DERI

Linked Broken Data?

Dr Axel Polleres

Digital Enterprise Research Institute, National University of Ireland, Galway

Based on joint work with Aidan Hogan, Andreas Harth, Renaud Delbru, Giovanni Tummarello, Stefan Decker

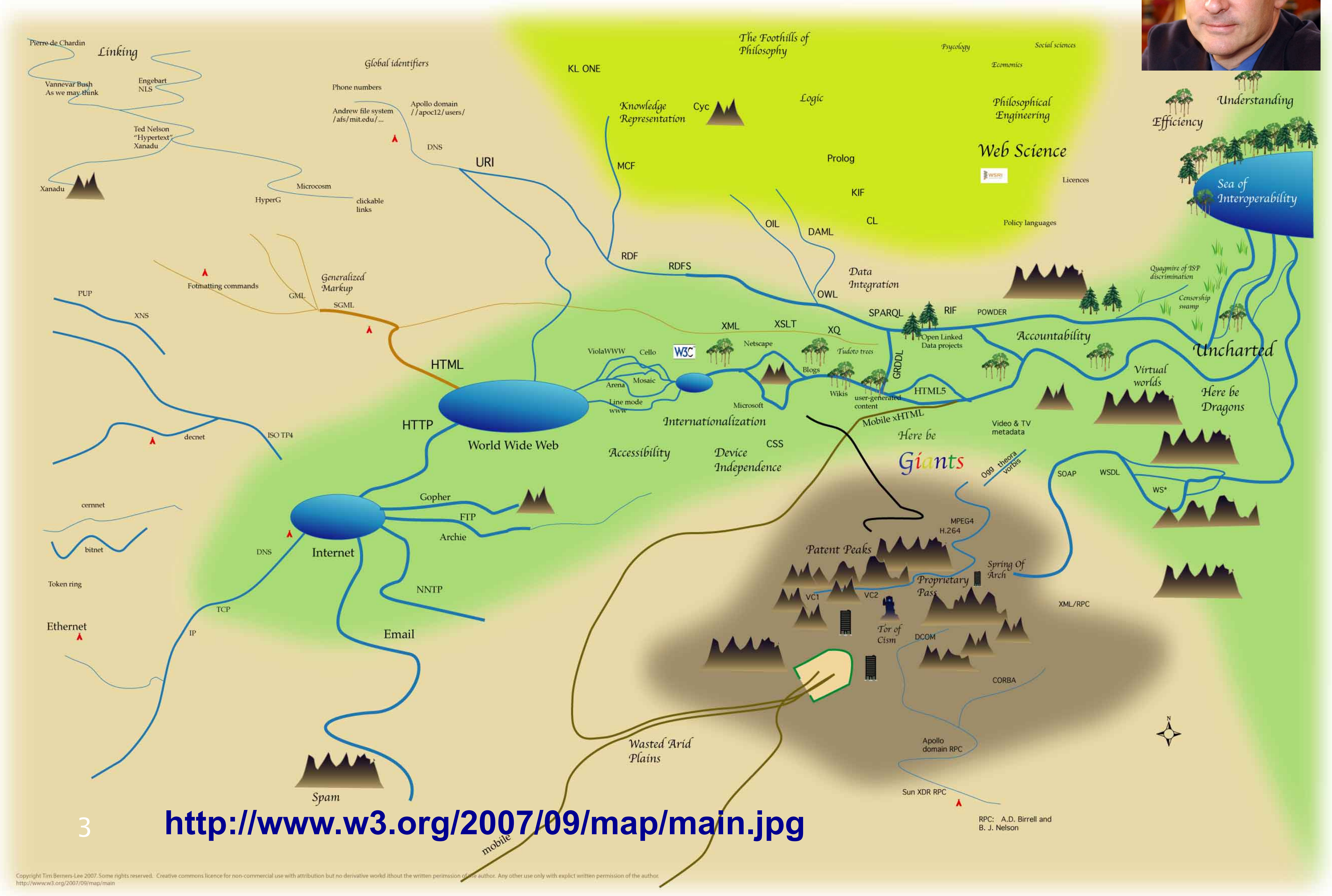
OÉ Gaillimh
NUI Galway

sfi
science foundation ireland
fondúireacht eolaíochta éireann

Today's talk is about…

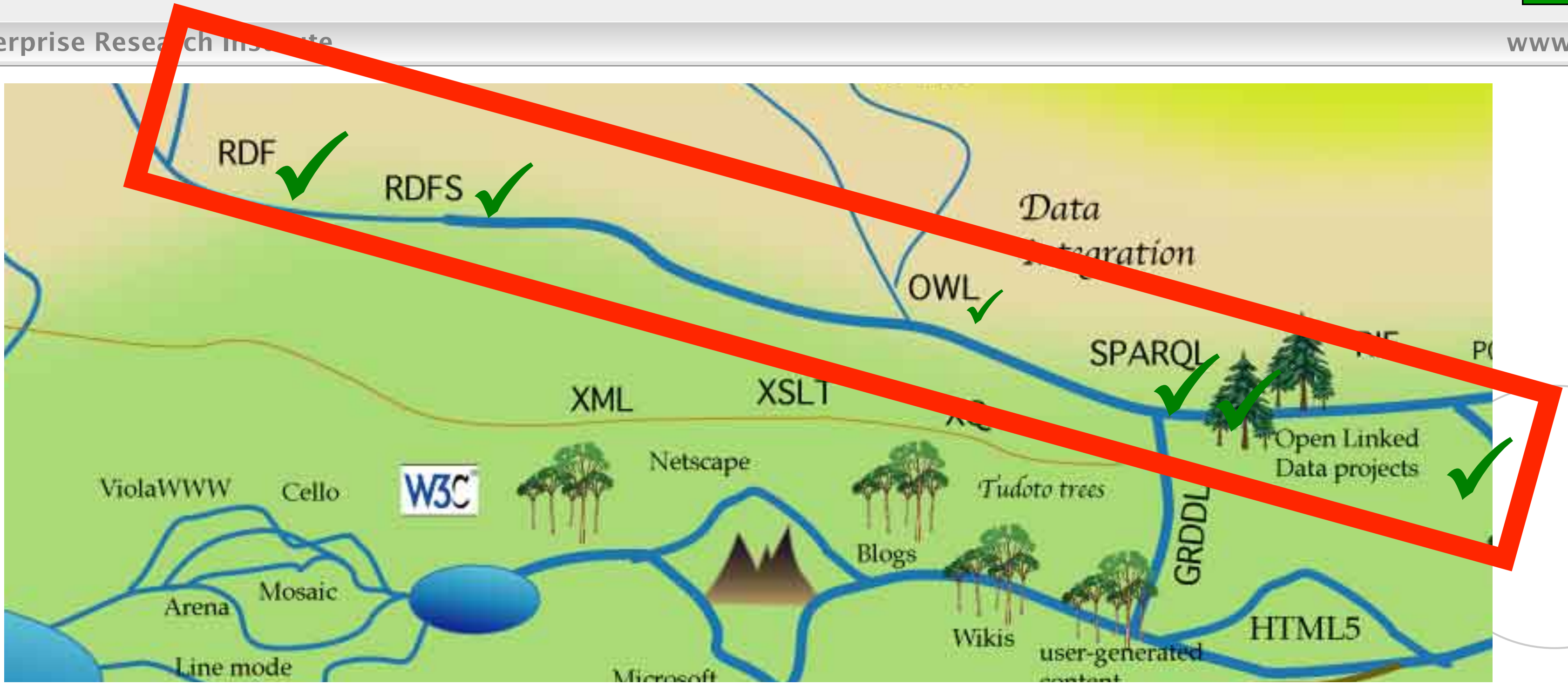
Reasoning **on** *today's Semantic Web…*

The Web map 2008 © Tim Berners-Lee



http://www.w3.org/2007/09/map/main.jpg

The Web map 2008 © Tim Berners-Lee



- more and more structured data (RDF) available on the Web thanks to …
- … vocabularies (RDFS+OWL) becoming established
- … exporters, (GRDDL, RDFa), Linked Open Data, etc.
- … In this talk: What can we do with it already in terms of Reasoning?

Outline

- Brief intro of RDF/OWL/Linked Open Data
- Reasoning over Web Data: Challenges
  - Inconsistencies
  - Common mistakes
- Reasoning over Web Data: Dealing with the challenges
  - Reasoning in **Sindice.com**
  - Reasoning in **SWSE.com**
- How to avoid common mistakes upfront:
  - **RDFAlerts, Pedantic-Web Group**
- What I'd hope you to take-home

Example: Finding experts/reviewers?

***Tim Berners-Lee**, Dan Connolly, Lalana Kagal, Yosi Scharf, Jim Hendler: **N3Logic: A logical framework for the World Wide Web**. Theory and Practice of Logic Programming (TPLP), Volume 8, p249-269*

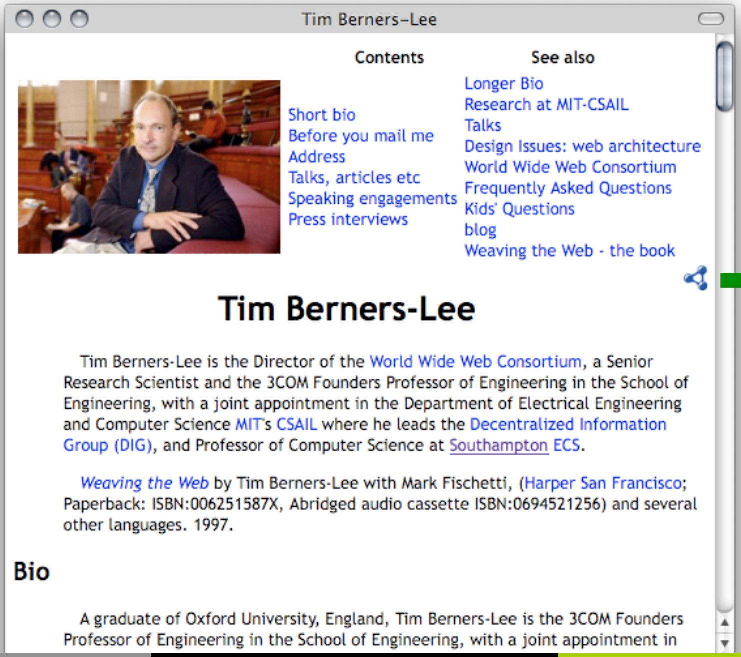
- Who are the right reviewers? Who has the right expertise?
- Which reviewers are in conflict?
- Observation: Most of the necessary data already on the Web!

- More and more of it follows the **Linked Data principles**, i.e.:
  1. Use URIs as names for things
  2. Use HTTP dereferenceable URIs so that people can look up those names.
  3. When someone looks up a URI, provide useful information.
  4. Include links to other URIs so that they can discover more things.

RDF on the Web

- **(i) directly by the publishers**
- (ii) by e.g. GRDDL transformations, D2R, RDFa exporters, etc.

FOAF/RDF linked from a home page: personal data (foaf:name, foaf:phone, etc.), relationships foaf:knows, rdfs:seeAlso)



Contents

- Short bio
- Before you mail me
- Address
- Talks, articles etc
- Speaking engagements
- Press interviews

See also

- Longer Bio
- Research at MIT-CSAIL
- Talks
- Design Issues: web architecture
- World Wide Web Consortium
- Frequently Asked Questions
- Kids' Questions
- blog
- Weaving the Web - the book

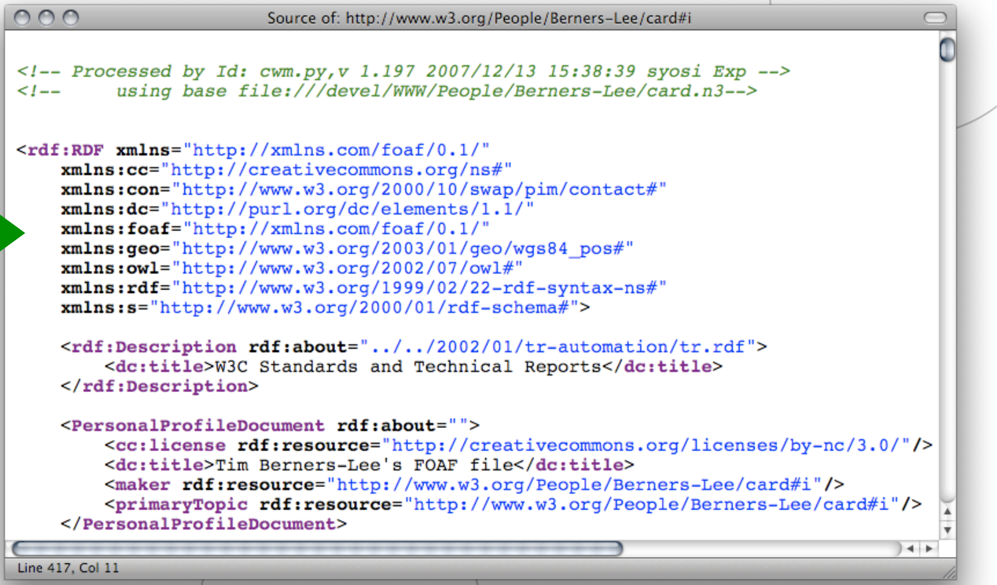
Tim Berners-Lee

Tim Berners-Lee is the Director of the World Wide Web Consortium, a Senior Research Scientist and the 3COM Founders Professor of Engineering in the School of Engineering, with a joint appointment in the Department of Electrical Engineering and Computer Science MIT's CSAIL where he leads the Decentralized Information Group (DIG), and Professor of Computer Science at Southampton ECS.

Weaving the Web by Tim Berners-Lee with Mark Fischetti, (Harper San Francisco; Paperback: ISBN:006251587X, Abridged audio cassette ISBN:0694521256) and several other languages. 1997.

