OBERÖSTERREICH

University of Applied Sciences

Web-Semantik-Technologien

Vorlesung WS 2019/20

thomas.kern@fh-hagenberg.at

Semantic Technologies

Semantic Knowledge Models, Semantic Web Stack

Semantics

- Semantics (from Ancient Greek: sēmantikós, "significant") is the **study of meaning**. It focuses on the relationship between signifiers—like words, phrases, signs, and symbols—and what they stand for, their denotation.
- **Linguistic semantics** is the study of meaning that is used for understanding human expression through language.
- **Other forms of semantics** include the semantics of programming languages, formal logics, and semiotics (signs and symbols).

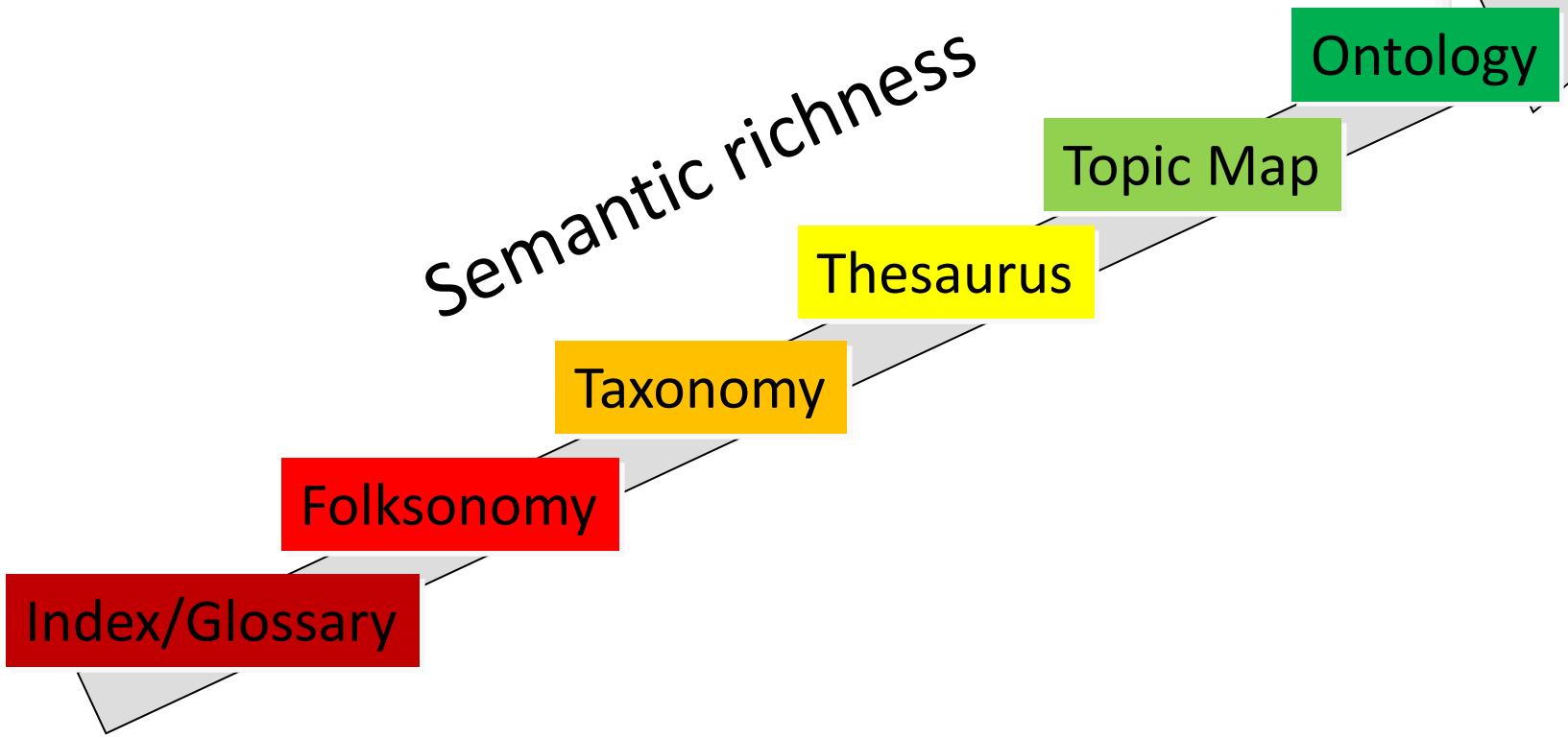
  (Definition from Wikipedia)

Characteristics of a common language

„People can´t share knowledge if they don´t speak a common language“

- Agreement on common symbols and terminology **(Index)**
- their meaning **(Glossary)**
- classifications of concepts **(Taxonomy)**
- Associations or crosslinks of concepts **(Thesaurus/Topic Map)**
- Rules and knowledge (experience) about which crosslinks are permitted and useful **(Ontology)**

Evolution of semantic knowledge models

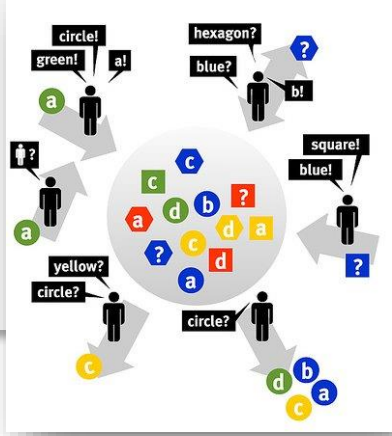


Index / Glossary

- **Index** (lat. „forefinger”, indicator, overview, title, table of contents, …)
  - https://en.wikipedia.org/wiki/Index

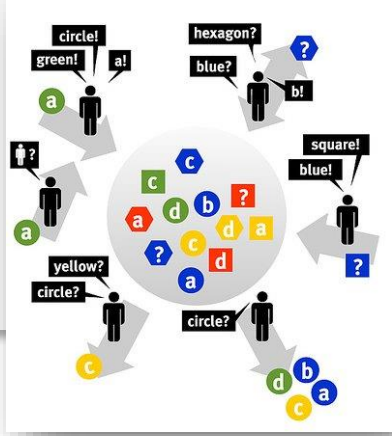
- **A glossary** (lat. „glossarium“ – tongue, language, strange word, …) is , also known as a vocabulary, or clavis, is an alphabetical list of terms in a particular domain of knowledge with the definitions for those terms.
  - https://en.wikipedia.org/wiki/Glossary

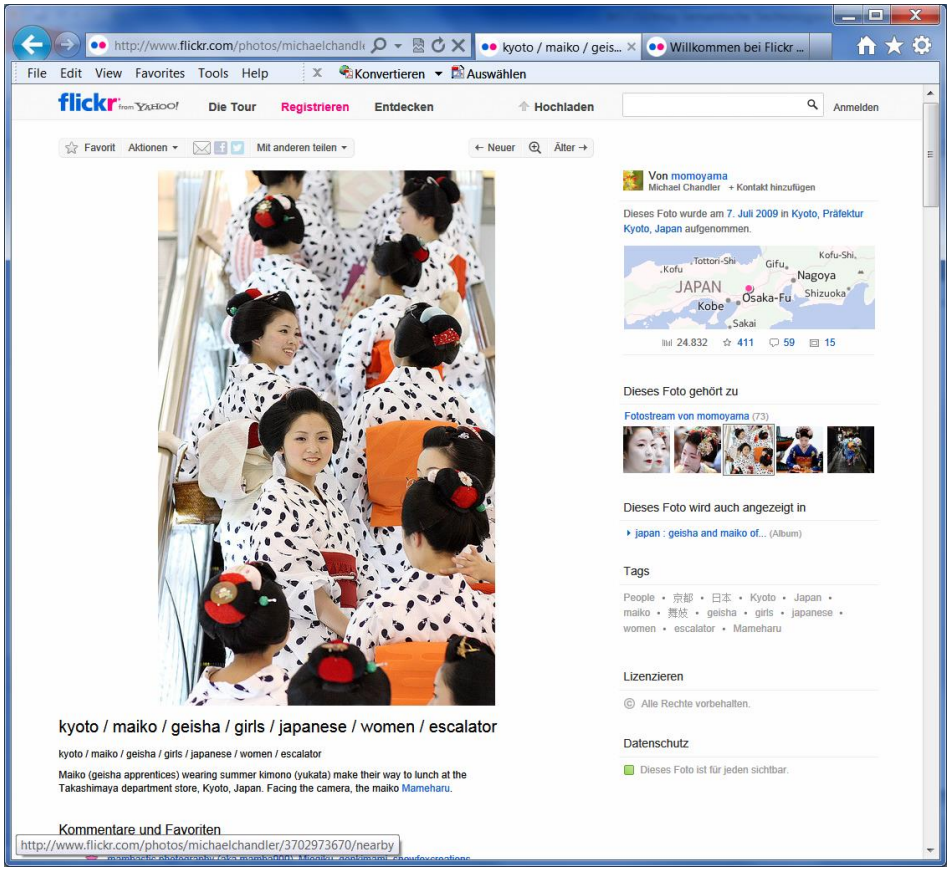
Folksonomy

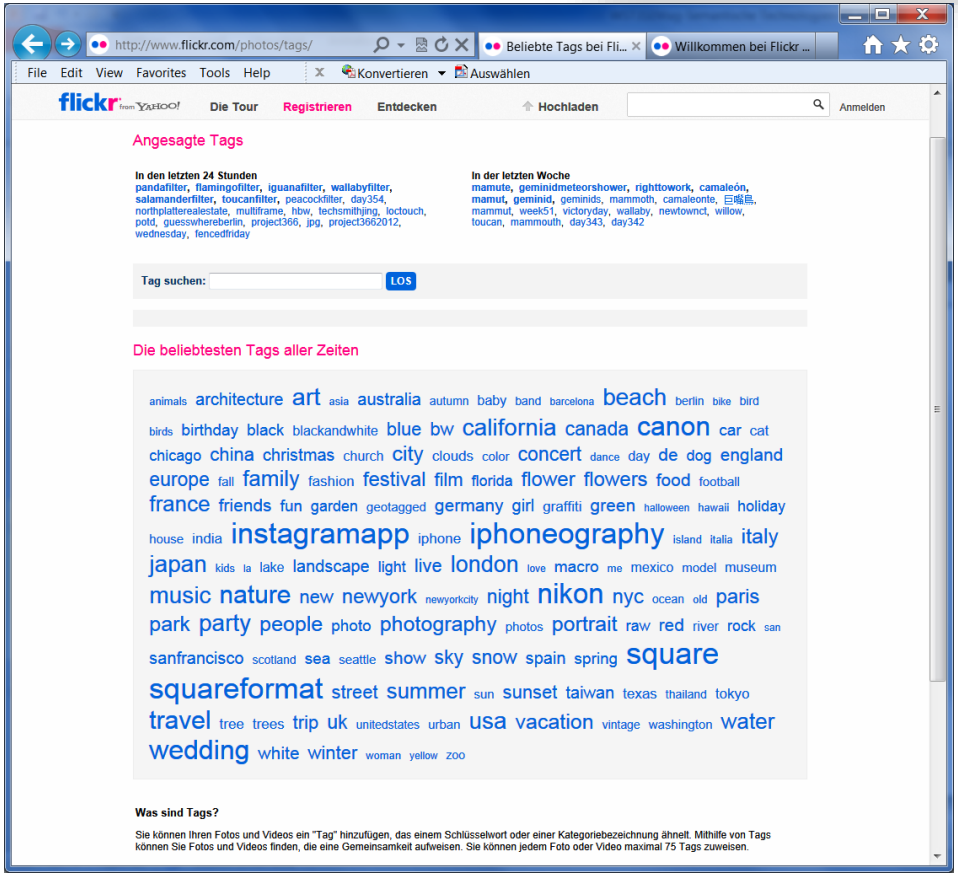


- Neologism for "**folk**" and "**taxonomy**“
- In contrast to the hierarchical classification in taxonomies **systematics** in folksonomies is **not a priori determined**.
- It is the organic result of **social tagging** in **social software**.
- **No hierarchy** of terms possible
- **No quality control** by experts
- **No sound theory**
  - Does the community agree on common terms?
- Application to photos, bookmarks, blogs
- Navigation via **Tag Clouds** possible
- Example: http://www.flickr.com/photos/tags/

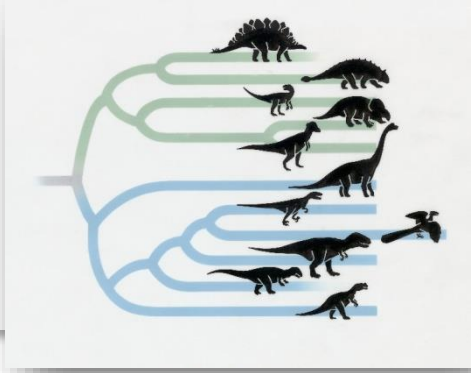
Examples of social tagging





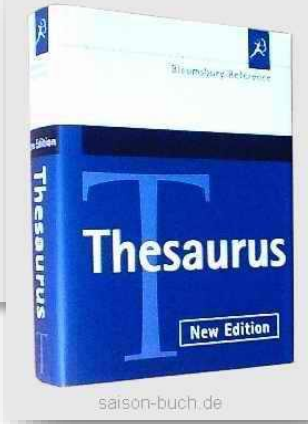


Taxonomy



- Taxonomies are **hierarchical classifications** (classes, subclasses etc.) of a subject area (for example, in biology: classification of living organisms).
- They define **classes of objects and relations** between them.
- But they **do not contain descriptions or definitions of objects** [see. Berners-Lee, Hendler, Lassila 2001].
- The concepts of a taxonomy are **arranged tree-like** so that with increasing depth their specificity increases.

