


Faceted Search over RDF and OWL2

Evgeny Kharlamov

Senior Research Fellow
Information Systems Group
Department of Computer Science
University of Oxford

Join work with

Marcelo Arenas
Bernardo Cuenca Grau
Sarunas Marciuska
Evgeny Sherkhonov
Dmitry Zheleznyakov

Semantic Data and Ontologies

RDF data and OWL ontologies are widely available in

- enterprises
- research labs
- governmental institutions

Examples

- Media: BBC
- Geographic: Ordnance Survey
- Governmental: Statistics.Data.gov.uk
- Industry: Siemens, Statoil, Aibel
- Knowledge Graphs: Google, DBpedia

How to Access Semantic Data?

SPARQL Queries

- Not for End Users
- Like programming language: takes time to learn
- Error prone
- Complex error messages

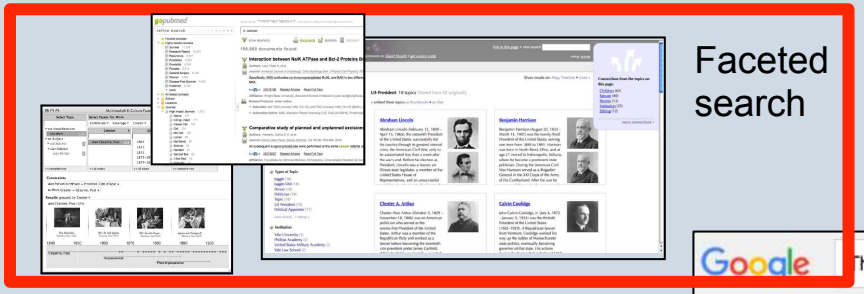
```
SELECT ?z
WHERE   {?x rdf:type “president”@en;
         ?x has_child ?z;
         {{?z is_graduated_from wiki:Harvard;}
         UNION
         {?z is_graduated_from wiki:Stanford;}
```

```
Virtuoso 37000 Error SP030:
SPARQL compiler, line 4: syntax error
at '.' before '}'
```

Many Tools are Available to Access Sem Data

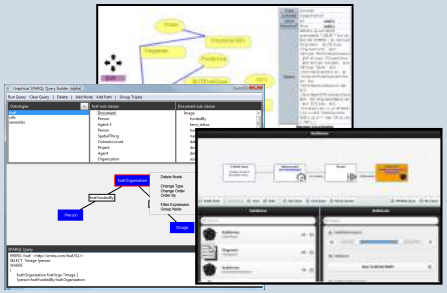
Query formulation tools

- Expressiveness
  - complex SPARQL queries ←→ exploration
- Control
  - precise ←→ ambiguous
- Usability
  - IT specialists ←→ end-user
- Adaption
  - Google ←→ research prototypes
- Domain of use
  - Oil&Gas ←→ Web

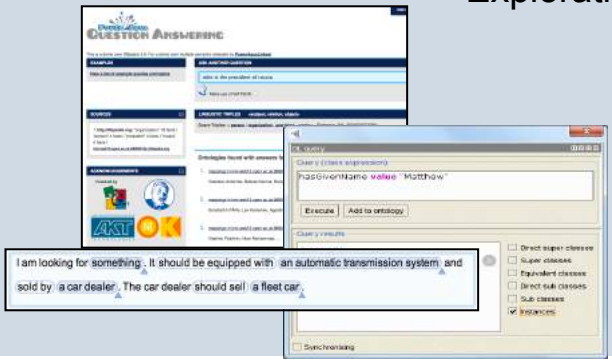


Faceted search

Exploration



Graphical query builders



Natural language interfaces

Faceted Search

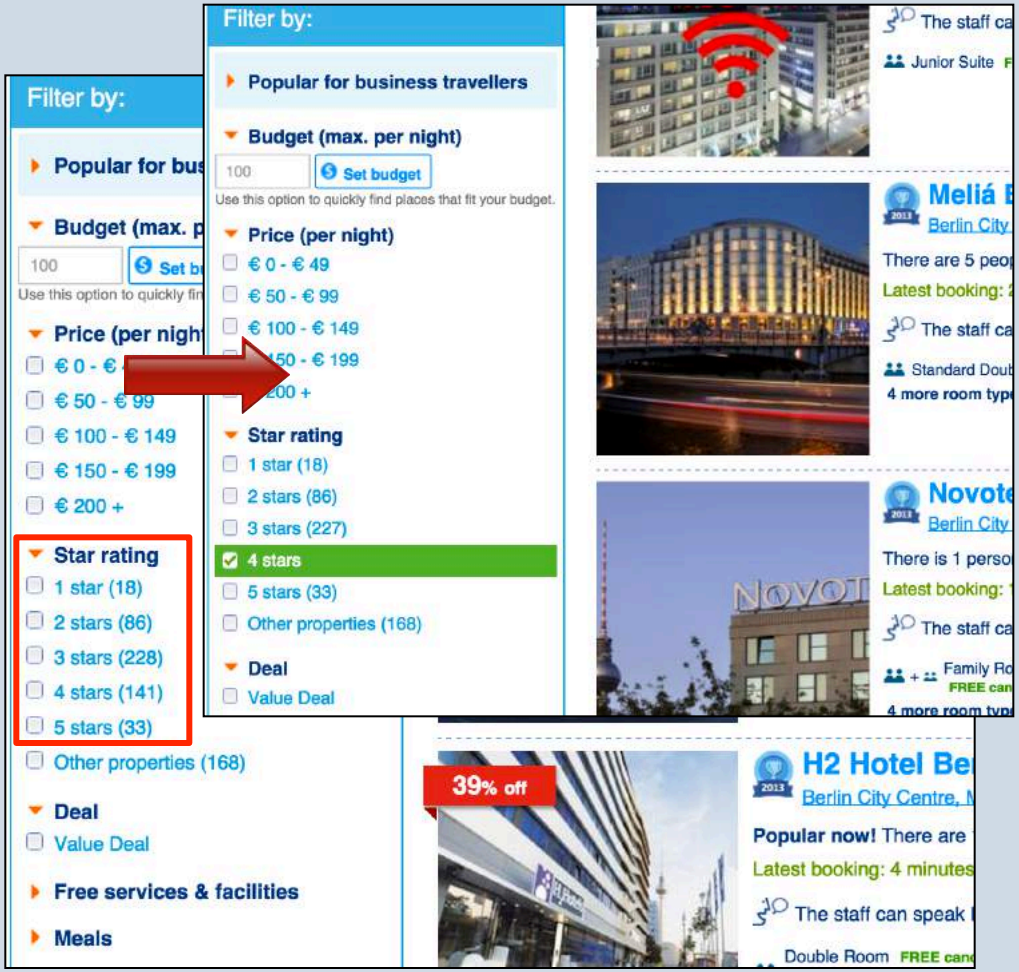
Query formulation paradigm for naïve users

