Explanation for the Semantic Web

Rakebul Hasan

Introduction

Existing work

Perspective

Work in progress

Demo (FOAFConnection)

Kolflow deliverables

Introduction

- Semantic Web
  - Interconnected and distributed data
  - Inferential capabilities

- Explanation
  - Understanding
    - how results are obtained
    - the flow of information
  - Trust

“Oh, yeah?” button to support users in assessing the reliability of information encountered on the Web

- Tim Berners-Lee

Consistent User Interface, W3C Design Issues, 1997

Contribution

- A brief review of the existing approaches to explanation in the Semantic Web
- Selection criteria
  - Semantic Web applications and publications that have contribution in the field of explanation
    - Google Scholar keyword search, cited by feature
    - ExaCt workshop series publications
    - ISWC series publications

Existing Work

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

Trust

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

Trust

What is explained?

- Information manipulation steps
  - Information manipulation operations
- Proof tree of derivations
- Provenance information such as How, When, Who, Where

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

Trust

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

Trust

Target

- Human users
  - Natural language explanation
  - Graphical explanation
- Software agents
  - Machine readable descriptions

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

Trust

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

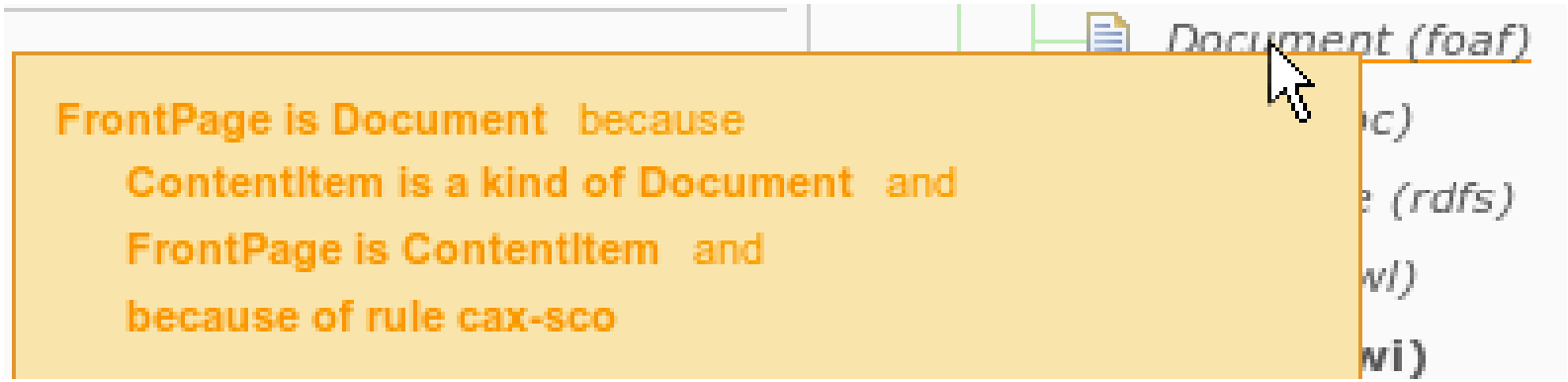
Trust

Presentation

```
1.  The triple:
2.
3.   Siemens AG has positive analyst report: "As Siemens agrees
4.   partnership with Novell unit SUSE ..."
5.
6.  fulfills the policy:
7.
8.   Only accept information from information providers who
9.   have received more positive than negative ratings.
10.
11. because:
12.
13.  The information was asserted by Peter Smith and
14.  Peter Smith received the following numbers of ratings:
15.        - 3 positive ratings (see detail 1)
16.        - 2 negative ratings (see detail 2)
17.
18.  Detail 1: Peter Smith received positive ratings from:
19.        - John Reynolds
20.        - Mary O'Conner
21.        - Elisa Armstoen
22.
23.  Detail 2: Peter Smith received negative ratings from:
24.        - Dave Berser
25.        - Colin Marwick
26.
```

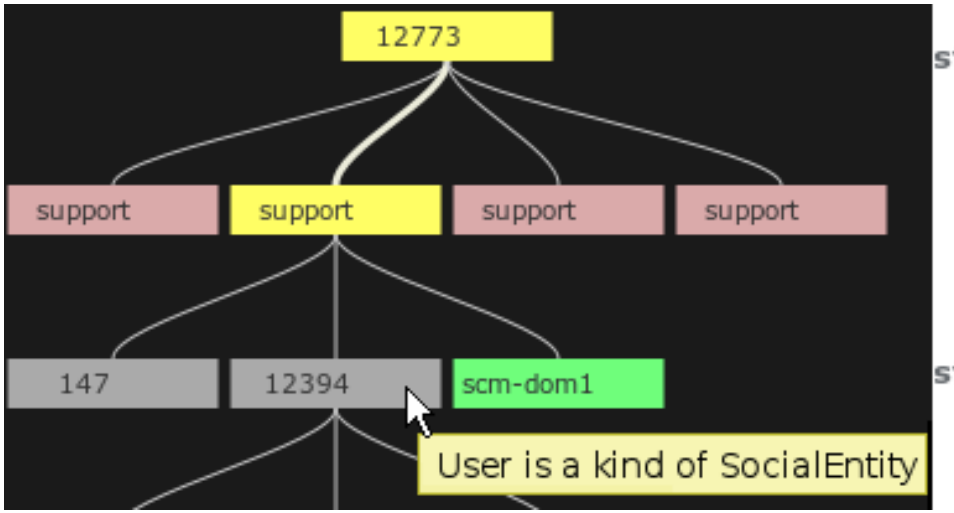
WIQA [Bizer, 2007]

Presentation



KiWi [Kotowski and Bry, 2010]

