Build it, and they will come:
Applications of semantic technology

Ian Horrocks
Information Systems Group
Department of Computer Science
University of Oxford
A Brief History of the Semantic Web
This work is available under a CC BY-SA license. This means you can remix, transform, and / or distribute it under the condition that you give proper attribution.

A Brief History of the Semantic Web
A Brief History of the Semantic Web

This work is available under a CC BY-SA license. This means you can use/modify/extend it under the condition that you give proper attribution.

A Brief History of the Semantic Web

[Diagram showing the history of the Semantic Web with timelines and technologies]
A Brief History of the Semantic Web

- **RDF Schema**: March 1999
- **RDF**: March 2002
- **OWL**: July 2002
- **SPARQL**: Oct 2004
- **XML**: Nov 1996, 1.0 Dec 2001, 1.1 Dec 2001
- **XQuery**: Feb 2001, 1.0 Jan 2006
- **XSLT**: Jul 1999, 1.0 Aug 1998, 2.0 April 2014
- **XPath**: Jul 1999, 1.0 Aug 1998, 2.0 Apr 2014
- **XML Schema**: May 2001, 1.0 Feb 2000, 1.1 Dec 2001
- **XQuery Update**: March 2013
- **RDF Schema**: May 2014, 1.1 Feb 2014
- **RDF**: 1.1 Aug 2011
- **RDFa**: 1.1 Apr 2011, 1.1 EV Apr 2011
- **RDFa 1.1/CT**: May 2013
- **JSON-LD 1.0**: April 2014
- **SKOS**
- **POWDER**
- **PROV**
- **RIF**
- **SAX**
- **SOAP**
- **WSDL**
- **RDB2RDF**
- **G2R**

This work is available under a CC BY-SA license. This means you can use/modify/extend it under the condition that you give proper attribution. Please cite as:

A Brief History of the Semantic Web

This work is available under a CC BY-SA license. This means you can use/modify/extend it under the condition that you give proper attribution. Please cite as:

A Brief History of the Semantic Web

Scientific American
Get the Idea?
(TOMORROW'S WEB WILL)

1998
2000
2008
2012
2010
2002
2004
2006
2014

This work is available under a CC BY-SA license. This means you can use/modify/extend it under the condition that you give proper attribution.

Please cite as:
A Brief History of the Semantic Web
A Brief History of the Semantic Web

This work is available under a CC BY-SA license. This means you can use/modify/extend it under the condition that you give proper attribution.

Please cite as:
Why the Angst?
Why the Angst?
Why the Angst?

“A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities”
Why the Angst?

Semantic Web
Why the Angst?
Why the Angst?
Why the Angst?

- Explicit KR sometimes needed, e.g., Knowledge Graph
  - Less rigorous treatment of semantics
  - Not using Semantic Web standards
Why the Angst?

- Explicit KR sometimes needed, e.g., Knowledge Graph
  - Less rigorous treatment of semantics
  - Not using Semantic Web standards
- **Hiring** Semantic Web people
Why the Angst?

Semantic Web
Data Access:
Data Access: Statoil Exploration

- Geologists & geophysicists use data from previous operations in nearby locations to develop stratigraphic models of unexplored areas
  - TBs of relational data
  - using diverse schemata
  - spread over 1,000s of tables
  - and multiple data bases
Data Access: Statoil Exploration

- Geologists & geophysicists use data from previous operations in nearby locations to develop stratigraphic models of unexplored areas
  - TBs of relational data
  - using diverse schemata
  - spread over 1,000s of tables
  - and multiple data bases

Data Access
- 900 geologists & geophysicists
- 30-70% of time on data gathering
- 4 day turnaround for new queries
Data Access: Statoil Exploration

- Geologists & geophysicists use data from previous operations in nearby locations to develop stratigraphic models of unexplored areas
  - TBs of relational data
  - using diverse schemata
  - spread over 1,000s of tables
  - and multiple data bases