A facet = query constructor

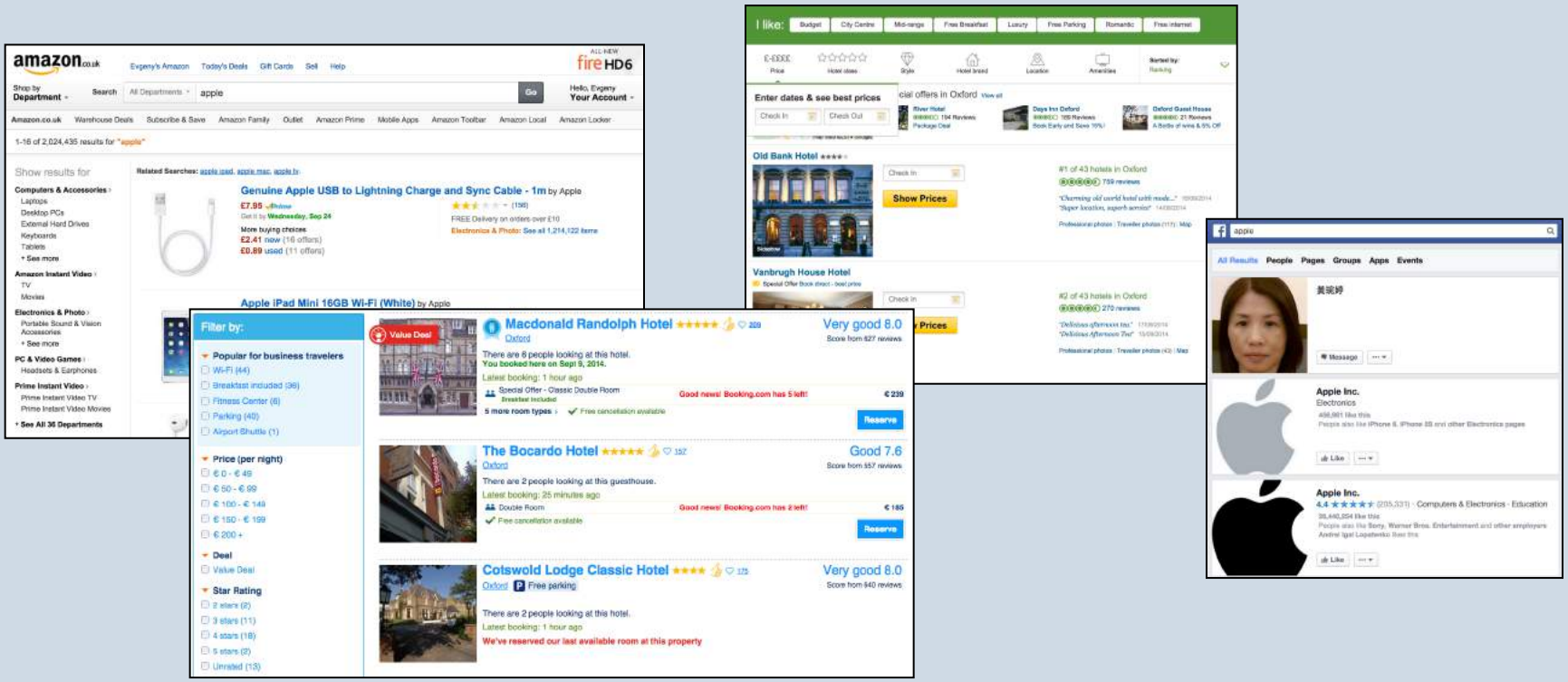
- Name
- Set of values

Facets in action

- Choose a value
- Restrict search result



Faceted Search: De Facto Web Query Standard



ebay amazon.com tripadvisor Booking.com Google

UNIVERSITY OF OXFORD

SemFacet

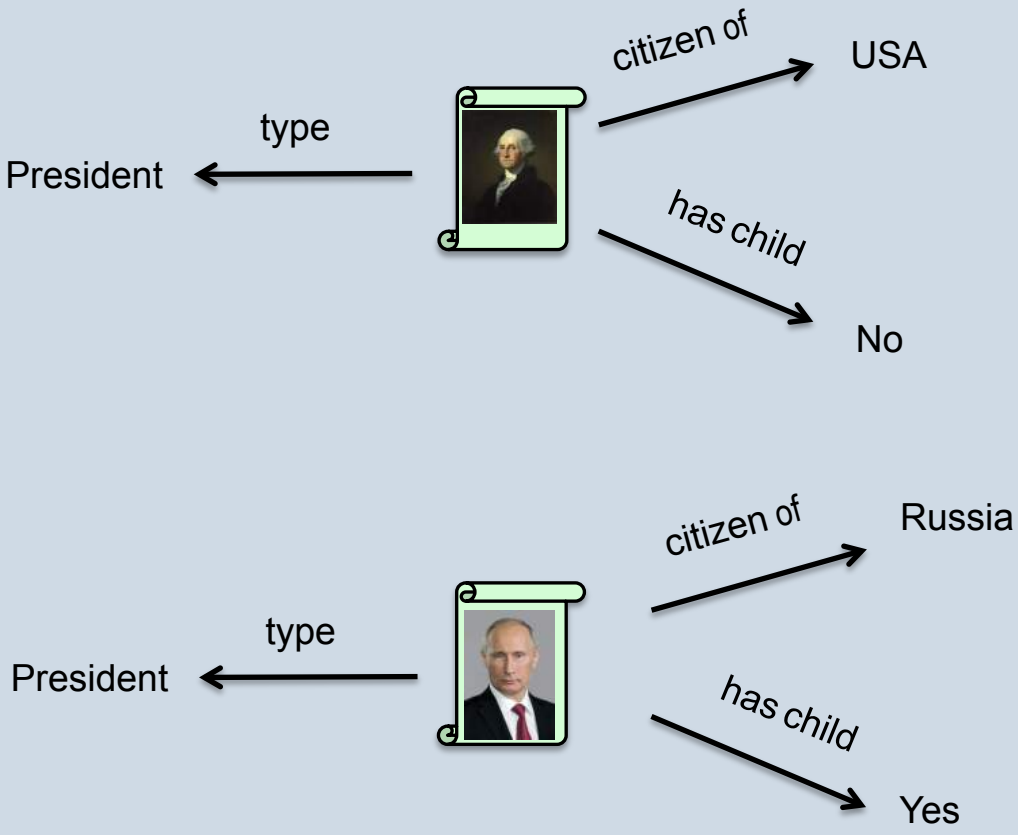
Classical Faceted Search: Data Model

Data

- Set of docs
- Annotated with strings

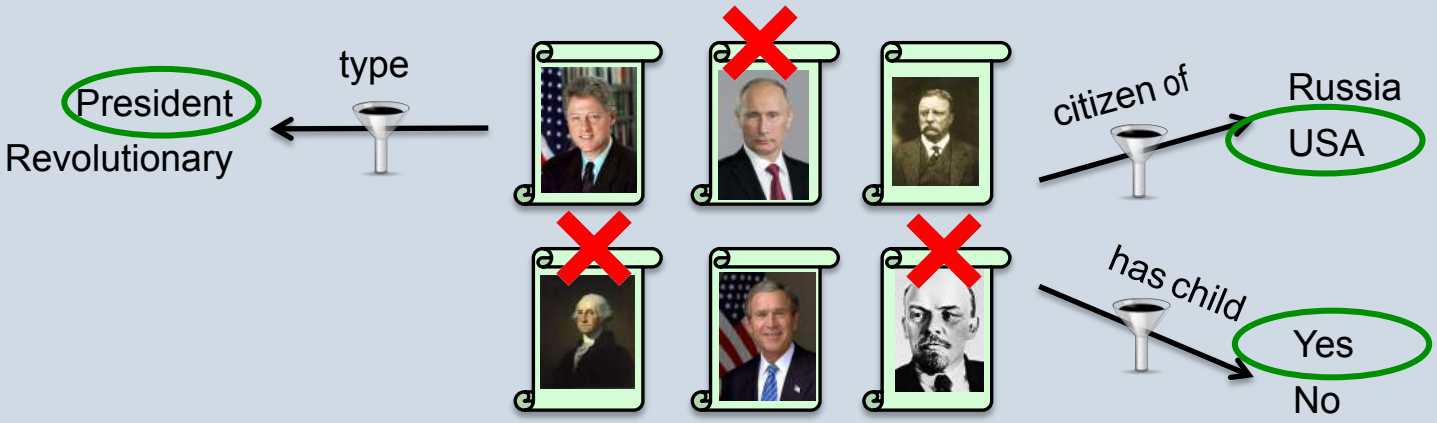
Example

- documents:
  - politician
- annotations:
  - type:
    - President, Revolutionary
  - citizen of
    - RUS, US
  - has child
    - Yes, No



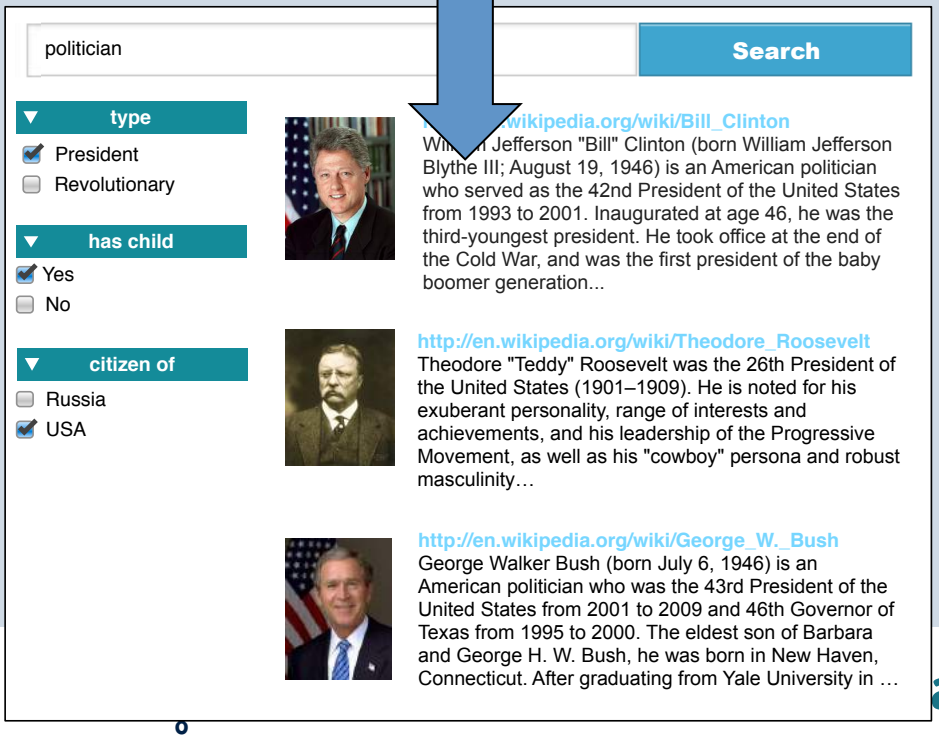
Classical Faceted Search: Search Process