Presentation



street domain SocialEntity because

street domain User and

User is a kind of SocialEntity and

because of rule scm-dom1

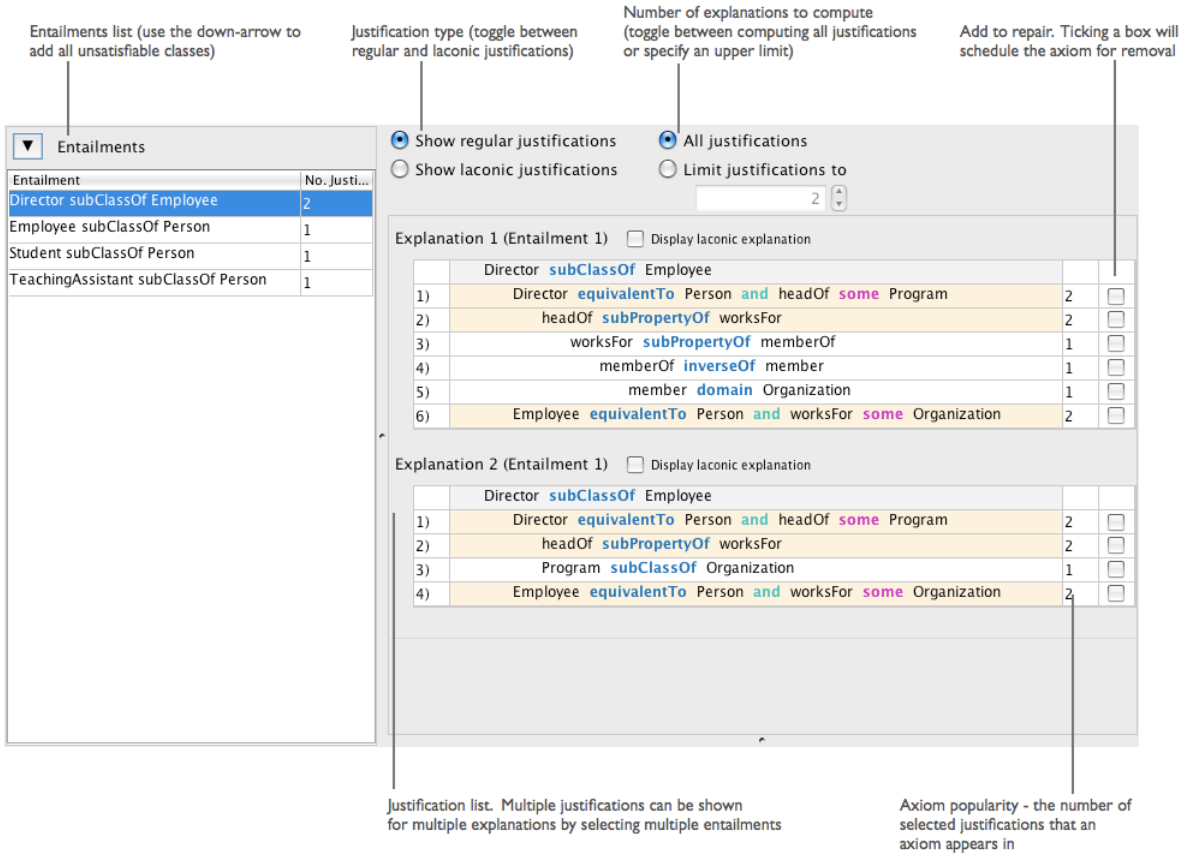
street domain SocialEntity because

street domain SocialEntity and

street subPropertyOf street and

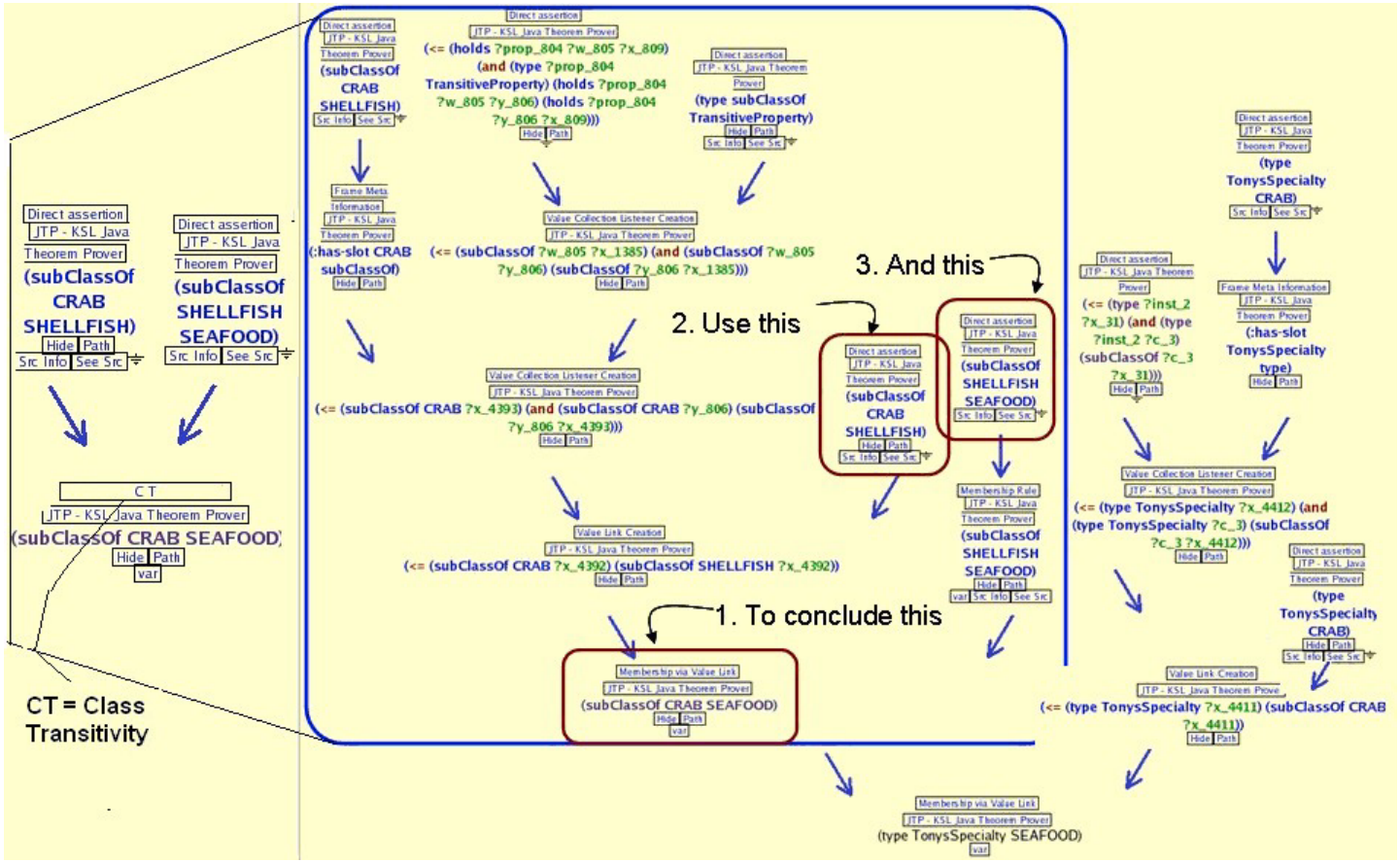
KiWi [Kotowski and Bry, 2010]

