Web Engineering

Social Networks and Semantic Web

Part II

Fernando Barraza A. Ms.C.
fbarraza@javerianacali.edu.co

Part II Content

- Modelling and Aggregating social Network Data
- Ontological representation
- Developing social semantic applications
  - Case studies

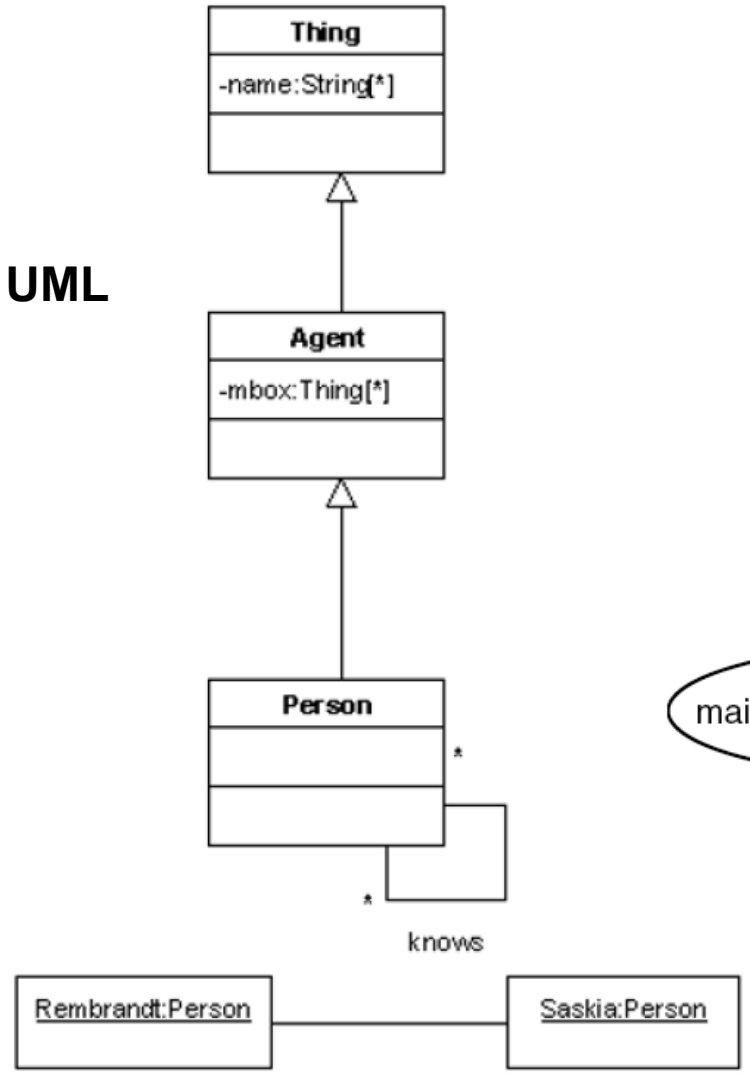
Why use Semantic Web Tech's

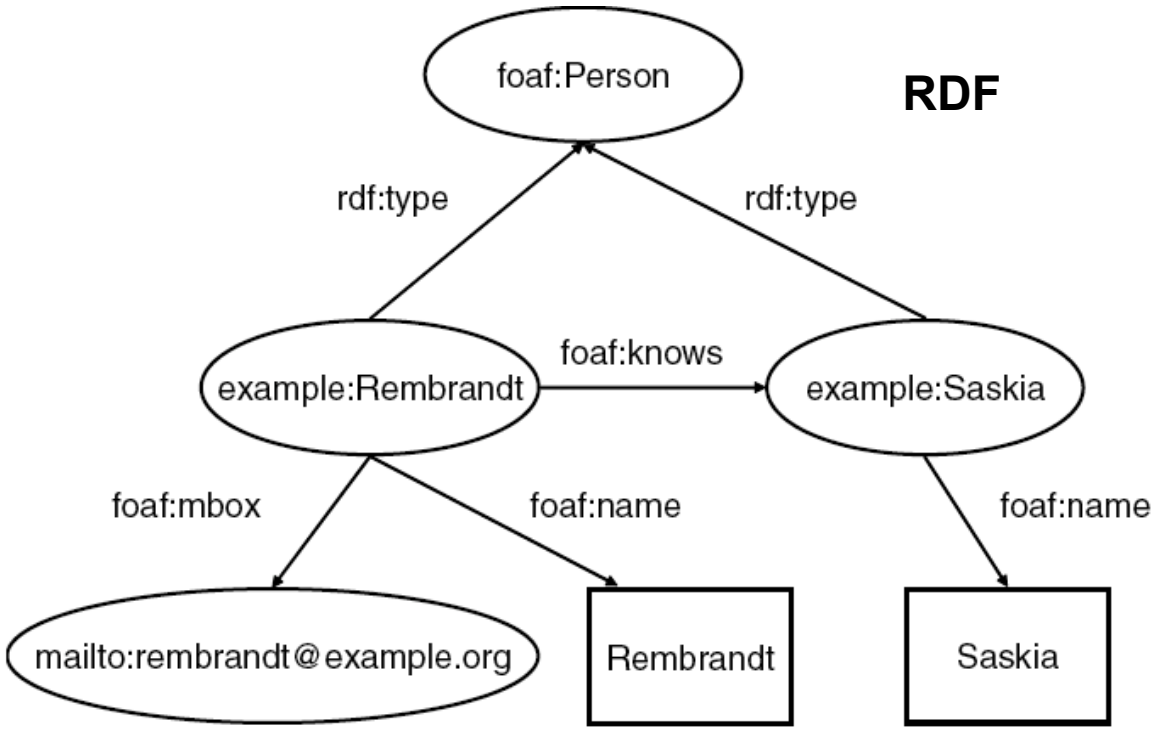
- To represent information on and about the current Web using formal languages that computers can process and reason with.
- Recapturing the information on the current Web and adding additional descriptions of Web resources (metadata) would allow machines to support in performin intelligent tasks, providing analysis by combining information from multiple sources.

Ontology-based Knowledge Representation

- Is for modelling the “domain model”
- Ontologies are expressed in formal languages with a well-defined semantics.
- Ontologies build upon a shared understanding within a *community*.
- Other approaches fails on above (UML, E/R, XML, etc.)