Query:
Find US presidents who got children



Search

- filters
  - via annotations
- search result
  - documents
  - a subset of the original doc. set



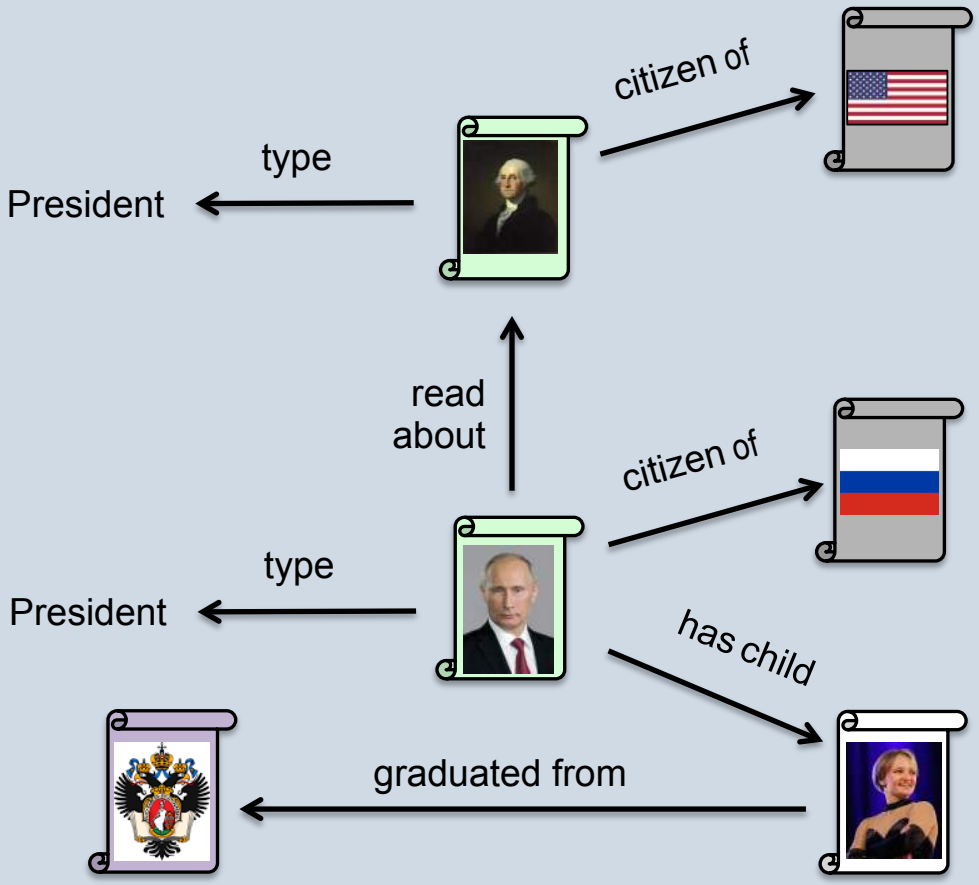
Faceted Search over RDF: Data Model

Data

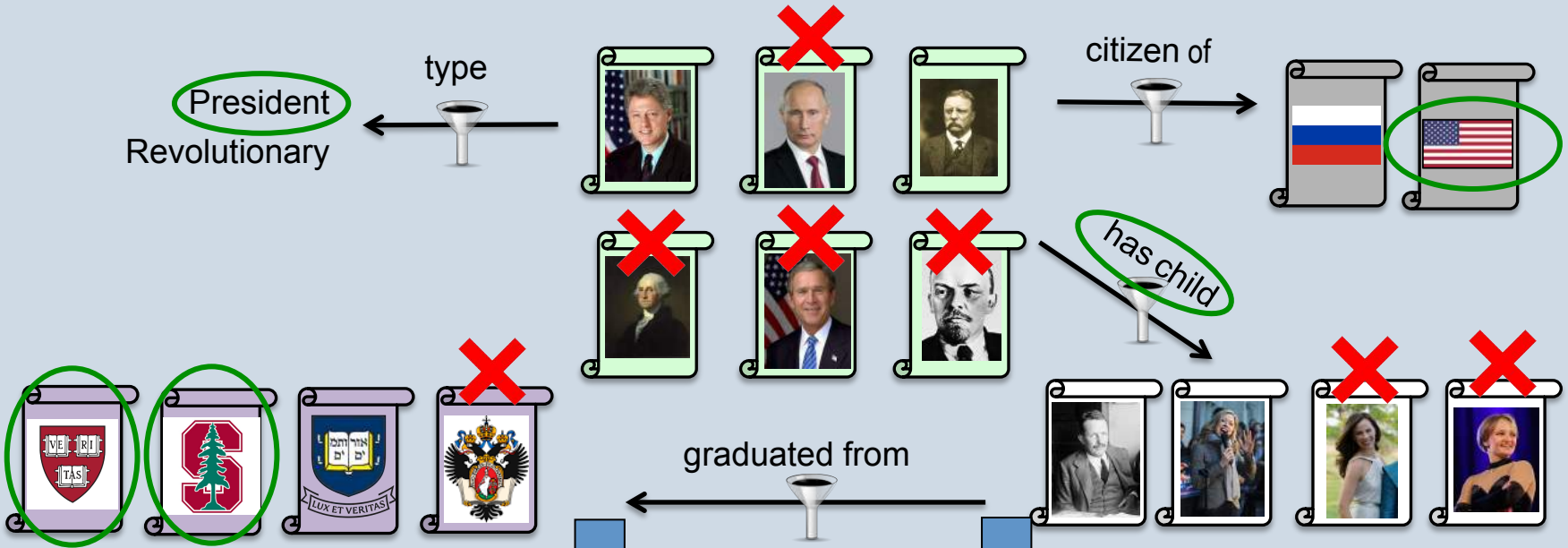
- Graph
- Documents (URIs)
  - of different kind
  - annotated with
    - strings
    - other URIs

