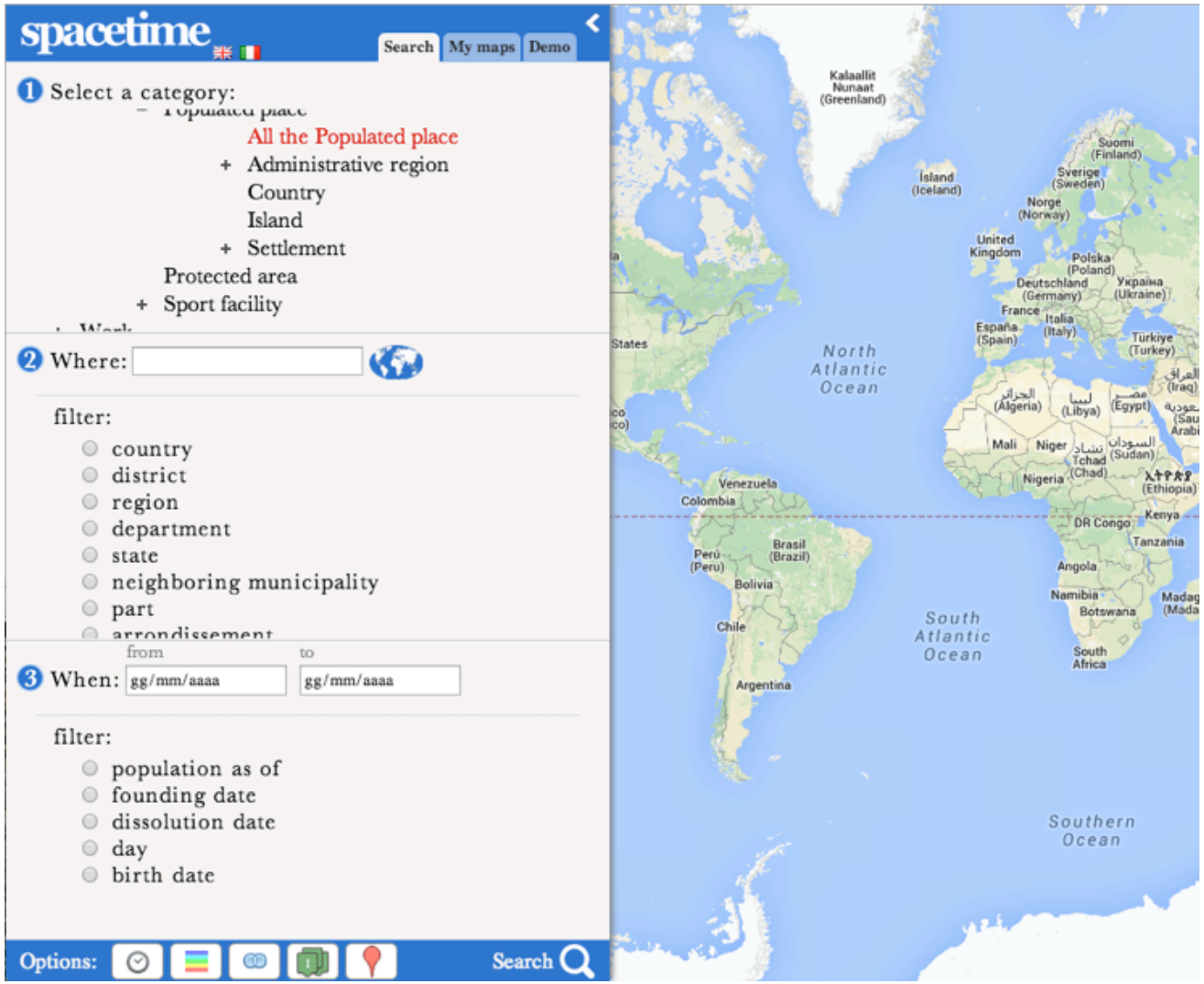
Our proposal: Spacetime

- WHAT? Set of DBPedia entities having spatial and temporal data

(*Organisation*, *Person*, *Event*, *Place*, *Species* and *Work*)

- WHERE? Attributes having geocoordinates
- WHEN? Attributes having data type *xsd:date*.