Bio

A graduate of Oxford University, England, Tim Berners-Lee is the 3COM Founders Professor of Engineering in the School of Engineering, with a joint appointment in

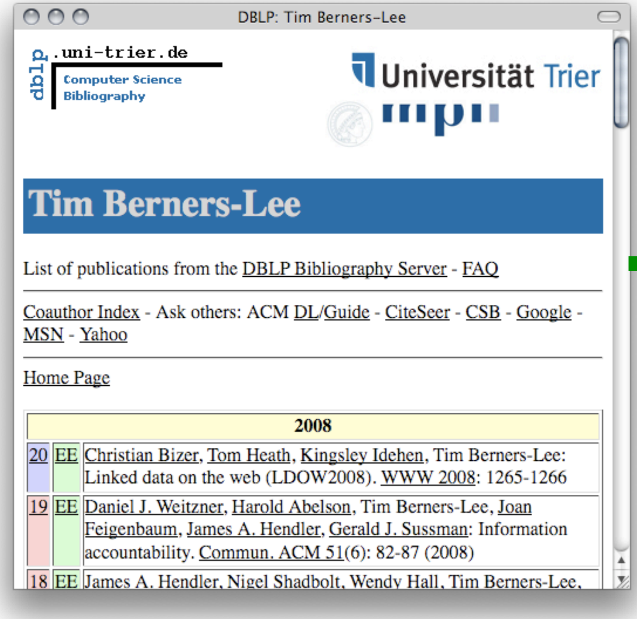


```
<!-- Processed by Id: cwm.py,v 1.197 2007/12/13 15:38:39 syosi Exp -->
<!--     using base file:///devel/WWW/People/Berners-Lee/card.n3-->


<rdf:RDF xmlns="http://xmlns.com/foaf/0.1/"
    xmlns:cc="http://creativecommons.org/ns#"
    xmlns:con="http://www.w3.org/2000/10/swap/pim/contact#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:s="http://www.w3.org/2000/01/rdf-schema#">

    <rdf:Description rdf:about="../../2002/01/tr-automation/tr.rdf">
        <dc:title>W3C Standards and Technical Reports</dc:title>
    </rdf:Description>

    <PersonalProfileDocument rdf:about="">
        <cc:license rdf:resource="http://creativecommons.org/licenses/by-nc/3.0/"/>
        <dc:title>Tim Berners-Lee's FOAF file</dc:title>
        <maker rdf:resource="http://www.w3.org/People/Berners-Lee/card#i"/>
        <primaryTopic rdf:resource="http://www.w3.org/People/Berners-Lee/card#i"/>
    </PersonalProfileDocument>
```

RDF on the Web

- (i) directly by the publishers
- **(ii) by e.g. GRDDL transformations, D2R, RDFa exporters, etc.**

e.g. L3S’ RDF export of the DBLP citation index, using FUB’s D2R (http://dblp.l3s.de/d2r/)



DBLP: Tim Berners-Lee

dblp.uni-trier.de Computer Science Bibliography

Universität Trier

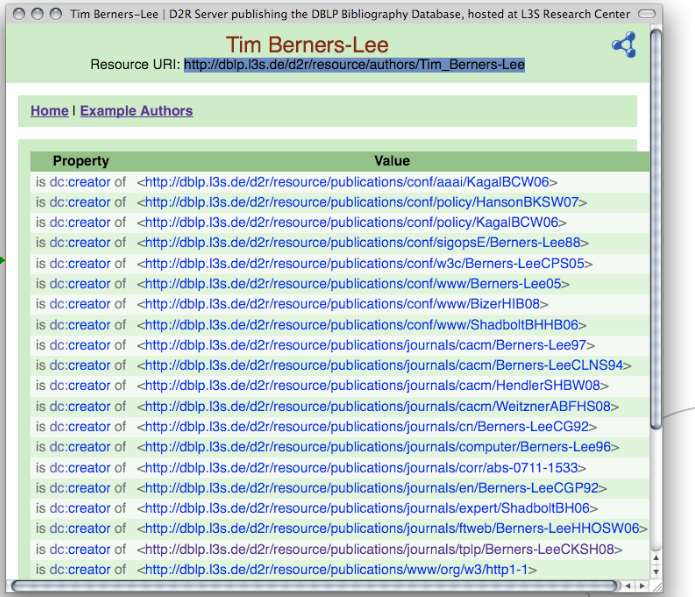
Tim Berners-Lee

List of publications from the DBLP Bibliography Server - FAQ

Coauthor Index - Ask others: ACM DL/Guide - CiteSeer - CSB - Google - MSN - Yahoo

Home Page

2008		
20	EE	Christian Bizer, Tom Heath, Kingsley Idehen, Tim Berners-Lee: Linked data on the web (LDOW2008). WWW 2008: 1265-1266
19	EE	Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James A. Hendler, Gerald J. Sussman: Information accountability. Commun. ACM 51(6): 82-87 (2008)
18	EE	James A. Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee,



Tim Berners-Lee | D2R Server publishing the DBLP Bibliography Database, hosted at L3S Research Center

Tim Berners-Lee

Resource URI: http://dblp.l3s.de/d2r/resource/authors/Tim_Berners-Lee

Home | Example Authors

Property	Value
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/conf/aaai/KagalBCW06>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/conf/policy/HansonBKSW07>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/conf/policy/KagalBCW06>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/conf/sigopsE/Berners-Lee88>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/conf/w3c/Berners-LeeCPS05>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/conf/www/Berners-Lee05>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/conf/www/BizerHIB08>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/conf/www/ShadboltBHHB06>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/cacm/Berners-Lee97>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/cacm/Berners-LeeCLNS94>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/cacm/HendlerSHBW08>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/cacm/WeitznerABFHS08>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/cn/Berners-LeeCG92>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/computer/Berners-Lee96>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/corr/abs-0711-1533>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/en/Berners-LeeCGP92>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/expert/ShadboltBH06>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/ftweb/Berners-LeeHHOSW06>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/journals/tplp/Berners-LeeCKSH08>
is dc:creator of	<http://dblp.l3s.de/d2r/resource/publications/www/org/w3/http1-1>

Gives unique URIs to authors, documents, etc. on DBLP! E.g.,

http://dblp.l3s.de/d2r/resource/authors/Tim_Berners-Lee,

http://dblp.l3s.de/d2r/resource/publications/journals/tplp/Berners-LeeCKSH08

Provides RDF version of all DBLP data + query interface!

RDF Data online: Example

- Data in RDF: **Triples**
  - □ DBLP:

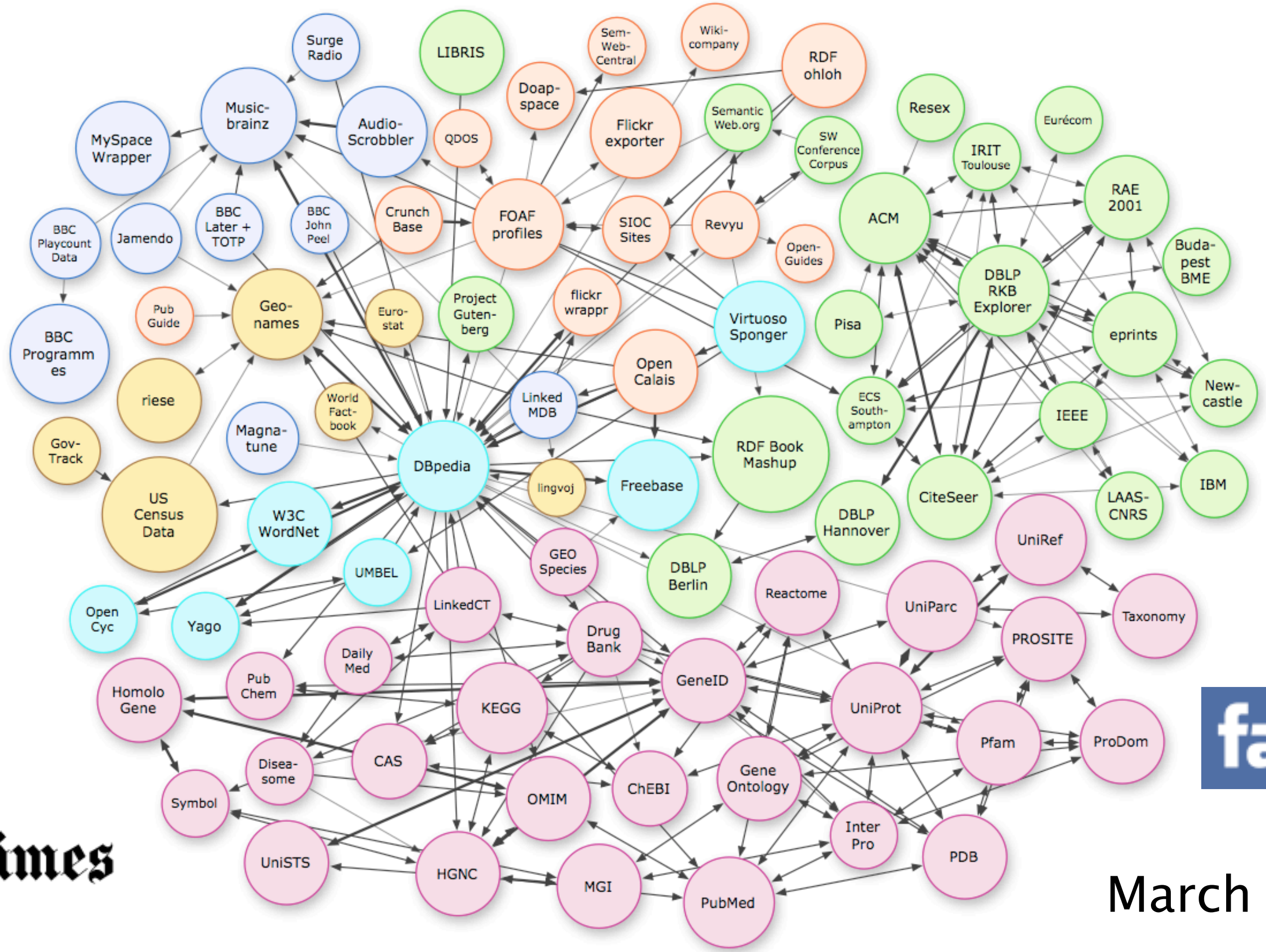
```
<http://dblp.l3s.de/…/journals/tplp/Berners-LeeCKSH08> rdf:type swrc:Article.
<http://dblp.l3s.de/…/journals/tplp/Berners-LeeCKSH08> dc:creator
     <http://dblp.l3s.de/d2r/…/Tim_Berners-Lee> .
  …
<http://dblp.l3s.de/d2r/…/Tim_Berners-Lee> foaf:homepage
       <http://www.w3.org/People/Berners-Lee/> .
…
<http://dblp.l3s.de/d2r/…/Dan_Brickley> foaf:name "Dan Brickley"^^xsd:string.
```

  - □ Tim Berners-Lee's FOAF file:

```
<http://www.w3.org/People/Berners-Lee/card#i> foaf:knows
    <http://dblp.l3s.de/d2r/…/Dan_Brickley> .
<http://www.w3.org/People/Berners-Lee/card#i> rdf:type foaf:Person .
<http://www.w3.org/People/Berners-Lee/card#i> foaf:homepage
    <http://www.w3.org/People/Berners-Lee/> .
```

Linked Open Data

dblp.uni-trier.de Computer Science Bibliography



March 2009

...

- Excellent tutorial here: http://www4.wiwiss.fu- berlin.de/bizer/pub/LinkedDataTutorial/

How can I query that data? SPARQL

- SPARQL – W3C approved standardized query language for RDF:
  - look-and-feel of “SQL for the Web”
  - allows to ask queries like
    - “All documents by Tim Berners-Lee”
    - “Names of all persons who co-authored with authors of http://dblp.l3s.de/d2r/…/Berners-LeeCKSH08 or known by co-authors”

    …

Example:

```
SELECT ?D
FROM <http://dblp.l3s.de/…/authors/Tim_Berners-Lee>
WHERE {?D dc:creator <http://dblp.l3s.de/…/authors/Tim_Berners-Lee>}
```

SPARQL more complex patters: e.g. UNIONs

- “**Names *of all persons who co-authored with authors of http://dblp.l3s.de/d2r/…/Berners-LeeCKSH08*** *or known by co-authors*”

```
SELECT ?Name WHERE
   { <http://dblp.l3s.de/d2r/resource/publication/journals/tplp/Berners-LeeCKSH08>
        dc:creator ?Author.
        ?D dc:creator ?Author.
        ?D dc:creator ?CoAuthor.
        ?CoAuthor foaf:name ?Name
   }
```

SPARQL more complex patters: e.g. UNIONs

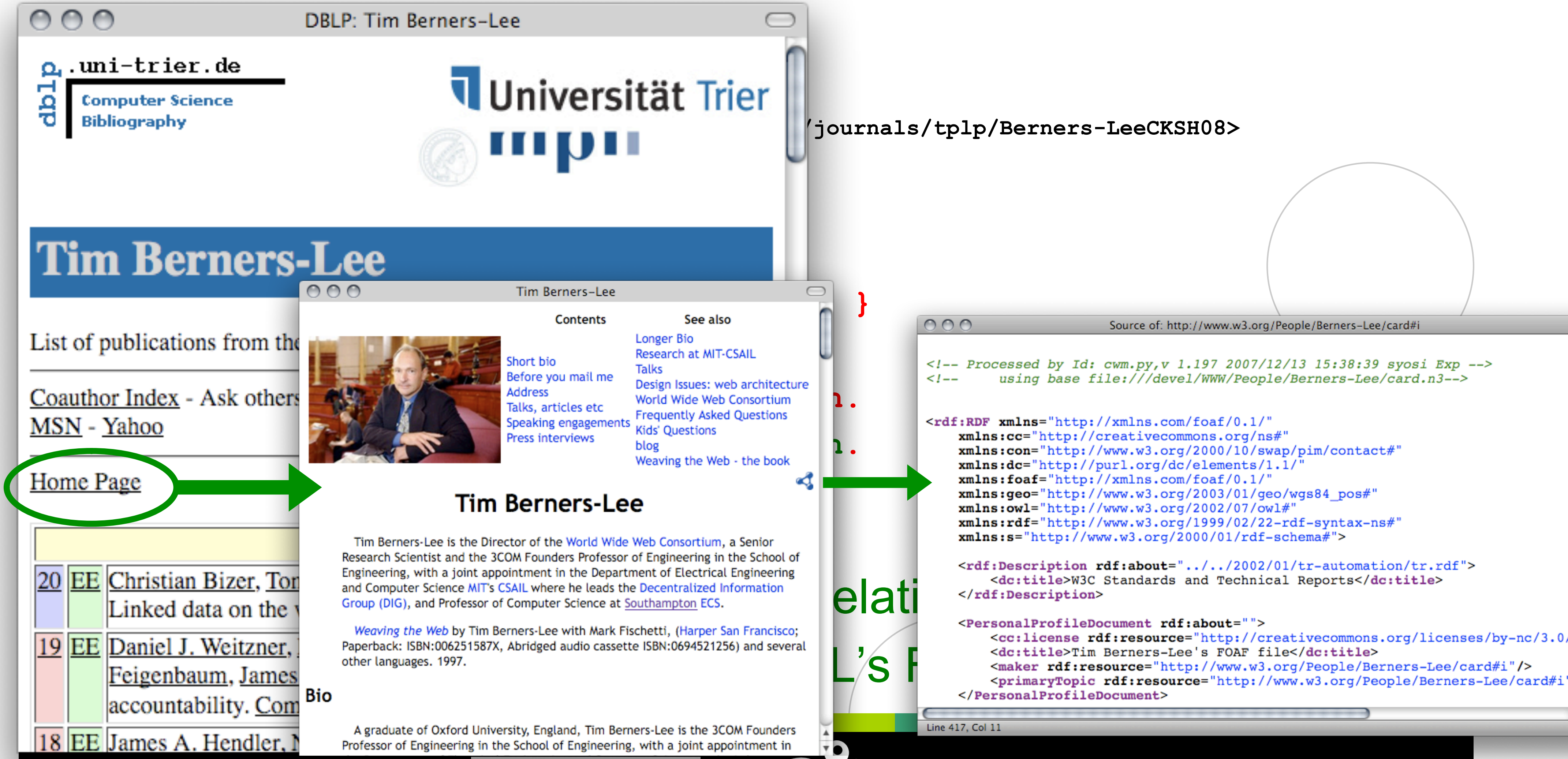
- "Names *of all persons who* ***co-authored*** *with authors of http://dblp.l3s.de/ d2r/…/Berners-LeeCKSH08* ***or known by co-authors***"

```
SELECT ?Name WHERE
    { <http://dblp.l3s.de/d2r/resource/publications/journals/tplp/Berners-LeeCKSH08>
          dc:creator ?Author.
          ?D dc:creator ?Author.
          ?D dc:creator ?CoAuthor.
          {  ?CoAuthor foaf:name ?Name . }
         UNION
          { ?CoAuthor foaf:knows ?Person.
            ?Person rdf:type foaf:Person.
           ?Person foaf:name ?Name }
    }
```

- Doesn't work… no foaf:knows relations in DBLP ☹
- Needs **Linked Data**! E.g. TimBL's FOAF file!