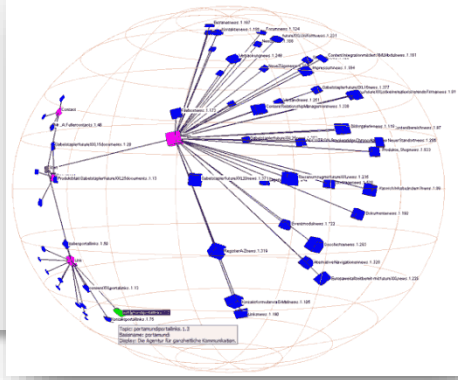
Thesaurus



- **Natural language is ambiguous**, diverse, and often does not indicate concept relationships.
- **A thesaurus is unambiguous**, normalized, standardized and shows term relationships explicitly.
- **It is a model that attempts to describe and represent a topic exactly.**
- It consists of a **systematically ordered collection of concepts** which are thematically related.
- A thesaurus is a **controlled vocabulary** (attribute value range) for each attribute to be described.
- Primarily **synonyms**, but also **broader and narrower terms** and **related terms** are managed.

  Examples:
  - Portrait (synonyms: image, image, mirror image),
  - Carpenter (broader term: craftsman),
  - Wedding (Related terms: bride, engagement, marriage)

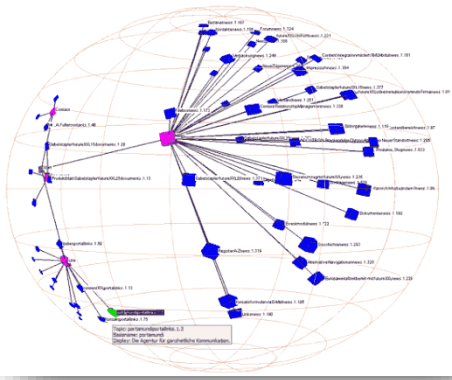
Maps



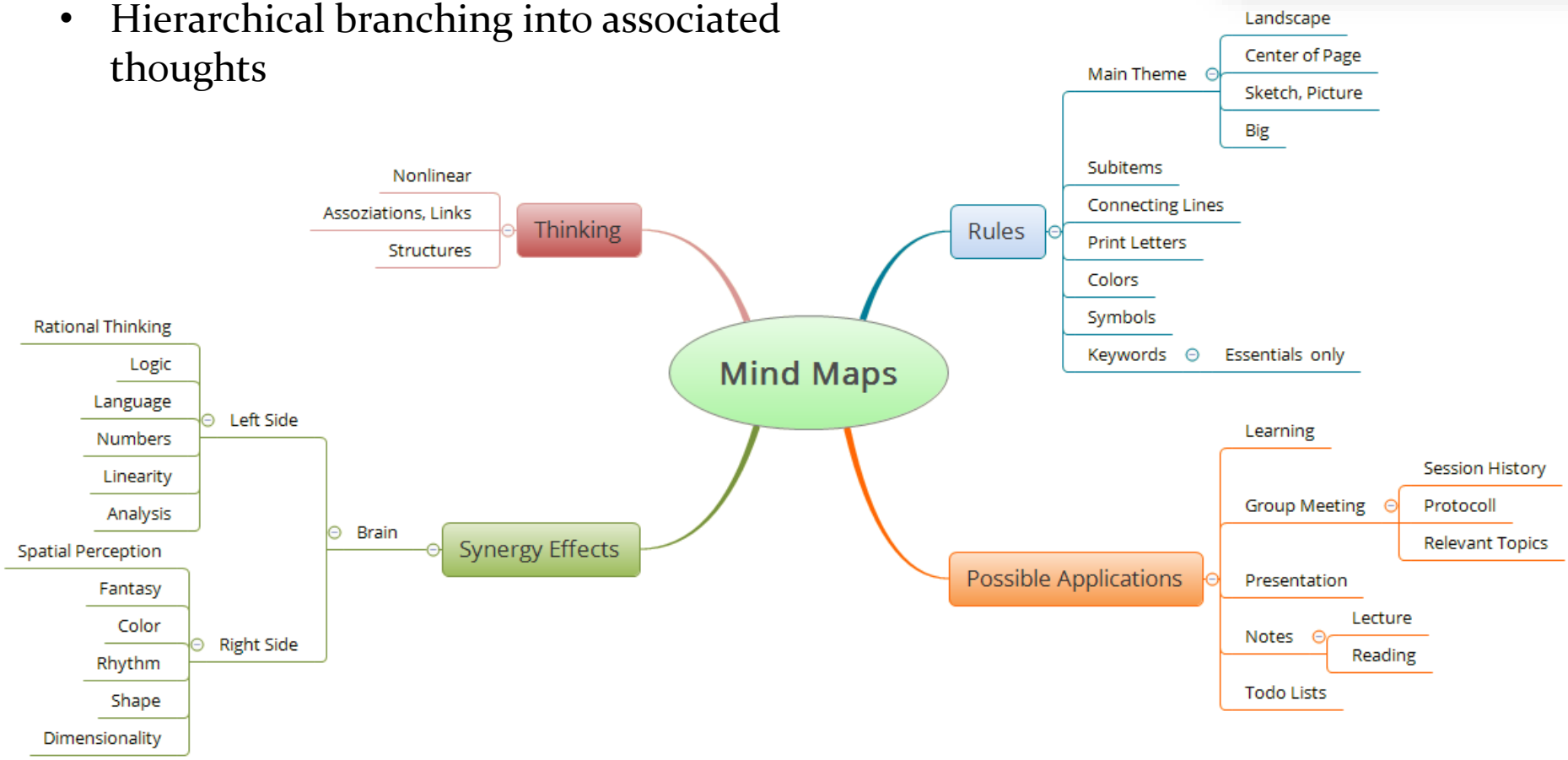
- **Maps are spatial visualizations of structure**
  - of content and causal structures
  - of ideas and knowledge
  - of information & resources

- The type of spatial representation (organization, representation) provides **information on the semantic meaning** of what is shown.

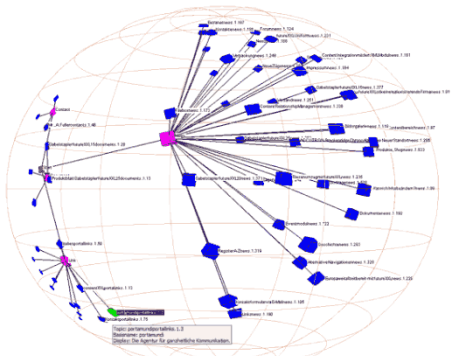
Example: Mindmap



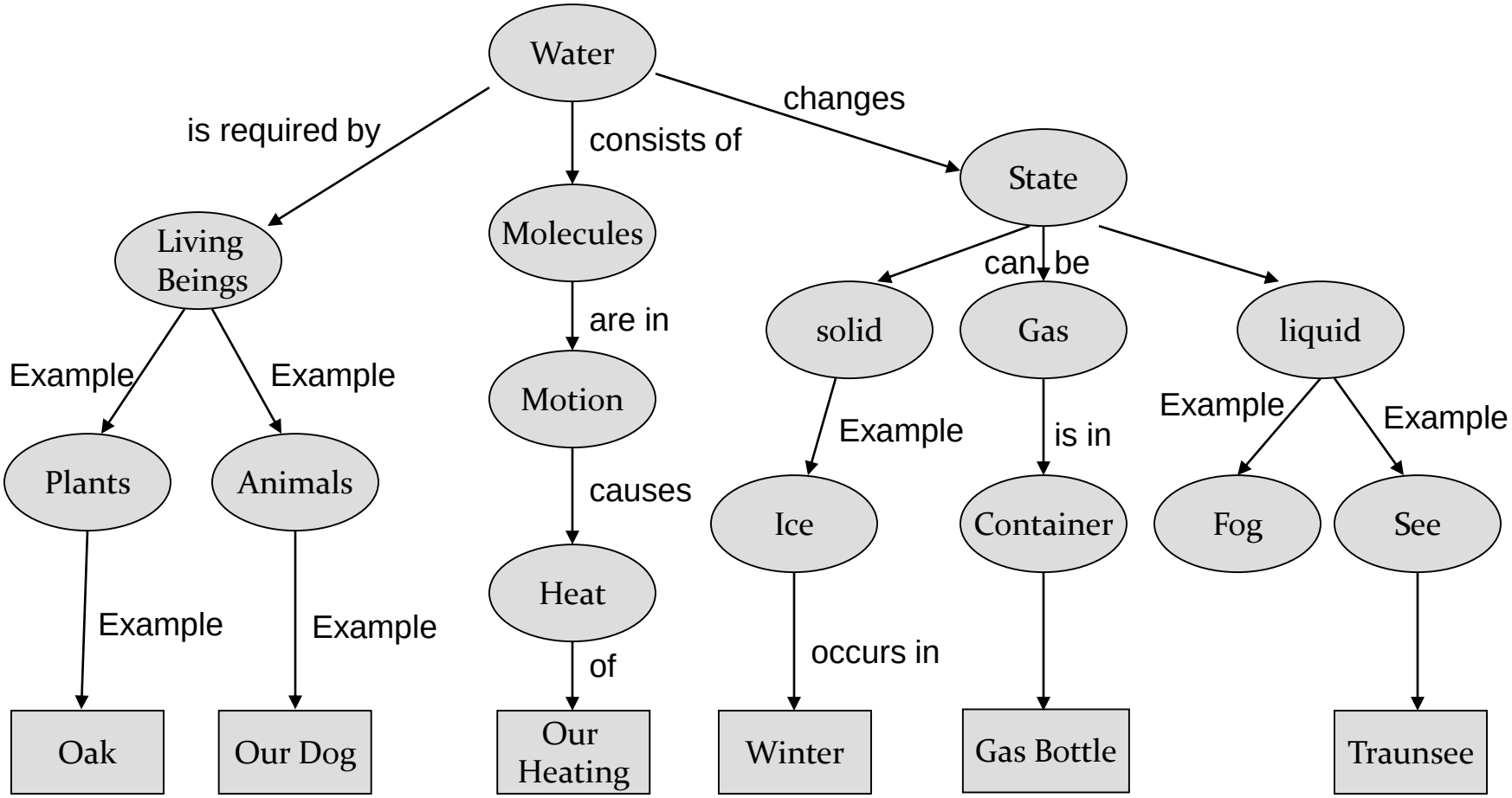
- One central main thought
- Hierarchical branching into associated thoughts



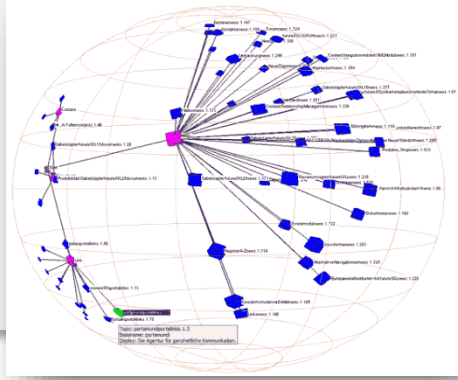
Example: Concept Map



- Concept map with terms and named relations.
- Relationship types are not defined strictly formal.

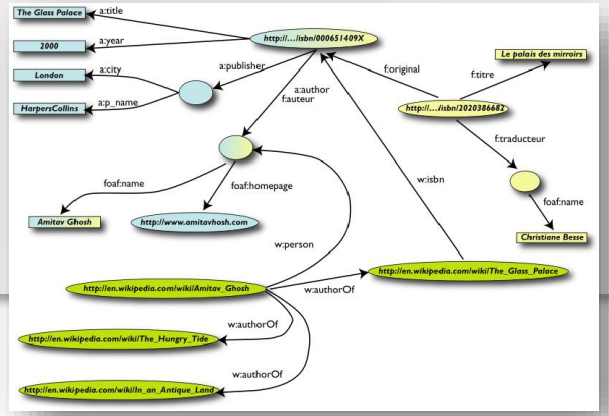


Topic Map



- A topic map is the collection of knowledge on **subjects** that are objects or topics of the description.
- Increased formalization compared to other mapping techniques (e.g .: mind maps, concept maps):
  - Information representation via TAO
  - **Topics** represent concepts (for example, person, country, organization),
  - **Associations** represent relationships between concepts,
  - **Occurrences** represent information resources on a specific topic.
- Topic Maps are standardized: ISO 13250, together with a XML-based serialization (XTM2)

Ontology Definition 1

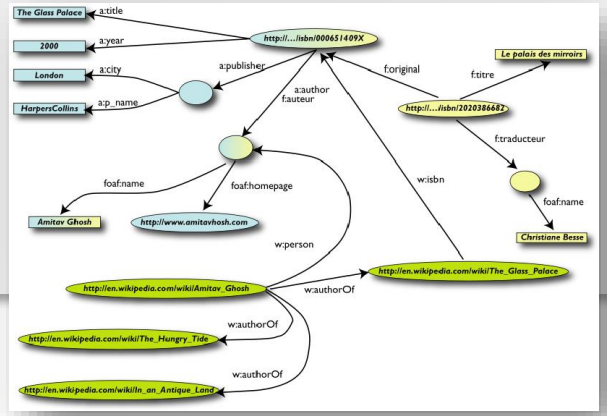


„An ontology is an explicit specification of a conceptualization.“

„Every Ontology is a treaty - a social agreement among people with some common motive in sharing.“

Tom Gruber, Standford University

Ontology Definition 2

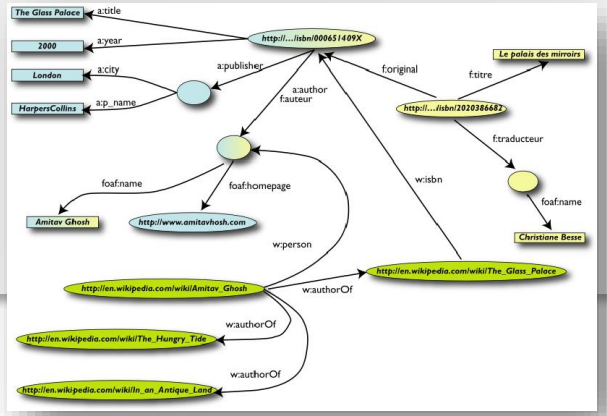


"An ontology defines a common vocabulary for researchers who need to share information in a domain."