FOAF representation

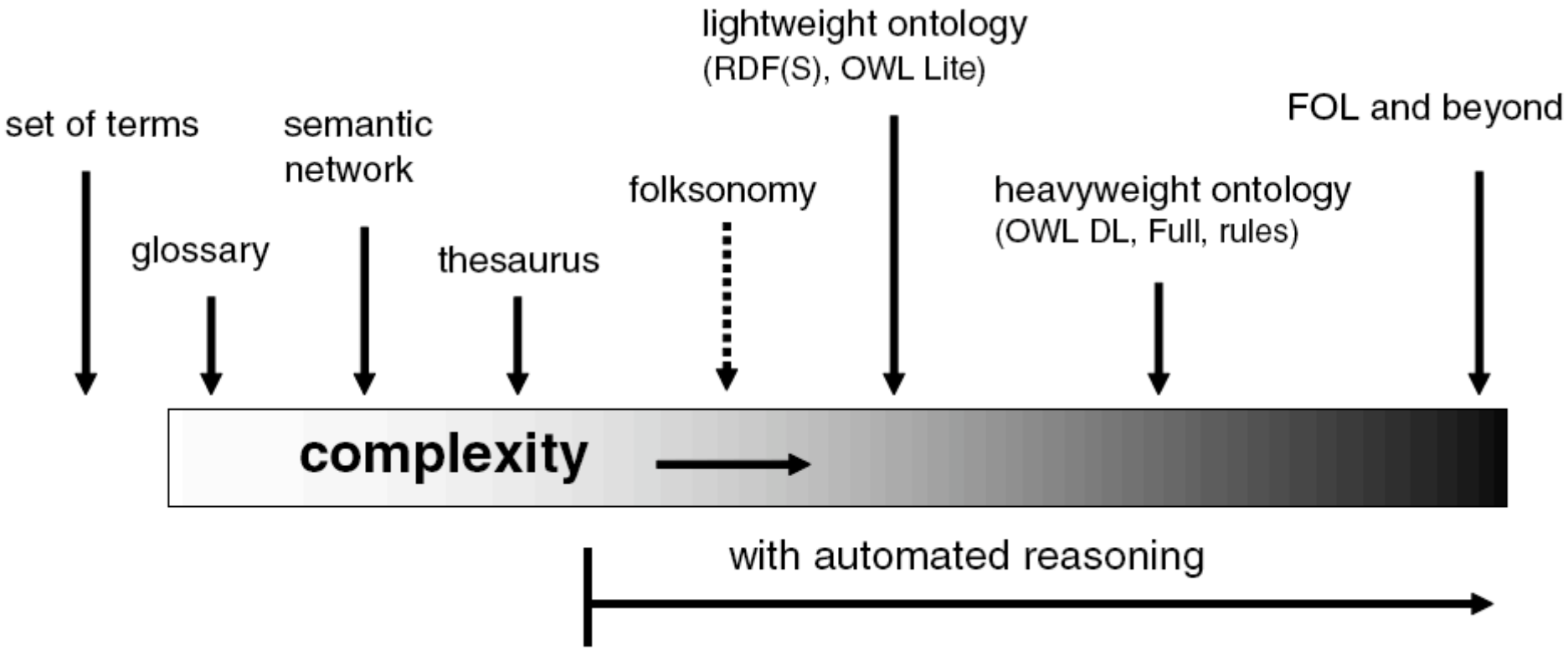




Languages comparison

	Origin	Application domain	Primitive	Expres-sivity	Dis-tributed represen-tation	Formal semantics
E/R	1976	Relational databases	Relation	•	no	no
UML	1995	OO software	Object	••	no	yes[33]
XML	1998	Text markup and data exchange	Entity	••	yes	no
RDF/OWL	2004	Resource markup and data exchange	Resource	• - • • •	yes	yes

Level of modelling detail



Ontologies can be organized according to complexity (informally, the level of semantics). [Smith and Welty]