Presentation



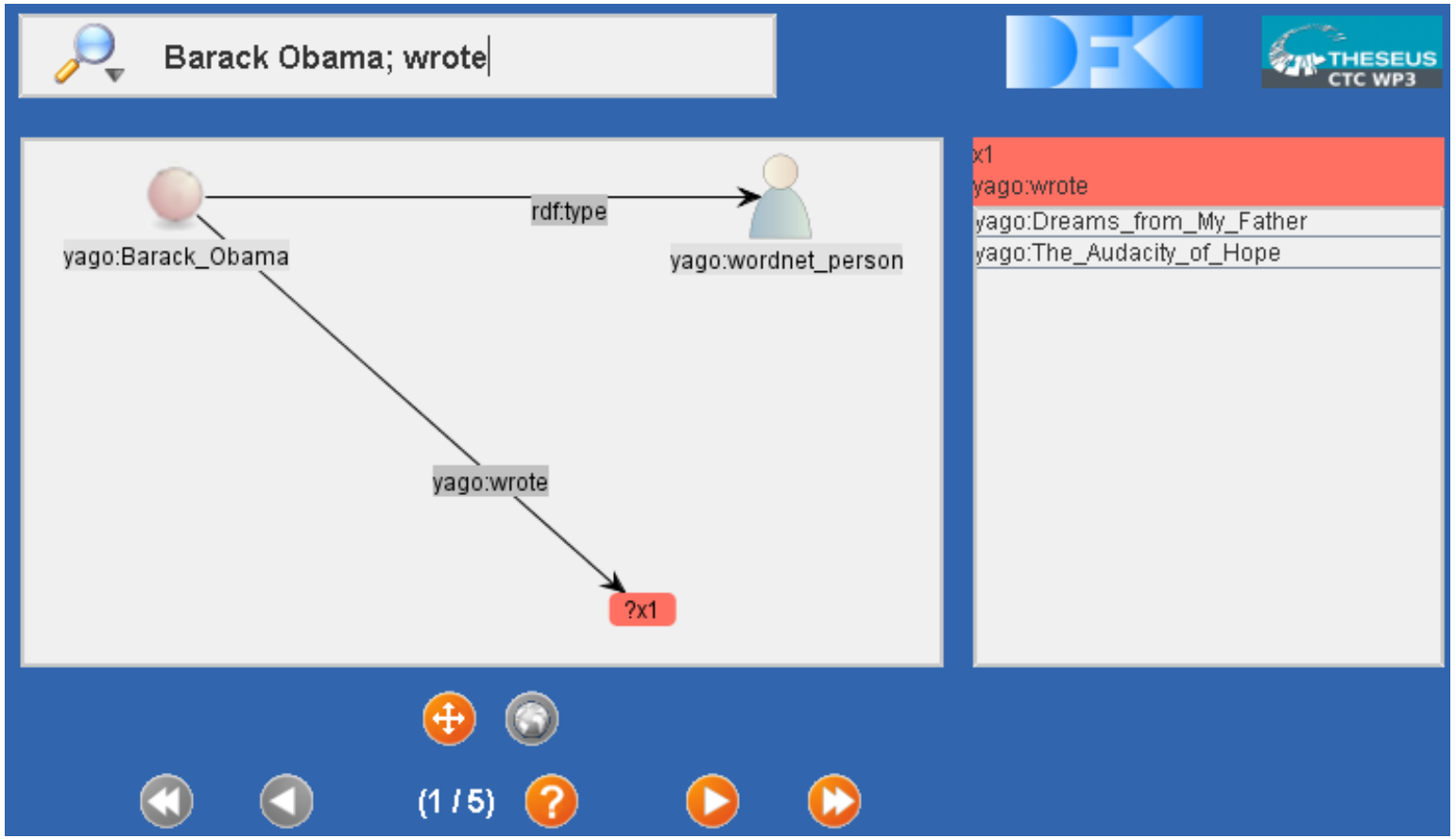
Explaining entailments in OWL ontologies [Horridge *et al.*, 2008]

Presentation



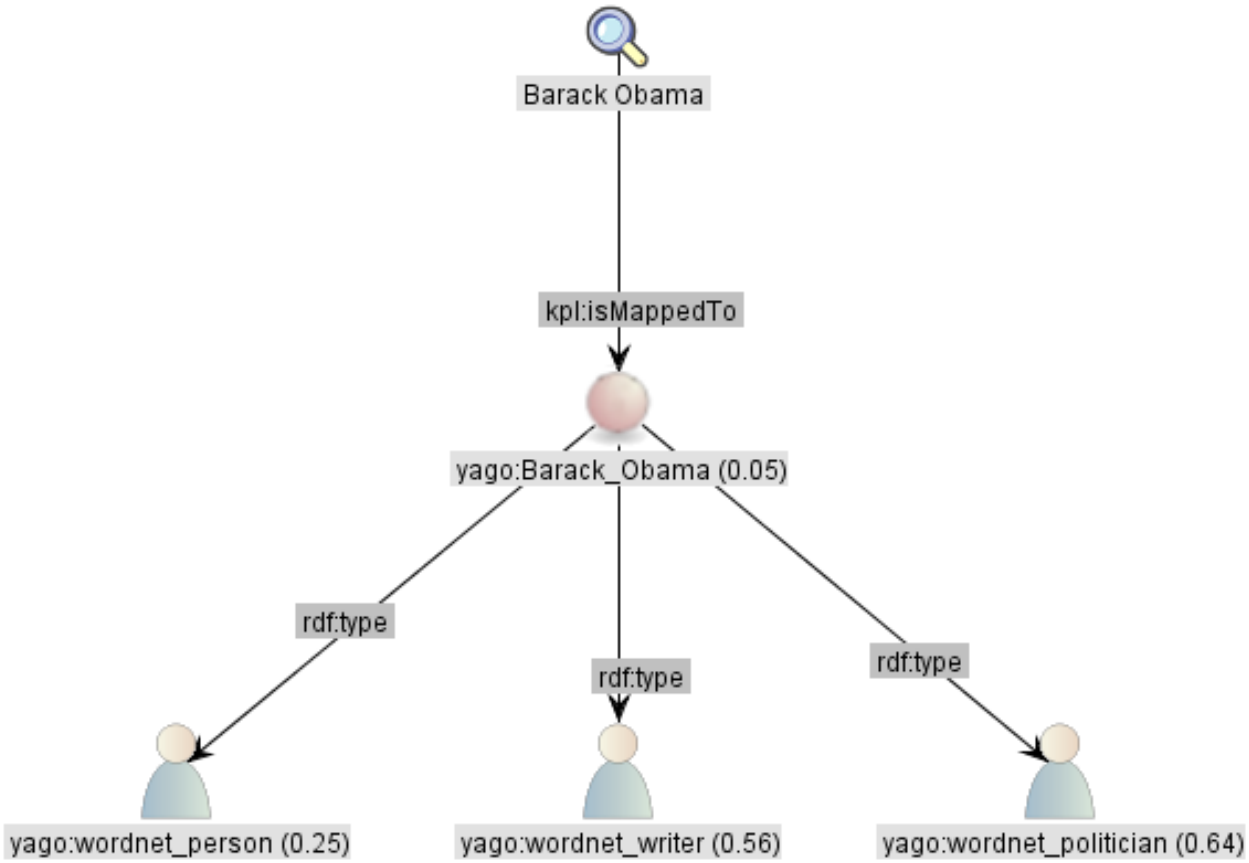
Inference Web [McGuinness *et al. (a), 2003*] [McGuinness *et al. (b), 2006*] [McGuinness *et al. (c), 2008*]

Presentation



EASD/KOIOS [Forcher *et al.*, 2010]

Presentation



EASD/KOIOS [Forcher *et al.*, 2010]