Data Access
- 900 geologists & geophysicists
- 30-70% of time on data gathering
- 4 day turnaround for new queries

Data Exploitation
- Better use of experts time
- Data analysis “most important factor” for drilling success
Service centres responsible for remote monitoring and diagnostics of 1,000s of gas/steam turbines

Engineers use a variety of data for visualization, diagnostics and trend detection:
- several TB of time-stamped sensor data
- several GB of event data
- data grows at 30GB per day

Service Requests
- 1,000 requests per center per year
- 80% of time used on data gathering

Diagnostic Functionality
- 2–6 p/m to add new function
- New diagnostics → better exploitation of data
Pipelines from oil facilities?
Q(?x) ← (?x, rdf:type, :Pipeline) \land
(?x, :fromFacility, ?y) \land
(?y, rdf:type, :OilFacility)
Pipelines from oil facilities?

\[ Q(\textit{x}) \leftarrow (\textit{x}, \text{rdf:type}, \text{Pipeline}) \land
(\textit{x}, \text{fromFacility}, \textit{y}) \land
(\textit{y}, \text{rdf:type}, \text{OilFacility}) \]
Pipelines from oil facilities?

\( Q(\?x) \leftarrow (\?x, \text{rdf:type}, \text{Pipeline}) \land (\?x, \text{fromFacility}, \?y) \land (\?y, \text{rdf:type}, \text{OilFacility}) \)
Pipelines from oil facilities?

\[ Q(\text{?}x) \leftarrow (\text{?}x, \text{rdf:type, Pipeline}) \land (\text{?}x, \text{fromFacility, ?}y) \land (\text{?}y, \text{rdf:type, OilFacility}) \]
$Q(x) \leftarrow (x, \text{rdf:type}, \text{:Pipeline}) \land (x, \text{:fromFacility}, y) \land (y, \text{rdf:type}, \text{:OilFacility})$

SubClassOf($\text{:OilPipeline}$
ObjectIntersectionOf($\text{:Pipeline}$
ObjectSomeValuesFrom($\text{:fromFacility, OilFacility}$)))
\[
Q(\text{?x}) \leftarrow (\text{?x, rdf:type, Pipeline}) \land \\
(\text{?x, fromFacility, ?y}) \land \\
(\text{?y, rdf:type, OilFacility})
\]
Pipelines from oil facilities?

\[ Q(\exists x \leftarrow (\exists x, \text{rdf:type, Pipeline}) \land (\exists x, \text{fromFacility, } y) \land (\exists y, \text{rdf:type, OilFacility}) \right) \]

OWL 2 QL ontology

\[
\text{SubClassOf}(\text{OilPipeline} \ \text{ObjectIntersectionOf}(\text{Pipeline} \ \text{ObjectSomeValuesFrom}(\text{fromFacility} \ \text{OilFacility})))
\]
Pipelines from oil facilities?

\[ Q(\,x\,) \leftarrow (\,x, \,\text{rdf:type}, \,:\text{Pipeline}\,) \land \\
(\,x, \,:\text{fromFacility}, \,y\,) \land \\
(\,y, \,\text{rdf:type}, \,:\text{OilFacility}\,) \]

**Rewrite**

\[ Q'(\,x\,) \leftarrow (\,x, \,\text{rdf:type}, \,:\text{Pipeline}\,) \land \\
(\,x, \,:\text{fromFacility}, \,y\,) \land \\
(\,y, \,\text{rdf:type}, \,:\text{OilFacility}\,) \lor (\,x, \,\text{rdf:type}, \,:\text{OilPipeline}\,)
\]

**Ontology**

\[ \text{SubClassOf}(\,:\text{OilPipeline}\, \text{ObjectIntersectionOf}(\,:\text{Pipeline}\, \text{ObjectSomeValuesFrom}(\,:\text{fromFacility}\, \,:\text{OilFacility}))) \]
Pipelines from oil facilities?