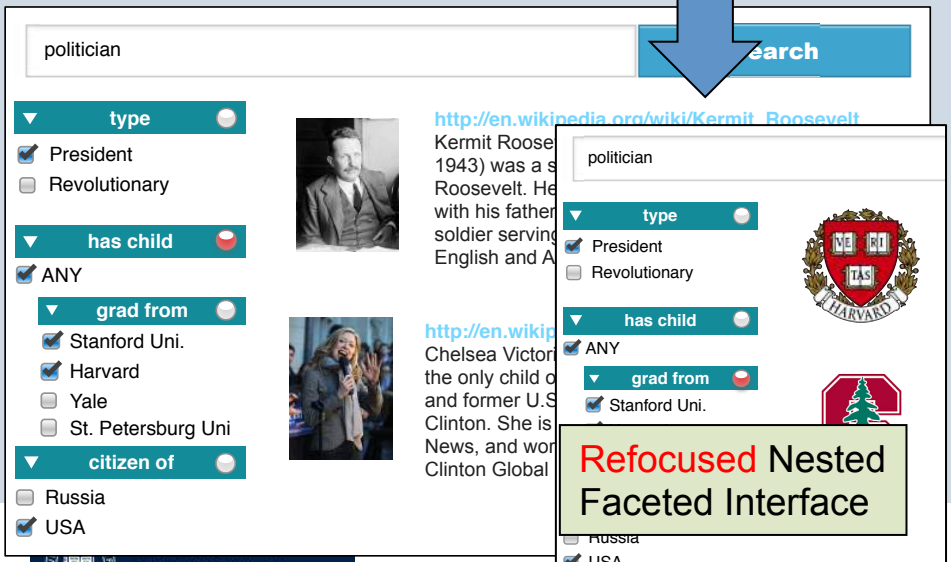
Example

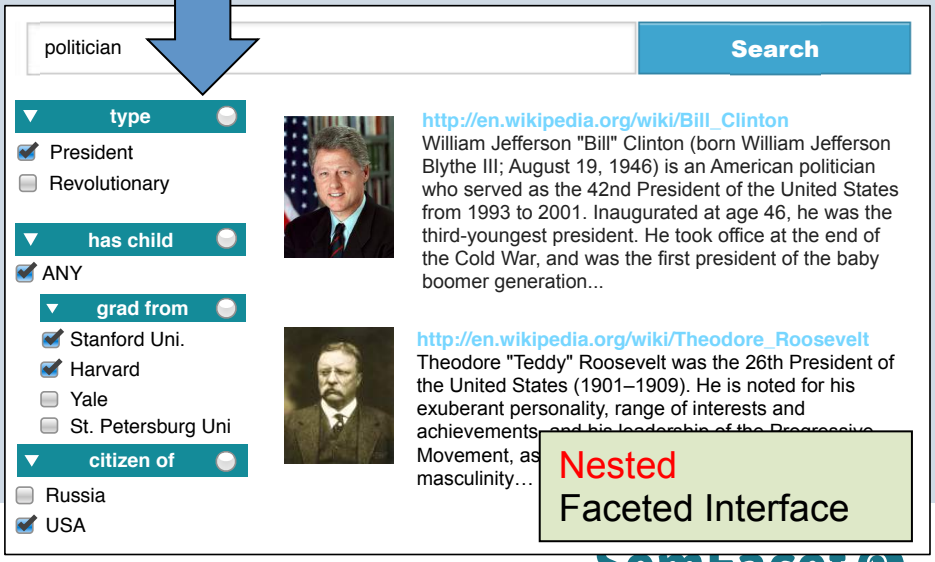
- documents:
  - Politician
  - Countries
  - Children
  - Universities
- annotations:
  - Type: President, Revolutionary



Faceted Search over RDF: Search Process







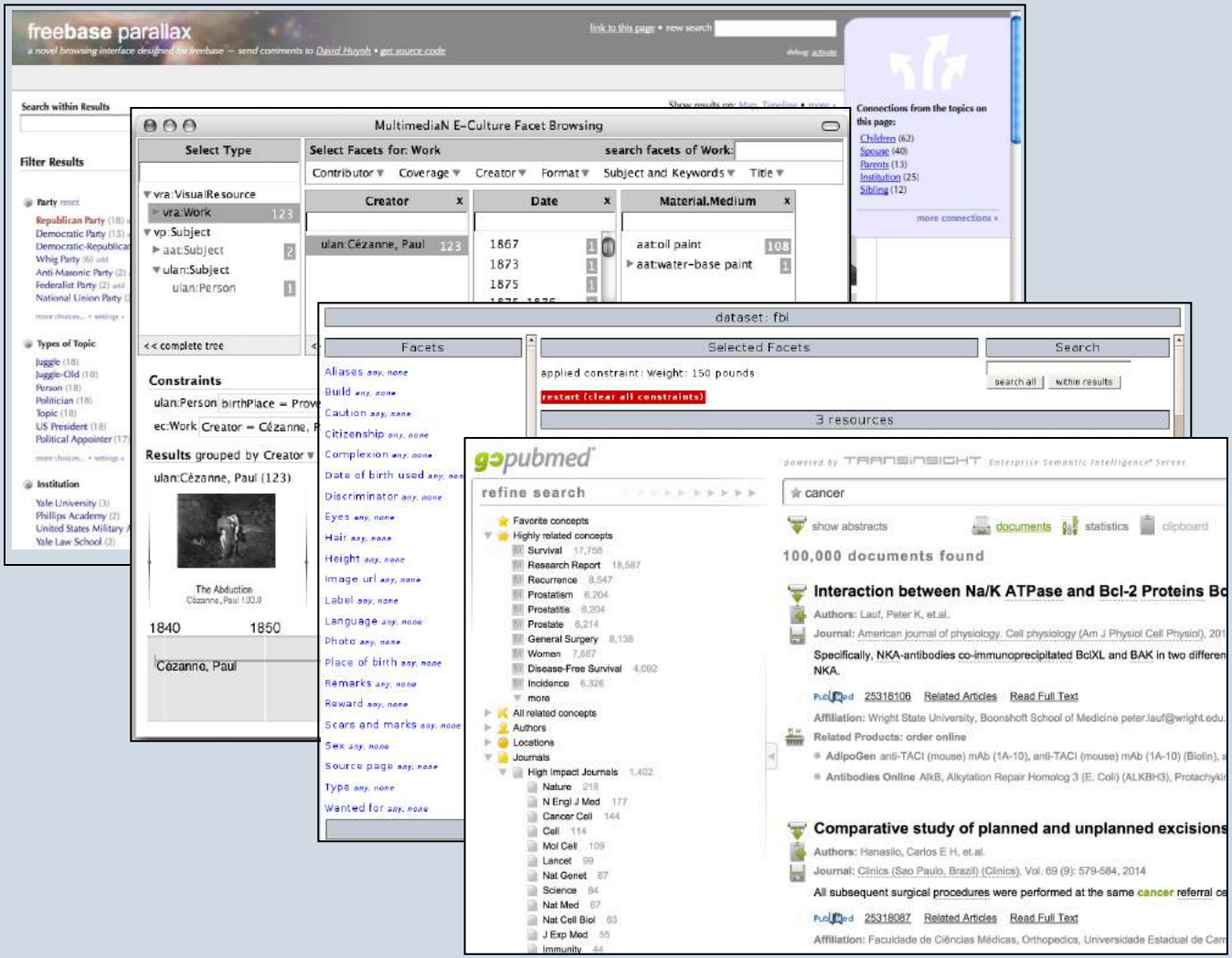
OXFORD

SemFacet

Existing Semantic Faceted Search Solutions

Systems

- Parallax
- /facet
- BroweRDF
- F-Search in Information Workbench
- mSpace
- Tabulator
- Humboldt
- GoPubMed
- gFacet
- …



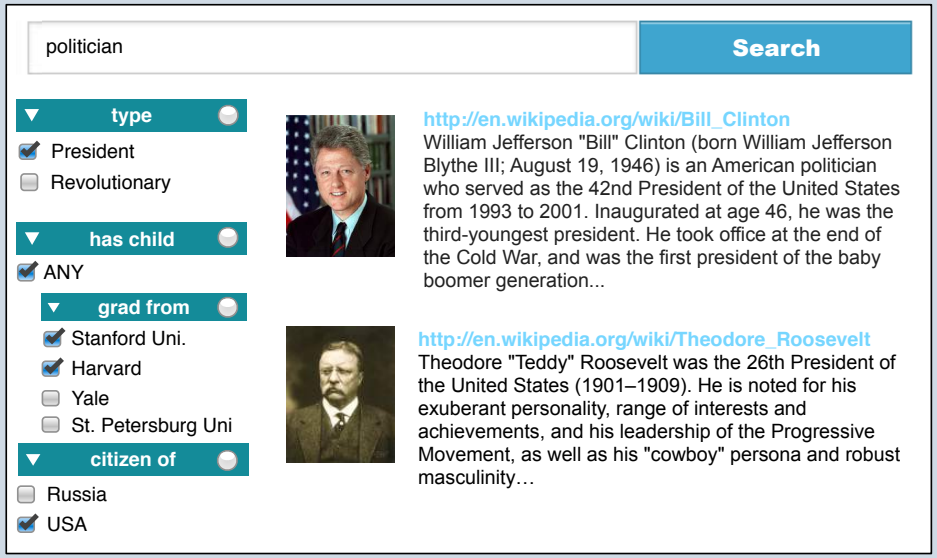
Were We Happy Existing Solutions?

Existing solutions (we aware of) are RDF data driven

- they (essentially) ignore ontological axioms
- they ignore complex (OWL 2) data statements

Theoretical underpinnings

- have received very little attention
- What fragments of SPARQL can be naturally captured?
- What is the complexity of answering such queries?
  - when data is enhanced with ontologies in OWL 2 profiles
- How can we formally capture interactive interface update?