Available Ontologies

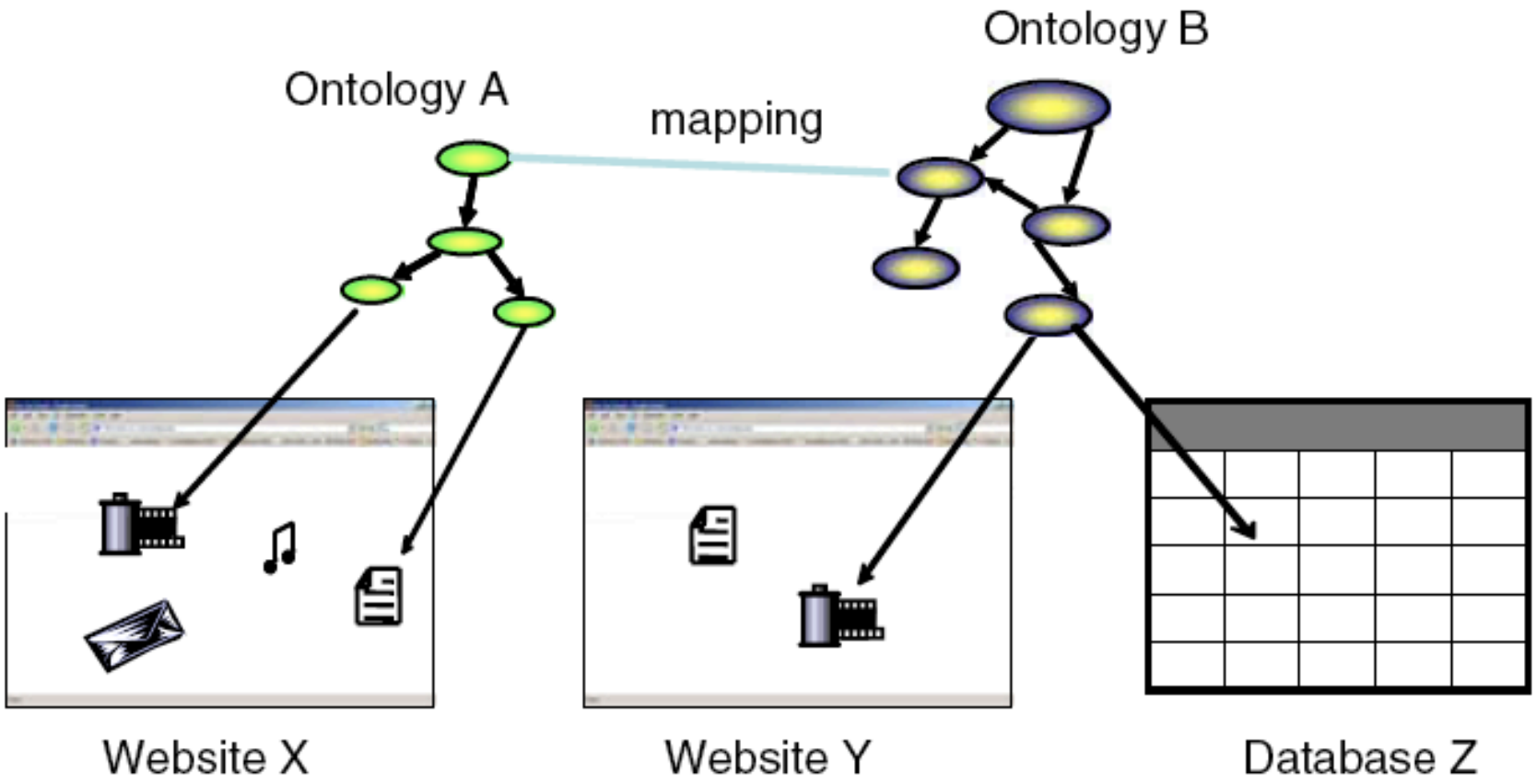
- Most common are all lightweight ontologies
  - Serving the needs of many applications with divergent goals.
  - Shared Web ontologies also tend to be small as they contain only the terms that are agreed on by a broad user base.

Available Ontologies (2)

- Heavyweight ontologies are more commonly found in targeted expert systems
  - Used in focused domains with a tradition of formalized processes and vocabularies such as the area of *life sciences* and *engineering*.

- But, different levels of ontologies can be combined …

Network of Ontologies



This architecture adds machine-processable semantics to existing web content, including Web resources and databases.

How to do it?

1. To identify concepts in ontologies using globally unique identifiers (URIs).
2. To use identifiers in data sources to point to a concept or relationship from an external, public ontology.
   - Similar to creating HTML pages and linking them to existing ones, anyone can create and publish an ontology, which may reference the concepts in remote ontologies.

How to do it? (2)

- Much like the hyperlinks among web pages, it is expected that these references will form a contiguous web linking all knowledge sources across the Web.

3. To use ontologies for reference existing Web resources and describe their characteristics (URLs are also URIs)

How to do it? (3)

4. The (meta)data and its schema can be made public either by publishing an RDF or OWL document online (which may be dynamically generated from a database or through a Web service invocation) or by providing access to a data source through implementing the standard query language and API Tool of the Semantic Web.