\[ Q(x) \leftarrow (x, \text{rdf:type}, \text{Pipeline}) \land (x, \text{fromFacility}, y) \land (y, \text{rdf:type}, \text{OilFacility}) \]

\[ Q'(x) \leftarrow (x, \text{rdf:type}, \text{Pipeline}) \land (x, \text{fromFacility}, y) \land (y, \text{rdf:type}, \text{OilFacility}) \lor (x, \text{rdf:type}, \text{OilPipeline}) \]

OWL 2 QL ontology

SubClassOf(:OilPipeline
ObjectIntersectionOf(:Pipeline
ObjectSomeValuesFrom(:fromFacility :OilFacility)))

PipeLines

<table>
<thead>
<tr>
<th>ID</th>
<th>Oil</th>
<th>From</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>N</td>
<td>f1</td>
</tr>
<tr>
<td>p2</td>
<td>Y</td>
<td>f2</td>
</tr>
<tr>
<td>p3</td>
<td>Y</td>
<td>Null</td>
</tr>
</tbody>
</table>
Pipelines from oil facilities?

\[
Q(?x) \leftarrow (?x, \text{rdf:type}, \text{Pipeline}) \land 
(?x, \text{fromFacility}, ?y) \land 
(?y, \text{rdf:type}, \text{OilFacility})
\]

\[
Q'(?x) \leftarrow (?x, \text{rdf:type}, \text{Pipeline}) \land 
(?x, \text{fromFacility}, ?y) \land 
(?y, \text{rdf:type}, \text{OilFacility}) 
\lor (?x, \text{rdf:type}, \text{OilPipeline})
\]

```
:OilPipeline = select ID from Pipeline 
where Oil = "Y"
```

\[\text{Pipeline}\]
<table>
<thead>
<tr>
<th>ID</th>
<th>Oil</th>
<th>From</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>N</td>
<td>f1</td>
</tr>
<tr>
<td>p2</td>
<td>Y</td>
<td>f2</td>
</tr>
<tr>
<td>p3</td>
<td>Y</td>
<td>Null</td>
</tr>
</tbody>
</table>

OWL 2 QL ontology

SubClassOf(:OilPipeline
ObjectIntersectionOf(:Pipeline
ObjectSomeValuesFrom(:fromFacility :OilFacility))))

(R2RML) mappings

\[\text{map}\]
Pipelines from oil facilities?

\[ Q(\{x\}) \leftarrow (\{x\}, \text{rdf\_type}, :\text{Pipeline}) \land (\{x\}, \text{fromFacility}, ?y) \land (\{y\}, \text{rdf\_type}, :\text{OilFacility}) \]

\[ Q'(\{x\}) \leftarrow (\{x\}, \text{rdf\_type}, :\text{Pipeline}) \land (\{x\}, :\text{fromFacility}, ?y) \land (\{y\}, \text{rdf\_type}, :\text{OilFacility}) \lor (\{x\}, \text{rdf\_type}, :\text{OilPipeline}) \]

\begin{align*}
\text{SubClassOf}(:\text{OilPipeline} &\quad \text{ObjectIntersectionOf}(:\text{Pipeline} \\
&\quad \text{ObjectSomeValuesFrom}( :\text{fromFacility} :\text{OilFacility})))
\end{align*}

(O2RML) mappings

\begin{align*}
:p1 &\quad \text{rdf\_type} &\quad :\text{Pipeline} \\
:p1 &\quad :\text{fromFacility} &\quad :f1 \\
:f1 &\quad \text{rdf\_type} &\quad :\text{OilFacility} \\
p2 &\quad \text{rdf\_type} &\quad :\text{OilPipeline} \\
p2 &\quad :\text{fromFacility} &\quad :f2 \\
f2 &\quad \text{rdf\_type} &\quad :\text{OilFacility} \\
p3 &\quad \text{rdf\_type} &\quad :\text{OilPipeline}
\end{align*}

select Pipeline.ID from Pipeline, \\ 
where Oil = "Y"

select Pipeline.ID from Pipeline, \\ 
where Pipeline.From = Facility.ID and \\ 
UNION \\ 
select ID from Pipeline \\ 
where Oil = "Y"

\begin{tabular}{|c|c|c|}
\hline
ID & Oil & From \\
\hline
p1 & N & f1 \\
p2 & Y & f2 \\
p3 & Y & Null \\
\hline
\end{tabular}
Pipelines from oil facilities?