"It includes machine-interpretable definitions of basic concepts in the domain and relations among them."

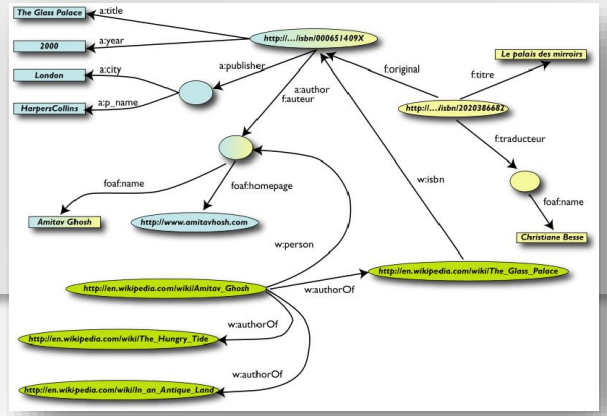
N. Noy & D. McGuiness, Stanford University

Ontology Purpose



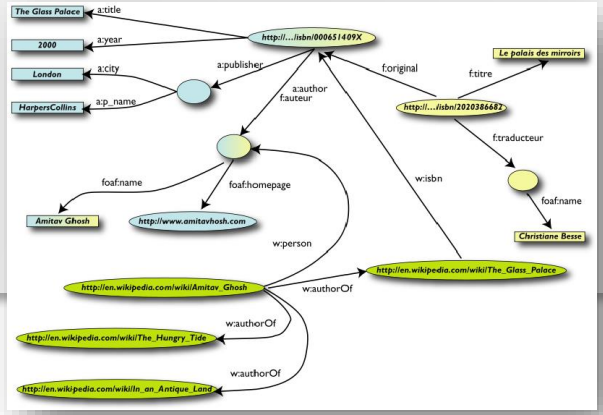
- Structuring
- Communication
- Interoperability
- Data Exchange
- Systems Engineering: specification, reliability, reusability

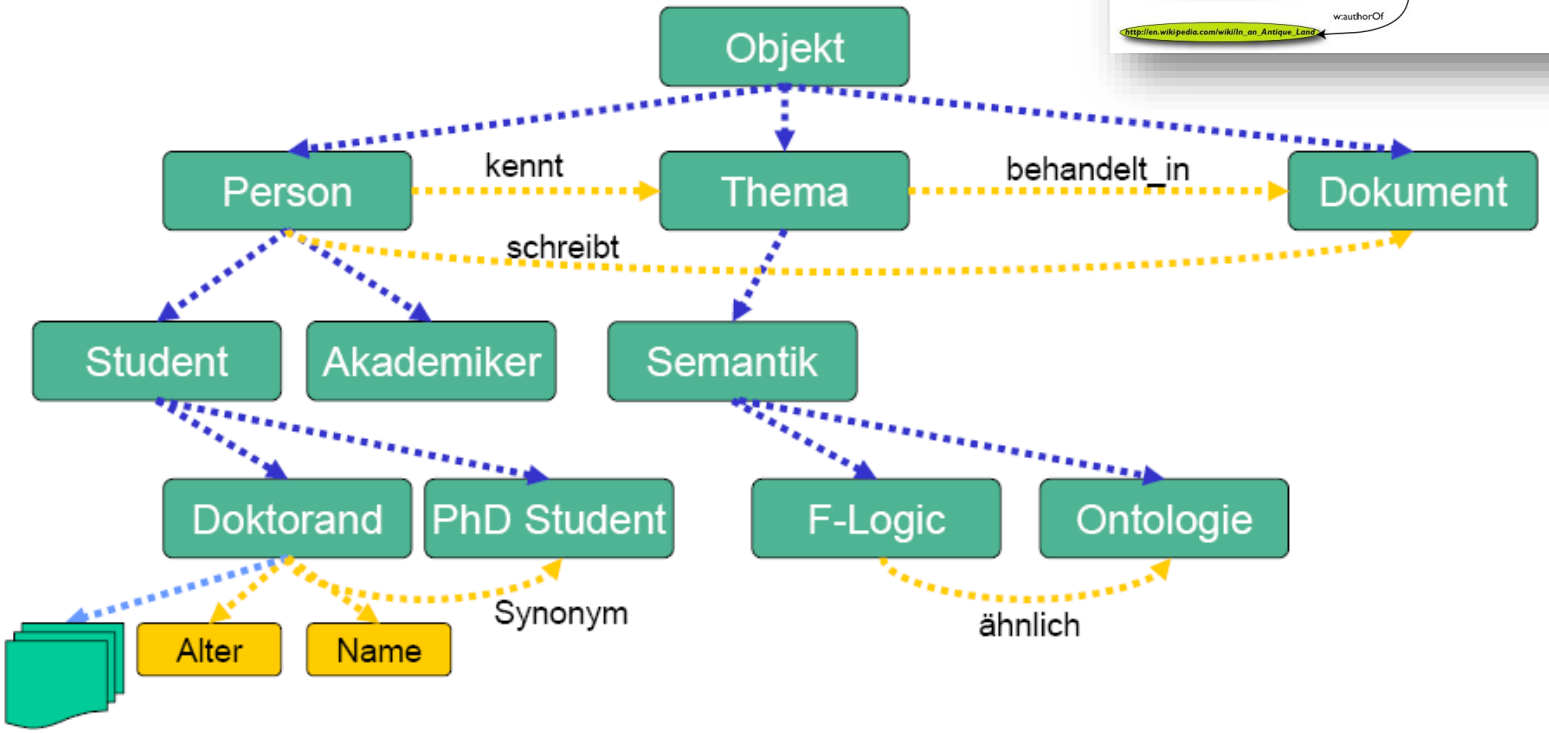
Ontology Components



- Concepts
- Types
- Individuals
- Properties
- Inheritance (Subclasses, Subproperties)
- Axioms
  - True statements within an ontology for the representation of knowledge that can not be inferred from other concepts.

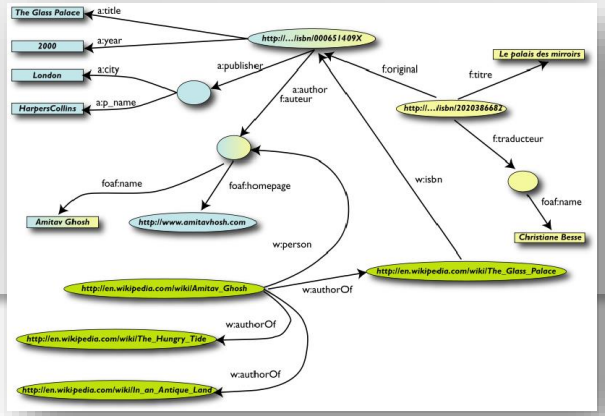
Ontology as a rule system for automatic reasoning

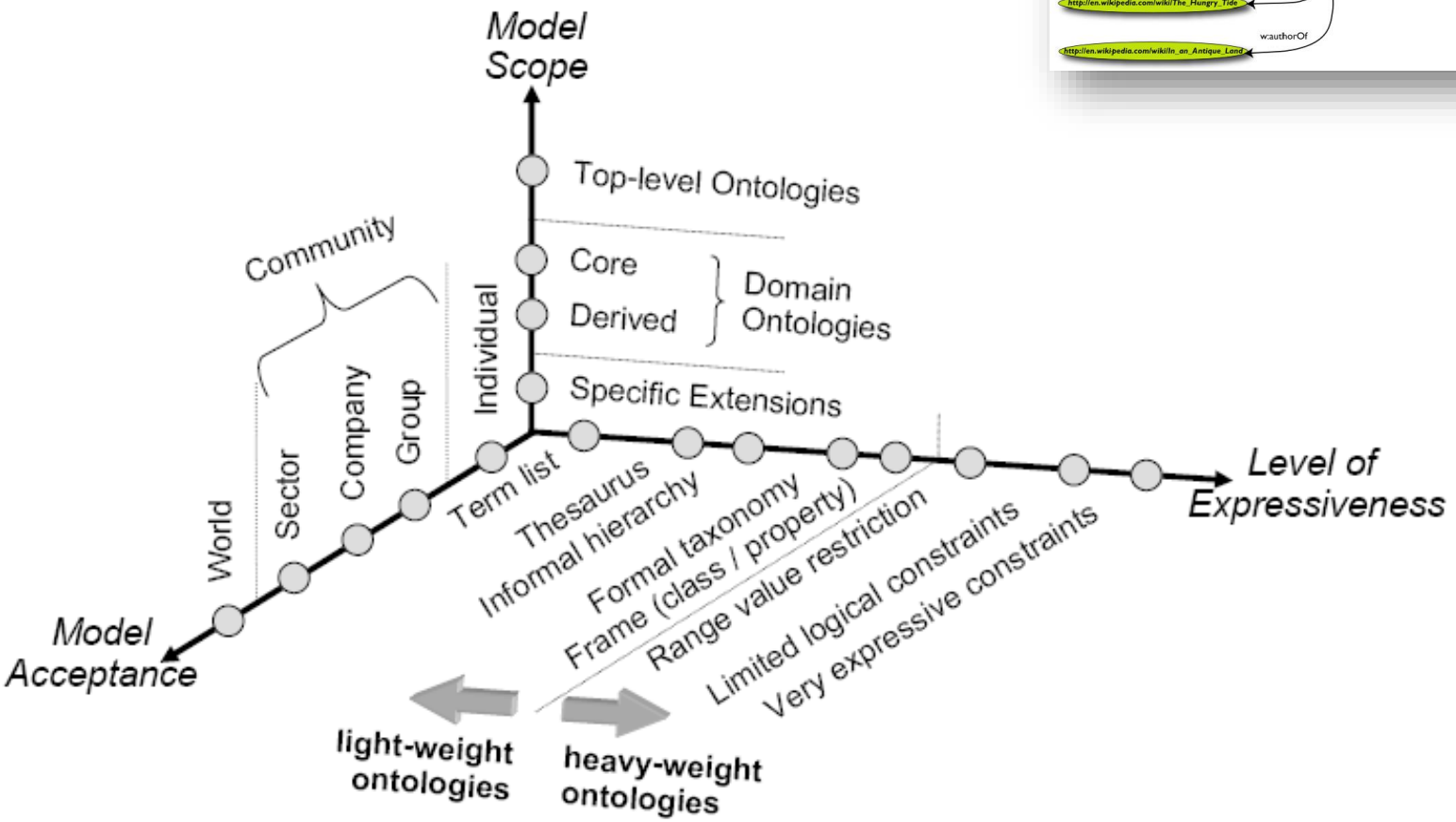




Rule Example:
If a person writes a document, then the person is familiar with the topic that is discussed with in the document.

Ontology Dimensions





© Gruber, Schaffert, Westenthaler, Salzburg Research

The Web Today …

- The current Web represents information
  - in natural language
  - graphics, multimedia, page layout

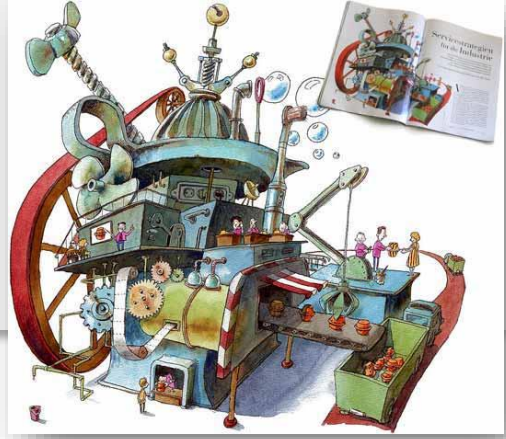
- People can easily process this information
  - People can infer facts from partial information
  - People can develop mental associations
  - People can deal with different kinds of sensory information

- The Hypertext Markup Language (HTML) is the standard format for web pages
  - Web browsers interpret HTML tags and show the HTML page accordingly.
  - Incorrect HTML statements (e.g.: unclosed tags) do not play a major role.

