# SemFacet: Semantic Faceted Search

B.Cuenca Grau, E. Kharlamov, E. Sherkhonov, D. Zheleznyakov
M. Arenas
L. Pierre
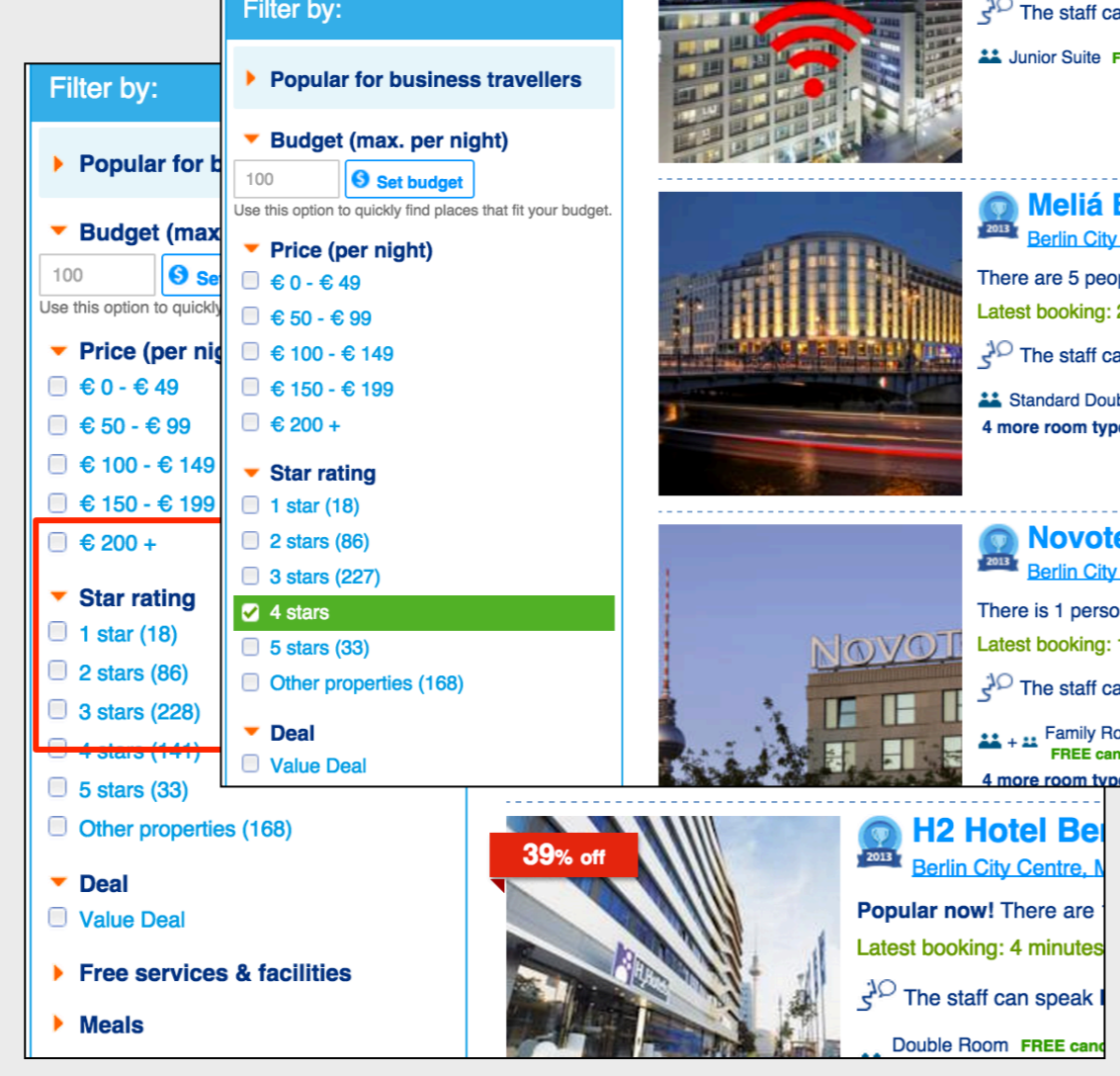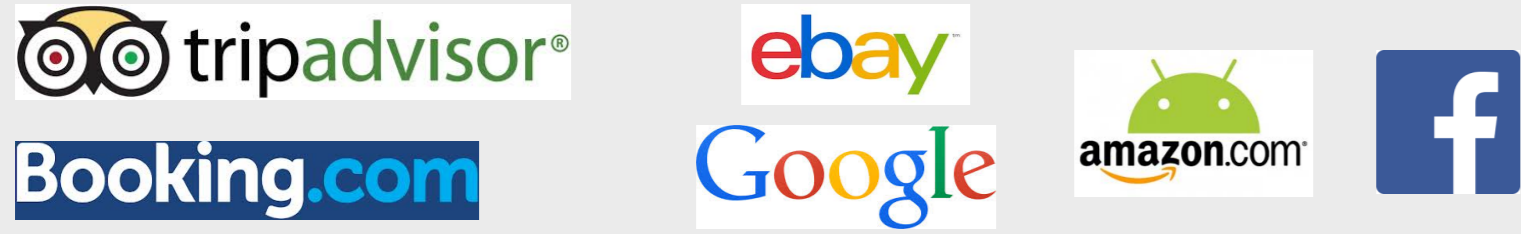
University of Oxford
Catholic University of Chile
Electicite de France