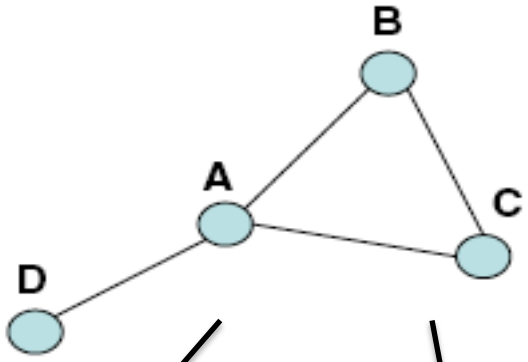
Modelling Social Networks

- Maintaining the semantics of social network data is crucial for aggregating social network information, due to
  - Heterogeneous environments where the individual sources of data are under diverse control.
  - Data are already available in an electronic format, which is typically the case of online communities.

Modelling Social Networks

- Social network data can be modelled by a graph where,
  - The nodes represent individuals and the edges represent binary social relationships. (Less commonly, higher-arity relationships may be represented using hyper-edges, i.e. edges connecting multiple nodes.)
  - Social network studies build on attributes of nodes and edges, which can be formalized as functions operating on nodes or edges.

Format Graph



UCINET DL

```
dl
n = 4
labels embedded
format = edgelist
data:
A B
A C
A D
B C
```

Pajek.NET

```
*Vertices 4
1 "A"
2 "B"
3 "C"
4 "D"
*Edges
1 1
1 2
1 3
1 4
2 3
```

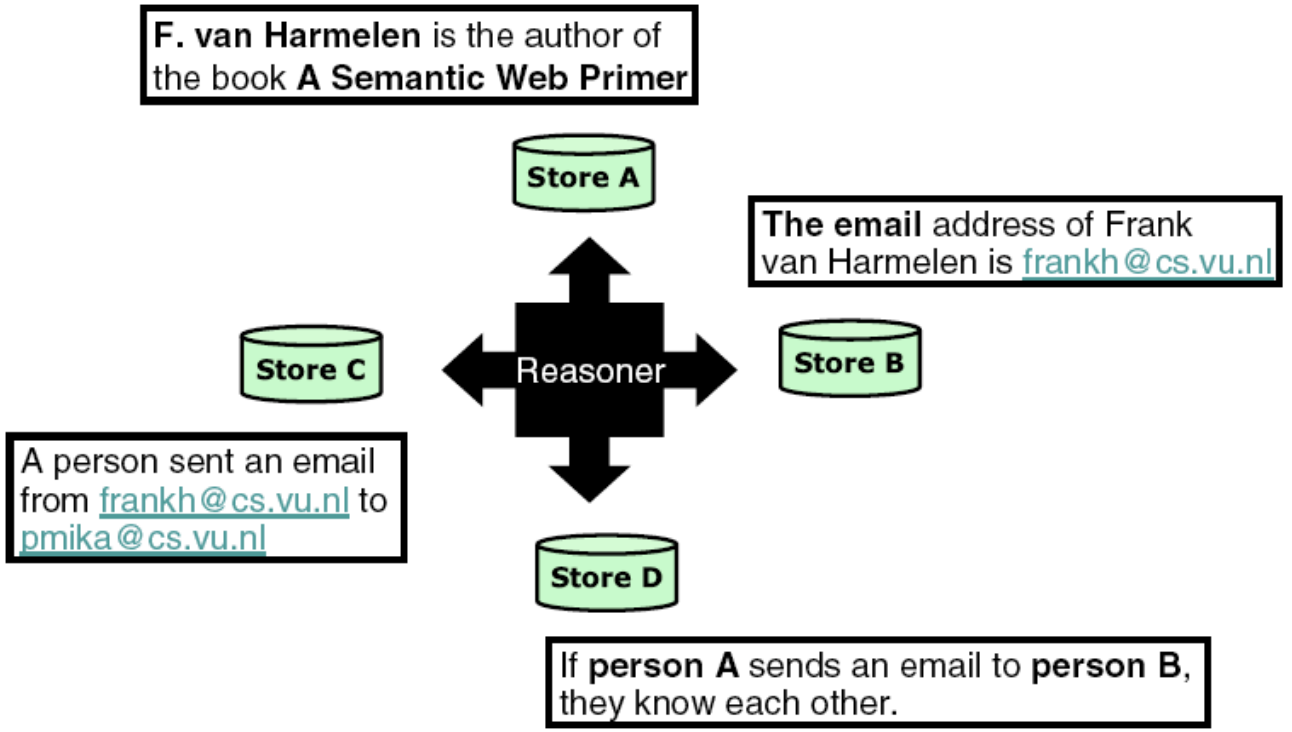
GraphML

```
<?xml version="1.0" encoding="UTF-8"?>
<graphml xmlns="http://graphml.graphdrawing.org/xmlns">
  <graph id="G" edgedefault="undirected">
    <node id="a"/>
    <node id="b"/>
    <node id="c"/>
    <node id="d"/>
    <edge source="a" target="b"/>
    <edge source="a" target="c"/>
    <edge source="a" target="d"/>
    <edge source="b" target="c"/>
  </graph>
</graphml>
```