Tasks on the Web …

- **Usual activities often require a combination of data on the Web**
  - Hotel and travel information can come from different sites
  - Search in various digital libraries
  - etc.
- **People can easily combine this information.**
  - even if different terminologies are used

Machines are ignorant

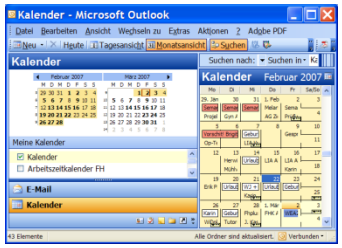


- **Incomplete information is useless.**
- **E.g.: It is difficult**
  - … to extract meaning from a figure,
  - … to infer on analogies automatically,
    … to combine information
    - Example: is <foo: creator> the same as <bar: author>?
  - How to combine different XML hierarchies?
    … etc. …

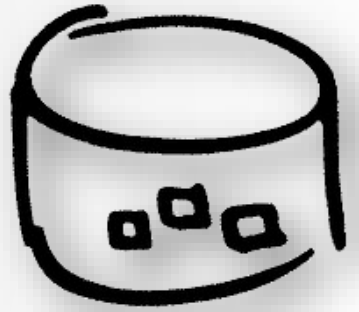
Example: Automatic flight reservation

- **Vision: An automatic Airline Reservation**
  - … knows our preferences
  - … Builds up a knowledge base from our past
  - … can combine local knowledge with remote services:
    - Airline preferences
    - Menu dietary requirements
    - Schedule integration

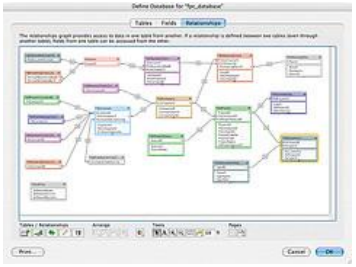
      … etc. …
- **… communicates with remote information (e.g.: in the web)**

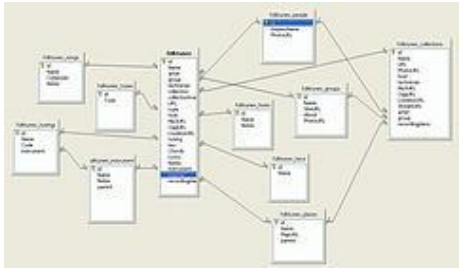


Example: System or Data(base) Integration



- **Databases of existing systems are very different in structure and content, even if they have similar functions (e.g .: ERP systems)**
  - e.g.: Integration of legacy systems by company mergers
- **Many applications require the management of various databases**
  - e.g.: combination of administrative data for e-Government
- **Many large data sets are available on the web (not necessarily public)**
  - e.g.: data from biochemical, genetic, pharmaceutical research, etc.
- **The semantics of these data(bases) should be known.**
  - How these semantics are mapped to internal structures is irrelevant.





Worldwide 2005 Vendor Revenue Estimates from RDBMS Software, Based on Total Software Revenue (Millions of Dollars)

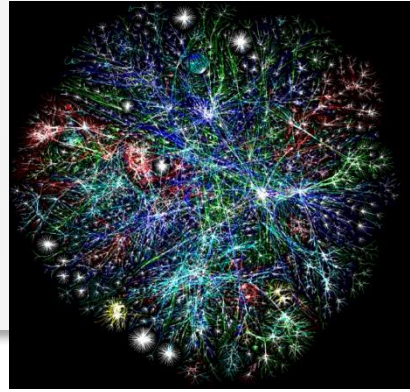
Company	2005	2005 Market Share (%)	2004	2004 Market Share (%)	2004-2005 Growth (%)
Oracle	6,721.10	48.6	6,234.10	48.9	7.8
IBM	3,040.70	22	2,860.40	22.4	6.3
Microsoft	2,073.20	15	1,777.90	13.9	16.6
Teradata	440.7	3.2	412.1	3.2	6.9
Sybase	407	2.9	382.8	3	6.3
Other Vendors	1,134.70	8.2	1,090.40	8.5	4.1
Total	**13,817.40**	**100**	**12,757.80**	**100**	**8.3**

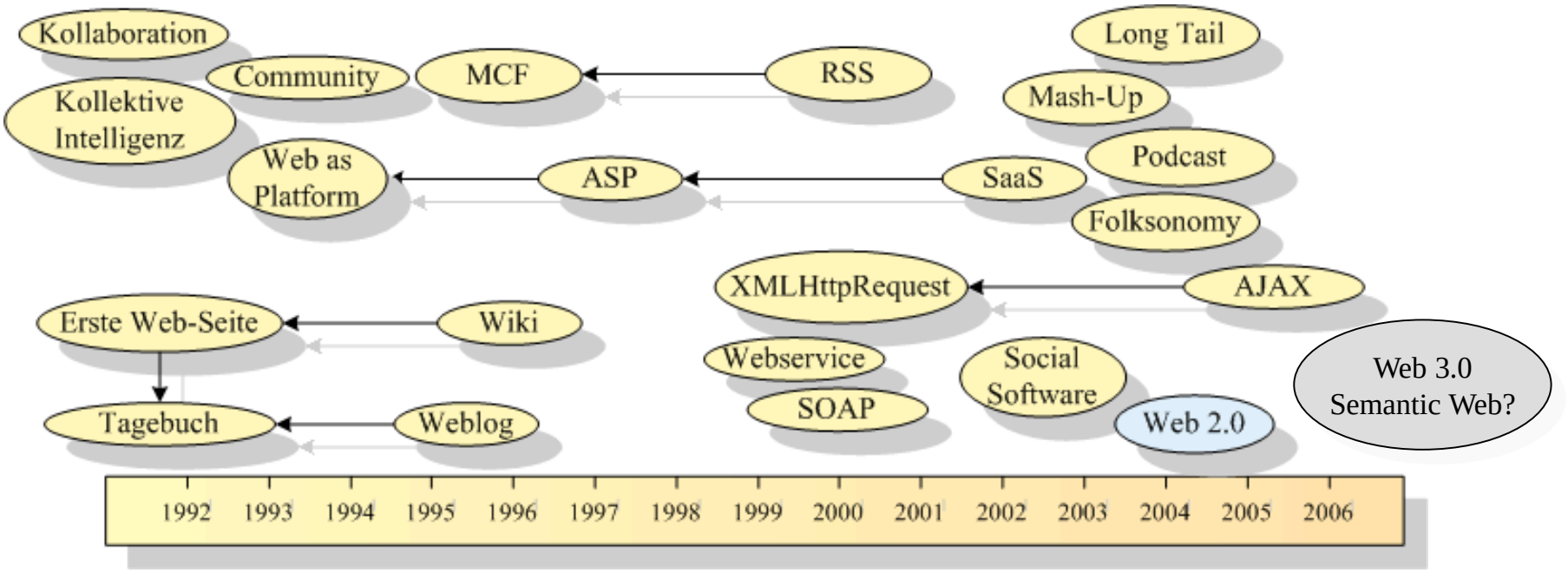
Source: Gartner Dataquest (May 2006)

Web 2.0

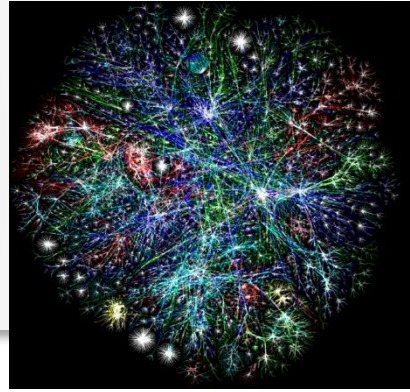
- The term Web 2.0 rather vaguely describes an altered perception and usage of the WWW.
  - **Users mutate from consumers to prosumers (producer + consumer).**
  - **Users create and/or edit the Internet Content increasingly themselves.**
  - Typical examples of technologies used are
    - **wikis, blogs, social networks, cloud computing, crowdsourcing, image and video sharing portals etc.**

Web 2.0 – Web 3.0 – Semantic Web?





Semantic Web – Definition (I)

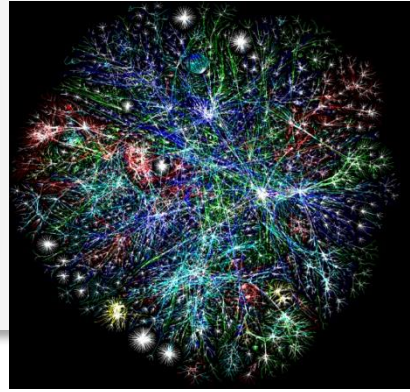


„The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.“

Tim Berners-Lee, Scientific American, May 2001

Sir Timothy John Berners-Lee ist ein britischer Informatiker, gilt als Begründer des World Wide Web und ist Leiter des W3C.

Semantic Web – Definition (II)



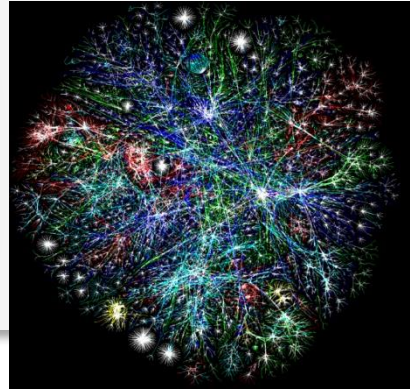
„The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.

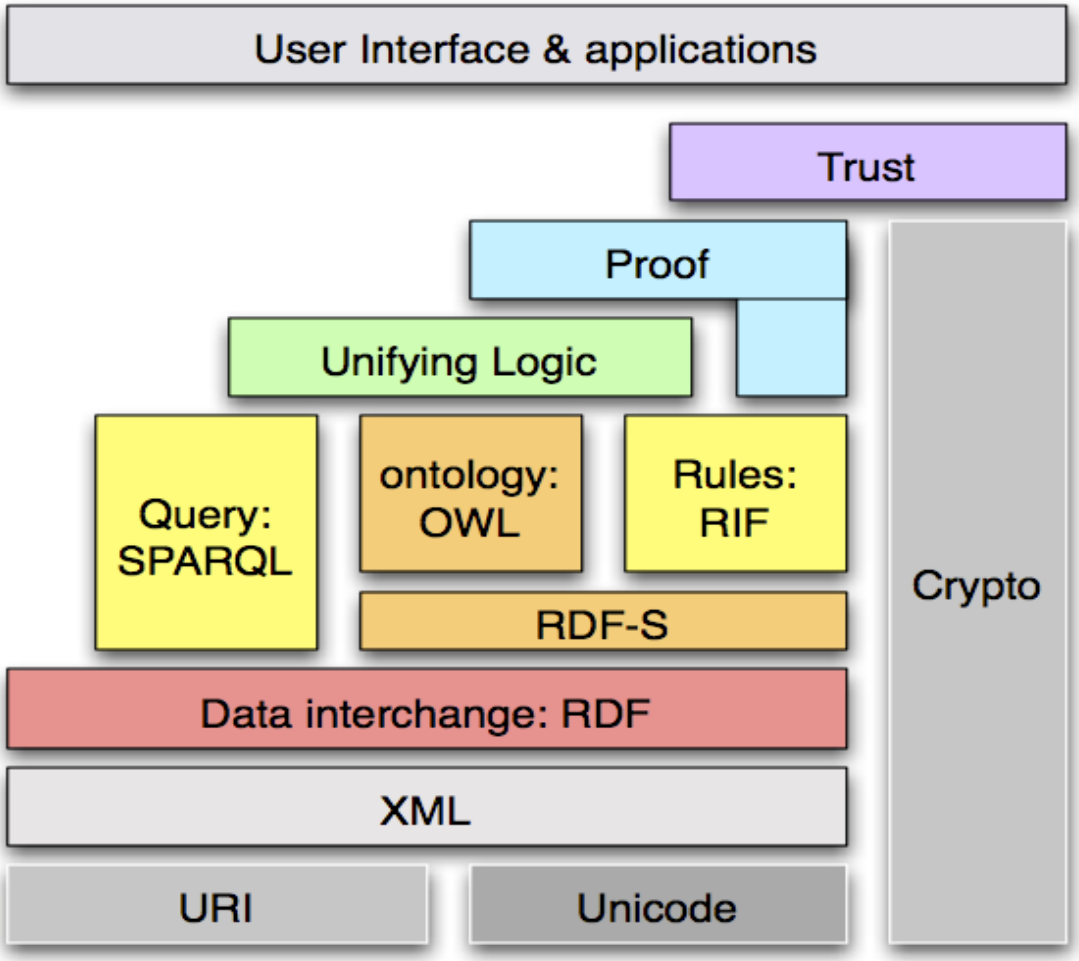
It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners.