## Faceted Search

**De-facto standard for e-comerce**
- Query formulation for naïve users
- A facet = query constructor
  - Name
  - Set of values
- Facets in action
  - Choose a value
  - Restrict search result

tripadvisor  ebay  Booking.com  Google  amazon.com  facebook

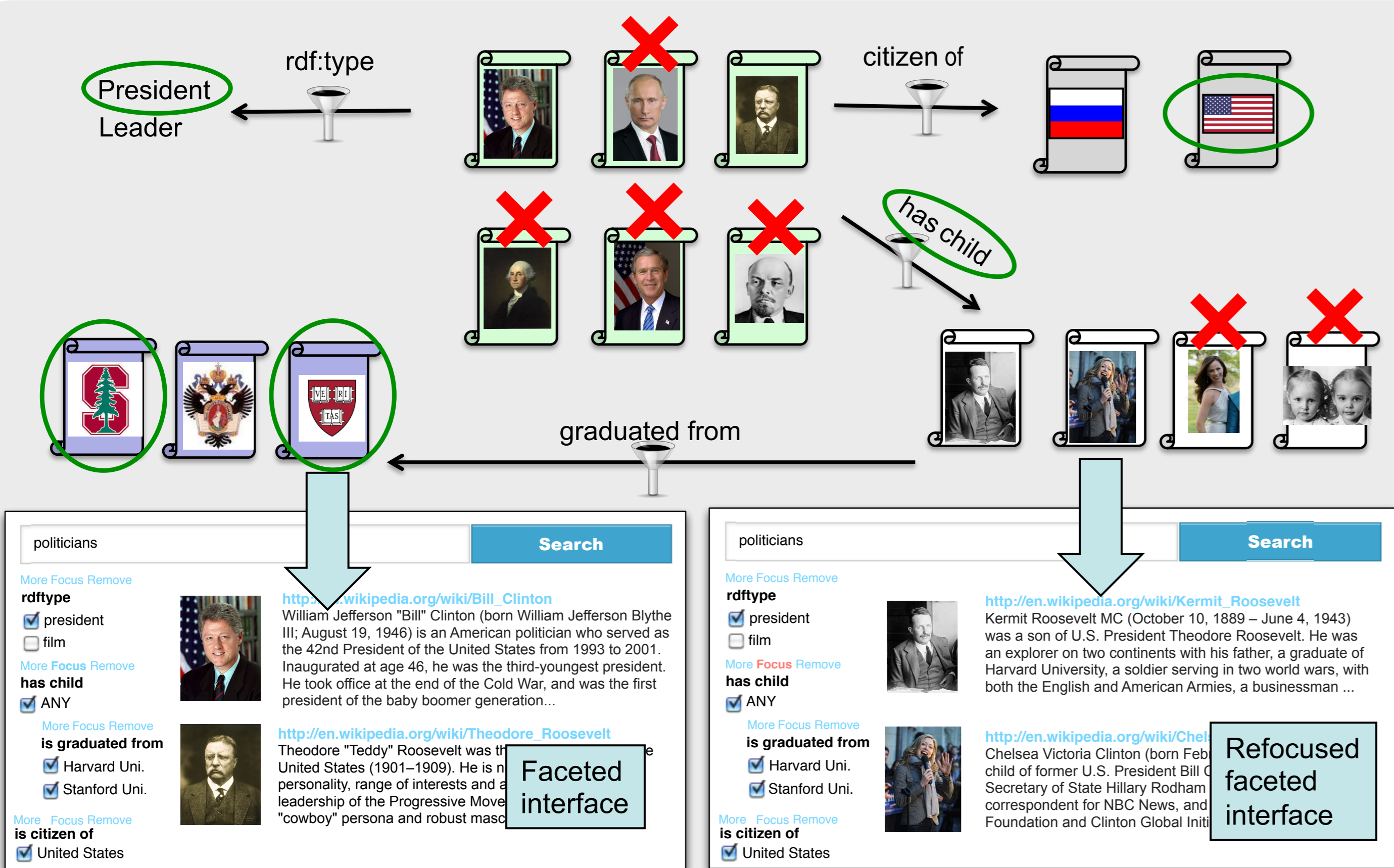## Access to Semantic Data Repositories

**Semantic Technologies**
- Gain importance in various domains
- OWL 2 Ontologies are used for
  - capturing industrial infor models ISA, ICE
  - modeling medical vocabularies
  - capturing structure of energy plants
- RDF data is used
  - as universal data exchange format
- SPARQL queries to access semantic data