SPARQL more complex patters: e.g. UNIONs

- “Names *of all persons who* **co-authored** *with authors of http://dblp.l3s.de/d2r/…/Berners-LeeCKSH08* **or known by co-authors**”



/journals/tplp/Berners-LeeCKSH08>

}

elati

L’s F

DBLP: Tim Berners-Lee

dblp.uni-trier.de Computer Science Bibliography

Universität Trier

mpi

Tim Berners-Lee

List of publications from the

Coauthor Index - Ask others

MSN - Yahoo

Home Page

20	EE	Christian Bizer, Tom Linked data on the
19	EE	Daniel J. Weitzner, Feigenbaum, James accountability. Com
18	EE	James A. Hendler,

Tim Berners-Lee

Contents

Short bio
Before you mail me
Address
Talks, articles etc
Speaking engagements
Press interviews

See also

Longer Bio
Research at MIT-CSAIL
Talks
Design Issues: web architecture
World Wide Web Consortium
Frequently Asked Questions
Kids' Questions
blog
Weaving the Web - the book

Tim Berners-Lee

Tim Berners-Lee is the Director of the World Wide Web Consortium, a Senior Research Scientist and the 3COM Founders Professor of Engineering in the School of Engineering, with a joint appointment in the Department of Electrical Engineering and Computer Science MIT's CSAIL where he leads the Decentralized Information Group (DIG), and Professor of Computer Science at Southampton ECS.

Weaving the Web by Tim Berners-Lee with Mark Fischetti, (Harper San Francisco; Paperback: ISBN:006251587X, Abridged audio cassette ISBN:0694521256) and several other languages. 1997.

Bio

A graduate of Oxford University, England, Tim Berners-Lee is the 3COM Founders Professor of Engineering in the School of Engineering, with a joint appointment in

Source of: http://www.w3.org/People/Berners-Lee/card#i

```
<!-- Processed by Id: cwm.py,v 1.197 2007/12/13 15:38:39 syosi Exp -->
<!--     using base file:///devel/WWW/People/Berners-Lee/card.n3-->


<rdf:RDF xmlns="http://xmlns.com/foaf/0.1/"
    xmlns:cc="http://creativecommons.org/ns#"
    xmlns:con="http://www.w3.org/2000/10/swap/pim/contact#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:s="http://www.w3.org/2000/01/rdf-schema#">

    <rdf:Description rdf:about="../../2002/01/tr-automation/tr.rdf">
        <dc:title>W3C Standards and Technical Reports</dc:title>
    </rdf:Description>

    <PersonalProfileDocument rdf:about="">
        <cc:license rdf:resource="http://creativecommons.org/licenses/by-nc/3.0/
        <dc:title>Tim Berners-Lee's FOAF file</dc:title>
        <maker rdf:resource="http://www.w3.org/People/Berners-Lee/card#i"/>
        <primaryTopic rdf:resource="http://www.w3.org/People/Berners-Lee/card#i"
    </PersonalProfileDocument>
```

Line 417, Col 11

Back to the Data:

- □ DBLP:

```
<http://dblp.l3s.de/…/journals/tplp/Berners-LeeCKSH08> rdf:type swrc:Article.
<http://dblp.l3s.de/…/journals/tplp/Berners-LeeCKSH08> dc:creator
     <http://dblp.l3s.de/d2r/…/Tim_Berners-Lee> .
  …
<http://dblp.l3s.de/d2r/…/Tim_Berners-Lee> foaf:homepage
        <http://www.w3.org/People/Berners-Lee/> .
```

- □ Tim Berners-Lee’s FOAF file:

```
<http://www.w3.org/People/Berners-Lee/card#i> foaf:knows
     <http://dblp.l3s.de/d2r/…/Dan_Brickley> .
<http://www.w3.org/People/Berners-Lee/card#i> foaf:homepage
     <http://www.w3.org/People/Berners-Lee/> .
```

- Even if I have the FOAF data, I cannot answer the query:
  - Different identifiers used for Tim Berners-Lee
  - Who tells me that Dan Brickley is a foaf:Person?
- Linked Data needs **Reasoning**!

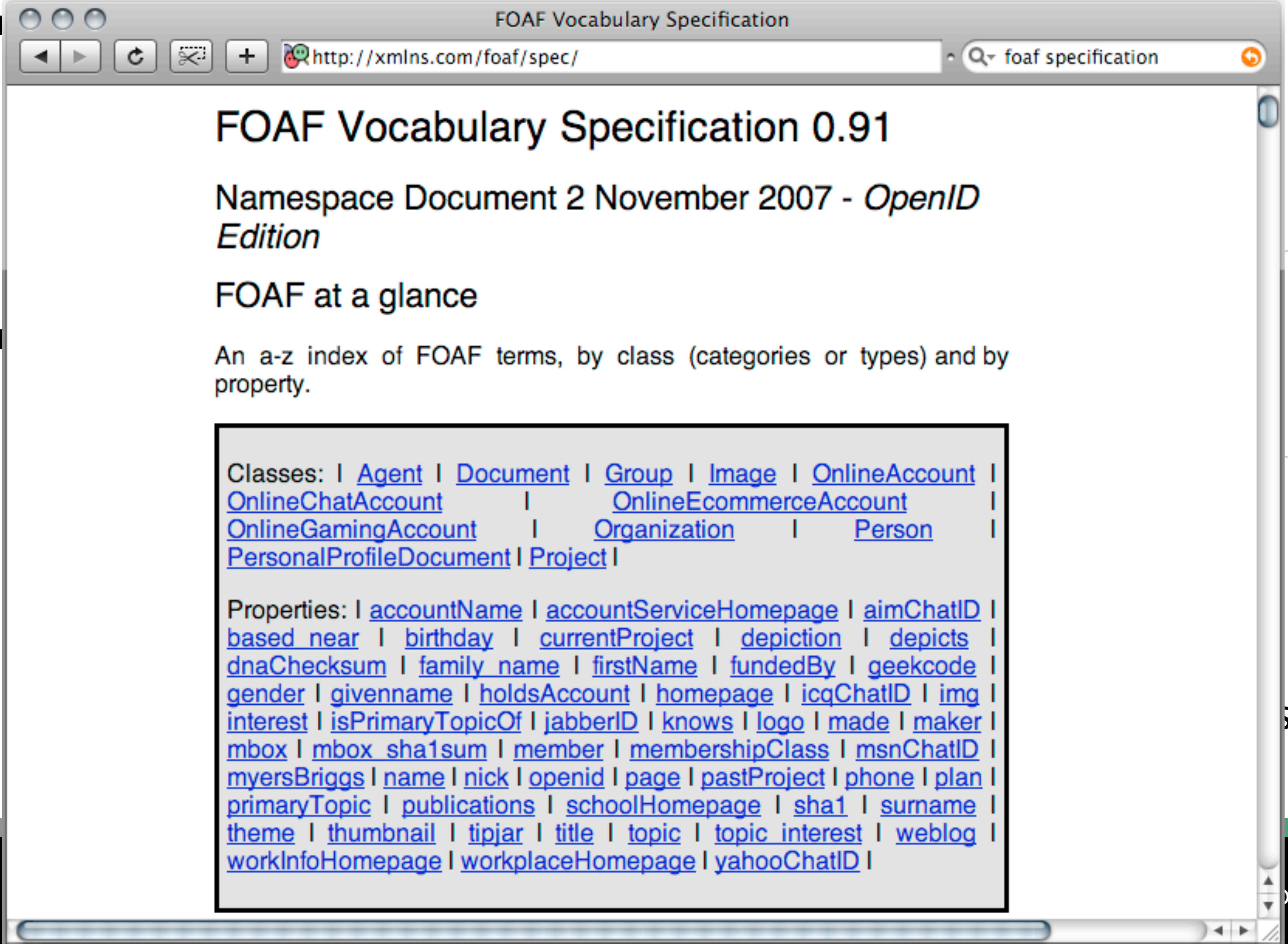
Reasoning on Semantic Web Data

- **Vocabularies** (i.e. collections of classes and properties that belong together, e.g. `foaf:`):
  - □ Properties: `foaf:name foaf:homepage, foaf:knows`
  - □ Classes: `foaf:Person, foaf:Document`

- Typically should have formal descriptions of their structure:
  - □ RDF Schema, and OWL
  - □ These formal descriptions often “called” **ontologies.**
  - □ Ontologies *add “semantics”* to the data.
  - □ Ontologies are themselves written in RDF, using special vocabularies (`rdf:, rdfs:, owl:`) with special semantics

  → Ontologies are themselves part of the Linked Data Web!

Reasoning on Semantic Web Data



FOAF Vocabulary Specification 0.91

Namespace Document 2 November 2007 - *OpenID Edition*

FOAF at a glance

An a-z index of FOAF terms, by class (categories or types) and by property.

Classes: | Agent | Document | Group | Image | OnlineAccount | OnlineChatAccount | OnlineEcommerceAccount | OnlineGamingAccount | Organization | Person | PersonalProfileDocument | Project |

Properties: | accountName | accountServiceHomepage | aimChatID | based_near | birthday | currentProject | depiction | depicts | dnaChecksum | family_name | firstName | fundedBy | geekcode | gender | givenname | holdsAccount | homepage | icqChatID | img | interest | isPrimaryTopicOf | jabberID | knows | logo | made | maker | mbox | mbox_sha1sum | member | membershipClass | msnChatID | myersBriggs | name | nick | openid | page | pastProject | phone | plan | primaryTopic | publications | schoolHomepage | sha1 | surname | theme | thumbnail | tipjar | title | topic | topic_interest | weblog | workInfoHomepage | workplaceHomepage | yahooChatID |

s

owledge.

Ontologies: Example FOAF

```
foaf:knows rdfs:domain foaf:Person
```

Everybody who knows someone is a Person

```
foaf:knows rdfs:range foaf:Person
```

Everybody who is known is a Person

```
foaf:Person rdfs:subclassOf foaf:Agent
```

Everybody Person is an Agent.

```
foaf:homepage rdf:type owl:inverseFunctionalProperty .
```

A homepage uniquely identifies its owner ("key" property)

…

RDFS+OWL inference by rules 1/2

- Semantics of RDFS can be partially expressed as (Datalog like) rules:

```
rdfs1: { ?S rdf:type ?C } :- { ?S ?P ?O . ?P rdfs:domain ?C . }
rdfs2: { ?O rdf:type ?C } :- { ?S ?P ?O . ?P rdfs:range ?C . }


rdfs3: { ?S rdf:type ?C2 } :- {?S rdf:type ?C1 . ?C1 rdfs:subclassOf ?C2 . }
```

cf. informative Entailment rules in [RDF-Semantics, W3C, 2004], [Muñoz et al. 2007]

RDFS+OWL inference by rules 2/2

- OWL Reasoning e.g. inverseFunctionalProperty can also (partially) be expressed by Rules:

```
owl1: { ?S1 owl:SameAs ?S2 } :-
           { ?S1 ?P ?O . ?S2 ?P ?O . ?P rdf:type owl:InverseFunctionalProperty }


owl2: { ?Y ?P ?O } :- { ?X owl:SameAs ?Y . ?X ?P ?O }
owl3: { ?S ?Y ?O } :- { ?X owl:SameAs ?Y . ?S ?X ?O }
owl4: { ?S ?P ?Y } :- { ?X owl:SameAs ?Y . ?S ?P ?X }
```

cf. pD* fragment of OWL, [ter Horst, 2005], or, more recent: OWL2 RL

RDFS+OWL inference by rules: Example:

- By rules of the previous slides we can infer additional information needed, e.g.

TimBL's FOAF: `<…/Berners-Lee/card#i> foaf:knows <…/Dan_Brickley> .`

FOAF Ontology: `foaf:knows rdfs:range foaf:Person`

by rdfs2 → `<…/Dan_Brickley> rdf:type   foaf:Person.`

TimBL's FOAF: `<…/Berners-Lee/card#i> foaf:homepage <http://www.w3.org/People/Berners-Lee/> .`

DBLP: `<…/dblp.l3s.de/d2r/…/Tim_Berners-Lee> foaf:homepage <http://www.w3.org/People/Berners-Lee/> .`

FOAF Ontology: `foaf:homepage rdfs:type owl:InverseFunctionalProperty.`

by owl1 → `<…/Berners-Lee/card#i> owl:sameAs <…/Tim_Berners-Lee>.`

  - Who tells me that Dan Brickley is a foaf:Person? → solved!
  - Different identifiers used for Tim Berners-Lee → solved!

RDFS+OWL inference, what's missing?

- Note: Not all of OWL Reasoning can be expressed in Datalog straightforwardly, e.g.:

```
foaf:Person owl:disjointWith foaf:Organisation
```

Can be written/and reasoned about with FOL/DL reasoners:

$$\forall X. Person(X) \equiv \neg Organisation(X)$$

$$Person \sqcap Organisation \sqsubseteq \bot$$

Problem: Inconsistencies! Complete FOL/DL reasoning is not necessarily suitable for Web data…

Why is complete reasoning non-optimal anyways?

- Our use case: Search the Semantic Web!

  - □ Hypothetically: The explosive semantics of inconsistencies in DL/FOL reasoning would spoil our results.

  - □ What if we throw all into one big KB? one inconsistency…

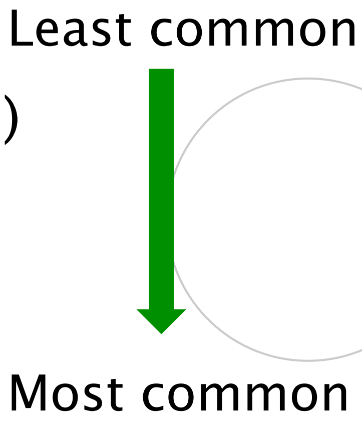
    a owl:differentFrom a .
    :me ex:age “old”^^xs:integer.

  … would make everything true.

Inconsistencies/wrong inferences on Web Data

4 main reasons

- □ Publishers deliberately publish spoilt data (“SPAM”)
- □ Opinions differ
- □ “URI-sense” ambiguities
- □ **Accidently** wrong/inconsistent



Least common

Most common

Publishers deliberately publish spoilt data ("SPAM")

- Examples:
  - □ a owl:differentFrom a .
  - □ http://www.polleres.net/nasty.rdf
- Can occur for "testdata" being published, deliberate SPAM can become an issue, as the SW grows!