It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming.“

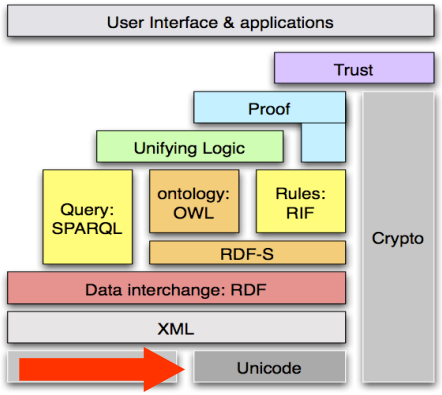
World Wide Web Consortium (W3C)

Semantic Web Stack des W3C



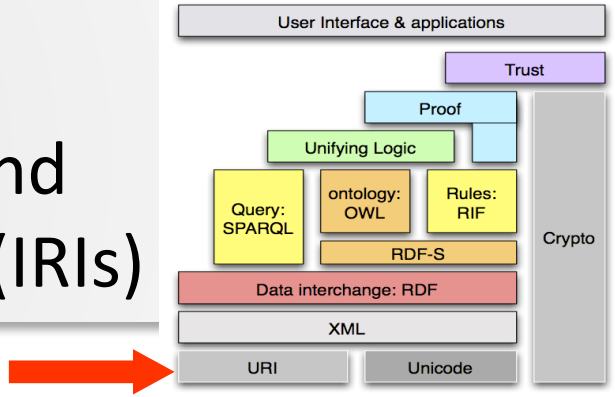


Character Set: Unicode



- **Basic coding standard for data** in the Semantic Web
- International character set, standard for the unamigous identificatoin of characters
- Combines different character sets (Arabic, Japanese, Cyrillic, Latin)
- URL: http://www.unicode.org/

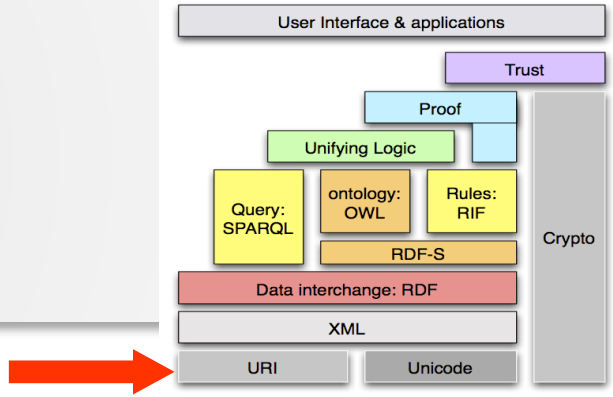
Uniform Ressource Identifiers (URIs) and Internationalized Resource Identifiers (IRIs)



- Unambigous identification of resources
- Only ASCII characters
- Syntax according to schemes, e.g.: http, ftp, mailto, pop
- *„URIs have global scope. Associating an URI with a resource means that anyone can link to it, refer to it, or retrieve a representation of it.“*
- **URIs are the standard for the identification of (online) resources**
  - Specified in RFC 3986 by the W3C
  - Generalization of the Uniform Resource Locators (URL, default of web addresses) and Uniform Resource Names (URN, identifier for a resource by a name in a namespace; e.g.: ISBN)
- **Resources are objects** (or things) one wants to say something about.
  - In the Semantic Web, they are identified by URIs
  - Examples: publisher, book, author, place, person, museum, campus
- URL: http://www.w3.org/Addressing/

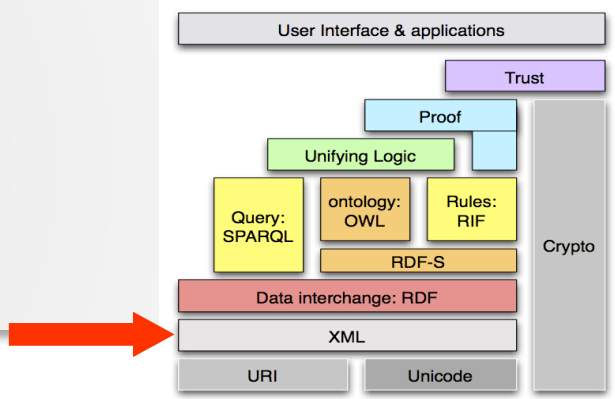
- **While URIs are limited to a subset of the ASCII character set, IRIs may contain characters from the Universal Character Set (UNICODE).**

Examples
Unicode & URIs



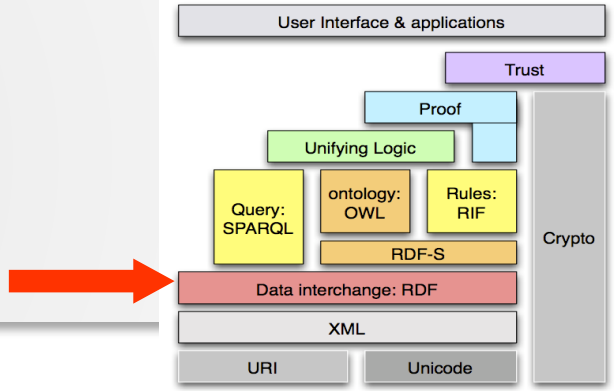
http://de.wikipedia.org/wiki/URI
ftp://ftp.is.co.za/rfc/rfc1808.txt
file://C:\UserName.HostName\Projects\Wikipedia_Articles\URI.xml
ldap://[2001:db8::7]/c=GB?objectClass?one
gopher://gopher.floodgap.com
mailto:John.Doe@example.com
sip:911@pbx.mycompany.com
news:comp.infosystems.www.servers.unix
data:text/plain;charset=iso-8859-7,%be%fg%be
tel:+1-816-555-1212
telnet://192.0.2.16:80/
urn:oasis:names:specification:docbook:dtd:xml:4.1.2
doi:10.1007/s003390201377

Syntax:
XML and Namespaces



- **Extended Markup Language**
  - Metalanguage AND markup language
    - Meta: Description of structure
    - Markup: Description of Information
  - **But: No specification of semantics**
  - **XML is the syntactic basis.**
  - It allows to store structured data.
  - Tags (Opening & Closing) structure content; Parsers permit the reading of well-formed XML files
  - Modeling languages allow you to set the structure (XML Schema). An XML document that follows a schema is valid.
- **XML Namespaces**
  - Simple method for unambiguous determination of element and attribute names in XML documents
  - **Namespaces allow the use of the same tag name in different contexts**
  - Identification by URI references (**URI** + # + **fragment identifier**)
  - Shortening by prefixes
  - Default Namespaces are possible

Ressource Description Framework (RDF)

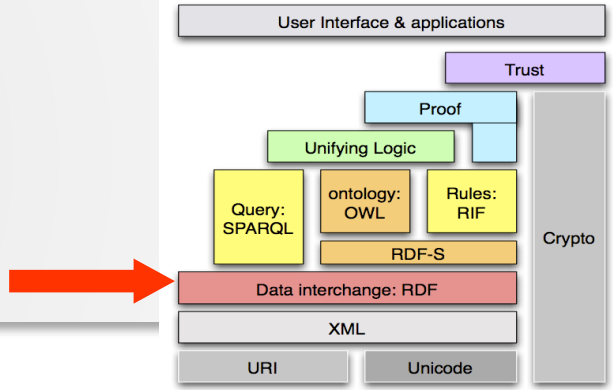


„RDF assigns specific Universal Resource Identifiers (URIs) to its individual fields.“

- While XML provides a mechanism to structure data, **RDF provides a mechanism to assert something about data** (metadata, data about data).
- RDF is not a language definition as XML but a "model" **(keyword RDF graph)**
- **RDF is machine processable.** Metadata can be exchanged without loss of meaning between systems.
- **RDF is defined by a set of triplets.** A triplet consists of subject, predicate and object (Statement)

```
Daniel Wilfing -- is student of --  BMI
   (subject)        (predicate)     (object)
```

Ressource Description Framework (RDF)

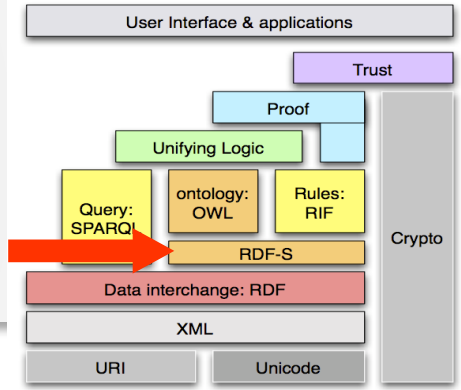


- **Resources are identified by URIs**
- **Predicates are defined by the RDF vocabulary**
  - (Dublin Core, FOAF, ...)
- **Namespaces shorten URIs**

- **RDF / XML is a representation of RDF in XML:**

```
<rdf:Description about='http://…/Why-RDF.html'>
     <Author>Tim Bray</Author>
     <Home-Page rdf:resource='http://…'/>
 </rdf:Description>
```

Taxonomies: RDF Schema

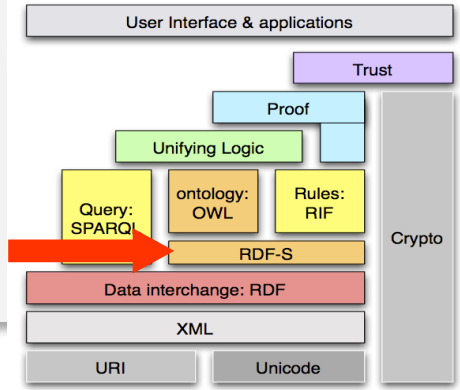


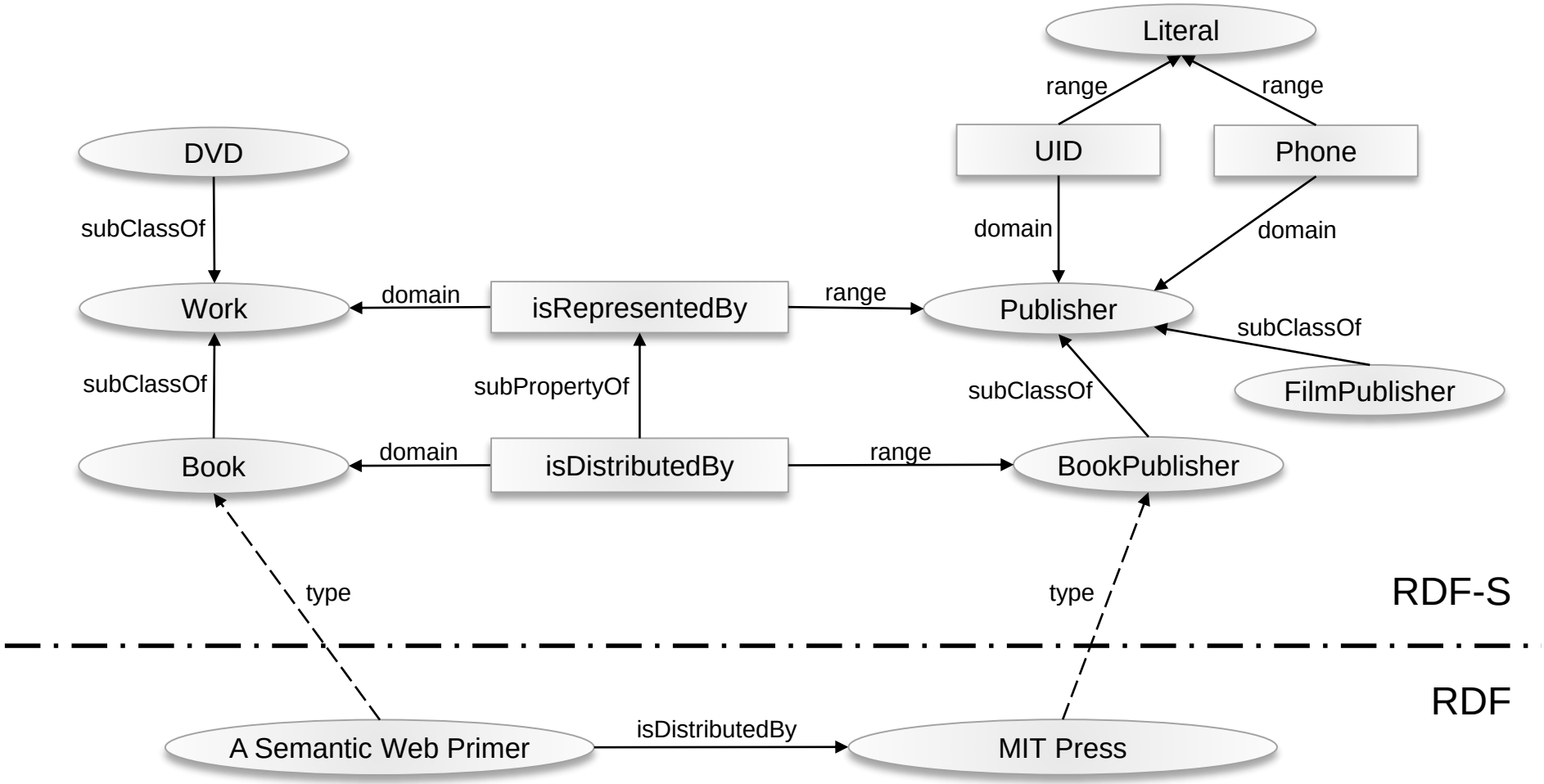
- **RDF Schema (RDFS) is a simple ontology language.**

  It allows the **creation of vocabularies**.