**End User**

**Challenge:** End user access semantic data

## Semantic Faceted Search



**Faceted Search**

**Semantic Faceted Search**

**Data model**
- Annotated documents
- No interconnection

**Structure**
- Hierarchies on facets

**RDF and OWL 2 data model**
- Graph data
- Complex data statements

**OWL 2 ontologies**
- Hierarchies
- Complex axioms

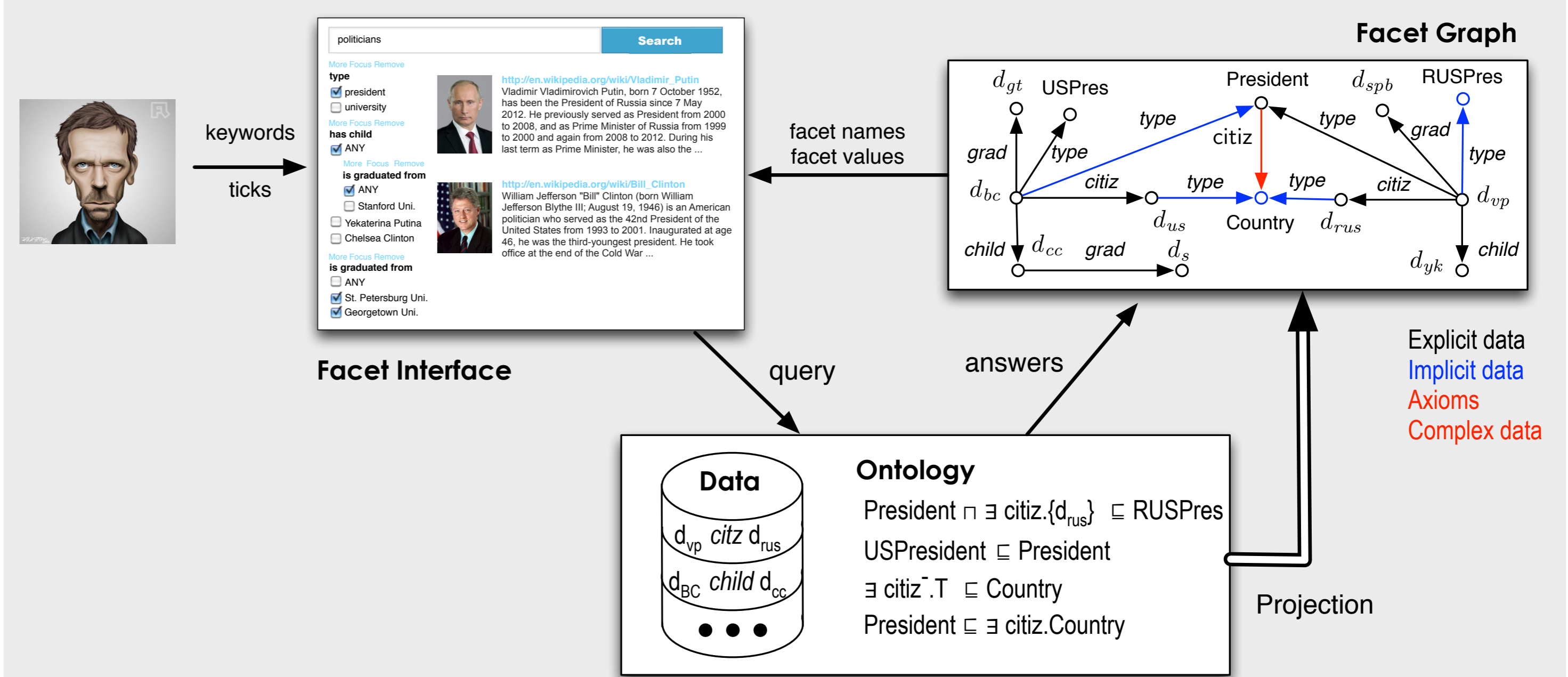Find US Presidents whose Children graduated from Harvard or Stanford

## Research Contributions

**Solid foundation for Semantic F-Search**
- Formal conceptual understanding of faceted search over
  - OWL 2 data and ontologies
- Proposed the notion of facet-graphs to project on them
  - (complex) data and ontological axioms

**Scalable algorithms**
- Generate and update faceted interface
- Evaluate faceted queries over semantic data
  - Exploits bottom up query evaluation



Explicit data
Implicit data
Axioms
Complex data

Facet Interface — query — answers — Facet Graph

Data — Ontology

$President \sqcap \exists citiz.\{d_{rus}\} \sqsubseteq RUSPres$
$USPresident \sqsubseteq President$
$\exists citiz^-.\top \sqsubseteq Country$
$President \sqsubseteq \exists citiz.Country$

Projection

## Collaboration with EDF

**Analytics**
- Specialised handling of datatypes: dates, numbers
- Support for aggregation in queries

**Answer visualisation**
- Answers are interconnected
- Snippets do not reveal answer structure
- EDF developed graph based visualisation of RDF data