Faceted Q. Interface

Query Answers

Our Contributions

Formalise faceted interfaces tailored towards RDF and OWL

- Capture key functionalities in implemented systems
- Abstract from GUI-specific considerations

Study of expressive power and complexity of faceted queries

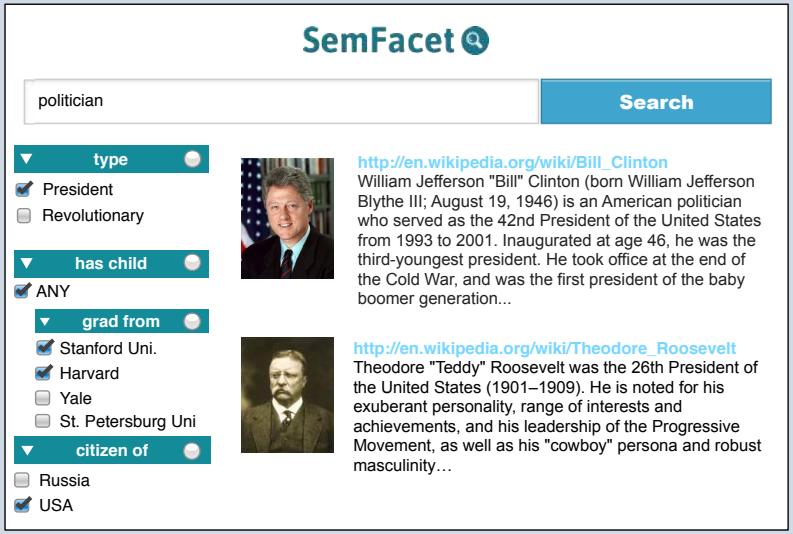
- Identify a fragment of FO sufficient to capture such queries
- Study complexity of query evaluation wrt ontologies in the OWL 2 profiles

Study interface generation and update

- Lift existing approaches RDF → OWL 2
- Propose generic algorithms

Develop SemFacet system

- Allows to generate faceted search interfaces over RDF and OWL 2
- Scales to millions of triples



Facet: Formally

What is Facet?

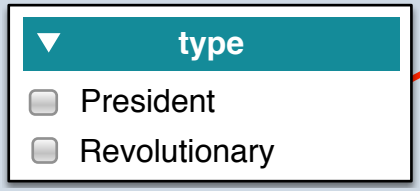
- an elementary building block of faceted interfaces
- it says what can be selected

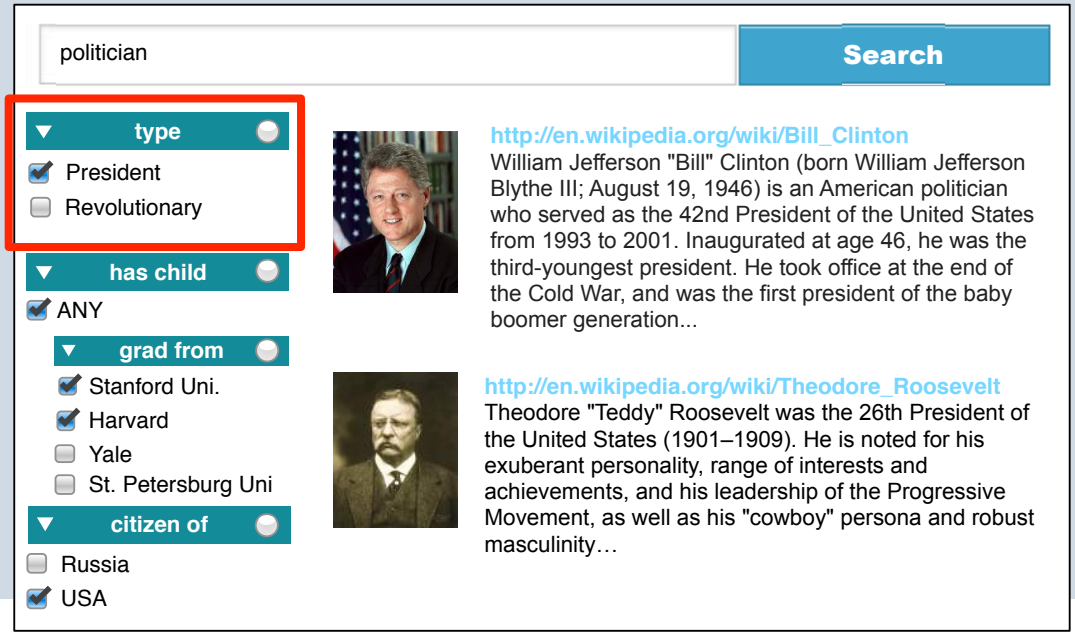
Facet consists of three components

- Property
- Logical OR or AND
- Set of values,
  - e.g., classes, objects, strings

Facets:

$$F_1 = (\text{type}, \vee\{\text{Pres}, \text{Revol}\})$$
$$F_2 = (\text{haschild}, \vee\{\text{any}, d_{kr}, d_{cc}, d_{bb}, d_{ep}\})$$
$$F_3 = (\text{gradFrom}, \vee\{d_{hu}, d_{su}, d_{yu}, d_{spu}\})$$
$$F_4 = (\text{citizenOf}, \vee\{d_{ru}, d_{us}\})$$





Faceted interface: Formally

Faceted Interface

- Basic Faceted Interface
  - Facet + Set of selected values
- General Faceted Interface:
  - Bool. combination of nested BIs
  - According to a grammar

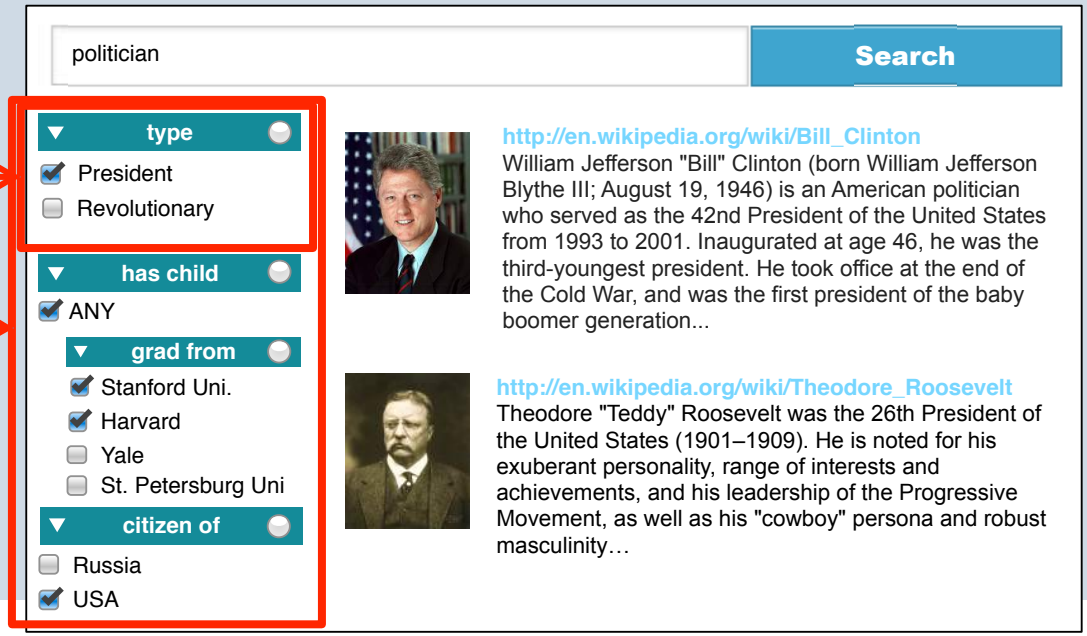
Facets:

$$F_1 = (\text{type}, \vee\{\text{Pres}, \text{Revol}\})$$
$$F_2 = (\text{haschild}, \vee\{\text{any}, d_{kr}, d_{cc}, d_{bb}, d_{ep}\})$$
$$F_3 = (\text{gradFrom}, \vee\{d_{hu}, d_{su}, d_{yu}, d_{spu}\})$$
$$F_4 = (\text{citizenOf}, \vee\{d_{ru}, d_{us}\})$$

$$I ::= path \mid (path \wedge path) \mid (path \vee path),$$
$$path ::= I_0 \mid (I_1/I).$$

Nested Faceted Interface:

$$(F_1, \{\text{Pres}\})$$
$$\wedge$$
$$(F_2, \{\text{any}\})/(F_3, \{d_{su}, d_{hu}\})$$
$$\wedge$$
$$(F_4, \{d_{us}\})$$