\[ Q(\bar{x}) \leftarrow (\bar{x}, \text{rdf:type, :Pipeline}) \land \\
(\bar{x}, \text{:fromFacility, ?y}) \land \\
(?y, \text{rdf:type, :OilFacility}) \]

\[ Q'(\bar{x}) \leftarrow (\bar{x}, \text{rdf:type, :Pipeline}) \land \\
(\bar{x}, \text{:fromFacility, ?y}) \land \\
(?y, \text{rdf:type, :OilFacility}) \\
\lor (\bar{x}, \text{rdf:type, :OilPipeline}) \]

\[ \text{:OilPipeline} = \text{select ID from Pipeline where Oil = "$Y"} \]

\[ \text{select Pipeline.ID from Pipeline} \quad \ldots \\
\text{where Pipeline.From = Facility.ID and \ldots} \\
\text{UNION} \\
\text{select ID from Pipeline where Oil = "$Y"} \]

\[ \begin{array}{|c|c|c|} 
\hline 
\text{ID} & \text{Oil} & \text{From} \\
\hline 
p1 & N & f1 \\
p2 & Y & f2 \\
p3 & Y & \text{Null} \\
\hline 
\end{array} \]
Optique Architecture

Diagram showing the architecture of Optique with components such as Application, Query Formulation, Ontology & Mapping Management, Query Transformation, Query Planning, Stream Adapter, Query Execution, and streaming data.
Query rewriting:
• uses ontology & mappings
• computationally hard
• ontology & mappings small

Query evaluation:
• ind. of ontology & mappings
• computationally tractable
• data sets very large
Optique Architecture

Query rewriting:
- uses ontology & mappings
- computationally hard
- ontology & mappings small

Query evaluation:
- ind. of ontology & mappings
- computationally tractable
- data sets very large

Other features:
support for query formulation
Optique Architecture

Query rewriting:
• uses ontology & mappings
• computationally hard
• ontology & mappings small

Query evaluation:
• ind. of ontology & mappings
• computationally tractable
• data sets very large

Other features:
• support for query formulation
  “Bootstrapping” Ontology & mappings
Research Issues
Research Issues

- **Expressive power:**
  - OWL QL (necessarily) has (very) restricted expressive power
  - Could use mappings to translate data into triples and use OWL RL

- **Scalability:**
  - Query size may increase exponentially in size of ontology
  - Rewritten queries may be hard for existing DBMSs
  - Extensive optimisation required [Bagosi et al]

- **Ontology and mapping engineering:**
  - Ontology engineering known to be hard
  - Less known about mappings, but likely to require similar tool support
Data Analysis: Kaiser Permanente
Data Analysis:

- **HEDIS**\(^1\) is a Performance Measure specification issued by **NCQA**\(^2\)
  - E.g., all diabetic patients must have annual eye exams
- Meeting HEDIS standards is a requirement for government funded healthcare (**Medicare**)
- Checking/reporting is difficult and costly
  - Complex specifications & annual revisions
  - Disparate data sources
  - Ad hoc schemas including *implicit information*

---

\(^1\) Healthcare Effectiveness Data and Information Set
\(^2\) National Committee for Quality assurance
Semantic Technology Solution
Semantic Technology Solution

- Capture HEDIS diabetic care spec using **OWL RL & SWRL**
  - 174 axioms/rules
- Load data into **RDFox** triple store
  - Data from **466k patients** (Georgia region); approx 100M records
  - Translated into **548M triples** (32GB RAM)
- Use **materialisation** and **SPARQL queries** to identify relevant patients and check HEDIS conformance
  - Entire process takes ≈ 7,000s on a commodity server
  - Extends graph to **731M triples** (43GB)
RDFox OWL RL Engine