**Multi-users**
- Simultaneous access by several users

**Information overflow**
- Too many answers and too many facets
- Autocompleting of facets based on keywords
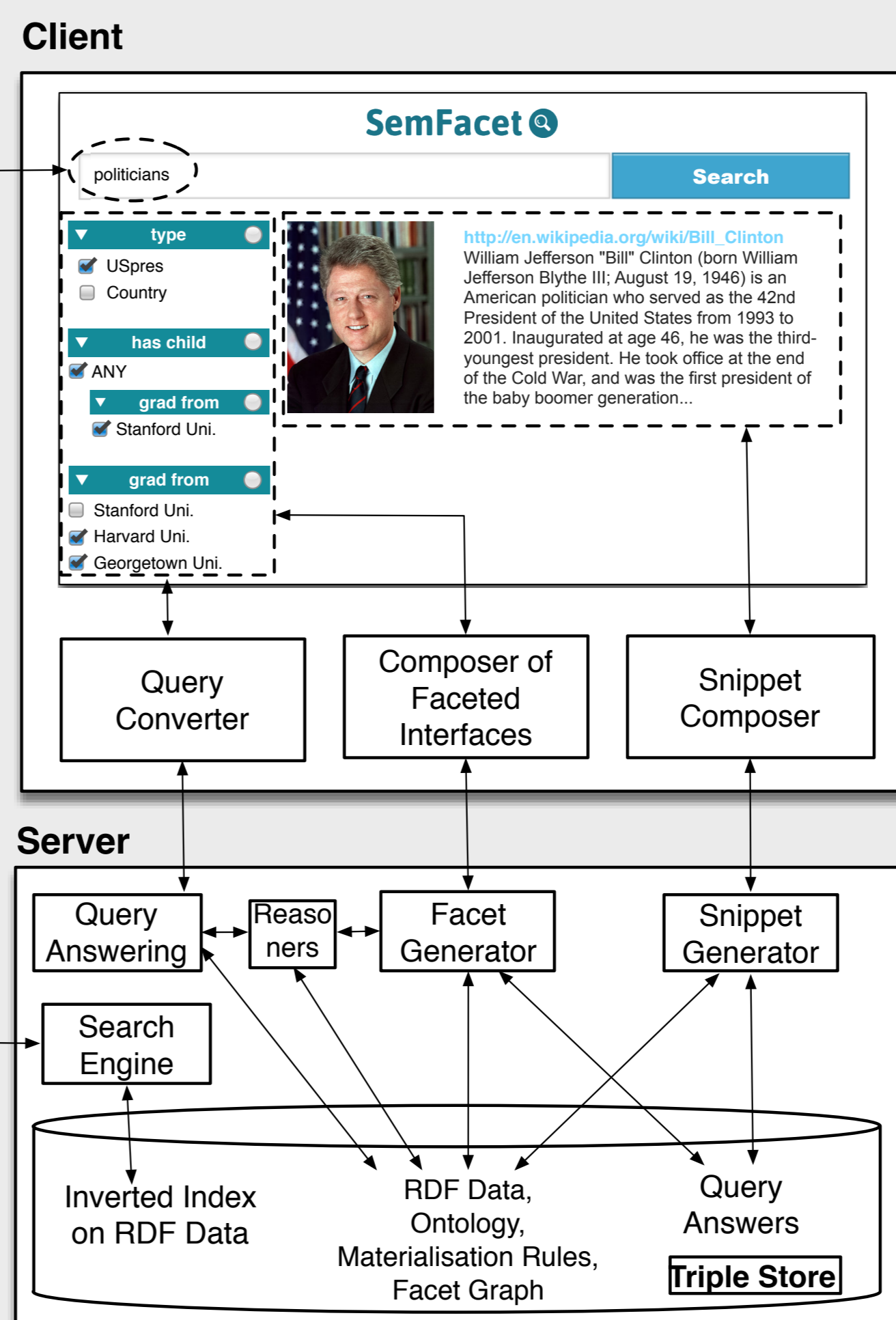- Smart indexes to scale facet / answer computation