Faceted interface with Refocusing

How we formalized refocusing

- Essentially:
  special value with
  special treatment

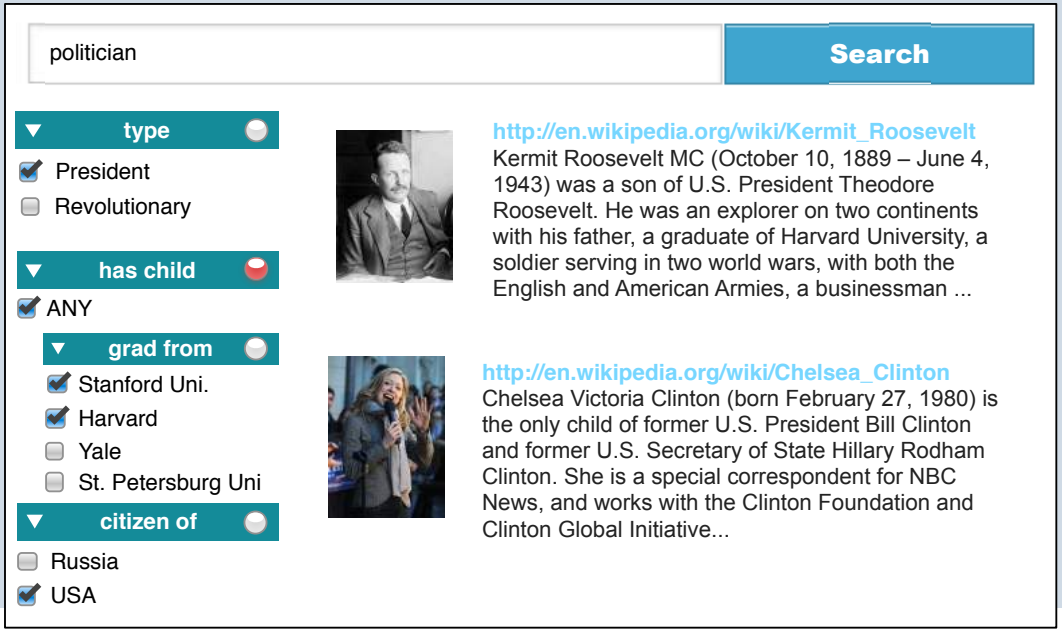
Facet with *focus*:

$$F_2 = (\text{haschild}, \vee\{\text{any}, d_{kr}, d_{cc}, d_{bb}, d_{ep}\})$$

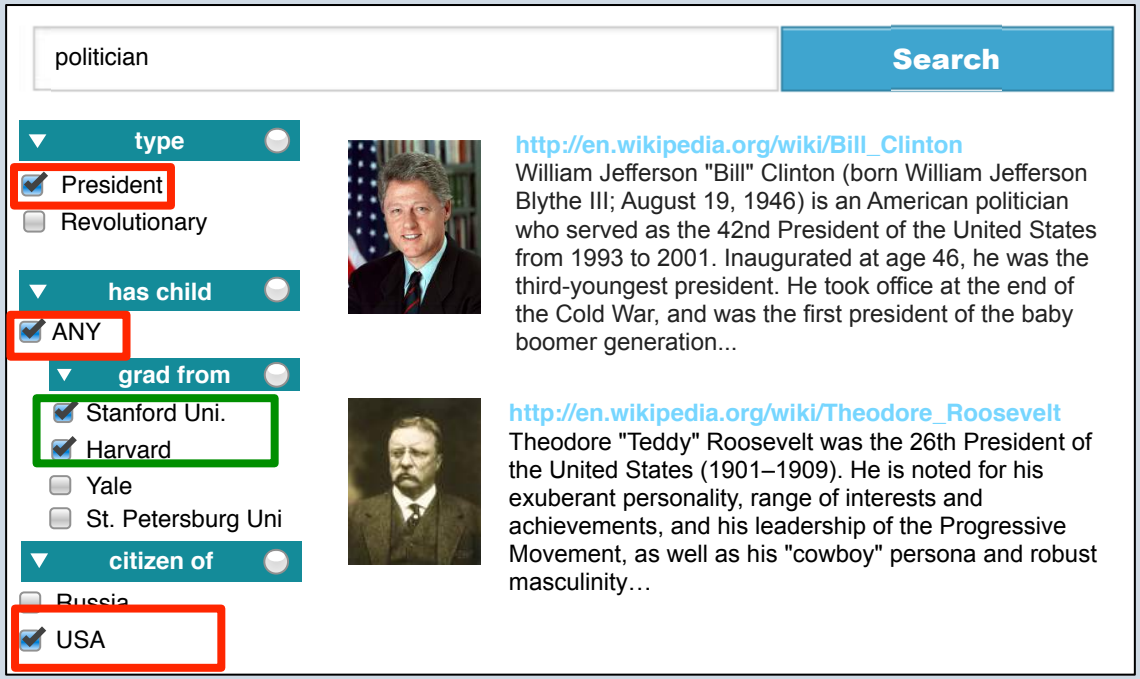
$$F_5 = (\text{haschild}, \vee\{\text{focus}, d_{kr}, d_{cc}, d_{bb}, d_{ep}\})$$

Nested Faceted Interface:

$$(F_1, \{\text{Pres}\})$$
$$\wedge$$
$$(F_5, \{\text{focus}\})/(F_3, \{d_{su}, d_{hu}\})$$
$$\wedge$$
$$(F_4, \{d_{us}\})$$



Turning Nested Faceted Interfaces into Queries



$$\begin{aligned}
&\mathsf{Pres}(x) \\
&\wedge (\exists y(\mathsf{hasChild}(x, y)) \\
&\quad \wedge (\exists z_1(\mathsf{gradFrom}(y, z_1) \wedge z_1 = d_{su}) \vee \\
&\qquad \exists z_2(\mathsf{gradFrom}(y, z_2) \wedge z_2 = d_{hu})) \\
&\wedge (\exists w(\mathsf{citizeOf}(x, w) \wedge w = d_{us}))
\end{aligned}$$

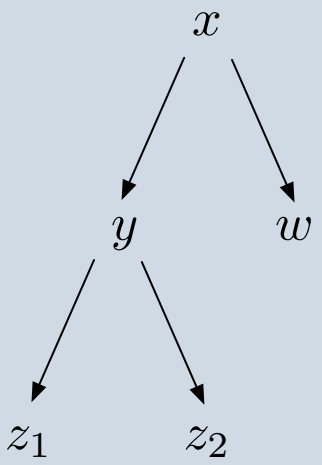
Faceted Queries

Faceted queries are FO formulas of restricted shape

- Positive existential formulas
- Monadic
- Variables form a directed tree rooted at the free variable
- Every two formulae disjunctively connected share one variable

$$\mathsf{Pres}(x) \wedge (\exists y(\mathsf{hasChild}(x,y) \wedge (\exists z_1(\mathsf{gradFrom}(y,z_1) \wedge z_1 = d_{su}) \vee \exists z_2(\mathsf{gradFrom}(y,z_2) \wedge z_2 = d_{hu})) \wedge (\exists w(\mathsf{citizeOf}(x,w) \wedge w = d_{us}))$$

Dependency Tree for Variables



Faceted query answering

Over RDF datasets

- Tractable
- Tree shaped queries with restricted disjunction
- Bottom up evaluation
- as opposed to NP-Hard for unrestricted positive existential queries

Over RDF datasets enhance[illegible]ntolo[illegible]

- Complexity depends on
  - OWL 2 Profile
  - Semantics

Compute data closure, answer w/o ontology

Hardness: known Membership RL: 3vars

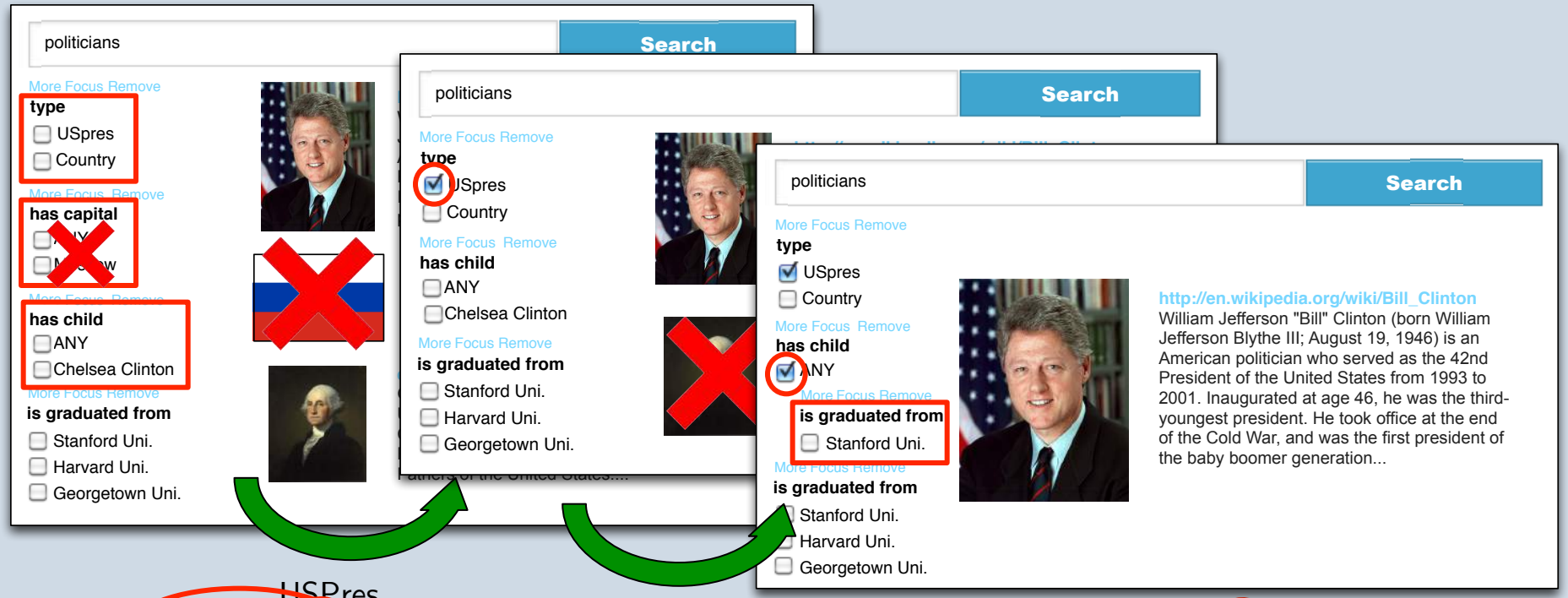
From hardness of evaluation for acyclic CQs

