



FactForge

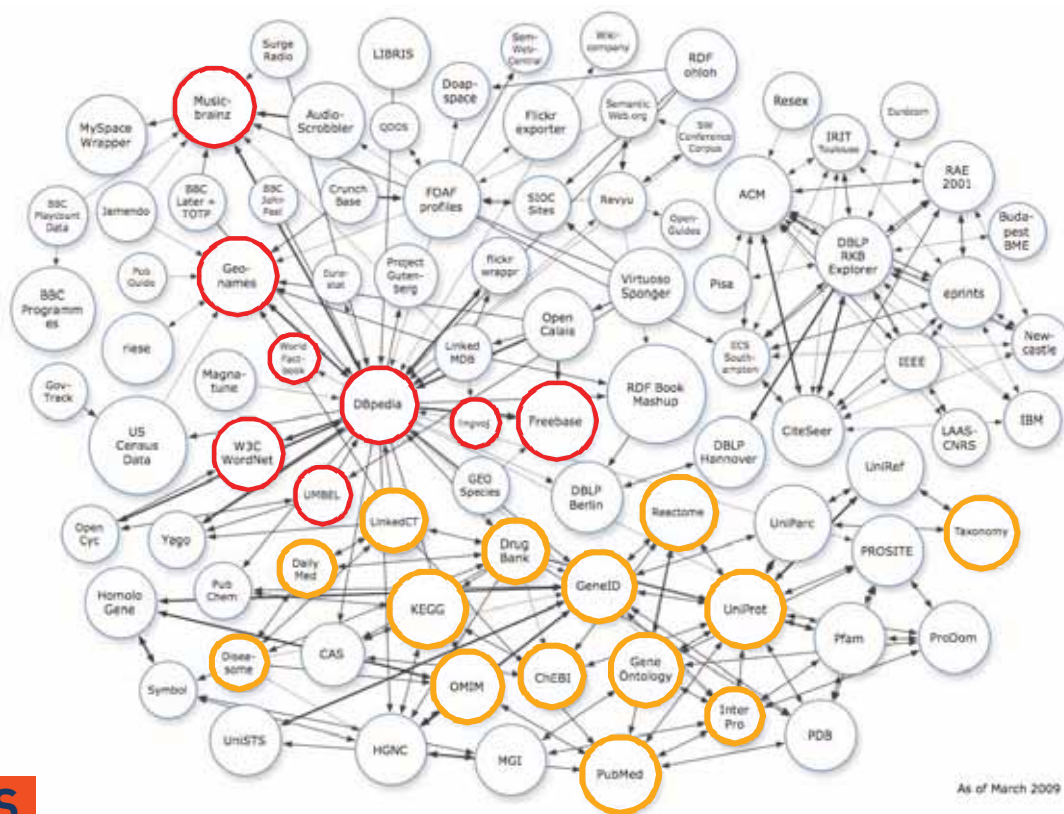
The Fast Track To The
Centre of the Data Web

THE DATA

Datasets included: DBPedia, Geonames, UMBEL, Wordnet, Freebase, CIA World Factbook, Lingvoj, MusicBrainz (marked in red on the LOD map below)

Ontologies: DC, SKOS, FOAF, RSS and the dataset's proprietary ontologies

Size: 1.2B explicit and 0.8B inferred statements were indexed;
the total number of retrievable statements is **10 billion**



THE ACCESS

Go-to-resource: incremental URI auto-suggest

Keyword search: RDF Search, returning a ranked list of RDF snippets

Exploration: traversing the data, one resource at a time

Structured queries: evaluation of queries in SPARQL

Remote server access: SPARQL end-point

THE MODIGLIANI TEST

FactForge (previously known as LDSR) was the first system to provide a solution to the test defined by Richard MacManus, the founder of the ReadWriteWeb:

“...the **tipping point for the Semantic Web** may be when one can ... deliver – using Linked Data – a comprehensive list of locations of original Modigliani art works around the world”

http://www.readwriteweb.com/archives/the_modigliani_test_for_linked_data.php

THE QUERY

```
PREFIX ...
SELECT DISTINCT ?painting_1 ?owner_1 ?city_fb_con ?city_db_loc ?city_db_cit
WHERE {
  ?p fb:visual_art.artwork.artist dbpedia:Amedeo_Modigliani ;
  fb:visual_art.artwork.owners [
    fb:visual_art.artwork_owner_relationship.owner ?ow ] ;
  ff:preferredLabel ?painting_1.

  ?ow ff:preferredLabel ?owner_1 .
  OPTIONAL { ?ow fb:location.location.containedby [ ff:preferredLabel ?city_fb_con ] } .
  OPTIONAL { ?ow dbp-prop:location ?loc.
    ?loc rdf:type umbel-sc:City ; ff:preferredLabel ?city_db_loc }
  OPTIONAL { ?ow dbp-ont:city [ ff:preferredLabel ?city_db_cit ] }
}
```

THE RESULT



RDF Search and Explore | SPARQL Query | Refinder | About | Contact

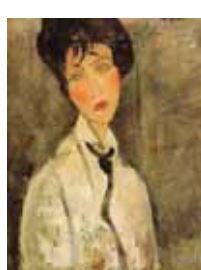
SPARQL Query

Results for PREFIX fb: <http://rdf.fr...> (8)

View as [Exhibit](#) | Download in [JSON](#) | [SPARQL Results in XML](#) | [SPARQL Results in JSON](#)

painting_1	owner_1	city_fb_con	city_db_loc	city_db_cit
Head@en	Museum of Modern Art	New York City		
Anna Zborowska@en	Museum of Modern Art	New York City		
Portrait of Diego Rivera@en	The São Paulo Museum of Art@en		São Paulo	
Woman with a Necklace@en	School of the Art Institute of Chicago@en			Chicago
Portrait of a Woman@en	School of the Art Institute of Chicago@en			Chicago
Reclining Nude@en	Museum of Modern Art	New York City		
Madam Pompadour@en	School of the Art Institute of Chicago@en			Chicago
Jeanne Hébuterne@en	Barnes Foundation@en	Philadelphia		

© 2009-2010 Ontotext AD. All rights reserved.



Unique Query Capabilities

THE MOST POPULAR ENTERTAINER

FactForge allows one to ask general questions against a dataset integrating various types of knowledge

THE QUESTION: Who is the most popular entertainer born in Germany?

The corresponding SPARQL query is available as a sample at <http://FactForge.net/sparql>

THE RESULT:

FactForge RDF Search and Explore | SPARQL Query | Refinder | About | Contact

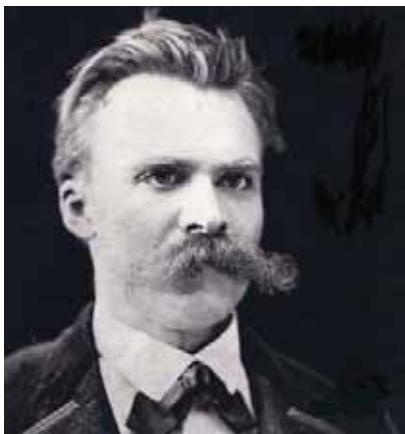
SPARQL Query

Results for PREFIX rdf: <http://www.w... (100) View as [Exhibit](#) Download in [JSON](#) | [SPARQL Results in XML](#) | [SPARQL Results in JSON](#)

Person	BirthPlace	RR
dbpedia:F. W. Nietzsche	dbpedia:Röcken	0.21
dbpedia:Heidi Klum	dbpedia:Bergisch Gladbach	0.20
dbpedia:Josef Beuys	dbpedia:Krefeld	0.20
dbpedia:Leni Riefenstahl	dbpedia:Bereullin	0.20
dbpedia:Rainer Werner Fassbinder	dbpedia:Bad Wörishofen	0.20
dbpedia:Wolf Biermann	dbpedia:Free City of Hamburg	0.17

RESULT ANALYSIS:

- **MusicBrainz** brings the little known fact that Nietzsche was a composer and piano player
- **DBPedia** brings the place of birth, **GeoNames** - the nesting of the locations
- **UMBEL** links MusicArtist from Musicbrainz with the class hierarchy in **UpperCyc**
- **Inference** over types, classes, **owl:sameAs** and transitive relationships is required
- Popularity is judged through the **RDFRank** in the FactForge integrated dataset
- We are asking for the most popular person, who qualifies as entertainer, not for the one who is most popular as an entertainer
- Without a reason-able view, such as FactForge, answering such queries in real time is impossible



Friedrich Nietzsche!