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

Trust

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

Trust

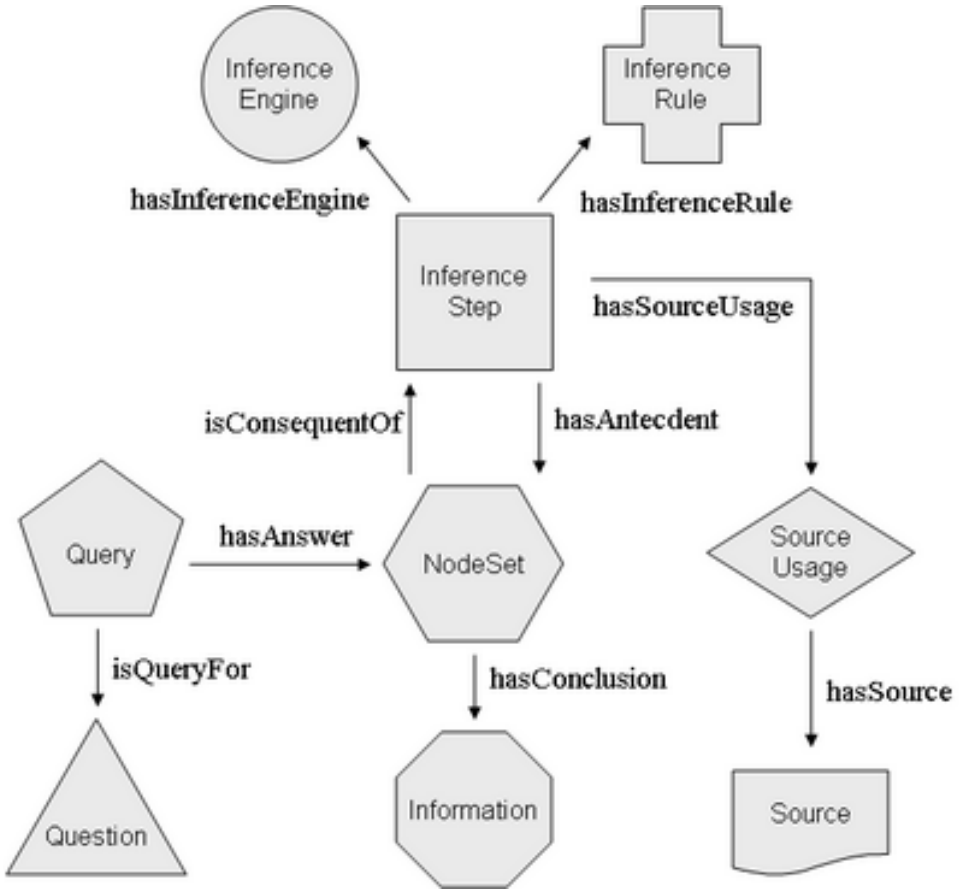
Metadata representation

- Proof trees for answers
- Operations used to compute answers
- Different types of provenance information
- Models for how explanations should be presented to human users
- Trust related information

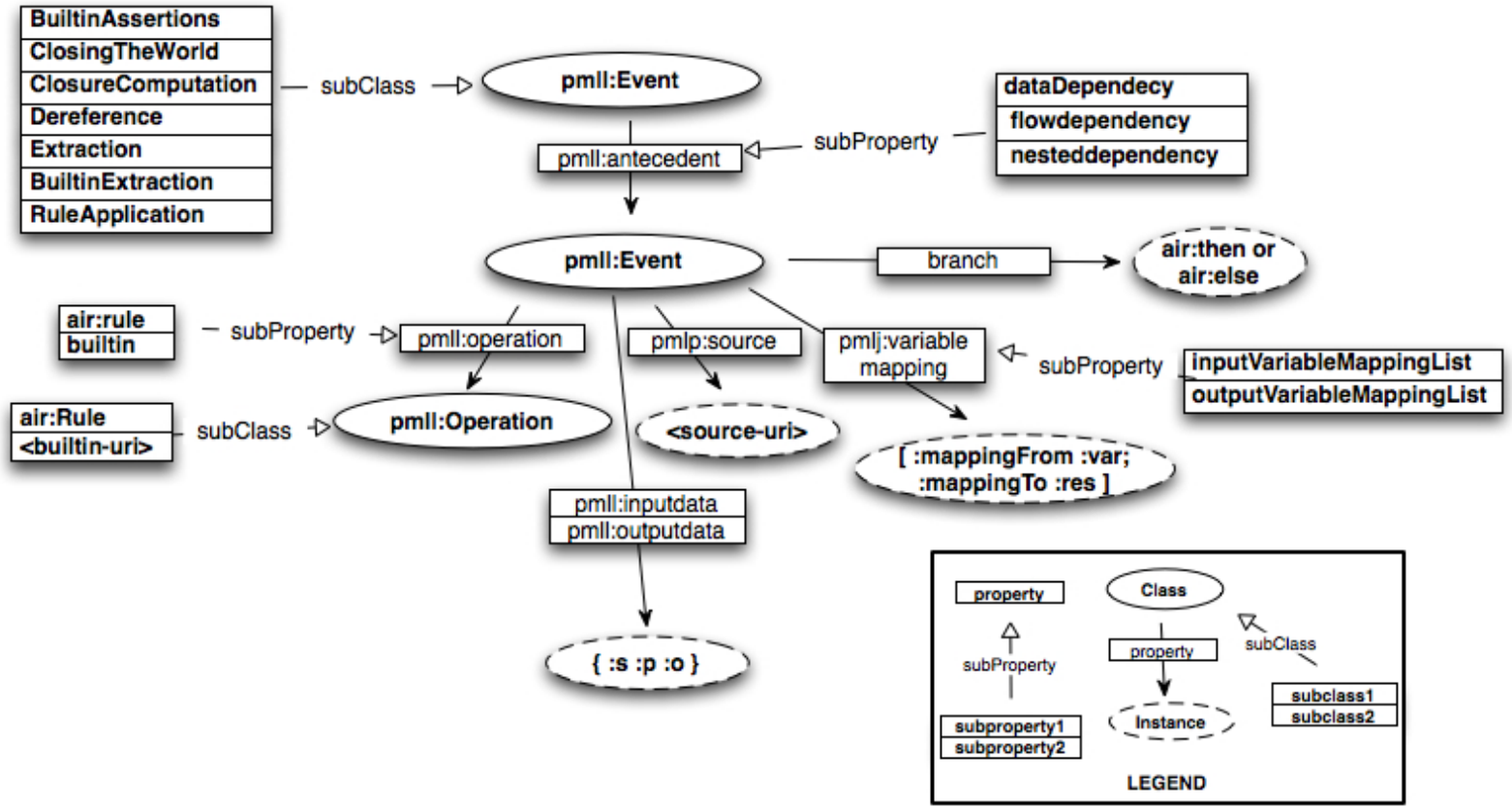
Metadata representation

- Proof Markup Language (PML) Ontology
  - Proof interlingua
  - Justifications: information manipulation steps and operations
  - Provenance information
  - Trust information

Metadata representation



Metadata representation



Accountability In RDF (AIR) [Kagal *et al., 2011*]

Metadata representation

- KOIOS
  - KOIOS Process Language (KPL) for describing the behavior of KOIOS problem solver
  - The Mathematical Graph Language (MGL) for transforming the process model to a graph based view.
  - VGL for describing the visualization model

Metadata representation

EXPL: WIQA describes its explanation trees (parts and subparts of an explanation) using the Explanation (EXPL) Vocabulary