Opinions differ

- Fictitous Example Ontology:

```
Originofthings.example.org:
           o1:surpremePower owl:disjointWith o1:naturalPhenomenom.
           o1:originsFrom rdf:type owl:functionalProperty.
       o1:god rdf:type o1:surpremePower.
           o1:evolution rdf:type o1:naturalPhenomenom.

darwin.example.org:
     ex:mankind o1:originsFrom o1:evolution .


creationism.example.org:
     ex:mankind o1:originsFrom o1:god

FlyingSpaghettimonster.org
    fsm::theSpaghettiMonster rdf:type surpremePower.
    ex:mankind o1:originsFrom fsm:theSpaghettiMonster.
```

“URI-sense” ambiguities

```
<http://www.polleres.net>
    foaf:knows <http://apassant.net>
```

i.e., why do I have to use a different URI for myself and my homepage?

Many people don’t understand/like this and make mistakes.

But is this really a mistake or a design error?

Accidentially inconsistent data

```
:me ex:age "old"^^xs:integer.
```

can e.g. arise from an exporter, that collects age from a form

```
Source1 (faulty):
TimBL foaf:homepage <http://www.w3.org>
TimBL rdf:type foaf:Person.
W3.org:
W3C foaf:homepage <http://www.w3.org>
W3C rdf:type foaf:Organisation.
```

Did occur in our Web crawls at some point, people don't have the right semantics in mind!

- Suspiciously resembles problems with e.g. flawed HTML … browsers, normal search engines still have to deal with it

→ So do we!

Accidently wrong (non-inconsistent data)

- FOAF Ontology:

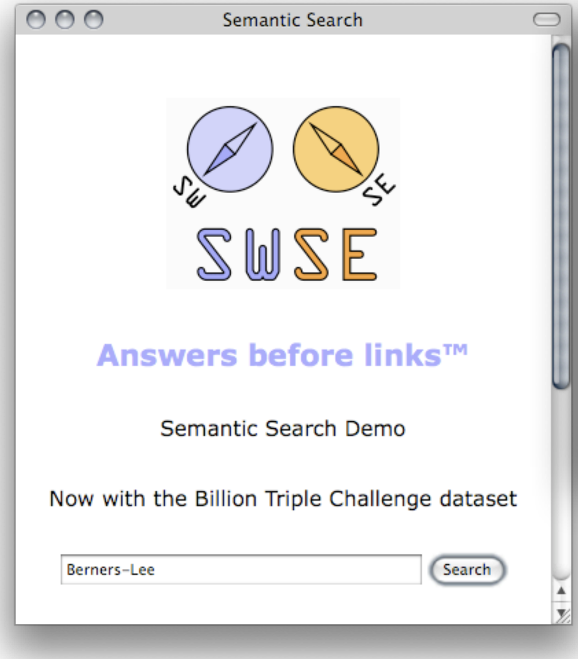
```
foaf:mbox rdf:type owl:InverseFunctionalProperty
```

- Careless FOAF exporters produce something like this for any empty email address:

```
ex:alice foaf:mbox “mailto:”
ex:bob foaf:mbox “mailto:”
…
```

IFP reasoning (Rules: owl1-4) on Web Data equates too many things! Dangerous!

How can I reason about Web Data in a Semantic Search Engine?

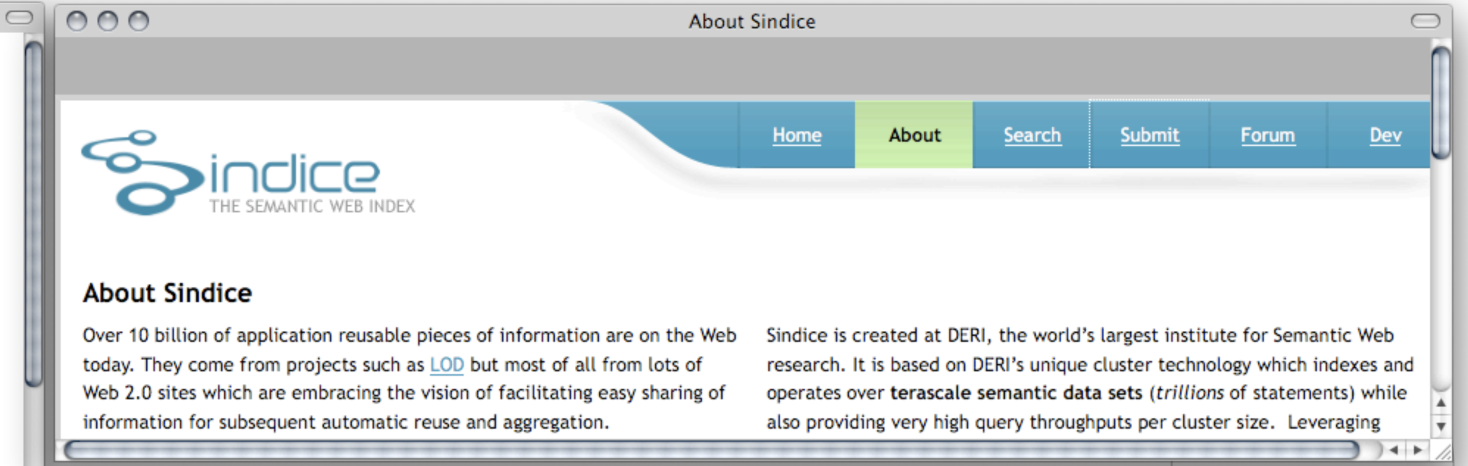


Answers before links™

Semantic Search Demo

Now with the Billion Triple Challenge dataset

Berners-Lee



About Sindice

Over 10 billion of application reusable pieces of information are on the Web today. They come from projects such as LOD but most of all from lots of Web 2.0 sites which are embracing the vision of facilitating easy sharing of information for subsequent automatic reuse and aggregation.

Sindice is created at DERI, the world's largest institute for Semantic Web research. It is based on DERI's unique cluster technology which indexes and operates over **terascale semantic data sets** (*trillions* of statements) while also providing very high query throughputs per cluster size. Leveraging

http://swse.deri.org
http://sindice.com

- Datawarehouse approach, e.g. SWSE
  - □ crawling, harvesting, SPARQL interface, RDFS+resricted OWL reasoning
- Search/Lookup indices for the Semantic Web, e.g. Sindice
  - □ Indexing RDF sources on the Web, go there and query yourself

Requirements:

- Scale
  - □ Both engines crawl millions, even billions of triples (rapidly increasing) … latest numbers talk about orders of 100B RDF triples online.
- “Humble” Inference
  - □ Both want to do at least limited inferencing to deliver valuable implicit information/connections
- Tolerance
  - □ Both should be tolerant/cautious against common faults
    - Filter if possible deliberate mess
    - Filter (repair?) Accidental errors
    - Keep inconsistencies local

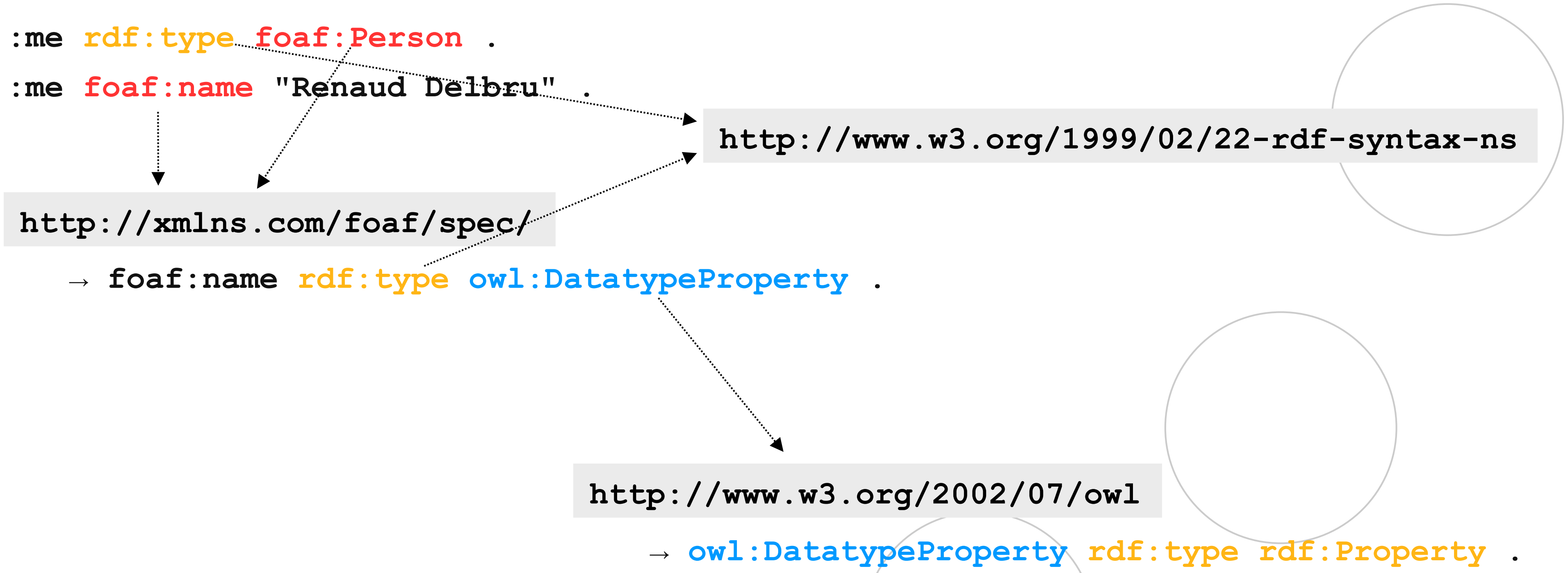
2 approaches

- Sindice:
  - □ Uses a standard rule-based OWL engine (OWLIM, ter Horst's pD* rules)
  - □ Inferencing "per document", only importing necessary ontologies
  - □ Keeps an "ontology cache" for all crawled ontologies for efficiency
  - □ No cross-document inferences

- SWSE+SAOR:
  - □ Works on whole crawl (huge file)
    - – Existing solutions, e.g. OWLIM don't work on that, infer too much
  - □ Our own reasoner: SAOR (scalable authoritative OWL reasoner)

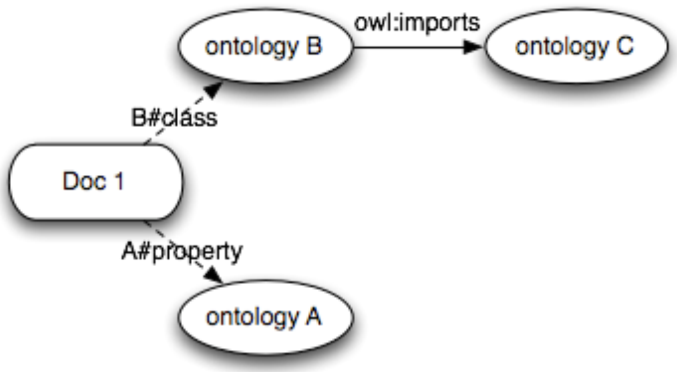
Reasoning in Sindice:

- Implicit import
  - □ Based on W3C best practices – Linked Data Principles
  - □ By dereferencing class or property URI



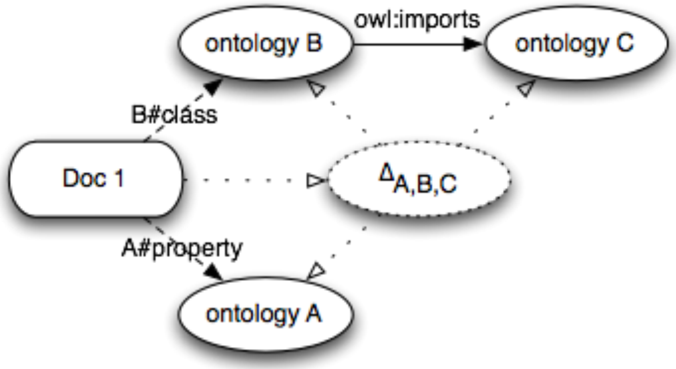
```
:me rdf:type foaf:Person .
:me foaf:name "Renaud Delbru" .

http://www.w3.org/1999/02/22-rdf-syntax-ns

http://xmlns.com/foaf/spec/
  → foaf:name rdf:type owl:DatatypeProperty .

http://www.w3.org/2002/07/owl
  → owl:DatatypeProperty rdf:type rdf:Property .
```

Reasoning in Sindice: Ontology Cache: Update Strategy



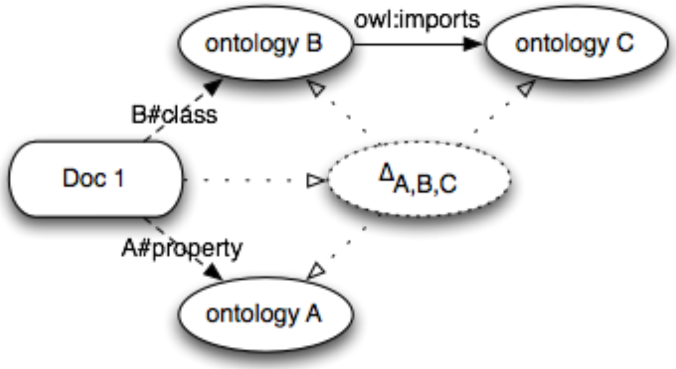
1. Import closure of Doc1 is materialised

Reasoning in Sindice: Ontology Cache: Update Strategy



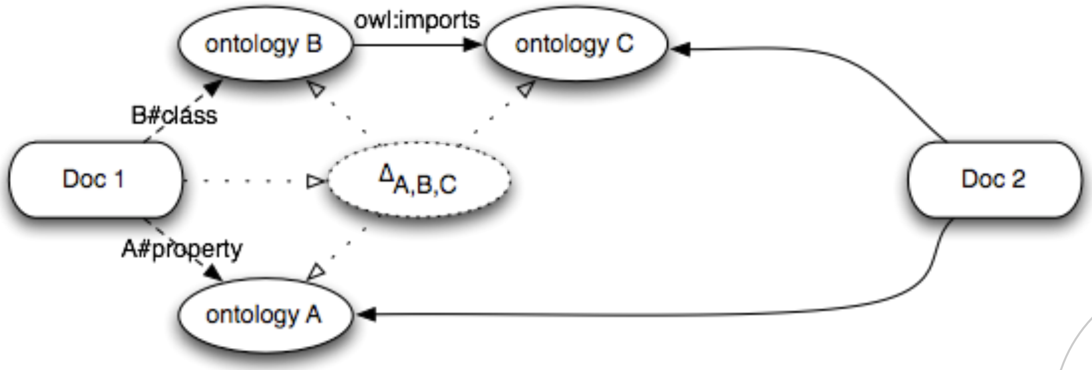
1. Import closure of Doc1 is materialised
2. Compute deductive closure of aggregate context O_A, O_B, O_C