**Project Goals:**
- Address analytics, multy-users, visusaliations, overflow
- Extend backend with new scalable algorithms
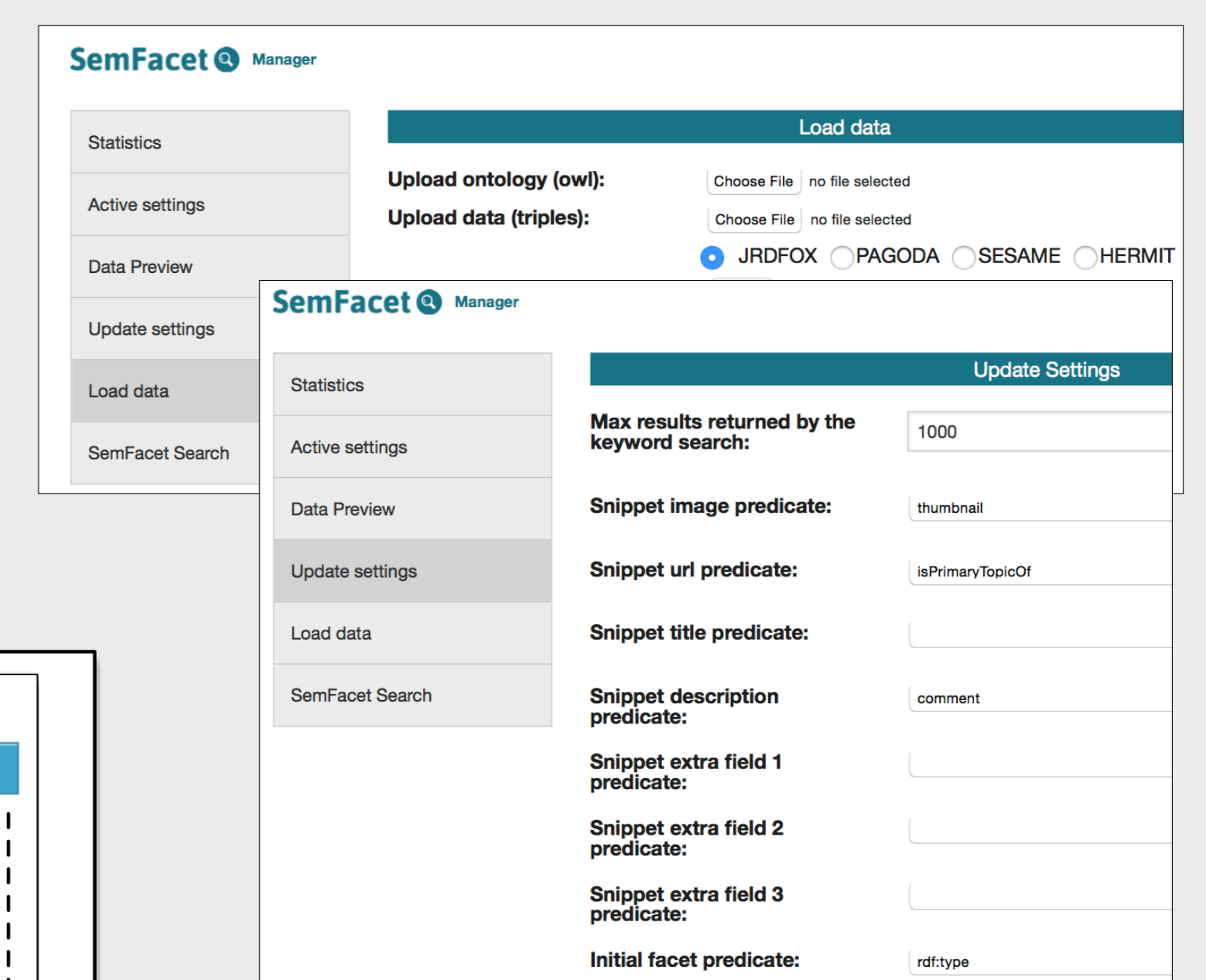- Perform user study

## SemFacet System

**SemFacet**

**Integration of**
- Keyword search
  to initiate search process
- Faceted search over
  RDF & OWL 2



**Main features**
- Automatic interface generation
- In memory
- Online and offline reasoning
- Efficient on millions of triples
- Backend: RDFOX, PAGOdA, Hermit, Sesame

**Flexible configuration**
- Interchangeable triple stores
- Configurable answers (snippets)
- Support of Or and And facets