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

Trust

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

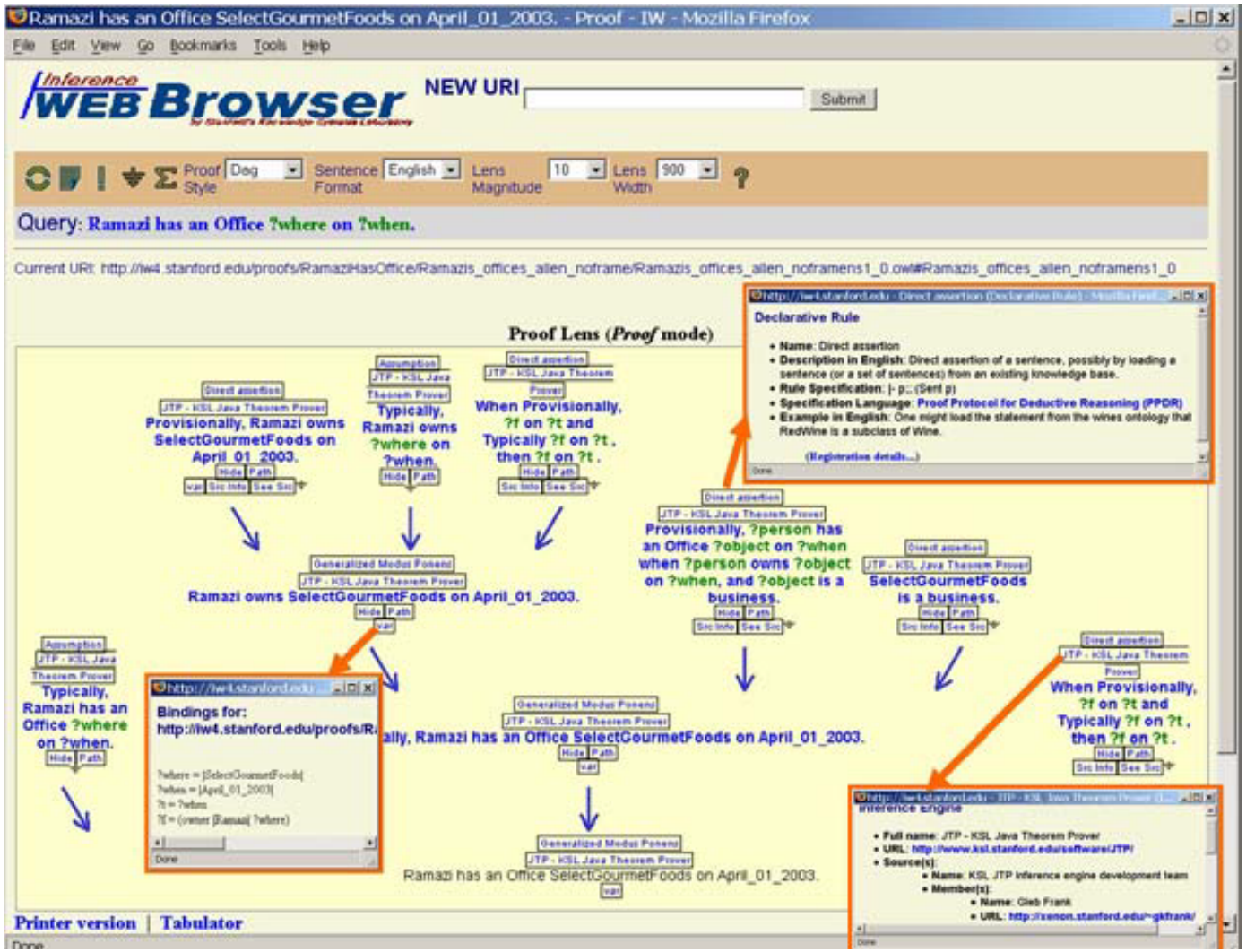
How human users interact with explanations?

Trust

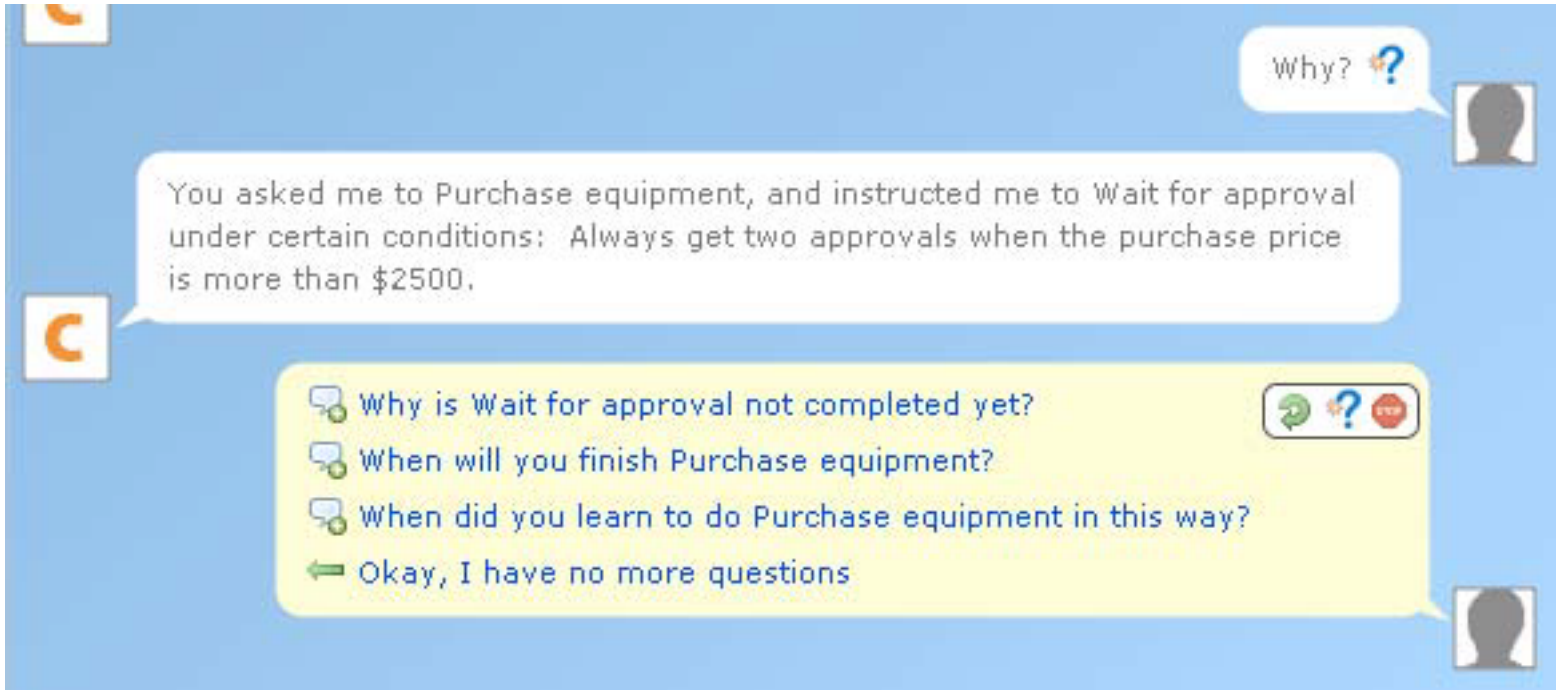
Interaction

- Navigation
- Follow-up

Interaction



Interaction



What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

How human users interact with explanations?

Trust

What is explained?

What are the targets?

How explanations are presented?

How explanations are represented for machines?