Reasoning in Sindice: Ontology Cache: Update Strategy



1. Import closure of Doc1 is materialised
2. Compute deductive closure of aggregate context O_A, O_B, O_C
3. Store $\Delta_{A,B,C}$ in a separate named RDF triple set

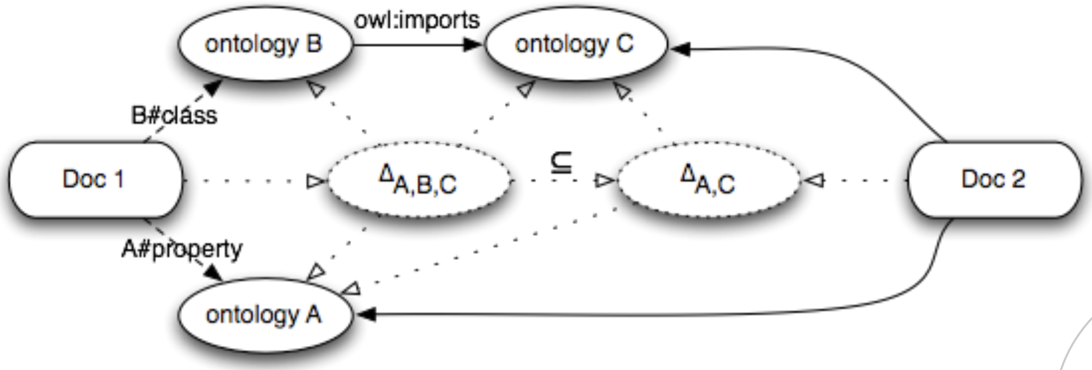
Reasoning in Sindice: Ontology Cache: Update Strategy



A new document is coming, importing only O_A and O_C :

1. Compute deductive closure of O_A and O_C

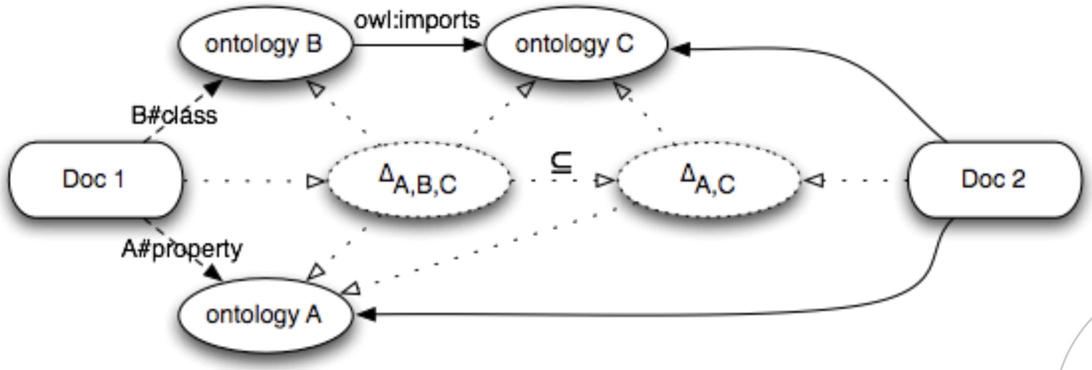
Reasoning in Sindice: Ontology Cache: Update Strategy



A new document is coming, importing only O_A and O_C :

1. Compute deductive closure of O_A and O_C
2. Store $\Delta_{A,C}$ in a separate named RDF triple set

Reasoning in Sindice: Ontology Cache: Update Strategy

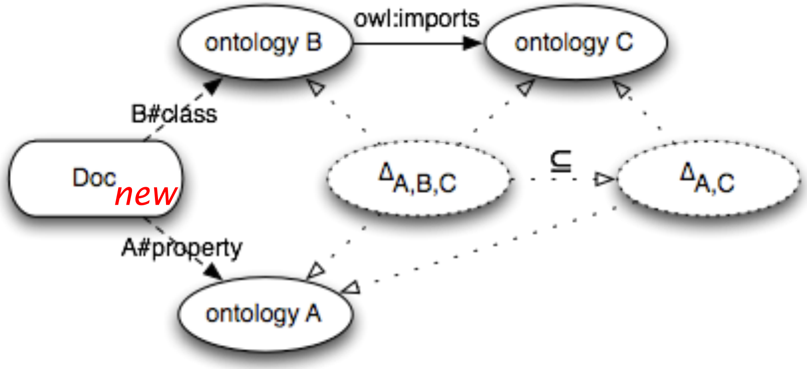


A new document is coming, importing only O_A and O_C :

1. Compute deductive closure of O_A and O_C
2. Store $\Delta_{A,C}$ in a separate named RDF triple set
3. Update deductive closure of O_A, O_B, O_C so that the inferred triples are never duplicated
   a) Substract $\Delta_{A,C}$ from $\Delta_{A,B,C}$
   b) add inclusion relation

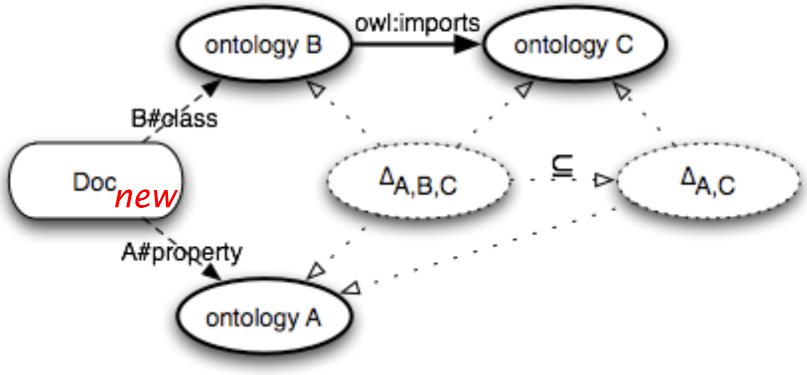
i.e., $\Delta_{A,B,C} := \Delta_{A,B,C} - \Delta_{A,C} + \Delta_{A,C}$ owl:imports $\Delta_{A,B,C}$

Reasoning in Sindice:
Ontology Cache: Querying Strategy



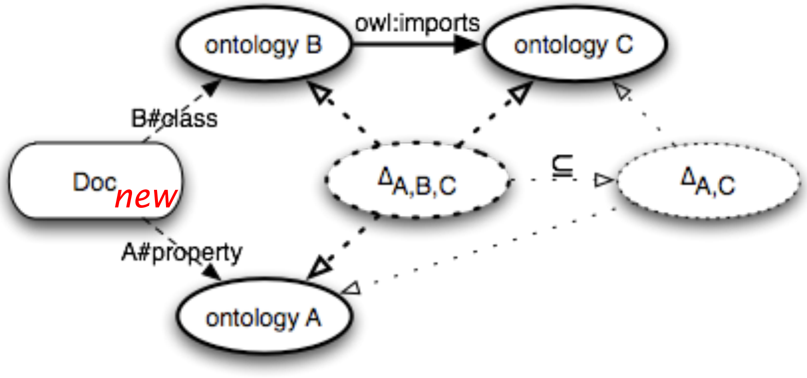
1. A document imports O_A and O_B

Reasoning in Sindice: Ontology Cache: Querying Strategy



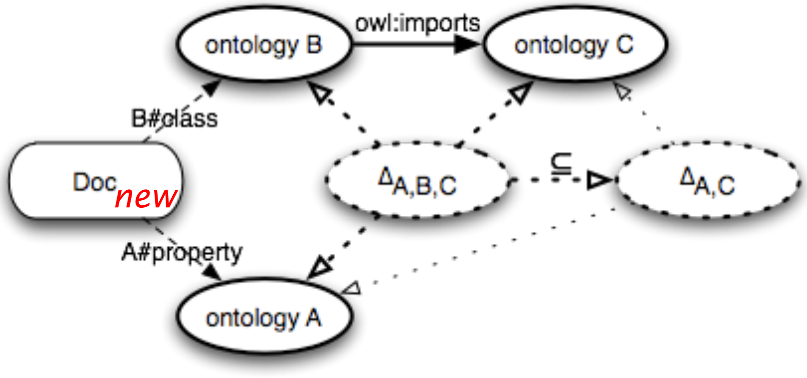
1. A document imports O_A and O_B
2. Import closure is derived, and corresponding ontology network activated

Reasoning in Sindice: Ontology Cache: Querying Strategy



1. A document imports O_A and O_B
2. Import closure is derived, and corresponding ontology network activated
3. The related $\Delta_{A,B,C}$ is derived and activated

Reasoning in Sindice: Ontology Cache: Querying Strategy



1. A document imports O_A and O_B
2. Import closure is derived, and corresponding ontology network activated
3. The related $\Delta_{A,B,C}$ is derived and activated
4. It is then found that $\Delta_{A,B,C}$ includes $\Delta_{A,C \text{ which}}$ is also activated

→ Our Observation: “caching” Tbox inferences makes indexing (mostly ABox) much faster

Reasoning in Sindice.com:

- Pros:
  - □ Works well, can be distributed
  - □ Stable against local inconsistencies/errors
  - □ Can use “off-the-shelf” reasoners (OWLIM is just the current choice)

- Cons:
  - □ might miss important inferences covering the “gist” of linked data e.g. ☹

```
Ontology o2:
    o2:hasAncestor rdf:type owl:transitiveProperty.
    o2:hasParent subPropertyOf ex:hasAncestor.
axel.rdf:
    <axel.rdf#me> o2:hasParent <mechthild.rdf#me>
mechthild.rdf:
    <mechthild.rdf#me> o2:hasParent <franz.rdf#me>
```

- Inference of ancestor relation between axel and franz needs both rdf datafiles!
  - □ Not covered by “ontology closure” alone
  - □ Extending “fetching closure” to instances too expensive…
  - □ … boils down to reasoning over the whole crawl … looses nice property of “keeping mess local”

SAOR - Reasoning for SWSE

http://swse.deri.org/

Objects before documents!

- Take the challenge to reason over the whole crawl dataset … HUGE!
- Approach:

  SAOR – **S**calable **A**uthoritative **O**WL **R**easoning

Idea

- Apply a subset of OWL reasoning using a **tailored ruleset.**
- Forward-chaining rule based approach based on [ter Horst, 2005], but tweaked.

- Reduced output statements for the SWSE use case…
  - □ Must be *scalable*, must be *reasonable*
- … incomplete w.r.t. OWL **BY DESIGN!**
  - □ **SCALABLE:** Tailored ruleset
    - – file-scan processing
    - – avoid joins
  - □ **AUTHORITATIVE:** Avoid Non-Authoritative inference (“hijacking”, “non-standard vocabulary use”)

Scalable Reasoning

- **Scan 1:**

  Scan all data (1.1b statements), separate T-Box statements, load T-Box statements (8.5m) into memory, perform authoritative analysis.

- **Scan 2:**

  Scan all data and join all statements with in-memory T-Box .

  □ Only works for inference rules with 0-1 A-Box patterns

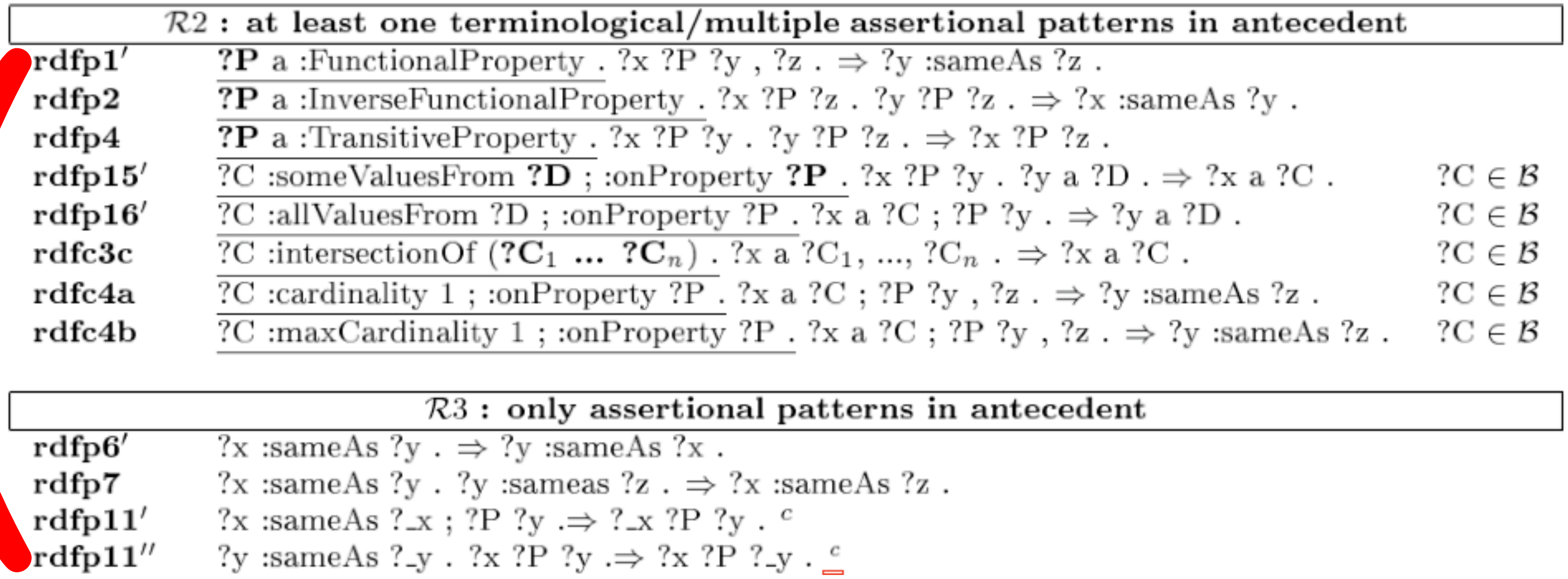
  □ No T-Box expansion by inference

  → Needs "tailored" ruleset

Rules Applied:
Tailored version of [ter Horst, 2005]