- Targets SOTA **main-memory, multil-core** architecture
  - Optimized in-memory storage with ‘mostly’ lock-free parallel inserts
  - Commodity server with 128 GB can store $>10^9$ triples
  - 10-20 x speedup with 32/16 threads/cores
  - LUBM 120K in 251s (20M triples/s) on T5-8 with 4TB/1024 threads
RDFox OWL RL Engine

- Targets SOTA main-memory, mulit-core architecture
  - Optimized in-memory storage with ‘mostly’ lock-free parallel inserts
  - Commodity server with 128 GB can store $>10^9$ triples
  - 10-20 x speedup with 32/16 threads/cores
  - LUBM 120K in 251s (20M triples/s) on T5-8 with 4TB/1024 threads

- Native equality reasoning (owl:sameAs) via rewriting
- Incremental reasoning (delete 5k triples from LUBM 50K in <1s)
Research Issues

- **Expressive power:**
  - Capturing HEDIS spec requires negation and aggregation
  - Current solution interleaves SPARQL queries & materialisation
  - Extending RDFox to support aggregation & stratified negation

- **Scalability:**
  - KP have approx 10M patients in total (20 times larger)
  - RDFox can store 10B triples on a 1TB machine [Nenov et al]
  - Working on (semantic) partitioning and distributed materialisation
Observations and Opinions
Language Design: Features/Bugs
Language Design: Features/Bugs

**RDF blank nodes**
- ❌ Existential semantics are not always intuitive or appropriate
- ❌ NP complexity for base layer (compared to $AC^0$ for DBs)
- ❌ Stack(s) is (are) broken
- ✔ Sometimes useful (of course)
- ✔ Endless entertainment/papers for theoreticians

**Open world semantics**
- ❌ Not always intuitive or appropriate (for DB people/applications)
- ❌ Difficult to combine with, e.g., defaults and NAF
- ✔ Appropriate for Web
- ✔ Appropriate (often) for data integration
Only unary and binary predicates
- “Incompatibility” with DBs
- Reification costly and problematical
  - Mapping from DBs is non-trivial
  - Canonical URI creation is critical

✓ Often sufficient in practice
✓ Simple(r) data structures and algorithms

RDF über alles
- Verbose, and difficult to fully specify/constrain/parse syntax
- Nonsensical statements (rdf:type, rdf:type: rdf:type)

✓ Single storage and querying infrastructure
✓ Uniform/combined data and schema queries
Language Design: Features/Bugs

Lacks/includes necessary/useless features

✗ Specification should have included ________*
✗ Specification should not have included ________*

THERE IS NO ONE PERFECT LANGUAGE

✓ Standardisation critical for infrastructure development and uptake
✓ RDF+OWL+SPARQL provides a **huge advantage/opportunity**

* Insert favourite/most-hated feature
Barriers to Uptake
Barriers to Uptake

Ontology (and mapping) engineering

- Critically dependent on good quality ontologies (and mappings)
- Sophisticated tools and methodologies available
- But it's still hard!
Barriers to Uptake

Ontology (and mapping) engineering

- Critically dependent on good quality ontologies (and mappings)
- Sophisticated tools and methodologies available
- But it's still hard!

Kudos to Protégé
Scalability - v- expressive power

- Users expect semantic technology features in addition to DB features
- RDF entailment already NP-complete
- OWL 2 DL entailment NP-Hard w.r.t. size of data and N2ExpTime-complete w.r.t. size of ontology+data
- OWL 2 profiles “more simply and/or efficiently implemented”
  - OWL 2 QL – AC⁰ data complexity (same as DBs)
  - OWL 2 EL – PTime-complete combined and data complexity
  - OWL 2 RL – PTime-complete combined and data complexity
- but at the cost of reduced expressive power
Scalability Beyond the Profiles?