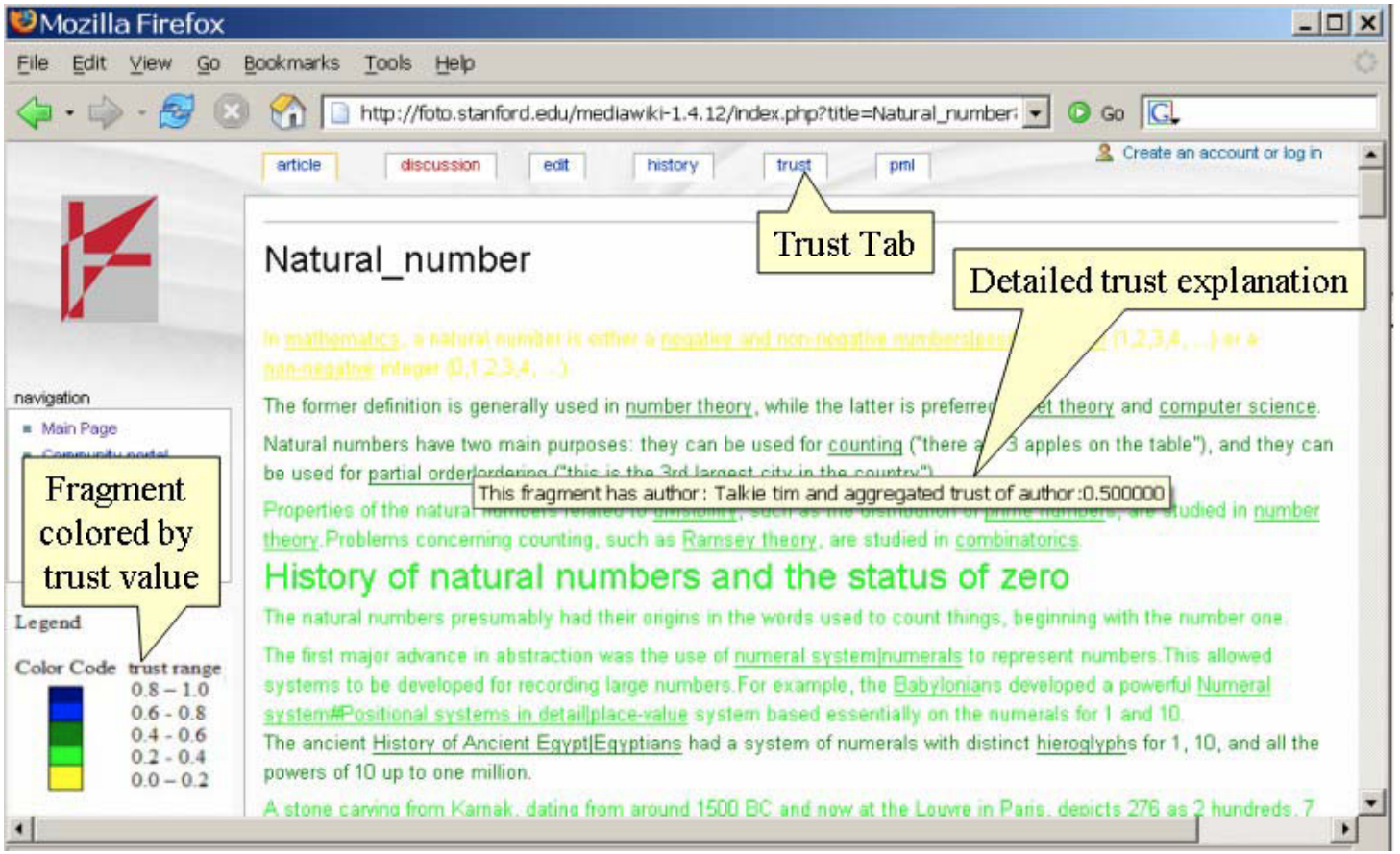
How human users interact with explanations?

Trust

Trust

- Inference Web
  - PML Trust vocabulary
  - Trust explanation

Trust


Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://foto.stanford.edu/mediawiki-1.4.12/index.php?title=Natural_number
Go
article
discussion
edit
history
trust
pml
Create an account or log in
Trust Tab
Detailed trust explanation
Natural_number
navigation
Main Page
Fragment colored by trust value
This fragment has author : Talkie tim and aggregated trust of author :0.500000
Legend
Color Code trust range
0.8 – 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 – 0.2
History of natural numbers and the status of zero

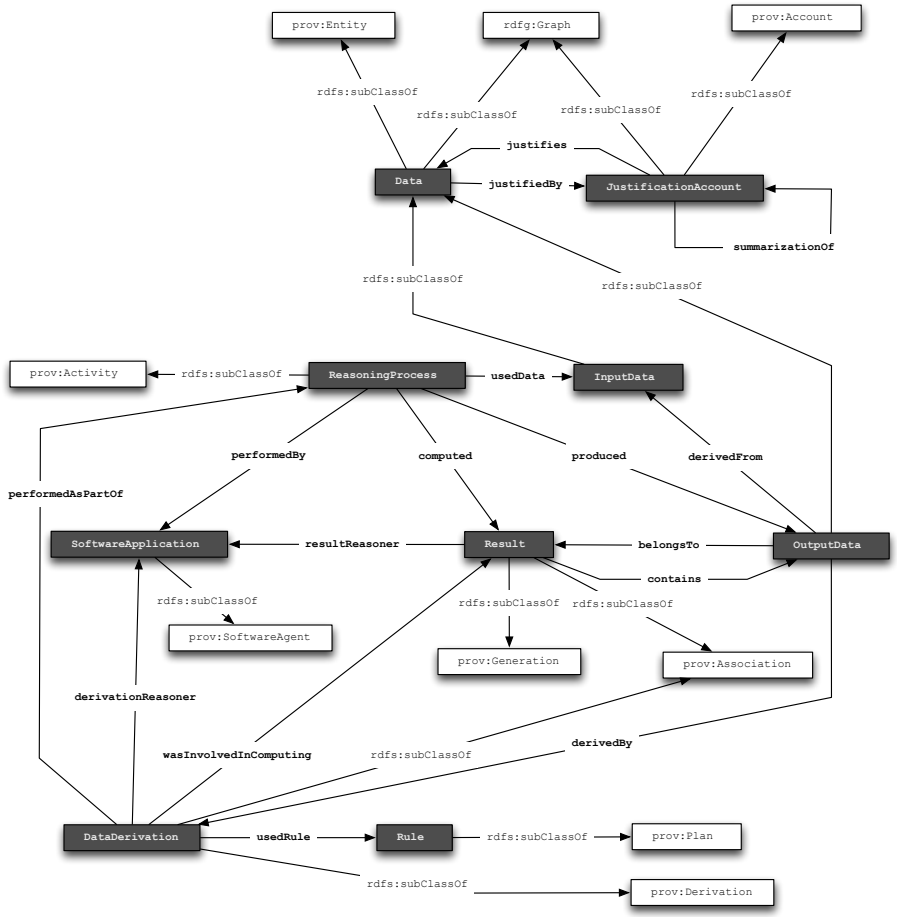
Perspective

Infrastructure

- Accommodating common data publishing principle
  - Publishing explanation metadata along with data using linked data principles
- Addressing heterogeneous and distributed nature of the Web promoting interoperability
  - W3C PROV-DM data model as an interchange data model

- Ratio4TA*, a lightweight vocabulary for encoding justifications.