#	DL Syntax	Rule	# Inferred
$\mathcal{G}0$: NO A-BOX PATTERNS IN ANTECEDENT			
00	$\{o_i ... o_n\}$	?C :oneOf (?o_1 ... ?o_n) . ⇒ ?o_1 ... ?o_n a ?C .	35,161
$\mathcal{G}1$: ONE A-BOX PATTERN IN ANTECEDENT			
01	$C \sqsubseteq D$	?C rdfs:subClassOf ?D . ?s a ?C . ⇒ ?s a ?D .	1,124,758,631
02_a	$C \equiv D$	?C :equivalentClass ?D . ?s a ?C . ⇒ ?s a ?D .	8,137,162
02_b		?C :equivalentClass ?D . ?s a ?D . ⇒ ?s a ?C .	90,372
03	$P \sqsubseteq Q$	?P rdfs:subPropertyOf ?Q . ?s ?P ?o . ⇒ ?s ?Q ?o .	156,462,399
04_a	$P \equiv Q$	?P :equivalentProperty ?Q . ?s ?P ?o . ⇒ ?s ?Q ?o .	5,667,464
04_b		?P :equivalentProperty ?Q . ?s ?Q ?o . ⇒ ?s ?P ?o .	6,642
05_a	$P \equiv P_0^-$	?P :inverseOf ?Q . ?s ?P ?o . ⇒ ?o ?Q ?s .	230,945,040
05_b		?P :inverseOf ?Q . ?s ?Q ?o . ⇒ ?o ?P ?s .	230,941,648
06	$\top \sqsubseteq \forall P^-.C$	?P rdfs:domain ?C . ?s ?P ?o . ⇒ ?s a ?C .	588,530,865
07	$\top \sqsubset \forall P.C$	?P rdfs:range ?C . ?s ?P ?o . ⇒ ?o a ?C .	528,995,909
08	$P \equiv P^-$	?P a :SymmetricProperty . ?s ?P ?o . ⇒ ?o ?P ?s .	560,460
09_a	$\exists P.x$	?C :hasValue ?x; :onProperty ?P . ?y ?P ?x . ⇒ ?y a ?C .	98,601
09_b		?C :hasValue ?x; :onProperty ?P . ?y a ?C . ⇒ ?y ?P ?x .	104,780
10	$C_1 \sqcup ... \sqcup C_n$	?C :unionOf (?C_1...?C_i...?C_n) . ?x a ?C_i . ⇒ ?x a ?C .	81,736,234
11	$(\geq 1P)$	?C :minCardinality 1; :onProperty ?P . ?x ?P ?y . ⇒ ?x a ?C .	65,283,322
12_a	$C_1 \sqcap ... \sqcap C_n$	?C :intersectionOf (?C_1 ... ?C_n) . ?y a ?C . ⇒ ?y a ?C_1, ..., ?C_n .	115,383
12_b	$C_1 \sqcap ... \sqcap C_n$	?C :intersectionOf (?C_1) . ?y a ?C_1 . ⇒ ?y a ?C .	42

Other SAOR rules with 2 or 3 Abox statements in the antecedent:



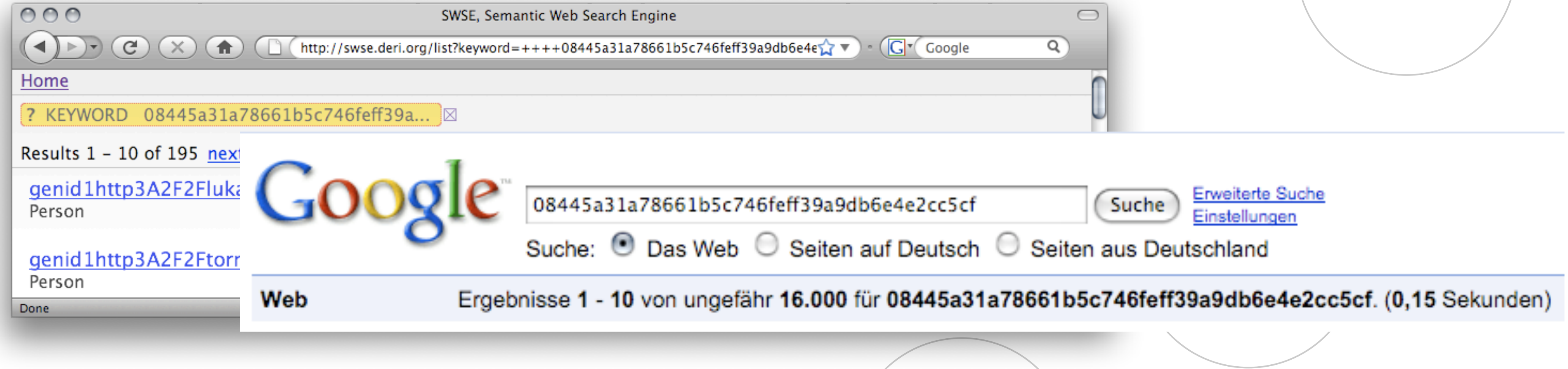
$\mathcal{R}2$: at least one terminological/multiple assertional patterns in antecedent		
rdfp1′	**?P** a :FunctionalProperty . ?x ?P ?y , ?z . ⇒ ?y :sameAs ?z .	
rdfp2	**?P** a :InverseFunctionalProperty . ?x ?P ?z . ?y ?P ?z . ⇒ ?x :sameAs ?y .	
rdfp4	**?P** a :TransitiveProperty . ?x ?P ?y . ?y ?P ?z . ⇒ ?x ?P ?z .	
rdfp15′	?C :someValuesFrom **?D** ; :onProperty **?P** . ?x ?P ?y . ?y a ?D . ⇒ ?x a ?C .	?C ∈ $\mathcal{B}$
rdfp16′	?C :allValuesFrom ?D ; :onProperty ?P . ?x a ?C ; ?P ?y . ⇒ ?y a ?D .	?C ∈ $\mathcal{B}$
rdfc3c	?C :intersectionOf (**?C**$_1$ **...** **?C**$_n$) . ?x a ?C$_1$, ..., ?C$_n$. ⇒ ?x a ?C .	?C ∈ $\mathcal{B}$
rdfc4a	?C :cardinality 1 ; :onProperty ?P . ?x a ?C ; ?P ?y , ?z . ⇒ ?y :sameAs ?z .	?C ∈ $\mathcal{B}$
rdfc4b	?C :maxCardinality 1 ; :onProperty ?P . ?x a ?C ; ?P ?y , ?z . ⇒ ?y :sameAs ?z .	?C ∈ $\mathcal{B}$
$\mathcal{R}3$: only assertional patterns in antecedent		
rdfp6′	?x :sameAs ?y . ⇒ ?y :sameAs ?x .	
rdfp7	?x :sameAs ?y . ?y :sameas ?z . ⇒ ?x :sameAs ?z .	
rdfp11′	?x :sameAs ?_x ; ?P ?y .⇒ ?_x ?P ?y . [c]	
rdfp11′′	?y :sameAs ?_y . ?x ?P ?y .⇒ ?x ?P ?_y . [c]	

- We avoid these for the moment in the real search engine…
  … experiments including these rules in **[Hogan et al. 2009, IJWSIS]** and also in our “pedantic-web” validator, more later.

Good “excuses” to avoid G2 rules

- The obvious:
  - □ $\mathcal{G}2$ rules would need joins, i.e. to trigger restart of file-scan,
  - □ Restricting to $\mathcal{G}0$, $\mathcal{G}1$ allows distribution again!
- The interesting one:
  - □ Take for instance IFP rule:

$$\top \sqsubseteq \forall \leq 1P^- \quad \underline{\textbf{?P}\ \text{a :InverseFunctionalProperty .}}\ \text{?x ?P ?o . ?y ?P ?o .} \Rightarrow \text{?x :sameAs ?y .}$$



SWSE, Semantic Web Search Engine

http://swse.deri.org/list?keyword=++++08445a31a78661b5c746feff39a9db6e4e

Home

? KEYWORD 08445a31a78661b5c746feff39a...

Results 1 – 10 of 195 next

genid1http3A2F2Fluk
Person

genid1http3A2F2Ftorr
Person

Done

08445a31a78661b5c746feff39a9db6e4e2cc5cf

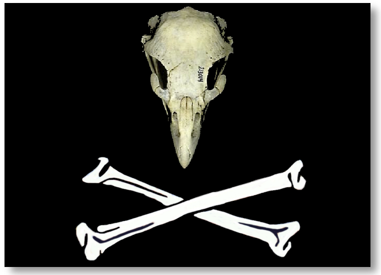
Suche: ● Das Web ○ Seiten auf Deutsch ○ Seiten aus Deutschland

Web Ergebnisse 1 - 10 von ungefähr 16.000 für 08445a31a78661b5c746feff39a9db6e4e2cc5cf. (0,15 Sekunden)

  - □ More experiments including $\mathcal{G}2$, $\mathcal{G}3$ rules in [Hogan, Harth, Polleres, ASWC2008]

Authoritative Reasoning

- Document **D** authoritative for concept **C** iff:
  - □ **C** not identified by URI
    - – OR
  - □ De-referenced URI of **C** coincides with or redirects to **D**
  - □ FOAF spec authoritative for `foaf:Person` ✓
  - □ MY spec not authoritative for `foaf:Person` ✘



Ontology Hijacking

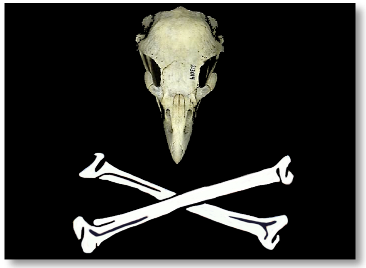
- Only allow extension in authoritative documents
  - □ `my:Person rdfs:subClassOf foaf:Person .` (MY spec) ✓
- BUT: Reduce obscure memberships
  - □ `foaf:Person rdfs:subClassOf my:Person .` (MY spec) ✘
- Similarly for other T-Box statements.
- In-memory T-Box stores authoritative values for rule execution

Rules Applied

#	DL Syntax	Rule	# Inferred
		$\mathcal{G}0$: NO A-BOX PATTERNS IN ANTECEDENT	
00	$\{o_i....o_n\}$	?C :oneOf (?o$_1$ ... ?o$_n$) . ⇒ ?o$_1$ ... ?o$_n$ a ?C .	35,161
		$\mathcal{G}1$: ONE A-BOX PATTERN IN ANTECEDENT	
01	$C \sqsubseteq D$	?C rdfs:subClassOf ?D . ?s a ?C . ⇒ ?s a ?D .	1,124,758,631
02_a	$C \equiv D$	?C :equivalentClass ?D . ?s a ?C . ⇒ ?s a ?D .	8,137,162
02_b		?C :equivalentClass ?D . ?s a ?D . ⇒ ?s a ?C .	90,372
03	$P \sqsubseteq Q$	?P rdfs:subPropertyOf ?Q . ?s ?P ?o . ⇒ ?s ?Q ?o .	156,462,399
04_a	$P \equiv Q$	?P :equivalentProperty ?Q . ?s ?P ?o . ⇒ ?s ?Q ?o .	5,667,464
04_b		?P :equivalentProperty ?Q . ?s ?Q ?o . ⇒ ?s ?P ?o .	6,642
05_a	$P \equiv P_0^-$	?P :inverseOf ?Q . ?s ?P ?o . ⇒ ?o ?Q ?s .	230,945,040
05_b		?P :inverseOf ?Q . ?s ?Q ?o . ⇒ ?o ?P ?s .	230,941,648
06	$\top \sqsubseteq \forall P^-.C$	?P rdfs:domain ?C . ?s ?P ?o . ⇒ ?s a ?C .	588,530,865
07	$\top \sqsubseteq \forall P.C$	?P rdfs:range ?C . ?s ?P ?o . ⇒ ?o a ?C .	528,995,909
08	$P \equiv P^-$	?P a :SymmetricProperty . ?s ?P ?o . ⇒ ?o ?P ?s .	560,460
09_a	$\exists P.x$	?C :hasValue ?x; :onProperty ?P . ?y ?P ?x . ⇒ ?y a ?C .	98,601
09_b		?C :hasValue ?x; :onProperty ?P . ?y a ?C . ⇒ ?y ?P ?x .	104,780
10	$C_1 \sqcup ... \sqcup C_n$	?C :unionOf (?C$_1$...?C$_i$...?C$_n$) . ?x a ?C$_i$. ⇒ ?x a ?C .	81,736,234
11	$(\geq 1P)$	?C :minCardinality 1; :onProperty ?P . ?x ?P ?y . ⇒ ?x a ?C .	65,283,322
12_a	$C_1 \sqcap ... \sqcap C_n$	?C :intersectionOf (?C$_1$ ... ?C$_n$) . ?y a ?C . ⇒ ?y a ?C$_1$, ..., ?C$_n$.	115,383
12_b	$C_1 \sqcap ... \sqcap C_n$	?C :intersectionOf (?C$_1$) . ?y a ?C$_1$. ⇒ ?y a ?C .	42

*The 17 rules applied including statements considered to be T-Box, elements which must be **authoritatively** spoken for (including for bnode OWL abstract syntax), and output count*

Authoritative Resoning covers rdfs: owl: vocabulary misuse



:rdfs :owl Hijacking

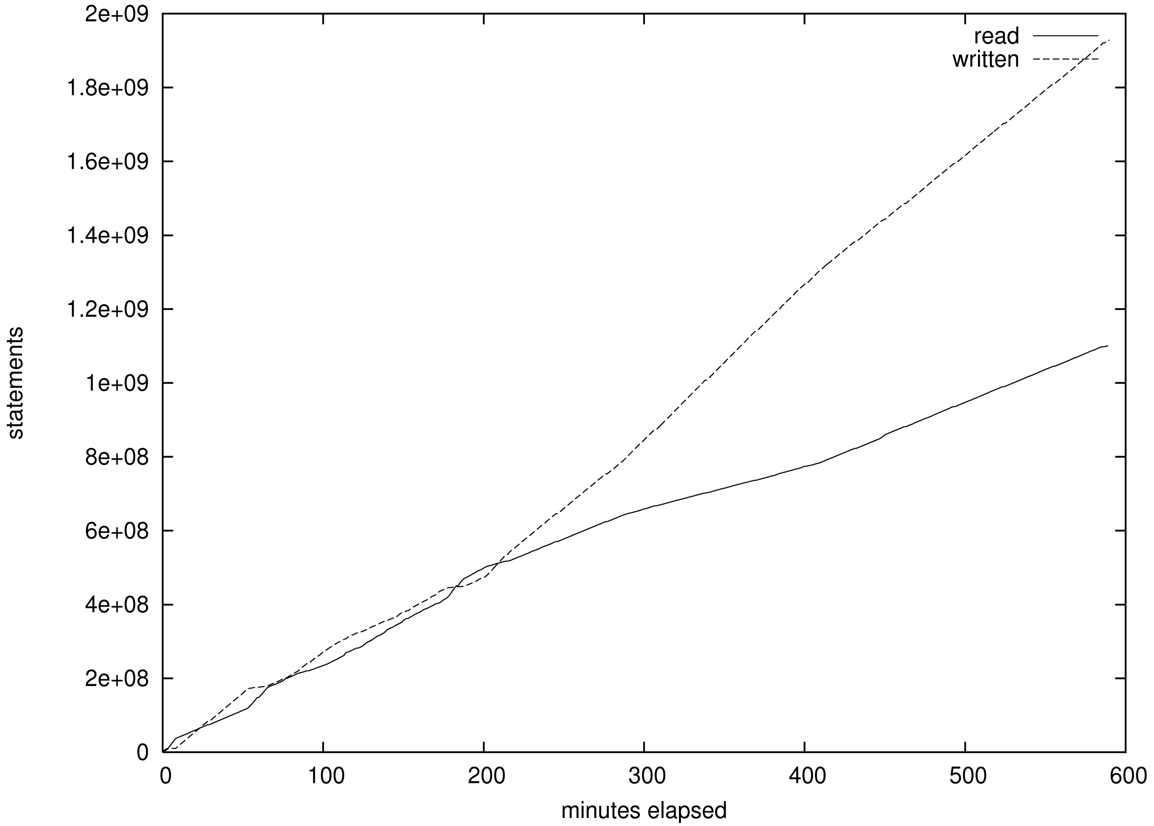
- http://www.polleres.net/nasty.rdf:

```
rdfs:subClassOf rdfs:subPropertyOf rdfs:Resource.
rdfs:subClassOf rdfs:subPropertyOf rdfs:subPropertyOf.
rdf:type rdfs:subPropertyOf rdfs:subClassOf.
rdfs:subClassOf rdf:type owl:SymmetricProperty.
```