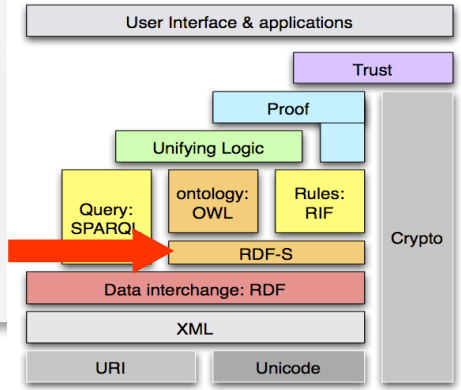
  - Definition of classes and properties
  - Domain and Range of properties
  - (Multiple) inheritance of classes and properties
  - Containers (Bag, Sequence,Alternative)

Example
RDF and RDF Schema



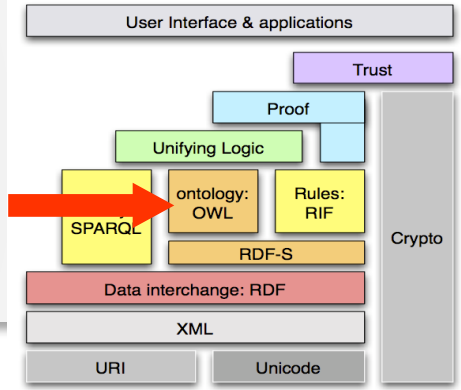


RDF-S Benefits?
E.g.: Inheritance



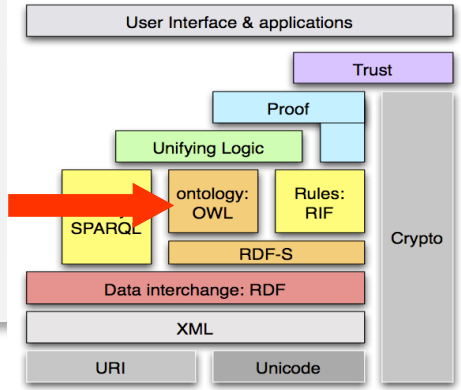
- Meaningless statements can be prevented by the use of classes
  - **`Franz Kafka -- is writen by -- Hermann Hesse`**
    - We can express that books may only be written by authors **(range restriction)**
    - Hermann Hesse is an author
    - He inherits the ability to write books from the class "authors“
    - This is done in RDF Schema by expressing the semantic relationship "is a subclass of“
    - The interpretation of "is a subclass of" by an RDF application is therefore unabigous.
  - **`Springer Verlag -- is written by -- Hermann Hesse`**
    - Only books can be written (no publishers!)
    - Limitations of those objects a property is allowed to be applied. **(domain restriction)**

Web Ontology Language



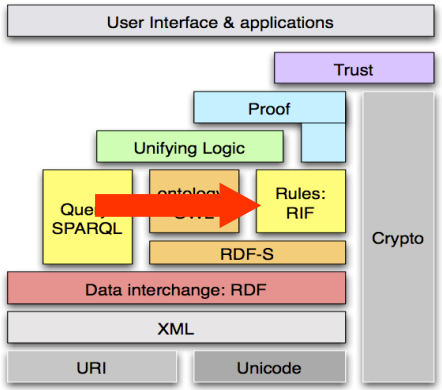
- The Ontology Layer is **used to build domain ontologies**.
- **OWL is based on RDF.** It has a higher expressiveness than RDF Schema and allows complex relationships between objects
  - Disjunction and boolean combination of classes
  - Specific characteristics of properties like functional, inverse functional, symmetric, transitive, cardinality
  - Based on predicate logic
  - etc.
- OWL is recommended by W3C
- URL: https://www.w3.org/TR/owl2-overview/

Ontologies in OWL



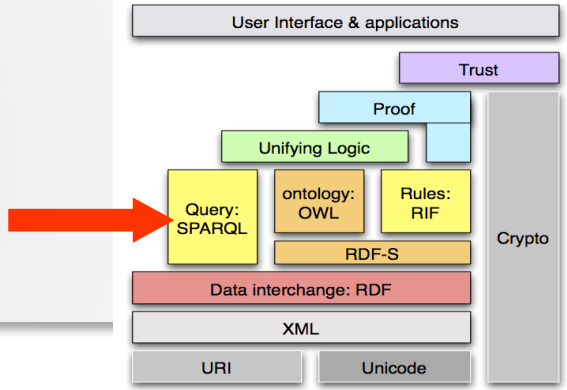
- **Ontologies describe a segment of the world from the perspective of a particular domain.**
- **Things = Concepts**
  - Classes (type description)
  - Individuals (specific instances)
- **Concepts have properties** or relationships.
- Properties and relationships follow **restrictions**.

Rules in SWRL und RIF



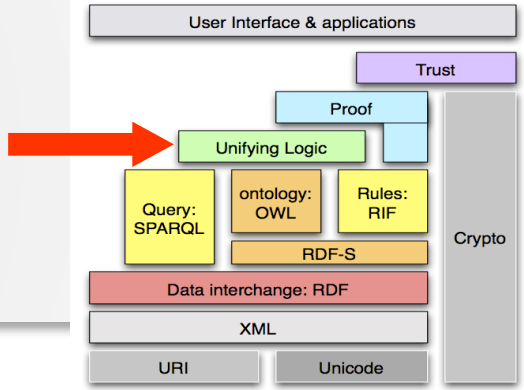
- **Rules** are executable parts of declarative knowledge.
  - Necessary to manage complex, dynamic operations.
- **Semantic Web Rule Language (SWRL)**
  - Combining OWL and Rule Markup Language (RuleML)
  - W3C Recommendation
- **Rule Interchange Format (RIF)**
  - Definition of a core rule language including extensions to enable translations between different rule languages.

Queries: SPARQL



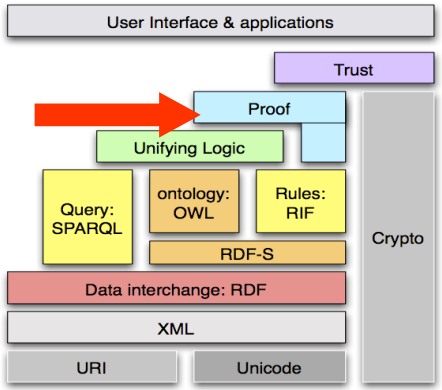
- Recursive acronym for **SPARQL Protocol and RDF Query Language**
  - SQL-like query language for RDF data
  - able **to retrieve and manipulate** data stored in Resource Description Framework (RDF)
- Three specifications
  - Query language
  - Presentation of results (XML format)
  - Access protocol (used WSDL 2.0)
- Allows linked queries using multiple RDF graphs
- Integration into various frameworks, e.g.: JENA (Java framework for SemWeb applications)
- W3C recommendation

Logic Framework



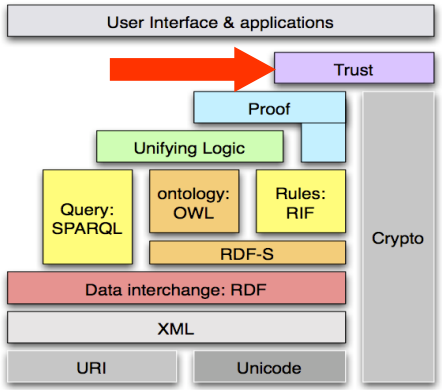
- An **inference engine** shall
  - **draw conclusions** out of rules and information of an ontology and
  - **derive new knowledge** from already existing knowledge.
    - Facts stored in ontologies serve as a basis for logical conclusions.
  - Logic is foundation for knowledge representation languages.
    - (Description) Logic is partially included in OWL (DL, Full)
    - Formal semantics assigns an unambiguous meaning to logical statements assigns
    - It has to define syntax AND semantics.

Proof



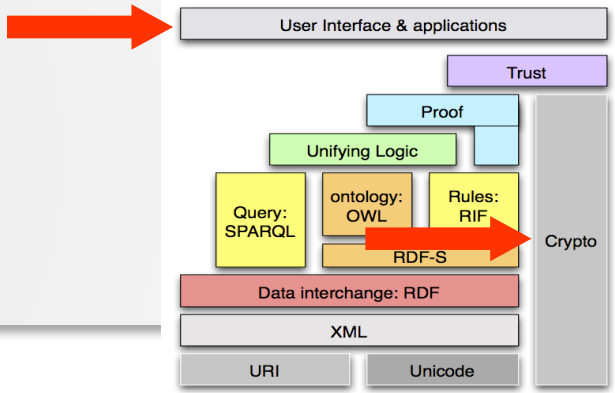
- **Determination of the validity of statements**
  - For a certain assertion an "Heuristic Engine“ shall search the Web for rules and ontologies as long an assertion can be confirmed or rejected.
  - Check the credibility (trust) by inference depending on the context
  - Critical for Semantic Web applications, still poorly researched
- **The Logic level applies rules and draws conclusions**
- **However, it is not sure if proof can be provided**

Trust



- **In the Semantic Web it is a requirement to check the validity of the information.**
  - This requires trust policies and authentication mechanisms
- **Confidence in Semantic Web applications**
  - Trustworthiness of the source
  - Authenticity of the source
  - Context dependency
- **Digital Signatures**
  - Electronic signature are used to identify sources and RDF statements unambiguously.
- **Web of Trust**
  - It is important to ensure and verify that semantic web statements are coming from trusted source.
  - With a few transitive steps a trust relationship between any two agents should be established

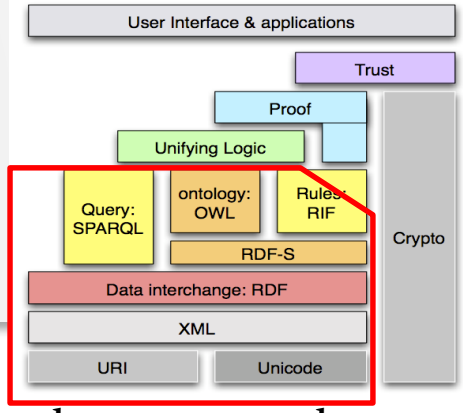
Cryptography and Applications



- **Cryptography**
  - Effective technologies are needed to ensure data security
  - XML Encryption Syntax and Processing (XMLEnc) is recommended by W3C

- **User Interface & Applications**
  - The final layer enables humans to use semantic web applications.
  - At present numerous frameworks, computer aided knowledge and software engineering tools (CAKE and CASE Tools), triple stores and inference engines (reasoners) are available.

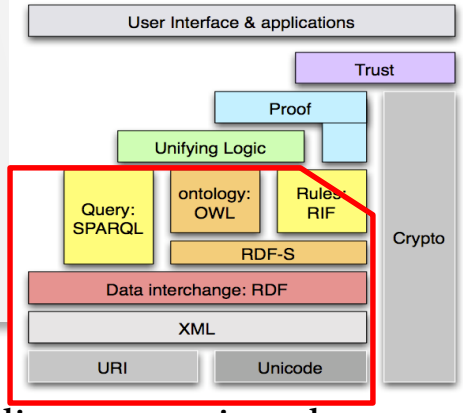
Semantic Web Today - Standardized Technologies



The bottom layers contain technologies that are well known from hypertext web and that without change provide basis for the semantic web.

- **Internationalized Resource Identifier (IRI), generalization of URI**, provides means for uniquely identifying semantic web resources. Semantic Web needs unique identification to allow provable manipulation with resources in the top layers.
- **Unicode** serves to represent and manipulate text in many languages. Semantic Web should also help to bridge documents in different human languages, so it should be able to represent them.
- **XML** is a markup language that enables creation of documents composed of structured data. Semantic web gives meaning (semantics) to structured data.
- **XML Namespaces** provides a way to use markups from more sources. Semantic Web is about connecting data together, and so it is needed to refer more sources in one document.

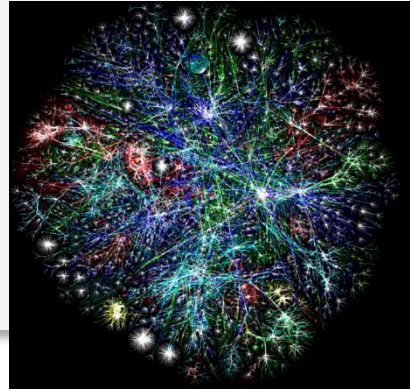
Semantic Web Today - Standardized Technologies



Middle layers contain technologies standardized by W3C to enable building semantic web applications.

- **Resource Description Framework (RDF)** is a framework for creating statements in a form of so-called triples. It enables to represent information about resources in the form of graph - the semantic web is sometimes called Giant Global Graph.
- **RDF Schema (RDFS)** provides basic vocabulary for RDF. Using RDFS it is for example possible to create hierarchies of classes and properties.
- **Web Ontology Language (OWL)** extends RDFS by adding more advanced constructs to describe semantics of RDF statements. It allows stating additional constraints, such as for example cardinality, restrictions of values, or characteristics of properties such as transitivity. It is based on description logic and so brings reasoning power to the semantic web.
- **SPARQL** is a RDF query language - it can be used to query any RDF-based data (i.e., including statements involving RDFS and OWL). Querying language is necessary to retrieve information for semantic web applications.
- **RIF** is a rule interchange format. It is important, for example, to allow describing relations that cannot be directly described using description logic used in OWL.

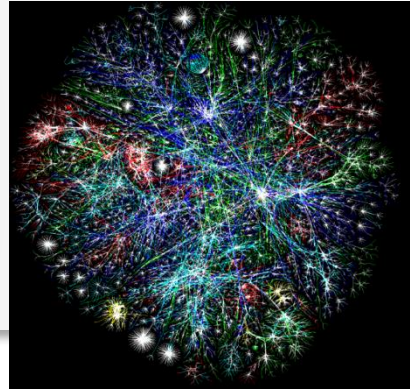
So what is the Semantic Web?