- A specialization of the W3C PROV ontology
  - interoperability



*http://ns.inria.fr/ratio4ta/

Target

- Level of user expertise should be taken into account while providing explanation
  - User profiling

What is explained?

- Semantic Web applications use distributed interconnected data in their reasoning process
  - Explaining network of data used in the reasoning process, flow of information
- How explanations exposing problem solving methods influence security and confidentiality?

Metadata representation

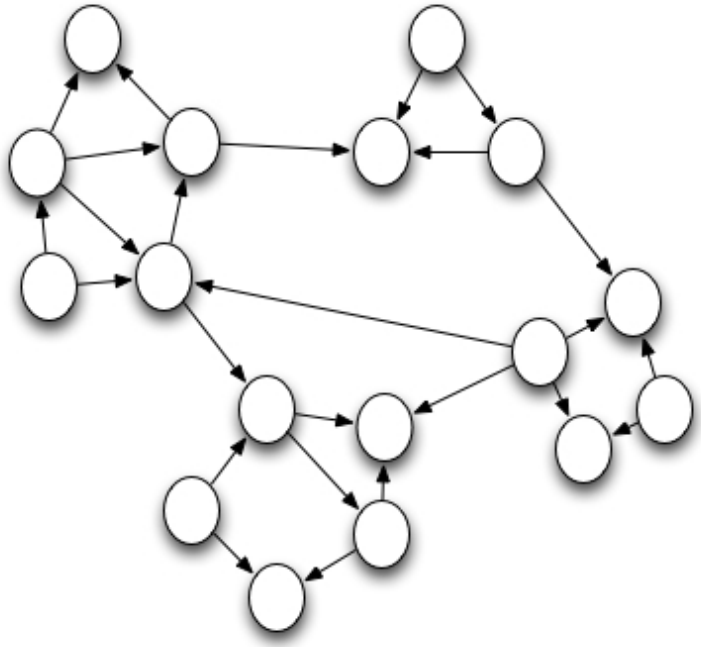
- Granularity
- Provenance
- Interoperability
- Compatibility with Linked Data

Presentation

- User expertise
  - What's useful and what's overwhelming?
  - Context-aware data consumption

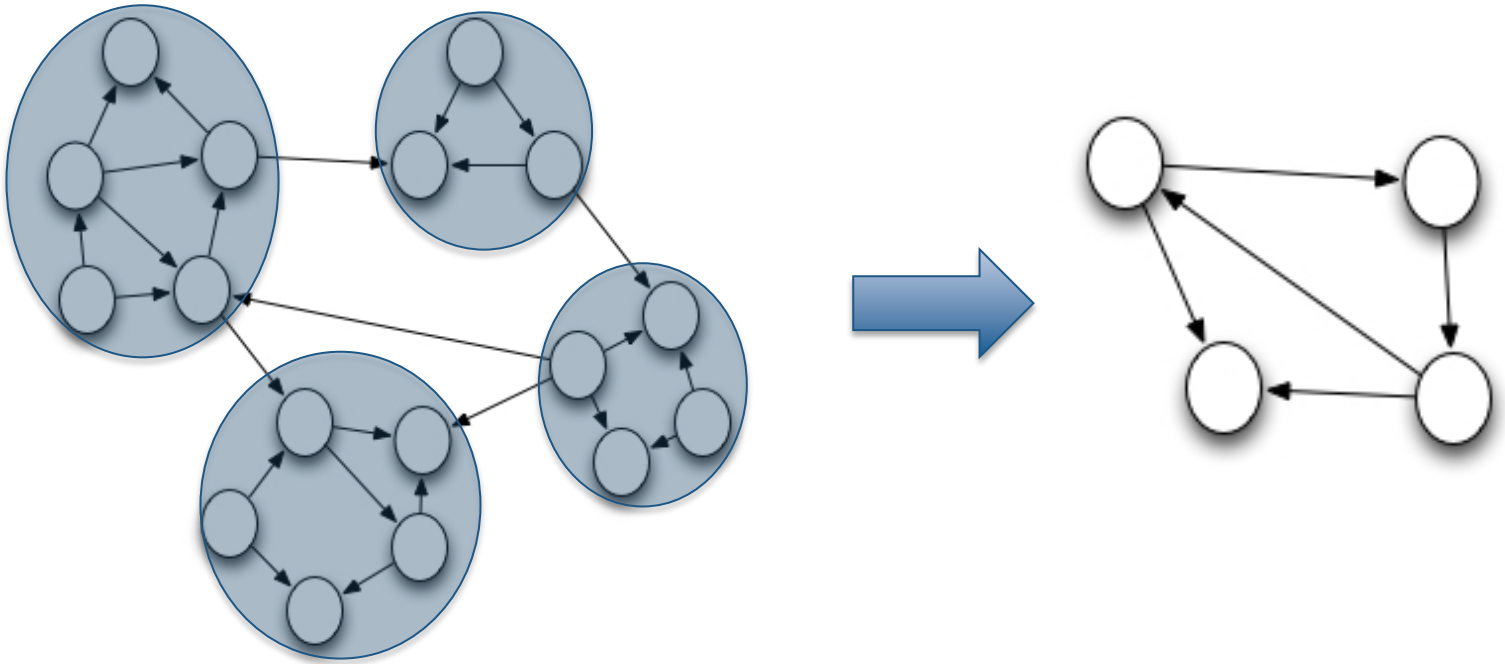
Presentation

Summarization



Presentation

Summarization



Interaction

- What kind of interactions are useful need to be understood
- How to interact established trust?

Trust

- How explanation influence trust in the Semantic Web?
- How to capture established trust and reason over it?
- Explaining trust itself

Work in Progress

Linked Justifications

- Ensuring trustworthiness in reasoning over Linked Data.

"Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/"

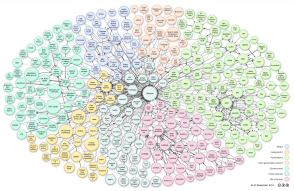
```
# From Dbpedia
lodapp:inputData1 {
  dbpedia:Philadelphia owl:sameAs geonames:4560349 .
  dbpedia:Philadelphia rdfs:label "Philadelphia"@en .
}
```

Consumes

Linked Data Application

Produces

```
# From GeoNames
lodapp:inputData2 {
  geonames:4560349 gn:parentFeature geonames:5205788.
  geonames:4560349 gn:name "Philadelphia"@en .
  geonames:5205788 gn:name "Philadelphia County"@en .
}
```



Published as

```
# Derived
lodapp:data1 {
  dbpedia:Philadelphia gn:parentFeature geonames:5205788.
}
```