Asking factual questions to a global knowledge base can bring unexpected and strange, but formally correct results
– one should make precise queries and consider the context



FactForge = Manageable Linked Data

Linking Open Data (LOD)

LOD is a W3C SWEO Community Project, which aims to facilitate the emergence of a web of linked data, by means of publishing and interlinking open data.

<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

Reason-able Views to LOD

The main **challenges** for the adoption of the linked data are that:

- LOD is **hard to comprehend**, making structured queries against 200 different schemata is tough
- LOD is generally **not reliable**, no consistency is guaranteed
- Querying data distributed on the Web is **slow**, because the “remote joins” are slow

The **reason-able views** concept:

- Makes it **easier and less risky** to use **part of the LOD data for specific purposes**
- Selects, cleans up and integrates selected datasets and ontologies in a compound dataset
- Loads the compound dataset in a **single semantic repository**

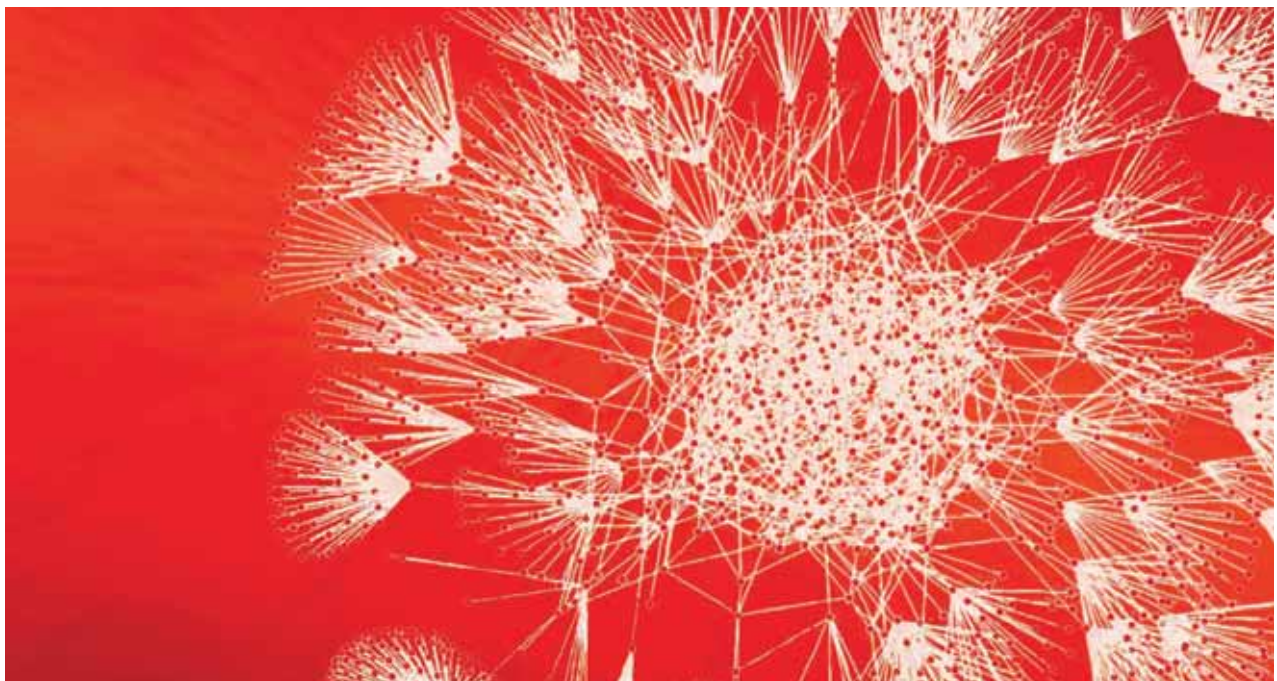
FactForge is a reason-able view including several of the central datasets of the LOD project.

OWL Horst reasoning is performed to “materialize” the facts that could be inferred from these data.

Sample Queries

An extensive set of sample queries available at FactForge aims to:

- **Guarantee data consistency** in the same way in which unit tests guarantee software quality
- **Lower the cost of entry**, demonstrating useful patterns of joining data from multiple datasets



The Digital Dandelion

Source: San Diego Supercomputer Center, UC San Diego

<http://www.FactForge.net>