- Naïve rules application would infer O(n^3) triples

- By use of authoritative reasoning SAOR/SWSE doesn’t stumble over these ☺

Performance



Graph showing SAOR's rate of input/output statements per minute for reasoning on 1.1b statements (ISWC 2009 Billion Triples challenge): reduced input rate correlates with increased output rate and vice-versa

Results

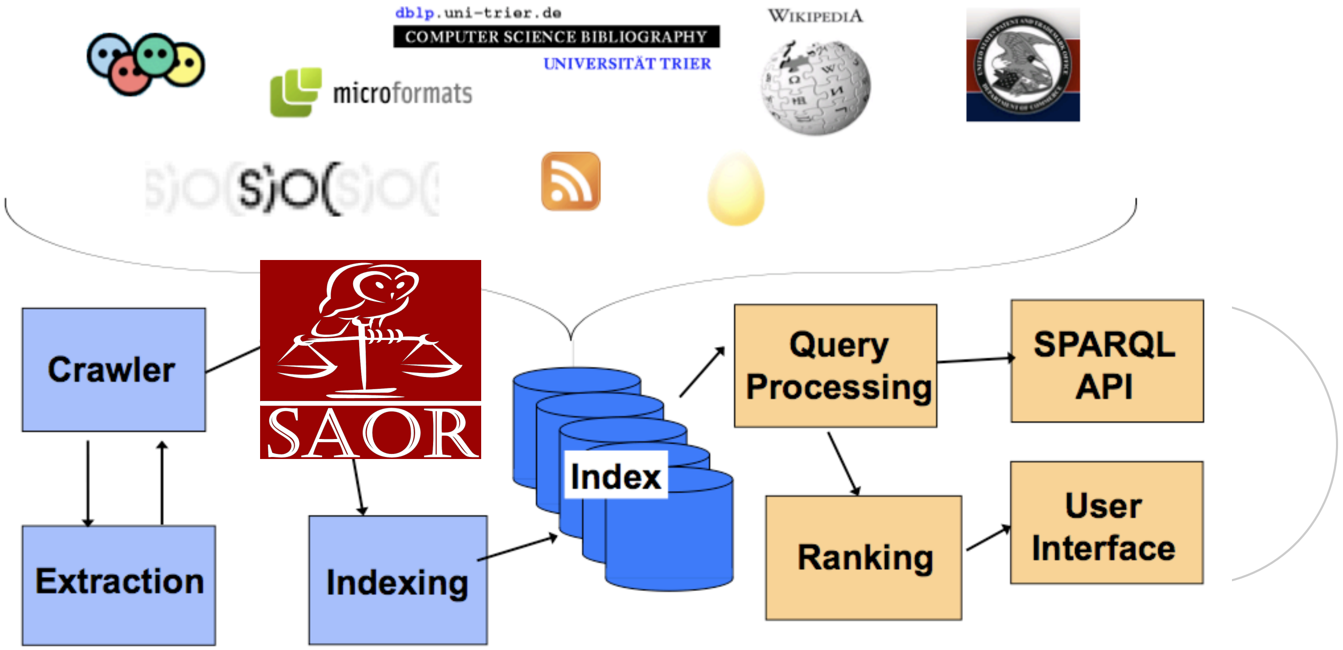
- **SCAN 1: 6.47 hrs**
  - □ In-mem T-Box creation, authoritative analysis:
- **SCAN 2: 9.82 hrs**
  - □ Scan reasoning – join A-Box with in-mem authoritative T-Box:
- **1.925b new statements inferred in 16.29 hrs**

1.1b + 1.9b inferred = 3 billion triples in SWSE

- **Other issues:**
  - □ **More valuable insights on our experiences from Web data…**
  - □ **Experiments involving G2 and G3 rules in [Hogan et al. 2009, IJWSIS]**
  - □ **Detailed comparison to OWL RL**
- **This is one machine,naïve approach… 2 related papers in this years' ISWC with similar approach but parallelisation show that you can do much faster with adding computing power.**

SWSE in one slide…

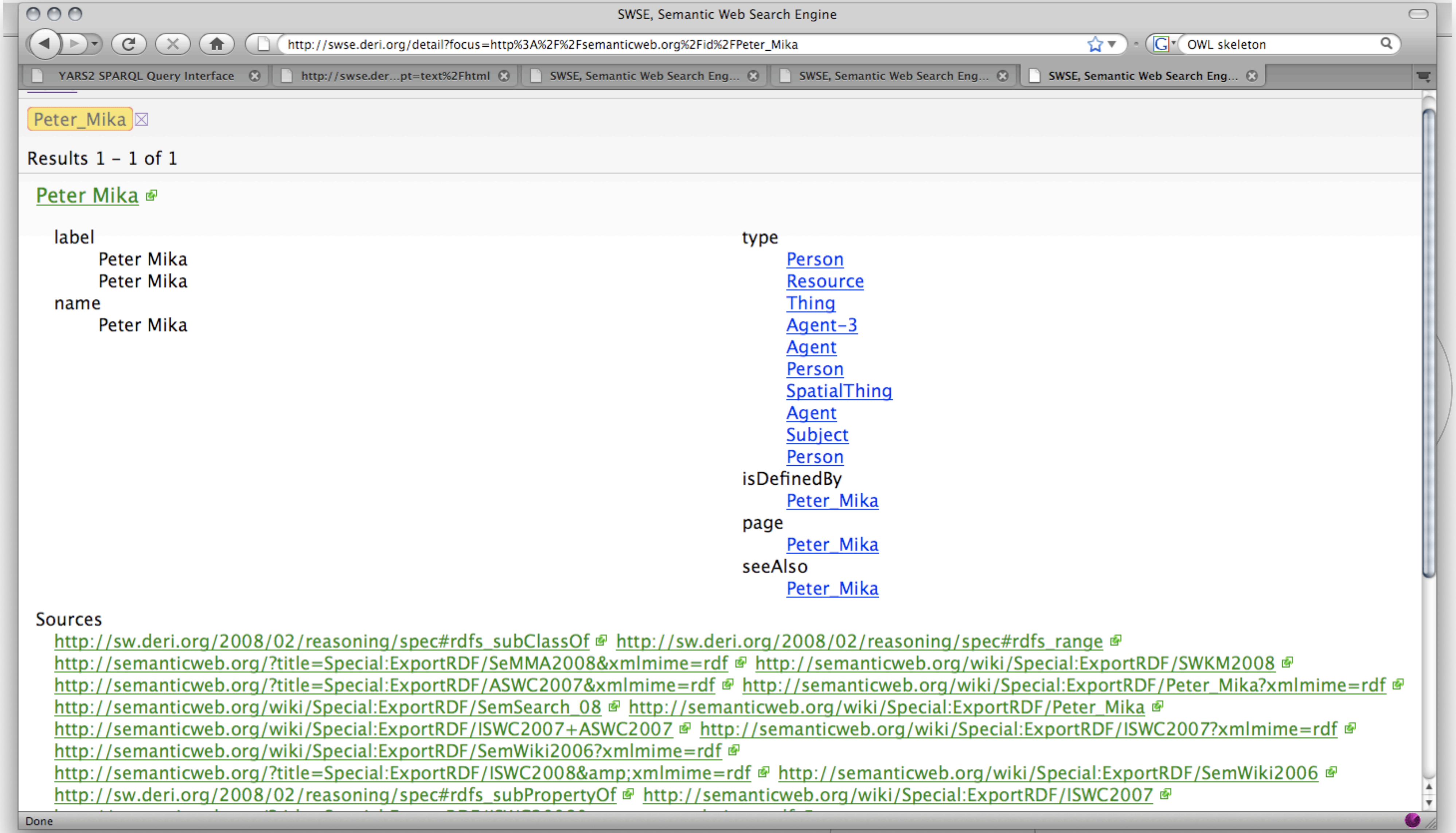
Enjoy the data…



GUI: http://swse.deri.org/

SPARQL interface: http://swse.deri.org/yars2/

Search result example:


SWSE, Semantic Web Search Engine
http://swse.deri.org/detail?focus=http%3A%2F%2Fsemanticweb.org%2Fid%2FPeter_Mika
OWL skeleton
YARS2 SPARQL Query Interface
http://swse.der...pt=text%2Fhtml
SWSE, Semantic Web Search Eng...
SWSE, Semantic Web Search Eng...
SWSE, Semantic Web Search Eng...
Peter_Mika
Done

Results 1 – 1 of 1

Peter Mika

label
- Peter Mika
- Peter Mika

name
- Peter Mika

type
- Person
- Resource
- Thing
- Agent-3
- Agent
- Person
- SpatialThing
- Agent
- Subject
- Person

isDefinedBy
- Peter_Mika

page
- Peter_Mika

seeAlso
- Peter_Mika

Sources

http://sw.deri.org/2008/02/reasoning/spec#rdfs_subClassOf http://sw.deri.org/2008/02/reasoning/spec#rdfs_range
http://semanticweb.org/?title=Special:ExportRDF/SeMMA2008&xmlmime=rdf http://semanticweb.org/wiki/Special:ExportRDF/SWKM2008
http://semanticweb.org/?title=Special:ExportRDF/ASWC2007&xmlmime=rdf http://semanticweb.org/wiki/Special:ExportRDF/Peter_Mika?xmlmime=rdf
http://semanticweb.org/wiki/Special:ExportRDF/SemSearch_08 http://semanticweb.org/wiki/Special:ExportRDF/Peter_Mika
http://semanticweb.org/wiki/Special:ExportRDF/ISWC2007+ASWC2007 http://semanticweb.org/wiki/Special:ExportRDF/ISWC2007?xmlmime=rdf
http://semanticweb.org/wiki/Special:ExportRDF/SemWiki2006?xmlmime=rdf
http://semanticweb.org/?title=Special:ExportRDF/ISWC2008&amp;xmlmime=rdf http://semanticweb.org/wiki/Special:ExportRDF/SemWiki2006
http://sw.deri.org/2008/02/reasoning/spec#rdfs_subPropertyOf http://semanticweb.org/wiki/Special:ExportRDF/ISWC2007

Insights/Lessons learned…:

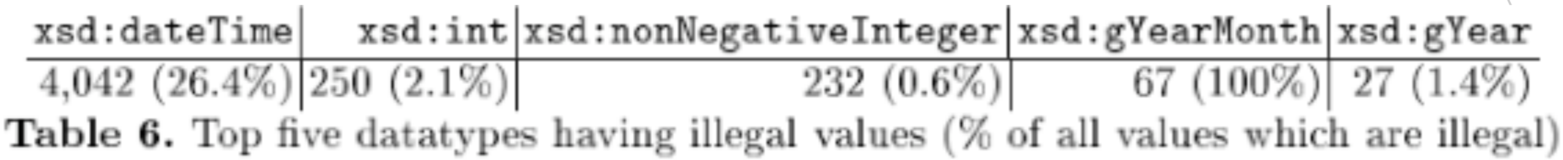
- Some more insights into our results on Reasoning with Web data:
  - □ Based on a crawl "6 hops from TimBL's FOAF file.
  - □ We did some in-depth analysis of common mistakes on that arguably representative SW crawl.

Data Analysis: Example

- Inconsistencies due to wrong/misused datatypes:

  e.g. :me ex:age “old”^^xs:integer.

- Common on the Web:



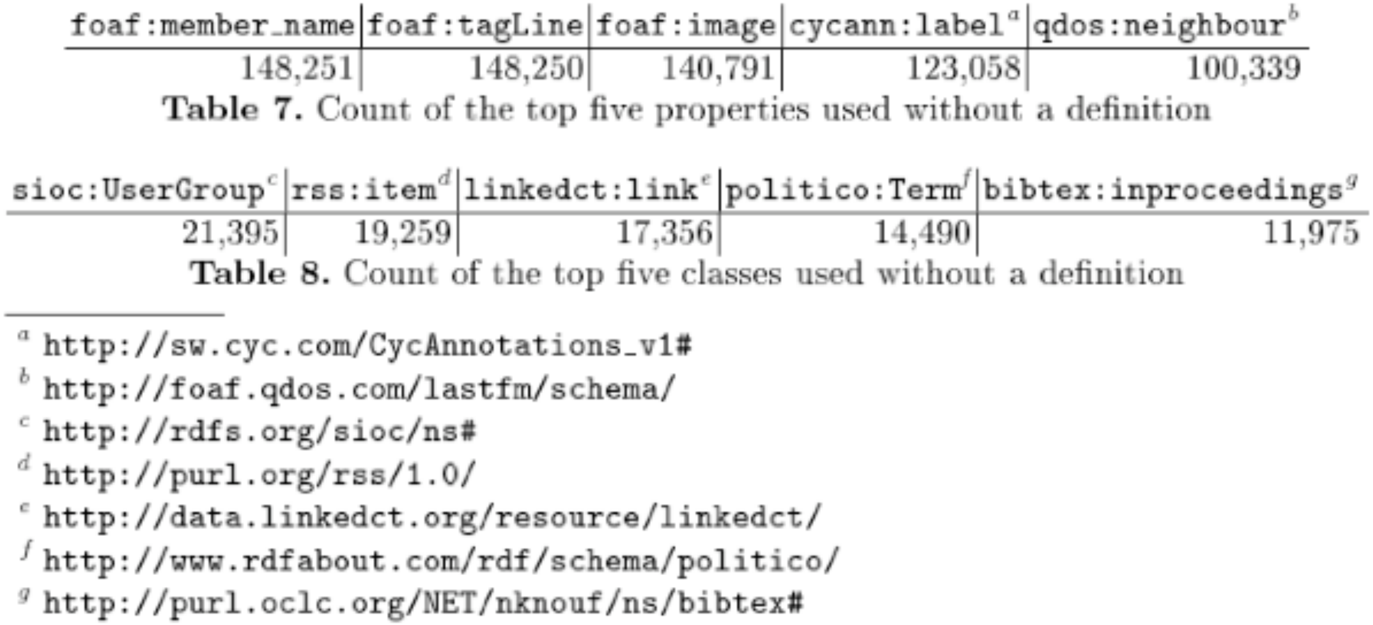
xsd:dateTime	xsd:int	xsd:nonNegativeInteger	xsd:gYearMonth	xsd:gYear
4,042 (26.4%)	250 (2.1%)	232 (0.6%)	67 (100%)	27 (1.4%)

Table 6. Top five datatypes having illegal values (% of all values which are illegal)

- Don’t affect SAOR reasoning so far, but we want to add Datatype support.