**LUBM ontology** includes the axioms

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :ResearchGroup)))

EquivalentClasses(:Employee
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :Organization)))

SubClassOf(:ResearchGroup :Organization)

and LUBM 1 data includes **547 instances of :ResearchAssistant**
Scalability Beyond the Profiles?

**LUBM ontology** includes the axioms

\[
\text{SubClassOf}(\text{:ResearchAssistant})
\text{ObjectIntersectionOf}(\text{:Person}
\text{ObjectSomeValuesFrom}(\text{:worksFor :ResearchGroup}))
\]

\[
\text{EquivalentClasses}(\text{:Employee}
\text{ObjectIntersectionOf}(\text{:Person}
\text{ObjectSomeValuesFrom}(\text{:worksFor :Organization})))
\]

\[
\text{SubClassOf}(\text{:ResearchGroup :Organization})
\]

and LUBM 1 data includes **547 instances of :ResearchAssistant**

**How many** of these RAs are in the answer to the following query?

\[
\text{SELECT } ?x \text{ WHERE } \{ ?x \text{ rdf:type :Employee } \}
\]
Scalability Beyond the Profiles?

**LUBM ontology** includes the axioms

```
SubClassOf(:ResearchAssistant
  ObjectIntersectionOf(:Person
    ObjectSomeValuesFrom(:worksFor :ResearchGroup)))
EquivalentClasses(:Employee
  ObjectIntersectionOf(:Person
    ObjectSomeValuesFrom(:worksFor :Organization)))
SubClassOf(:ResearchGroup :Organization)
```

and LUBM 1 data includes 547 instances of :ResearchAssistant

**How many** of these RAs are in the answer to the following query?

```
SELECT ?x WHERE { ?x rdf:type :Employee }
```

Answer computed by (most) RL reasoners: none of them
Combined Approach

- Combined approach allows RL reasoners to be used to support scalable query answering for all profiles.
Combined Approach

- Combined approach allows RL reasoners to be used to support scalable query answering for all profiles
- And even for many non-profile ontologies in RSA class

![Diagram showing property chains in PTime, RSA, EL, QL, and RL]
Combined Approach

- **Combined approach** allows RL reasoners to be used to support **scalable query answering** for all profiles
- And even for many **non-profile ontologies in RSA class**

- **How does it work?**
  - Overapproximate $\mathcal{O}$ (e.g., “Skolemise” RHS existentials) into $\mathcal{O}'$
    - i.e., $\mathcal{O}' \models \mathcal{O}$
  - Queries answered w.r.t. $\mathcal{O}'$ to give **complete but unsound** answers
  - Spurious answers are eliminated by a filtration step
Combined Approach

\( \mathcal{O} \quad \sim \quad \mathcal{O}' \)

\[ \text{SubClassOf}(:\text{ResearchAssistant} \quad \text{ObjectIntersectionOf}(:\text{Person} \quad \text{ObjectSomeValuesFrom}(:\text{worksFor} \quad :\text{ResearchGroup}))) \]

\[ \text{SubClassOf}(:\text{ResearchAssistant} \quad \text{ObjectIntersectionOf}(:\text{Person} \quad \text{ObjectHasValue}(:\text{worksFor} \quad :\text{RG1}) \quad (:\text{RG1}, \text{rdf:type}, :\text{ResearchGroup})) \)
Combined Approach

\[ O \]

SubClassOf(ResearchAssistant
  ObjectIntersectionOf(Person
  ObjectSomeValuesFrom(worksFor ResearchGroup)))

\[ \downarrow \]

\[ O' \]