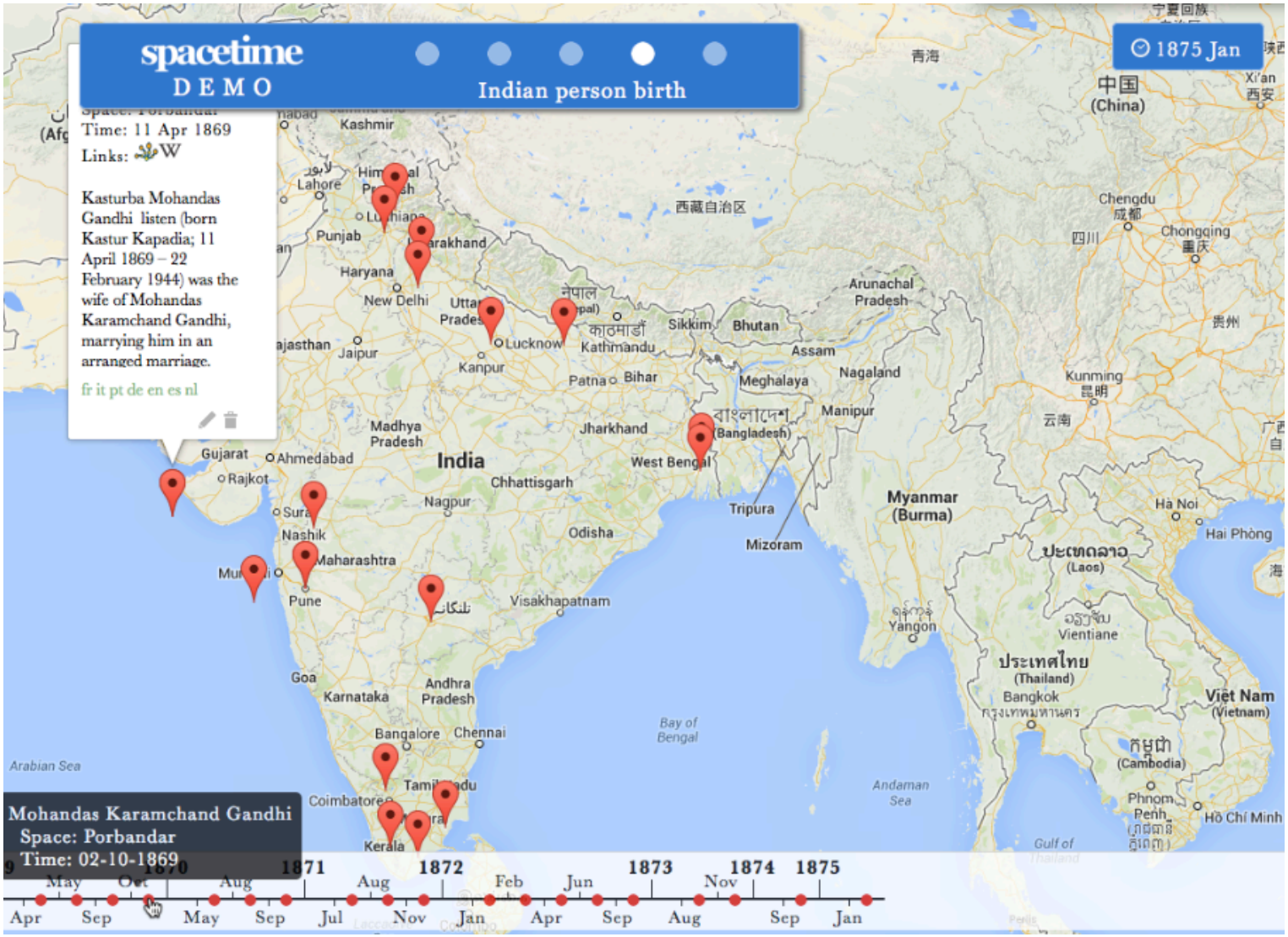
Formulating queries



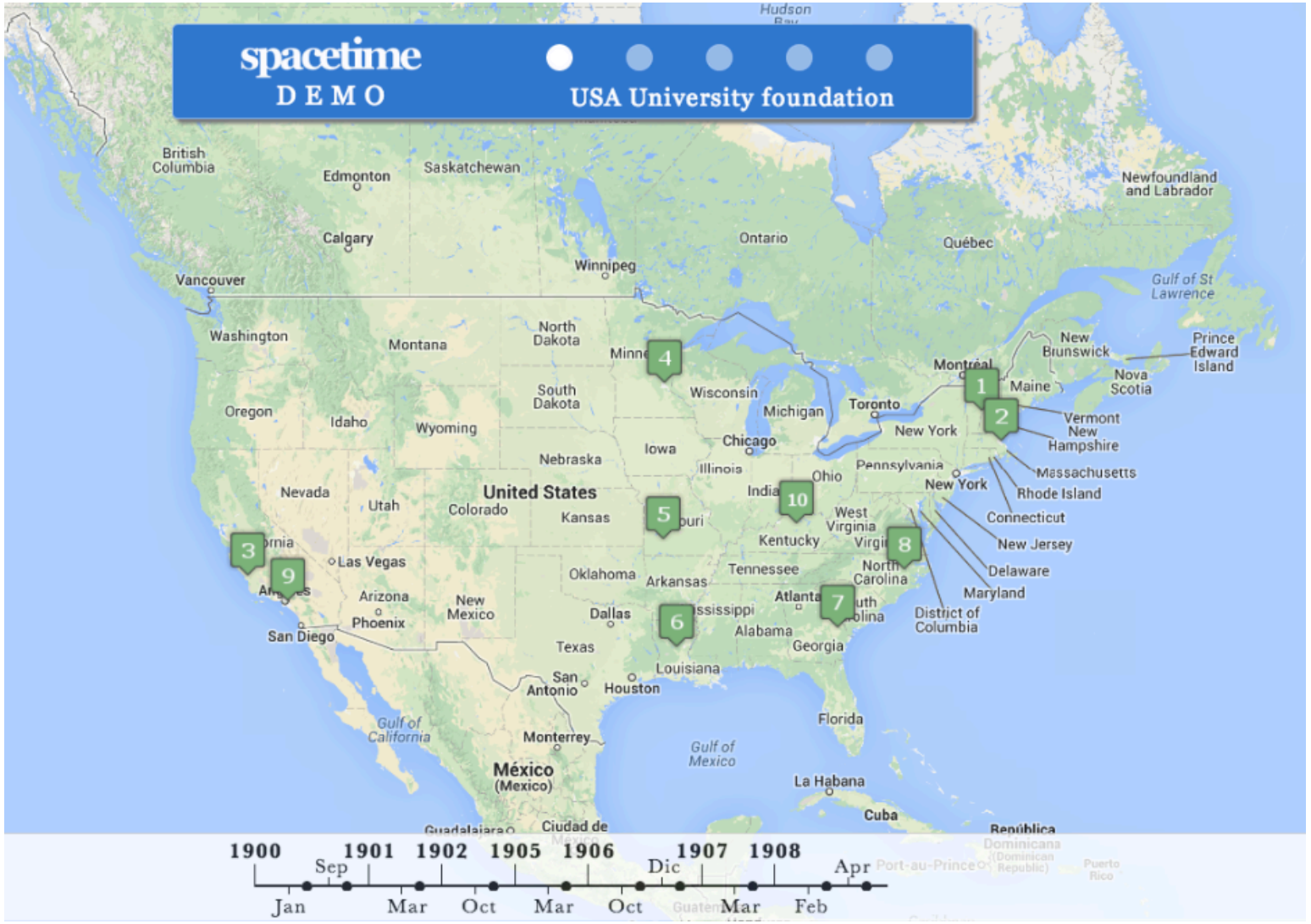
Geographical approximation

- Bounding box of the geoentity

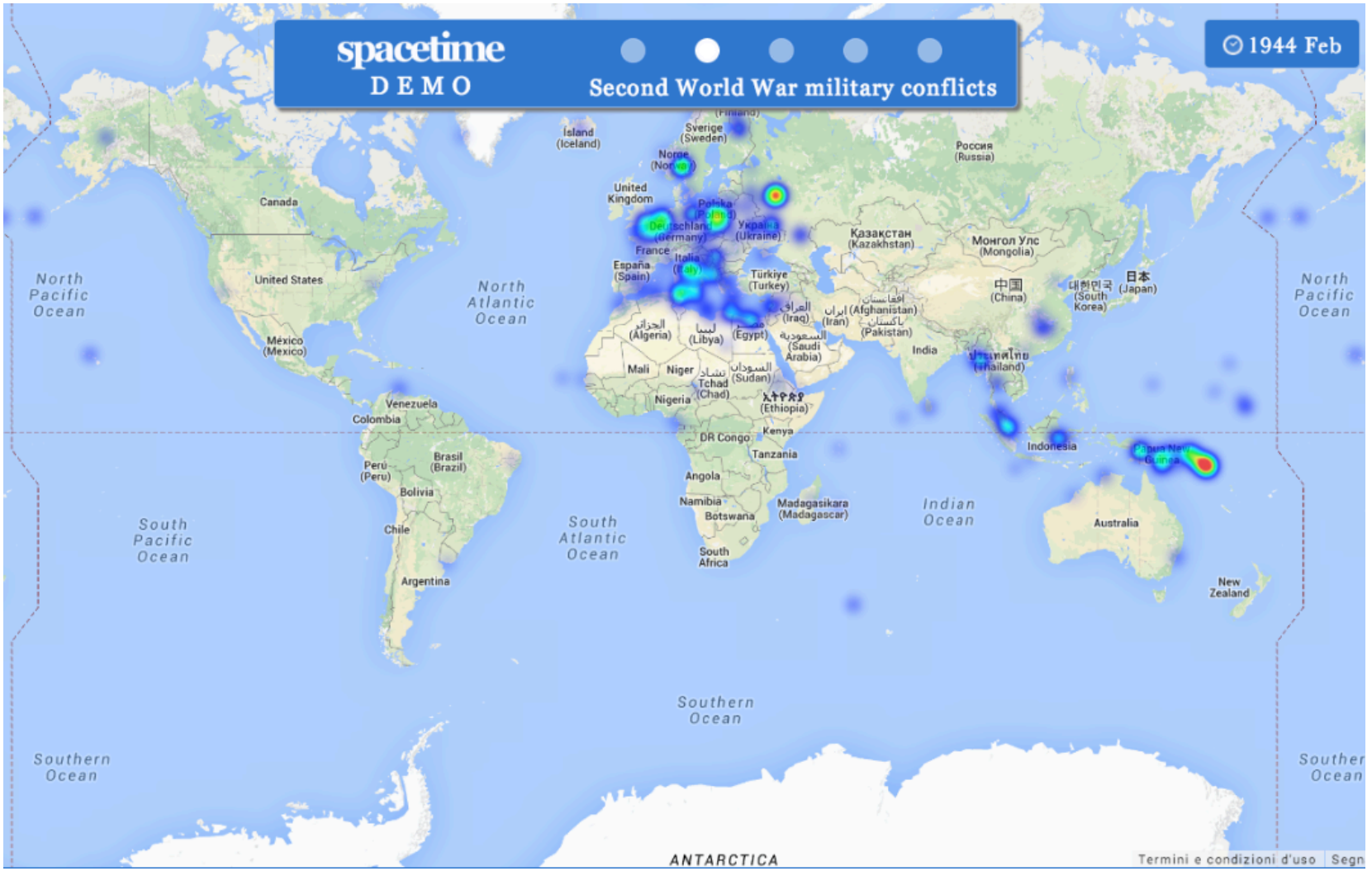
Query results



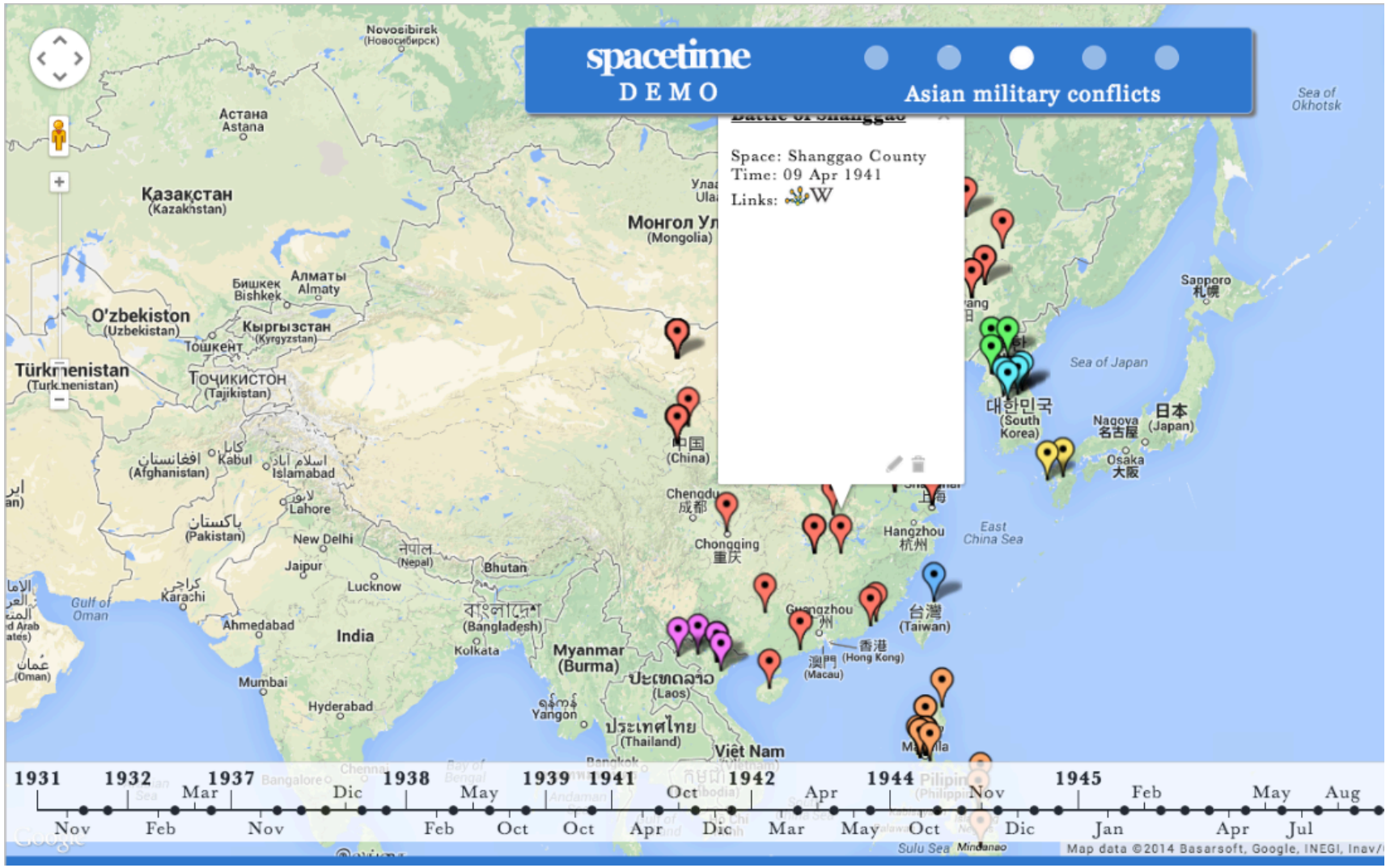
Sequence of results



Density maps



Grouping different queries