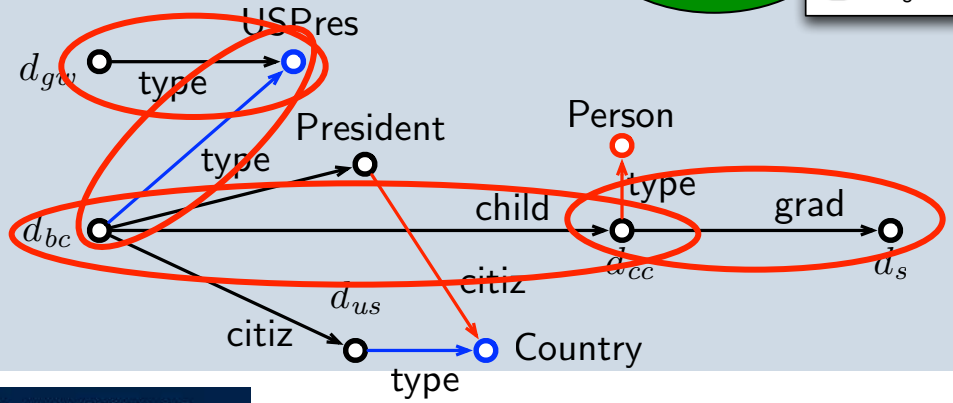
Semantics	RL	EL	QL
Active Domain	P-complete	P-complete	in P
Classical	P-complete	P-complete (guarded EL)	NP-complete

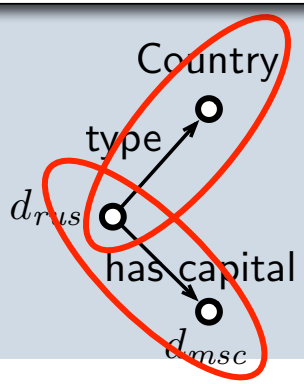
Similar techniques as in [1-2]

[1] G. Stefanoni, B. Motik, Answering Conjunctive Queries over EL Knowl- edge Bases with Transitive and Reflexive Roles, of AAAI, 2015.
[2] M. Kroetzsch, S. Rudolph, P. Hitzler, ELP: Tractable Rules for OWL 2, in: ISWC, 2008

Interface Generation & Update w/ F-graph



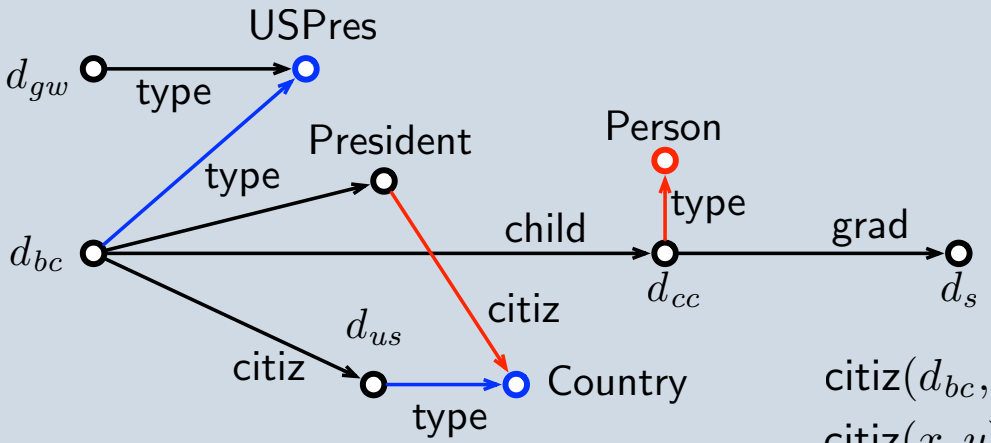




UNIVERSITY OF OXFORD

SemFacet

Navigation Graph



Legend:
- Explicit data
- Implicit data
- Axioms

$\text{citiz}(d_{bc}, d_{us})$

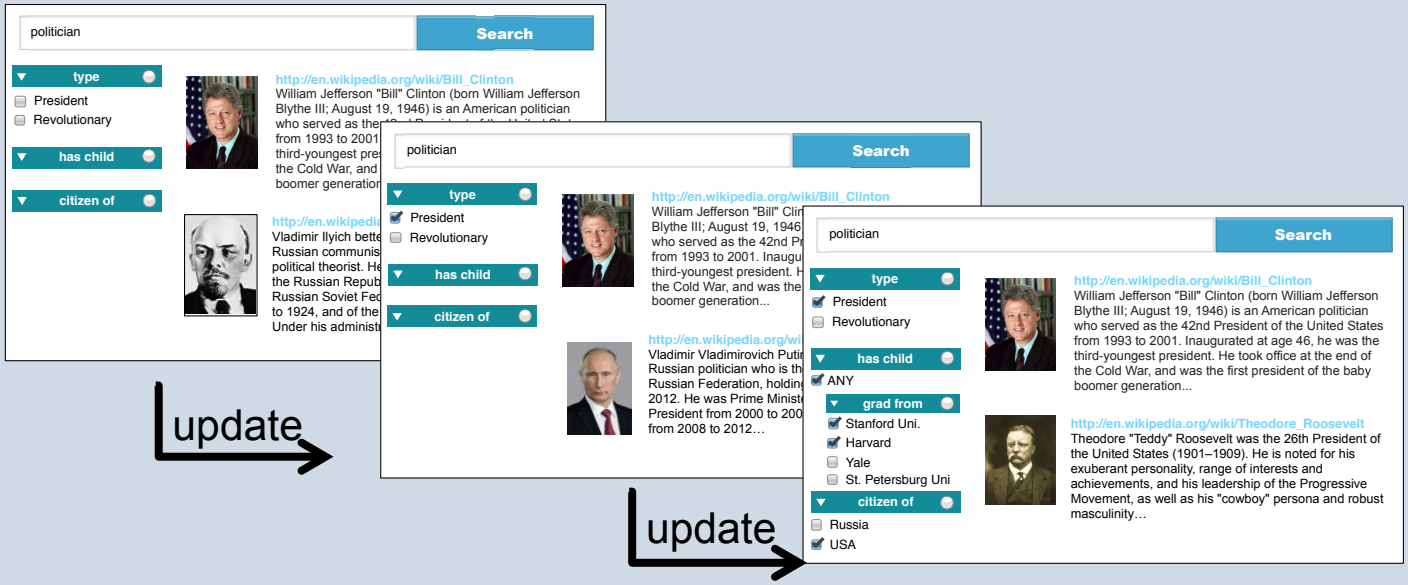
$\text{citiz}(x, y) \rightarrow \text{Country}(y)$

$\text{President}(x) \rightarrow \exists y.\text{citiz}(x, y) \wedge \text{Country}(y)$

Facet graph: our main "data structure" for interface gen. & update:

- Nodes are possible facet values (unary predicates and constants)
- Edges are possible facet names (binary predicates or type)
- Every edge must be justified by
  - An entailed fact
  - an entailed axiom
    - projection of axioms on a graph