SubClassOf(ResearchAssistant
  ObjectIntersectionOf(Person
  ObjectHasValue(worksFor RG1)))
  (RG1, rdf:type, ResearchGroup)
Combined Approach

\[ \mathcal{O} \quad \sim \quad \mathcal{O}' \]

\[ \text{SubClassOf}(\text{:ResearchAssistant}) \]
\[ \text{ObjectIntersectionOf}(\text{:Person}) \]
\[ \text{ObjectSomeValuesFrom}(\text{:worksFor :ResearchGroup})) \]

\[ \text{EquivalentClasses}(\text{:Employee}) \]
\[ \text{ObjectIntersectionOf}(\text{:Person}) \]
\[ \text{ObjectSomeValuesFrom}(\text{:worksFor :Organization})) \]

\[ \text{SubClassOf}(\text{:ResearchAssistant}) \]
\[ \text{ObjectIntersectionOf}(\text{:Person}) \]
\[ \text{ObjectHasValue}(\text{:worksFor :RG1})) \]
\[ (:RG1, \text{rdf:type}, :\text{ResearchGroup}) \]

\[ \text{SubClassOf}(\text{:Employee}) \]
\[ \text{ObjectIntersectionOf}(\text{:Person}) \]
\[ \text{ObjectHasValue}(\text{:worksFor :ORG1})) \]
\[ (:\text{ORG1}, \text{rdf:type}, :\text{Organization}) \]

\[ \text{SubClassOf}(\text{ObjectIntersectionOf}(\text{:Person}) \]
\[ \text{ObjectSomeValuesFrom}(\text{:worksFor :Organization})) \]
\[ (:\text{Employee}) \]
Combined Approach

\[ \mathcal{O} \quad \mathcal{O}' \]

$$\text{SubClassOf}(:\text{ResearchAssistant}$$
$$\quad \text{ObjectIntersectionOf}( :\text{Person}$$
$$\quad \text{ObjectSomeValuesFrom}( :\text{worksFor} :\text{ResearchGroup})$$
$$\quad \text{ObjectIntersectionOf}( :\text{Person}$$
$$\quad \text{ObjectSomeValuesFrom}( :\text{worksFor} :\text{Organization})$$
$$\quad \text{ObjectIntersectionOf}( :\text{Person}$$
$$\quad \text{ObjectHasValue}( :\text{worksFor} :\text{RG1})$$
$$\quad ( :\text{RG1}, \text{rdf:type}, :\text{ResearchGroup})$$
$$\quad \text{ObjectIntersectionOf}( :\text{Person}$$
$$\quad \text{ObjectHasValue}( :\text{worksFor} :\text{ORG1})$$
$$\quad ( :\text{ORG1}, \text{rdf:type}, :\text{Organization})$$
$$\quad \text{ObjectIntersectionOf}( :\text{Person}$$
$$\quad \text{ObjectSomeValuesFrom}( :\text{worksFor} :\text{Organization})$$
$$\quad :\text{Employee})$$
$$\quad \text{SubClassOf}( :\text{ResearchGroup} :\text{Organization})$$

$$\text{SubClassOf}( :\text{ResearchGroup} :\text{Organization})$$
Combined Approach

\[ \mathcal{O} \sim \mathcal{O}' \]

SubClassOf(\:ResearchAssistant
   ObjectIntersectionOf(\:Person
   ObjectSomeValuesFrom(\:worksFor \:ResearchGroup)))

EquivalentClasses(\:Employee
   ObjectIntersectionOf(\:Person
   ObjectSomeValuesFrom(\:worksFor \:Organization)))

SubClassOf(\:ResearchGroup \:Organization)

\[ \mathcal{O}' \cup \{(\:RA1, rdf:type, \:ResearchAssistant)\} \models (\:RA1, rdf:type, \:Employee) \]

SubClassOf(\:ResearchAssistant
   ObjectIntersectionOf(\:Person
   ObjectHasValue(\:worksFor \:RG1)))
   (\:RG1, rdf:type, \:ResearchGroup)

SubClassOf(\:Employee
   ObjectIntersectionOf(\:Person
   ObjectHasValue(\:worksFor \:ORG1)))
   (\:ORG1, rdf:type, \:Organization)

SubClassOf(ObjectIntersectionOf(\:Person
   ObjectSomeValuesFrom(\:worksFor \:Organization))
   \:Employee)