Data Analysis: Example

- There is a significant used of undefined (dereferencing doesn't give a definition) classes and properties:



foaf:member_name	foaf:tagLine	foaf:image	cycann:label[a]	qdos:neighbour[b]
148,251	148,250	140,791	123,058	100,339

Table 7. Count of the top five properties used without a definition

sioc:UserGroup[c]	rss:item[d]	linkedct:link[e]	politico:Term[f]	bibtex:inproceedings[g]
21,395	19,259	17,356	14,490	11,975

Table 8. Count of the top five classes used without a definition

[a] http://sw.cyc.com/CycAnnotations_v1#
[b] http://foaf.qdos.com/lastfm/schema/
[c] http://rdfs.org/sioc/ns#
[d] http://purl.org/rss/1.0/
[e] http://data.linkedct.org/resource/linkedct/
[f] http://www.rdfabout.com/rdf/schema/politico/
[g] http://purl.oclc.org/NET/nknouf/ns/bibtex#

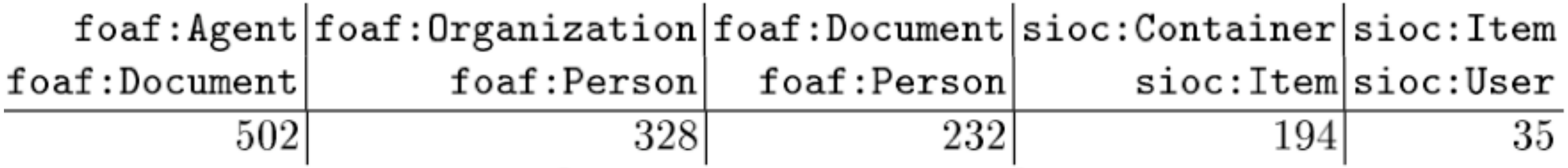
- Message: If you need a new property e.g. in FOAF, define your own new ontology and extend it, not just invent things in other's namespaces!

Data Analysis: Example

- Reasoning inconsistency:

```
TimBL rdf:type foaf:Person.
TimBL rdf:type foaf:Organisation.
foaf:Person owl:disjointWith foaf:Organisation.
```

- Common on the Web (after inference):



foaf:Agent	foaf:Organization	foaf:Document	sioc:Container	sioc:Item
foaf:Document	foaf:Person	foaf:Person	sioc:Item	sioc:User
502	328	232	194	35

Table 14. Top five instantiated pairs of disjoint classes

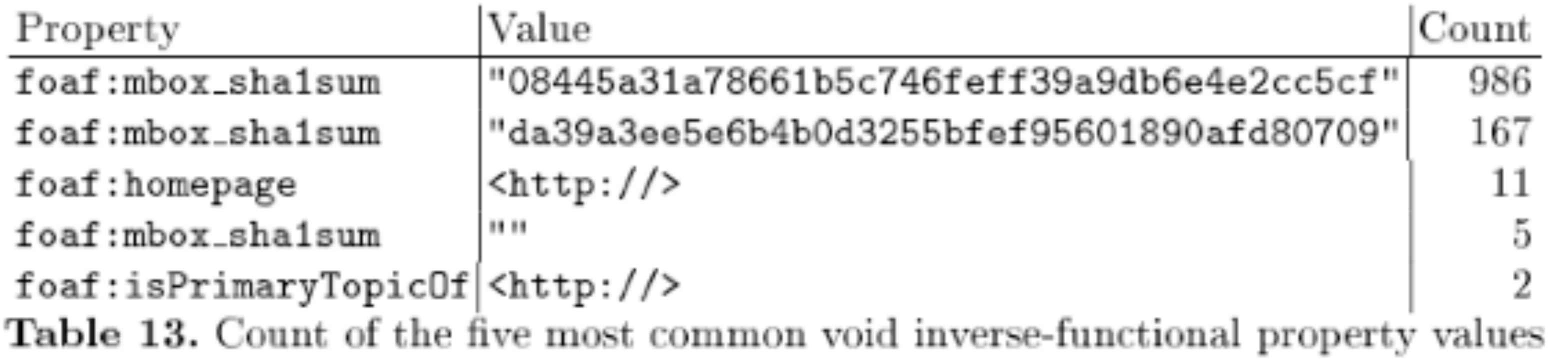
- Mostly from **exporters** which carelessly use properties with respective domains/ranges.

Data Analysis: Example

- Reasoning noise:

```
ex:alice foaf:mbox “mailto:”
ex:bob foaf:mbox “mailto:”
```

- Common on the Web:



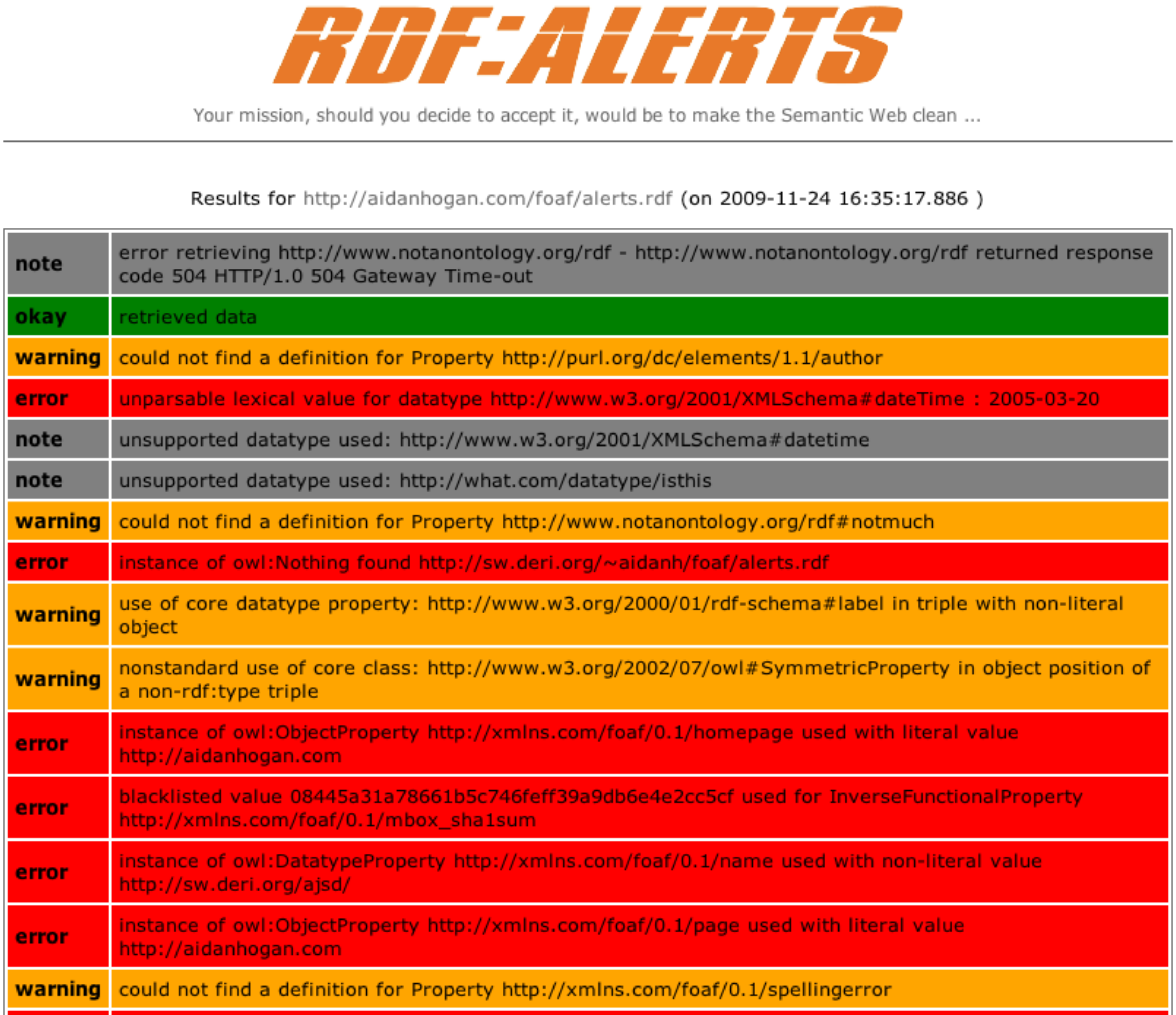
Property	Value	Count
foaf:mbox_sha1sum	"08445a31a78661b5c746feff39a9db6e4e2cc5cf"	986
foaf:mbox_sha1sum	"da39a3ee5e6b4b0d3255bfef95601890afd80709"	167
foaf:homepage	<http://>	11
foaf:mbox_sha1sum	""	5
foaf:isPrimaryTopicOf	<http://>	2

Table 13. Count of the five most common void inverse-functional property values

“Suspicious” IFP values can often been identified by heuristics (threshold of number of equated instances, etc.)

However, possibly expensive to evaluate.

Better: Make people aware, provide validation tools for checking their datasets!



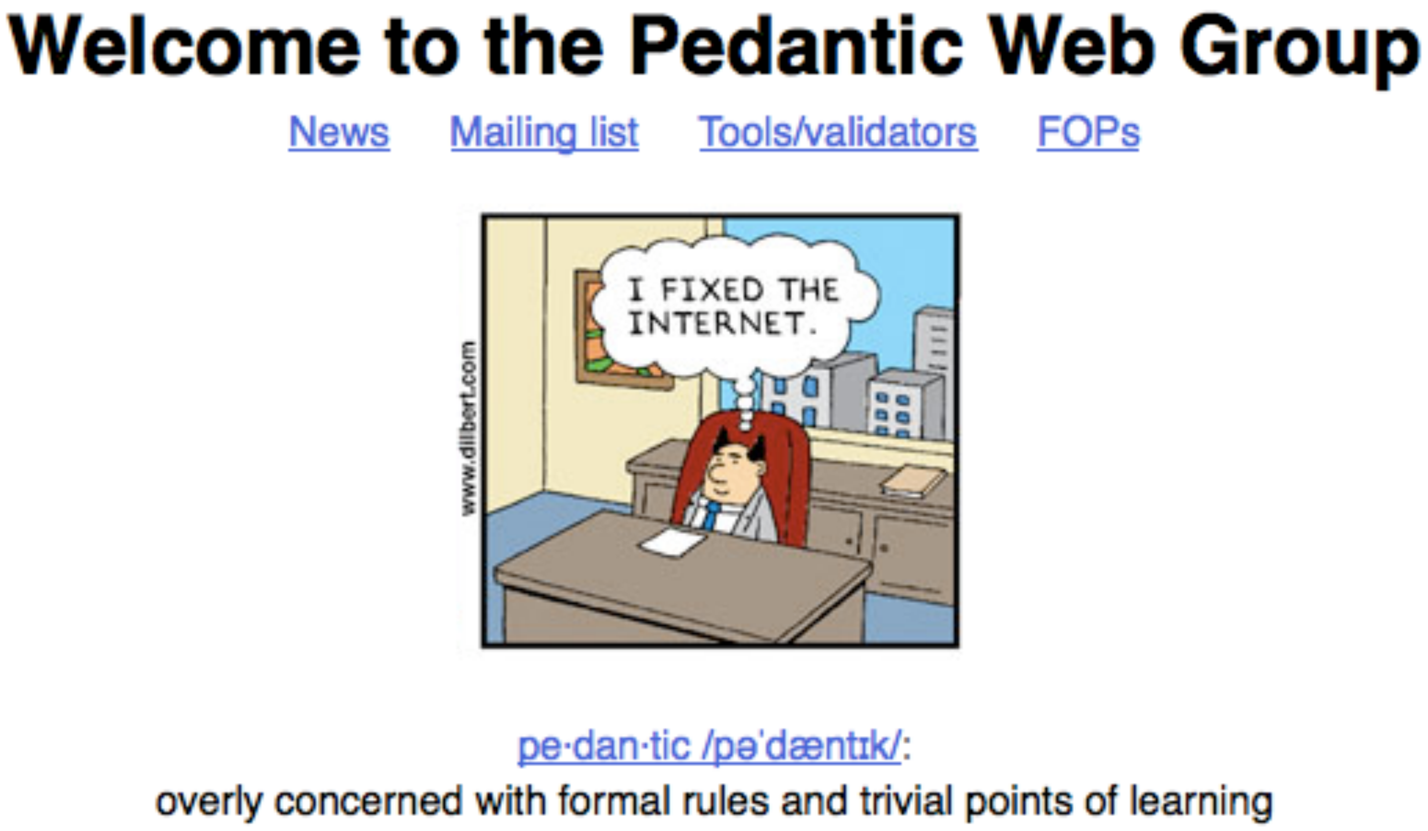
RDF:ALERTS

Your mission, should you decide to accept it, would be to make the Semantic Web clean ...

Results for http://aidanhogan.com/foaf/alerts.rdf (on 2009-11-24 16:35:17.886)

note	error retrieving http://www.notanontology.org/rdf - http://www.notanontology.org/rdf returned response code 504 HTTP/1.0 504 Gateway Time-out
okay	retrieved data
warning	could not find a definition for Property http://purl.org/dc/elements/1.1/author
error	unparsable lexical value for datatype http://www.w3.org/2001/XMLSchema#dateTime : 2005-03-20
note	unsupported datatype used: http://www.w3.org/2001/XMLSchema#datetime
note	unsupported datatype used: http://what.com/datatype/isthis
warning	could not find a definition for Property http://www.notanontology.org/rdf#notmuch
error	instance of owl:Nothing found http://sw.deri.org/~aidanh/foaf/alerts.rdf
warning	use of core datatype property: http://www.w3.org/2000/01/rdf-schema#label in triple with non-literal object
warning	nonstandard use of core class: http://www.w3.org/2002/07/owl#SymmetricProperty in object position of a non-rdf:type triple
error	instance of owl:ObjectProperty http://xmlns.com/foaf/0.1/homepage used with literal value http://aidanhogan.com
error	blacklisted value 08445a31a78661b5c746feff39a9db6e4e2cc5cf used for InverseFunctionalProperty http://xmlns.com/foaf/0.1/mbox_sha1sum
error	instance of owl:DatatypeProperty http://xmlns.com/foaf/0.1/name used with non-literal value http://sw.deri.org/ajsd/
error	instance of owl:ObjectProperty http://xmlns.com/foaf/0.1/page used with literal value http://aidanhogan.com
warning	could not find a definition for Property http://xmlns.com/foaf/0.1/spellingerror

Visit: http://pedantic-web.org/



Welcome to the Pedantic Web Group

News Mailing list Tools/validators FOPs

pe·dan·tic /pə'dæntɪk/:
overly concerned with formal rules and trivial points of learning

Already several successes in finding/fixing: FOAF, dbpedia, NYtimes, even W3C specs… etc.

Take home:

- **Practical reasoning over web data ≠ science fiction.**
- Linked Data & Linked Ontologies are as messy as the normal HTML Web

- We showed some ways to deal with them:
  - □ Rule-based Reasoning on Web Data typically gives good approximation…
  - □ … actually still too much, if not done cautiously

- Not all problems solved yet
  - □ Dropping sameAs reasoning, we'd miss some important inferences, heuristics might help (e.g. for controlled equality reasoning)
  - □ Important: Making data publishers aware to produce better quality data might help (RDFAlerts, pedantic-web)