Format Limitations

- Most common formats does not support the aggregation and reuse of electronic data.
- Reusing a number of data sources is important for
  - Describing the same set of individuals and their relationships (email archives and publication databases holding information about researchers).
  - Perform *triangulation* i.e. to use a variety of data sources and/or methods of analysis to verify the same conclusion.
- Aggregation for support a dynamic model like social networks are

Reuse and aggregating example



A case of identity reasoning: Based on a semantic representation, a reasoner would be able to conclude, for example, that Peter Mika knows the author of the book *A Semantic Web Primer*

Case study: FOAF

FOAF Basics
Agent
Person
name
nick
title
homepage
mbox
mbox_sha1sum
img
depiction (depicts)
surname
family_name
givenname
firstName

Personal Information
weblog
knows
interest
currentProject
pastProject
plan
based_near
workplaceHomepage
workInfoHomepage
schoolHomepage
topic_interest
publications
geekcode
myersBriggs
dnaChecksum

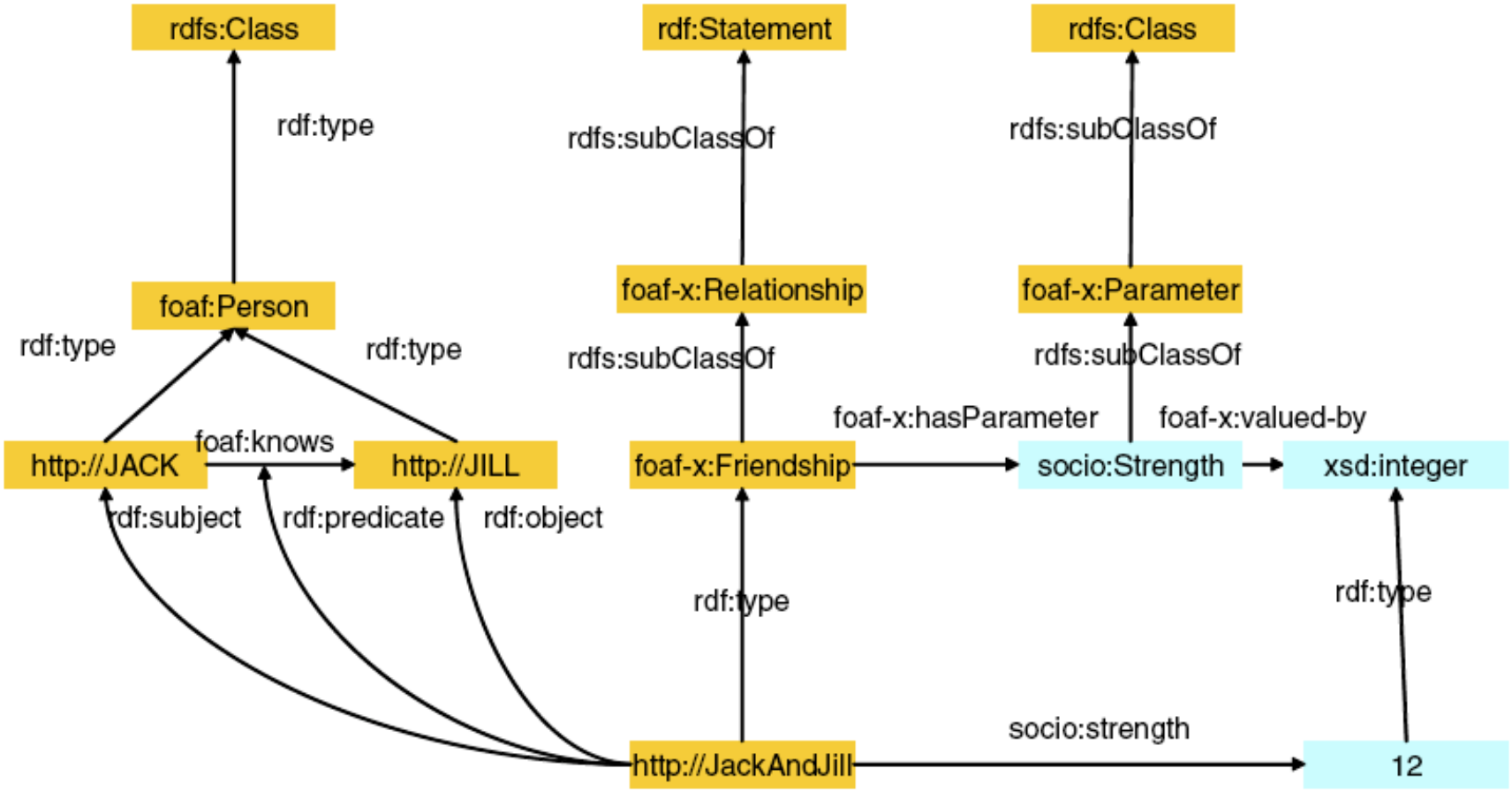
Online Accounts / IM
OnlineAccount
OnlineChatAccount
OnlineEcommerceAccount
OnlineGamingAccount
holdsAccount
accountServiceHomepage
accountName
icqChatID
msnChatID
aimChatID
jabberID
yahooChatID

Projects and Groups
Project
Organization
Group
member
membershipClass
fundedBy
theme

Documents and Images
Document
Image
PersonalProfileDocument
topic (page)
primaryTopic
tipjar
sha1
made (maker)
thumbnail
logo

The idea of FOAF was to provide a machine processable format for representing the kind of information that made the original Web successful, namely the kind of personal information described in homepages of individuals

RDF representation



Aggregating and reasoning with social network data

- First, convert the electronic data sets in traditional formats (RDBMS, excel, XML files etc.) into an RDF-based syntax, which allows,
  - Store the data in an ontology store and
  - Manipulate it with ontology-based tools.
- In this process is necessary,
  - Assign identifiers to resources and represent the data in terms of a shared ontology such as FOAF.
  - Use a *ontology mapping*

Ontology Mapping

- In case when data sets come from external sources it is often more natural to preserve their original schema.
- For example, in case of converting data from a relational database or Excel sheet it is natural to preserve the schema of the database or spreadsheet as represented by the table definitions or table headings.