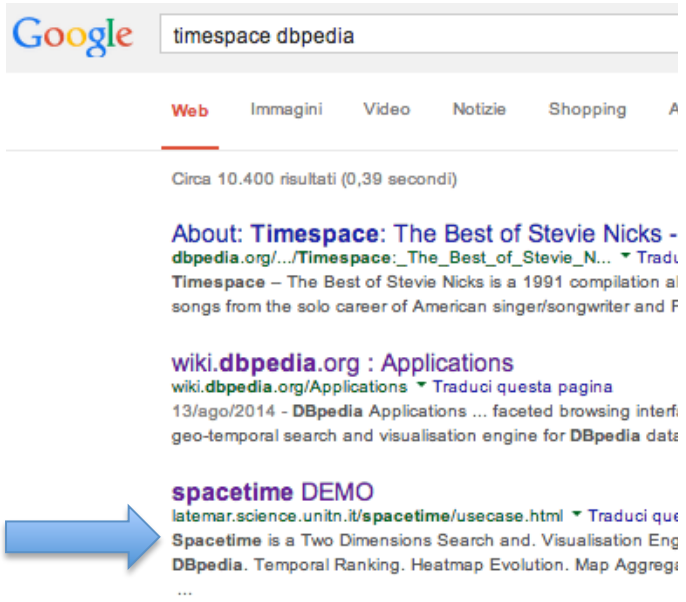
Creating a story

Technologies

- SPARQL: Queries are composed by the Javascript engine, and are executed through the SPARQL endpoint;
- JSON: the results of the SPARQL queries are returned as JSON strings;
- Google Maps JavaScript API v.3: the Google Maps API used for populating a map with the data extracted in the JSON file
- JavaScript and JQuery library: the scripting language and its library define a set of functions that are the core of the application JQuery allows the creation of animations inside Spacetime;
- AJAX: this technology is used to have a responsive user interface compliant with the Rich Internet Application paradigm;
- CSS: for designing the graphical aspect of Spacetime;
- HTML5: used for developing certain parts of the application, such as the map saving operation, and some graphical feature.

Conclusion

- The system is up and running
  - (google for Timespace DBPedia)



- Included in the DBPedia front end demos

  http://wiki.dbpedia.org/Applications

- No user assessment done (yet)