SubClassOf(\:ResearchGroup \:Organization)
Combined Approach

\[ \mathcal{O} \leftrightarrow \mathcal{O}' \]

- SubClassOf(:ResearchAssistant
  ObjectIntersectionOf(:Person
  ObjectSomeValuesFrom(:worksFor :ResearchGroup)))
- EquivalentClasses(:Employee
  ObjectIntersectionOf(:Person
  ObjectSomeValuesFrom(:worksFor :Organization)))
- SubClassOf(:ResearchGroup :Organization)

\[ \mathcal{O}' \cup \{ (:RA1, rdf:type, :ResearchAssistant) \} \models (:RA1, rdf:type, :Employee) \]

Scalable Query Answering for OWL DL?

Given an OWL DL ontology $\mathcal{O}$ dataset $\mathcal{D}$ and query $q$

- We can transform $\mathcal{O}$ into strictly stronger OWL RL ontology $\mathcal{O}_u$
  - Roughly speaking, Skolemise, and transform $\lor$ into $\land$
Scalable Query Answering for OWL DL?

Given an OWL DL ontology $\mathcal{O}$ dataset $\mathcal{D}$ and query $q$

- We can transform $\mathcal{O}$ into strictly stronger OWL RL ontology $\mathcal{O}_u$
  - Roughly speaking, Skolemise, and transform $\lor$ into $\land$
- RL reasoning w.r.t. $\mathcal{O}_u$ gives upper bound answer $U$

$$\text{ans}(q, \langle \mathcal{O}, \mathcal{D} \rangle) \subseteq \text{ans}(q, \langle \mathcal{O}_u, \mathcal{D} \rangle) = U$$
Scalable Query Answering for OWL DL?

Given an **OWL DL ontology** $\mathcal{O}$ dataset $\mathcal{D}$ and query $q$

- We can transform $\mathcal{O}$ into **strictly stronger OWL RL ontology** $\mathcal{O}_u$
  - Roughly speaking, Skolemise, and transform $\lor$ into $\land$

- RL reasoning w.r.t. $\mathcal{O}_u$ gives **upper bound answer** $U$

  $$\text{ans}(q, \langle \mathcal{O}, \mathcal{D} \rangle) \subseteq \text{ans}(q, \langle \mathcal{O}_u, \mathcal{D} \rangle) = U$$

- If $L = U$, then both answers are **sound and complete**

  $$L \subseteq \text{ans}(q, \langle \mathcal{O}, \mathcal{D} \rangle) \subseteq U$$
Scalable Query Answering for OWL DL?
Scalable Query Answering for OWL DL?

- If $L \neq U$, then $U \setminus L$ identifies a (small) set of "possible" answers
  - Delineates range of uncertainty
  - Can more efficiently check possible answers using, e.g., HermiT (but still infeasible if dataset is large)
  - Can use $U \setminus L$ to identify small(er) "relevant" subset of axioms/data sufficient to check possible answers (using proof tracing)

- Further optimisations
  - Use ELHO "combined" technique to tighten lower bound [Stefanoni et al]
  - Use "summarisation" technique to tighten upper bound [Dolby et al]
  - ...
PAGOdA System

Full reasoning
Extracting subsets
Computing query bounds
Materialisation
PAGOda System

(b) LUBM query processing
A Perspective on the Semantic Web
A Perspective on the Semantic Web
A Perspective on the Semantic Web

- **Graph DB** with rich and flexible schema
- Applications in **data integration and analysis** (as well as the Web)
- **Competing** with Graph/NoSQL DBs, Bigtable, HBase, ...
- **RDF+OWL+SPARQL** standards and technologies offer important advantages
A Perspective on the Semantic Web
A Perspective on the Semantic Web

We are here
A Perspective on the Semantic Web

- We have the (right) **languages**
- We have the (right) **technology**
- We have interest and even enthusiasm from (potential) **users**
- All(!) we need to do is **engage** and (continue to) **deploy**

We are here
Thank you for listening
Thank you for listening

Any questions?
References


References