The following slides contain excerpts of

http://www.w3.org/2006/Talks/0927-Berlin-IH/

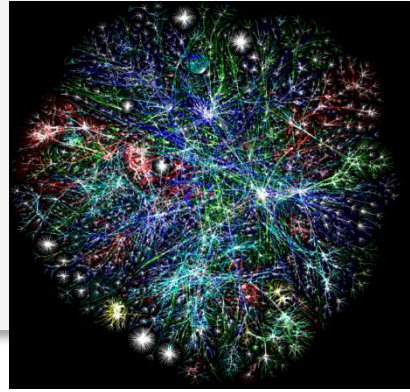
Is the Semantic Web AI on the Web?



NO!

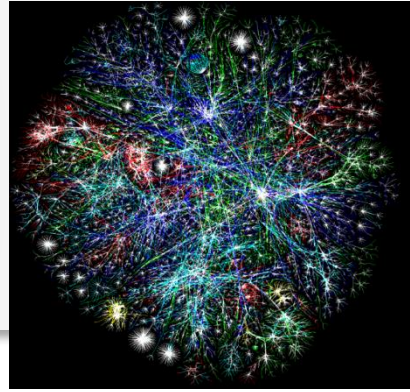
- Humans can easily "connect the dots" when browsing the Web but machines can't!
- The goal is to have a **Web of Data** to ensure smooth integration with data, too.
- It allows machines to "connect the dots".
- It provides a **common framework** to share data on the Web across application boundaries.
- And what is the relationship to AI?

And what is the relationship to AI?



- Some technologies in the Semantic Web has benefited from AI research and development.
- Semantic Web has also brought some new concerns, problems, use cases to AI.
- But AI has many many different problems that are not related to the Web at all (e.g.: image understanding).

What is RDF then?



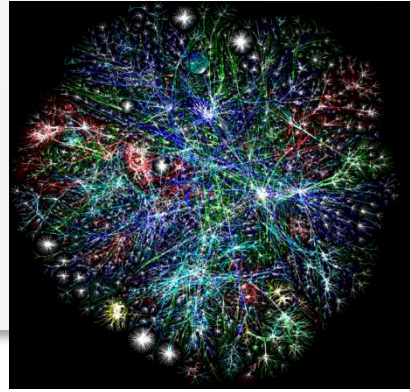
- For all applications listed above the issues are to **create relations among resources on the Web** and to **interchange those data**.

- Pretty much like (hyper)links on the traditional web, except that:
  - There is no notion of "current" document; ie, relationship is between any two resources.
  - A relationship must have a name: a link to my CV should be differentiated from a link to my calendar.
  - There is no attached user-interface action like for a hyperlink.

- RDF is a model for such relationships and interchange
  - to be a bit more techie: it is a model of **`(s p o)` triplets** with **`p`** naming the relationship between **`s`** and **`o`**.

- URIs (IRIs) are used as universal naming tools, including for properties.

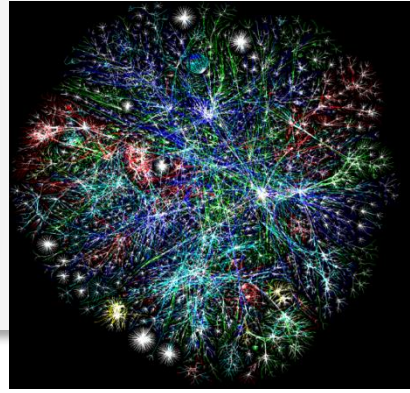
But isn't RDF simply an (ugly) XML application?

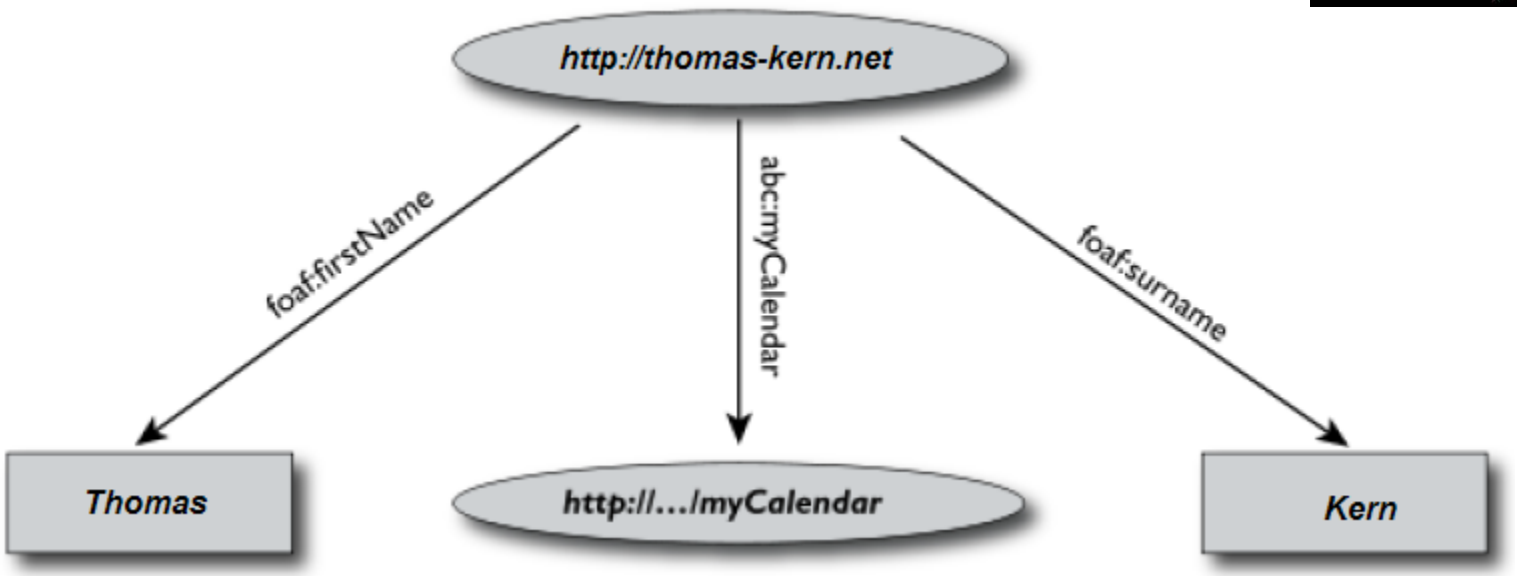


NO!

- **RDF is a graph representing a set of relationships.**
- An `(s,p,o)` triple can be viewed as a labeled edge in a graph
  - i.e., a set of RDF statements is a directed, labeled graph
    - the nodes represent the resources that are bound
    - the labeled edges are the relationships with their names
- This set must be serialized for machines; this can be done into XML (using RDF/XML), or to other formats (Turtle, N-Triples, TriX, …)
- Think in terms of graphs, the rest is syntactic sugar!

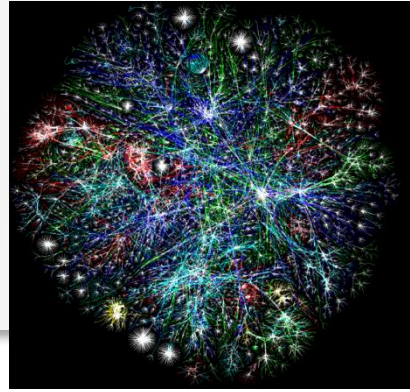
A Simple RDF Example






```
<rdf:Description rdf:about="http://www.thomas-kern.net">
  <foaf:firstName>Thomas</foaf:firstName>
  <abc:myCalendar rdf:resource="http://…/myCalendar"/>
  <foaf:surname>Kern</foaf:surname>
</rdf:Description>
```

RDF and data integration?



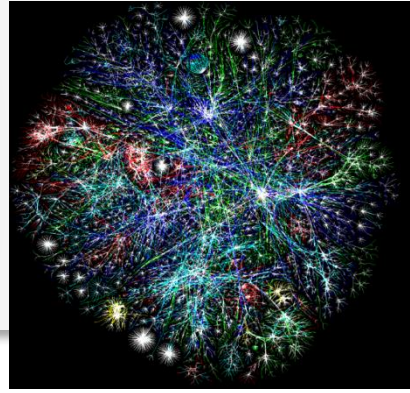
- Consider this (simplified) bookstore data set

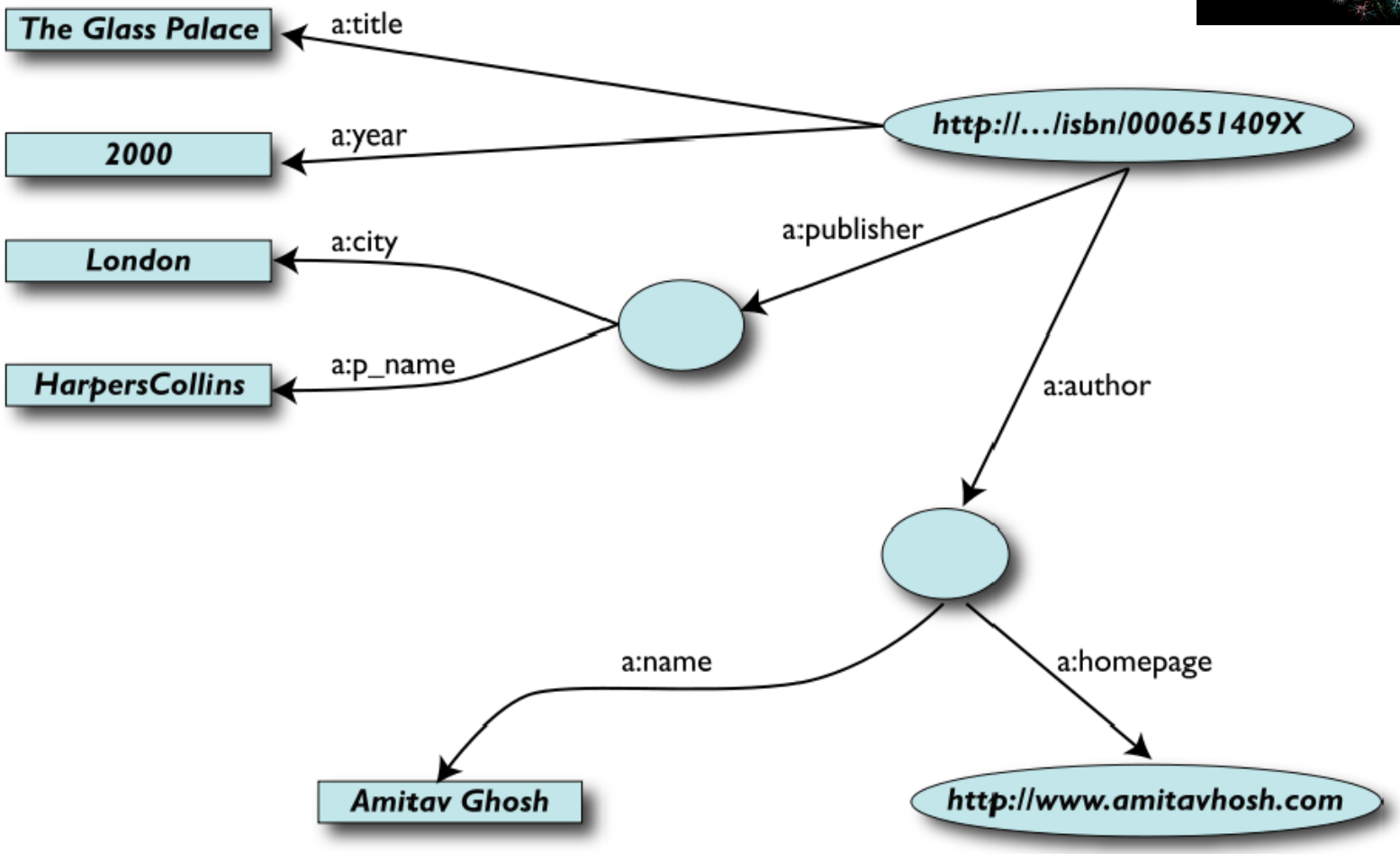
ID	Author	Title	Publisher	Year
ISBN 0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

ID	Name	Home page
id_xyz	Amitav Ghosh	http://www.amitavghosh.com/

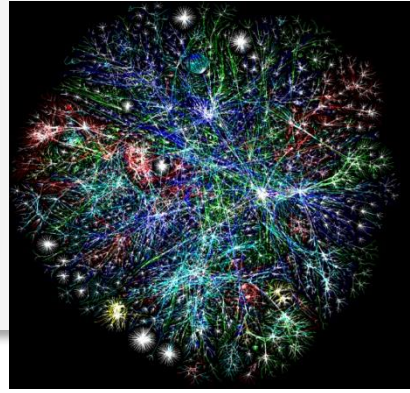
ID	Publisher Name	City
id_qpr	Harper Collins	London