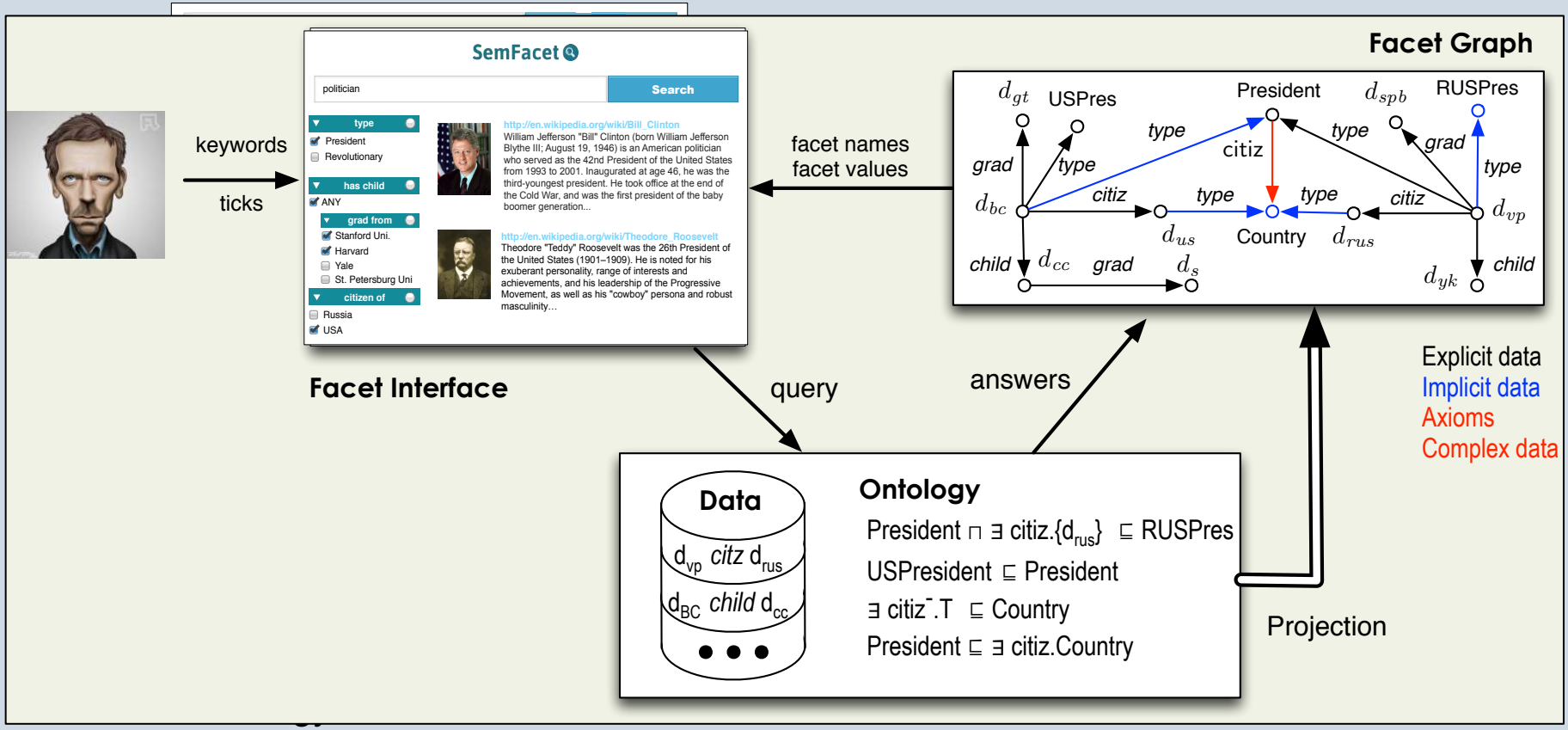
Interface Generation & Update



Our generation and update algorithms are

- "guided" by
  - data
  - ontology
- justified by entailments
  - each facet & value in initial interface
  - each interface update is justified by entailments

Interface Generation & Update



- justified by entailments
  - each facet & value in initial interface
  - each interface update is justified by entailments

SemFacet System

Integration of

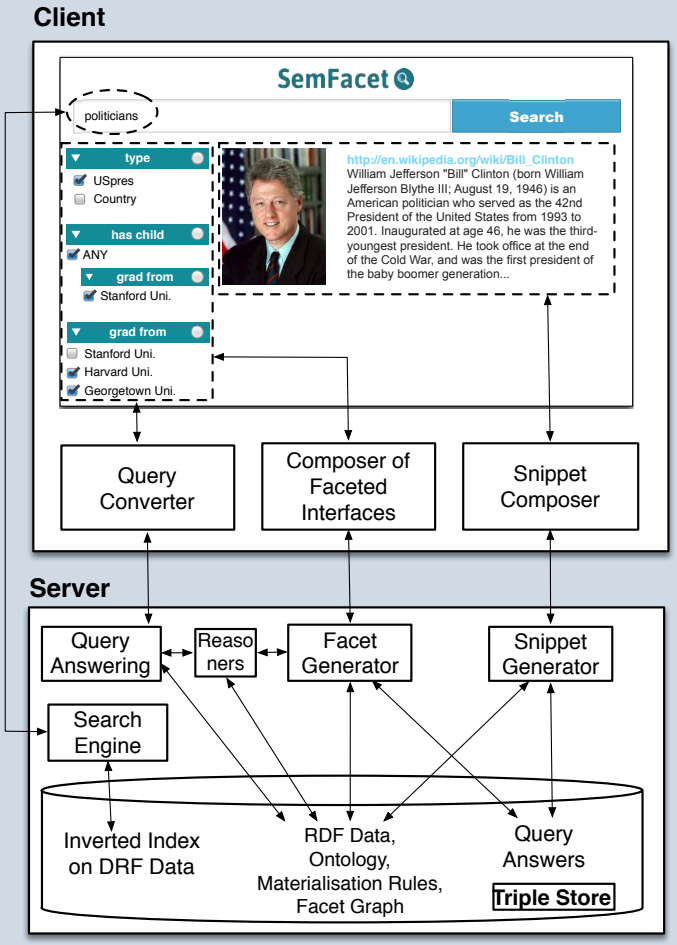
- Keyword search and
- faceted search over RDF & OWL 2

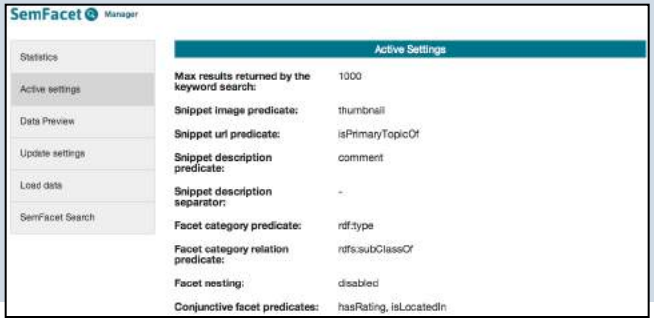
Main features

- Automatic generation of faceted interfaces
- In memory
- Online and offline reasoning
- Efficient on millions of triples
- Backend:
  RDFOX, PAGOdA, Hermit, Sesame

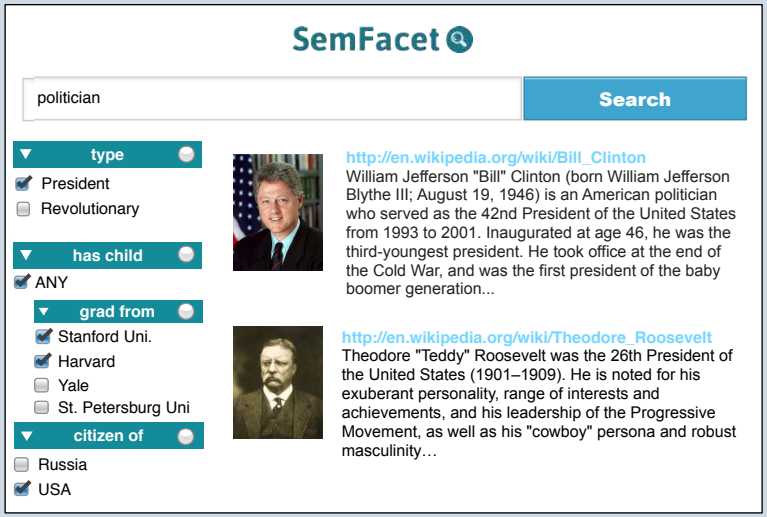
Flexible configuration

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





Contributions



Foundations

- Formalization of Semantic Faceted Search
- Complexity of query answering

Projection of ontologies on graphs

- Allows to incorporate ontologies into faceted search
- Better faceted interfaces: generate more facets / prune irrelevant facets

Scalable algorithms to

- Generate and update facets from data and ontologies
- Evaluate faceted queries over semantic data

System implementation

- SemFacet System