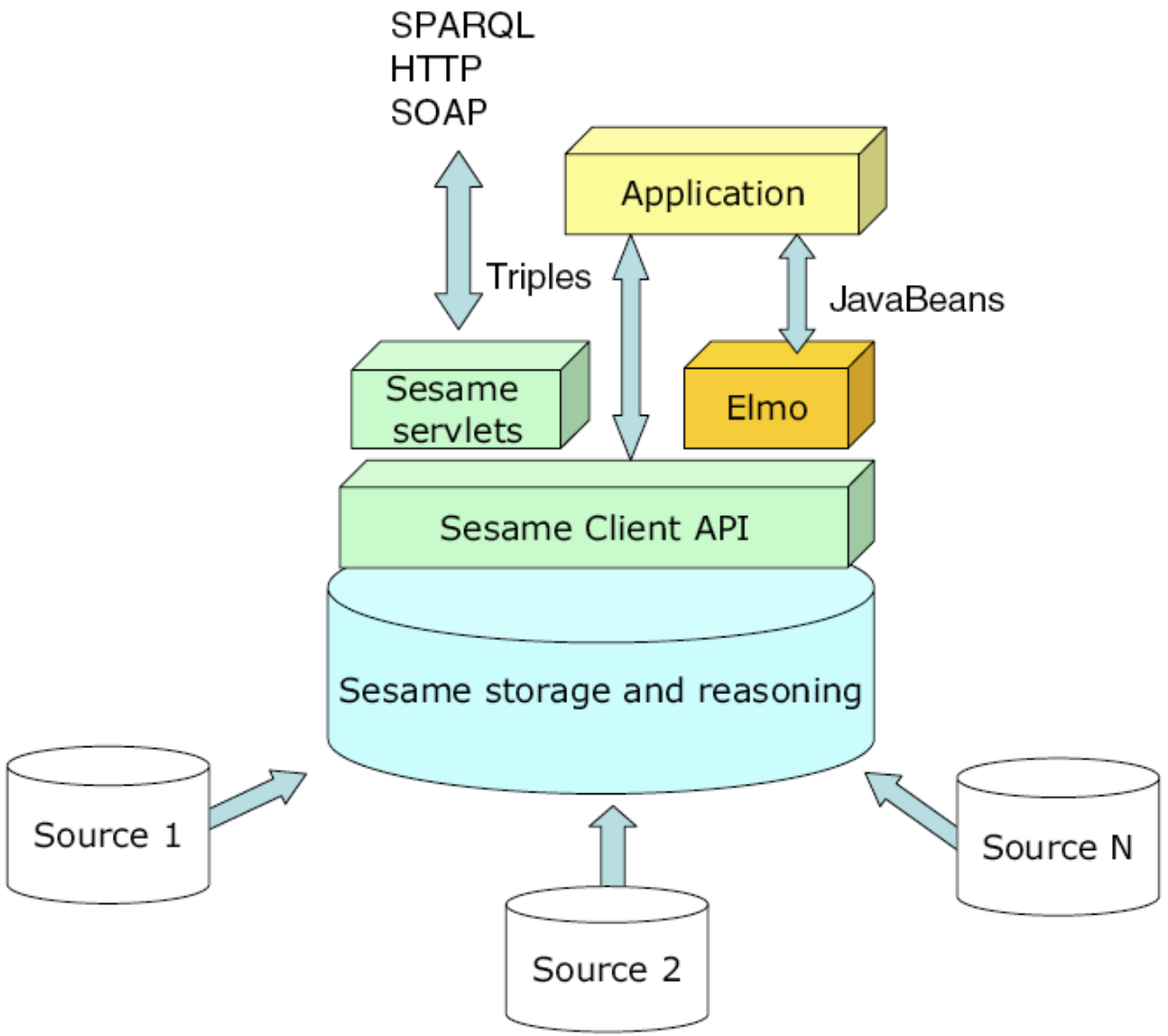
Aggregating

- To find identical resources across the data sets. This is a two step process:
  - It requires capturing the domain-specific knowledge of when to consider two instances to be the same
  - To carry out the actual instance unification or *smushing,* as a reasoning task, where we iteratively execute the rules or procedures that determine equality until no more equivalent instances can be found.

Developing Social Semantic App's

- Facts:
  - Most of the current data-driven applications populating the Web have been developed using relational databases as a back-end (e.g LAMP)
  - Web Semantic App's deal with a different kind of data (ontologies with formal semantics)
  - Semantic Web Technologies are oriented for distributing nature of applications.

Generic Architecture App



Credits

- **Peter Mika**. Social Networks and the Semantic Web. Springer.