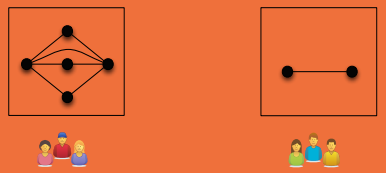
r4ta:justifies

```
# Justification
lodapp:justification1 {
  # Type declarations
  lodapp:justification1 r4ta:justifies lodapp:data1 .
  lodapp:justification1 rdf:type r4ta:JustificationAccount .
  lodapp:reasoningProcess1 rdf:type r4ta:ReasoningProcess .
  lodapp:corese rdf:type r4ta:SoftwareApplication .
  lodapp:geoFeatureRule rdf:type r4ta:Rule .
  lodapp:result1 rdf:type r4ta:Result .
  lodapp:data1 rdf:type r4ta:OutputData .
  lodapp:inputData1 rdf:type r4ta:InputData .
  lodapp:inputData2 rdf:type r4ta:InputData .
  lodapp:derivation1 rdf:type r4ta:DataDerivation .
  # Reasoning process
  lodapp:reasoningProcess1 r4ta:performedBy lodapp:corese .
  lodapp:reasoningProcess1 r4ta:usedData lodapp:inputData1.
  lodapp:reasoningProcess1 r4ta:usedData lodapp:inputData2.
  lodapp:reasoningProcess1 r4ta:computed lodapp:result1 .
  lodapp:reasoningProcess1 r4ta:produced lodapp:data1 .
  # Computed result
  lodapp:result1 r4ta:resultReasoner lodapp:corese .
  # Output data
  lodapp:data1 r4ta:derivedFrom lodapp:inputData1 .
  lodapp:data1 r4ta:derivedFrom lodapp:inputData2 .
  lodapp:data1 r4ta:belongsTo lodapp:result1 .
  lodapp:data1 r4ta:derivedBy lodapp:derivation1 .
  # Data derivation
  lodapp:derivation1 r4ta:usedRule lodapp:geoFeatureRule .
  lodapp:derivation1 r4ta:wasInvolvedInComputing lodapp:result1 .
  lodapp:derivation1 r4ta:derivationReasoner lodapp:corese .
  lodapp:derivation1 r4ta:performedAsPartOf lodapp:reasoningProcess1 .
}
```

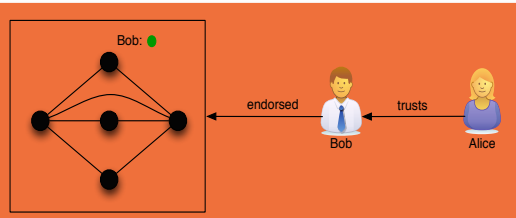
Natural Language Explanations

Graphical Explanations

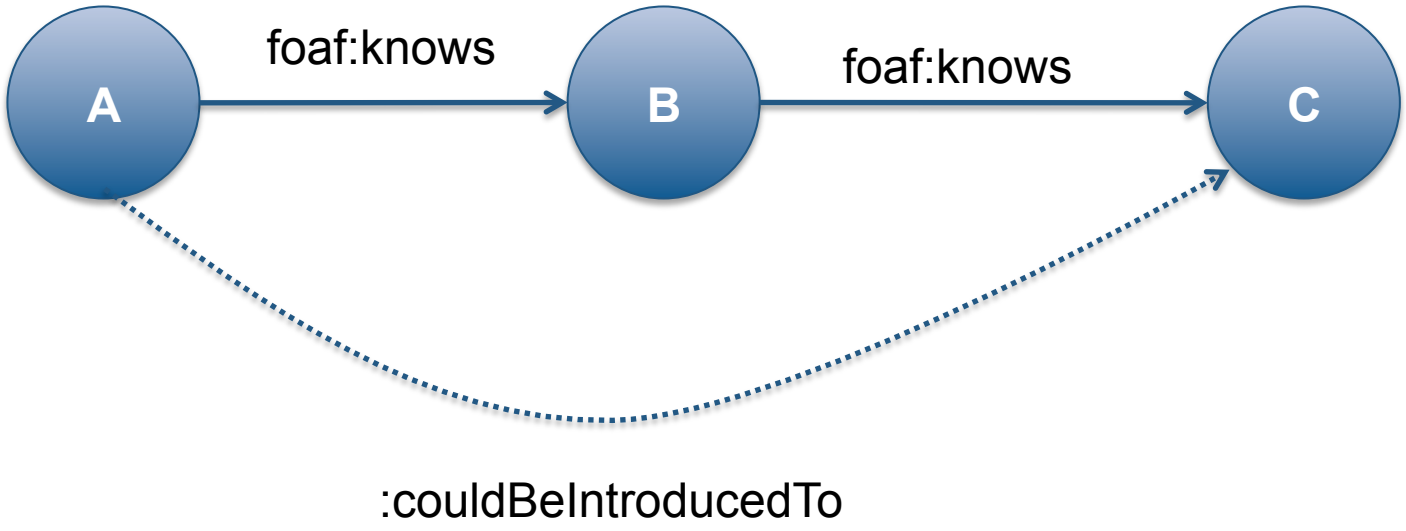
Navigation using follow-your-nose principle



Summarization & user profiling



FOAFConnection



Kolflow Deliverables

Task 4: Traces and explanations: documenting inferences, query solving and interactions

Kolflow Deliverables

No.	Description	Due (Months)	Delivered (PhD Months)	Due according to PhD start
D41	State of the art on the different topics addressed in this task	August 2011 (6)	May 2012 (11)	December 2011
D42	Algorithm to explain basic query mechanisms	December 2011 (10)	March 2012 (9)	April 2012
D43	Algorithm to provide performance and errors indicators	May 2012 (15)		September 2012
D44	Test and evaluation of the alter ego assistant with regard to the scenarios	August 2012 (18)		December 2012
D45	Algorithm to explain ontology-based processing	October 2012 (20)		February 2013
D46	Algorithm to suggest queries and changes to queries	February 2013 (26)		August 2013

Currently in the 14th month of my PhD

Kolflow Deliverables – Documenting Inferences

So far

- State of the art of explanation in the semantic web(D42)
- Linked Justifications (D41)
  - Ratio4TA vocabulary
  - Platform for publishing and consuming justifications

Work in progress

- Summarization (D45)
  - Finding patterns in justification RDF graphs
  - Partitioning justification RDF graphs for creating summarized graphs
  - User Interfaces with the support of visualizing summarized explanations (zooming in, zooming out)
  - Scenarios: DBPedia.fr (RDFS type inferences), FOAF

Kolflow Deliverables – Documenting Query Solving and Interactions

Future

- Explanation of ontology based processing (D45)
  - Explanation of RDFS inferences in SPARQL query results in Corese SPARQL Engine
- Performance (D43)
  - which part of the query failed most often
  - where is most time spent in solving a query
  - Looking into DB community work in query performance indicators
- Query suggestions (D46)
  - Use the performance indicators for suggesting improvements
- Interacting and propagating trust

Kolflow Deliverables

No.	Description	Due (Months)	Delivered (PhD Months)	Due according to PhD start
D41	State of the art on the different topics addressed in this task	August 2011 (6)	May 2012 (11)	December 2011
D42	Algorithm to explain basic query mechanisms	December 2011 (10)	March 2012 (9)	April 2012
D43	Algorithm to provide performance and errors indicators	May 2012 (15)	Delay (August 2013)	September 2012
D44	Test and evaluation of the alter ego assistant with regard to the scenarios	August 2012 (18)		December 2012
D45	Algorithm to explain ontology-based processing	October 2012 (20)	Next (December 2012 -> Summarization, April 2013 ->SPARQL)	February 2013
D46	Algorithm to suggest queries and changes to queries	February 2013 (26)		August 2013

Currently in the 14th month of my PhD

Thank you