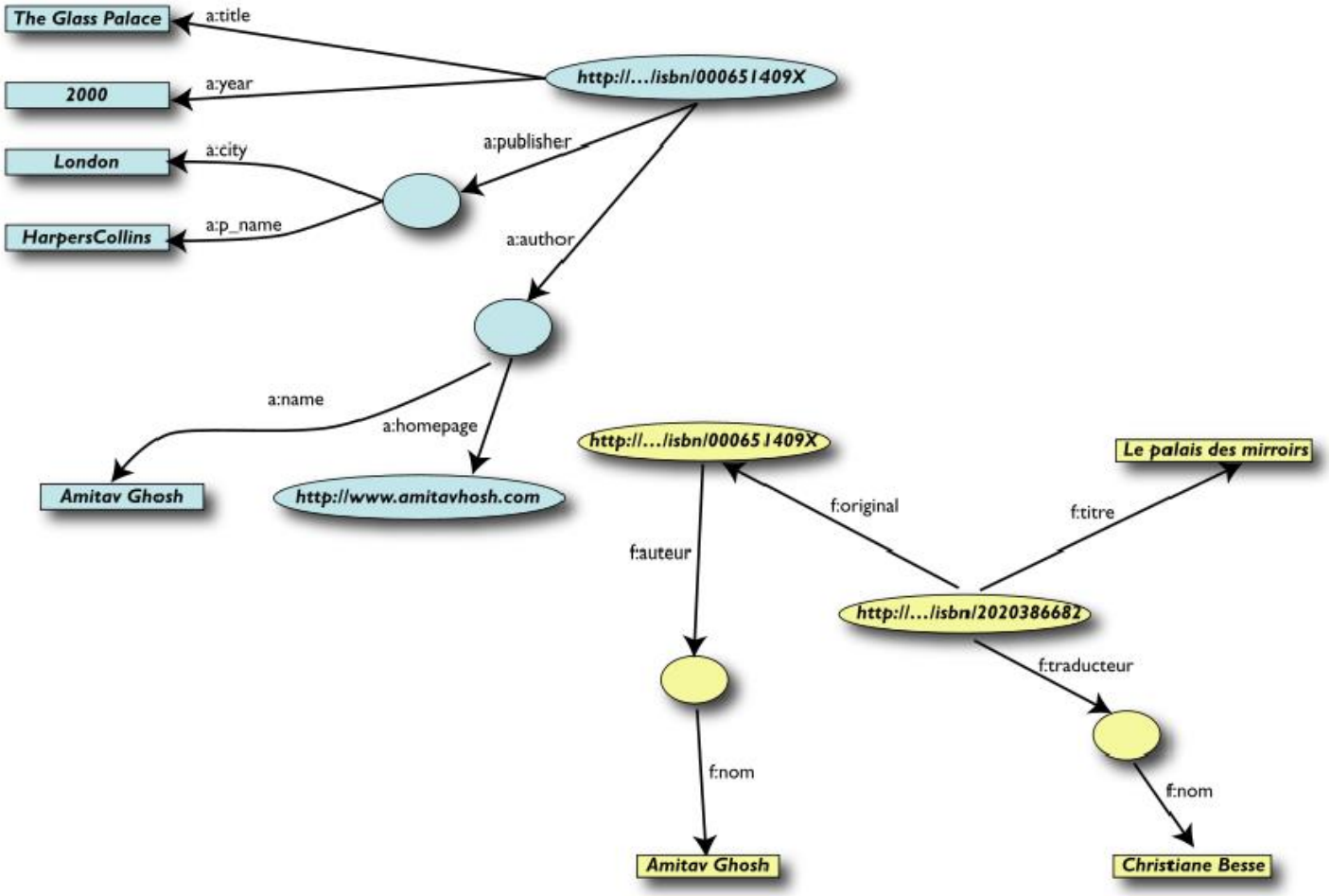
Export your data as a set of relations (in RDF) …



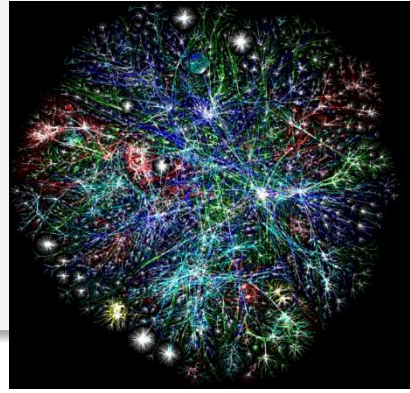


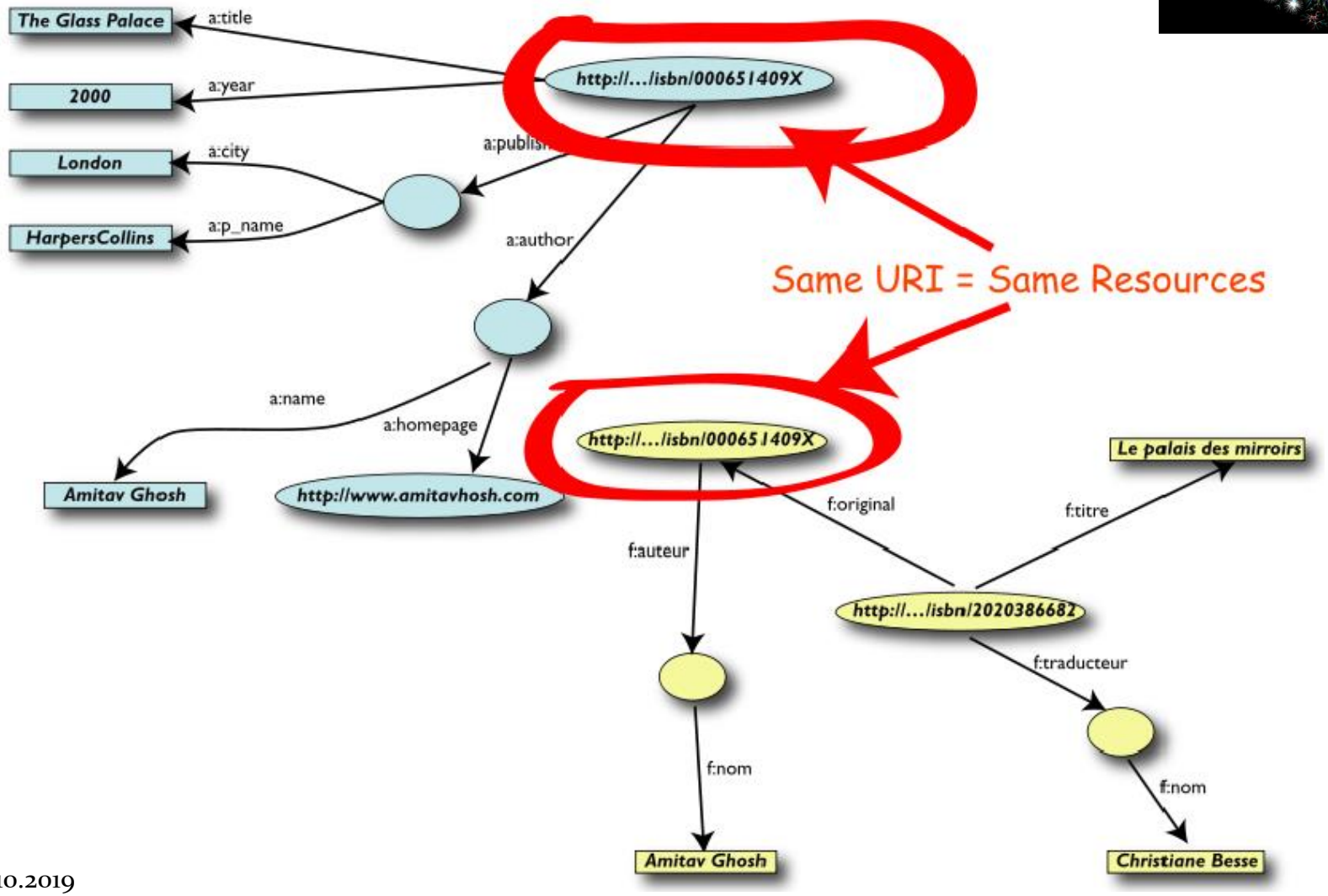
Add the data from another publisher…



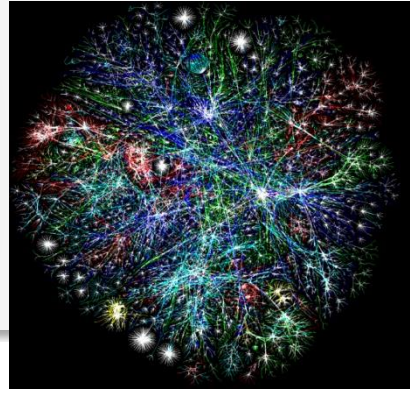


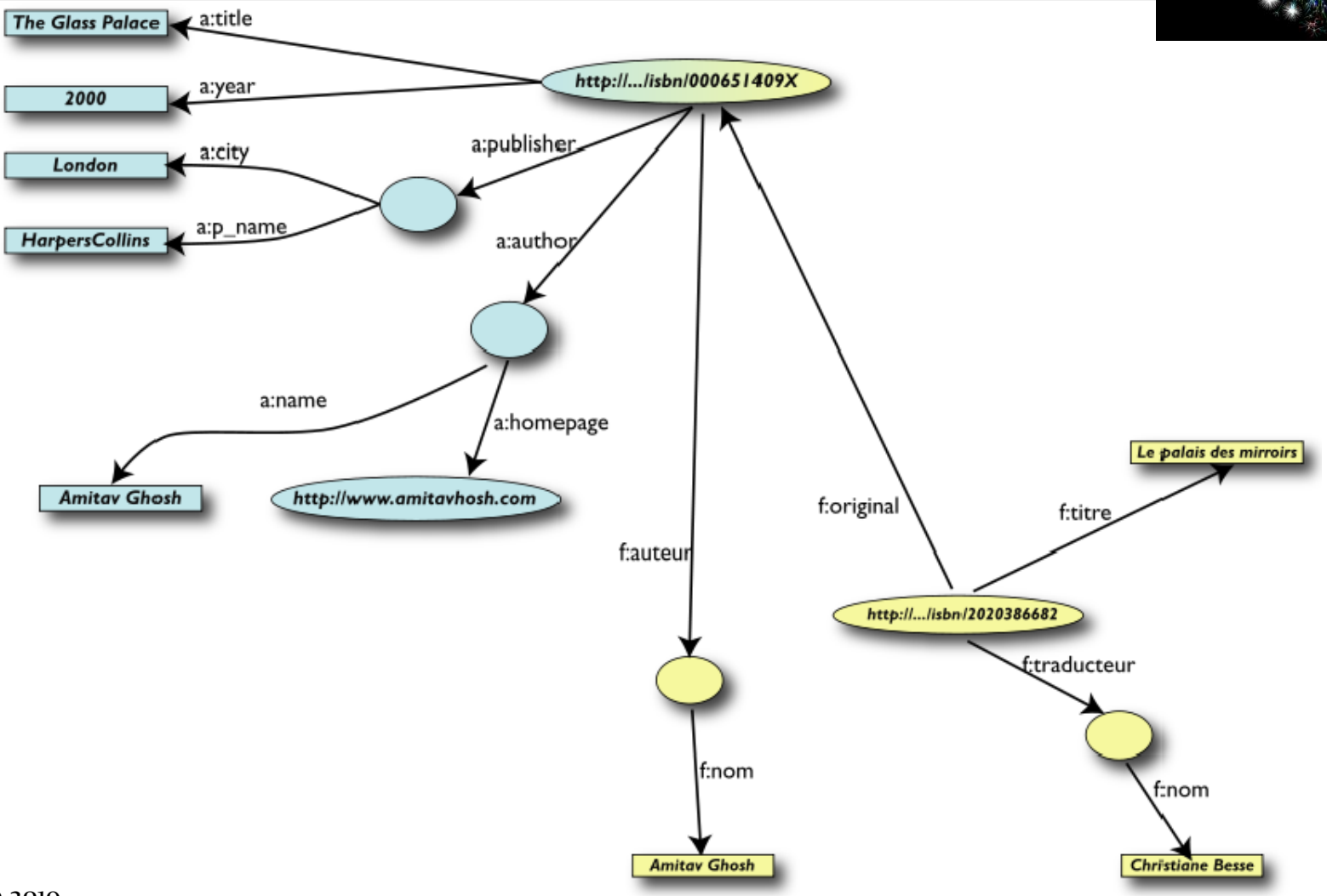
Start merging…



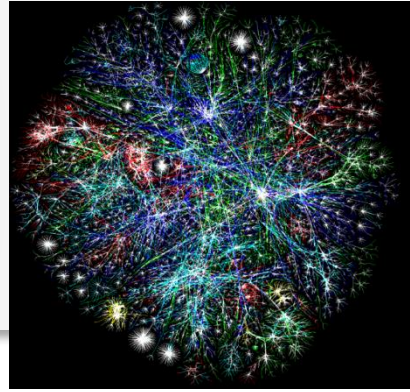


Simple integration…



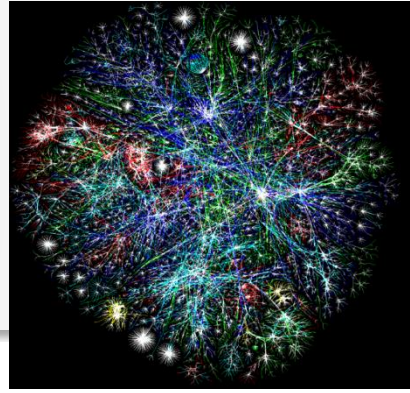


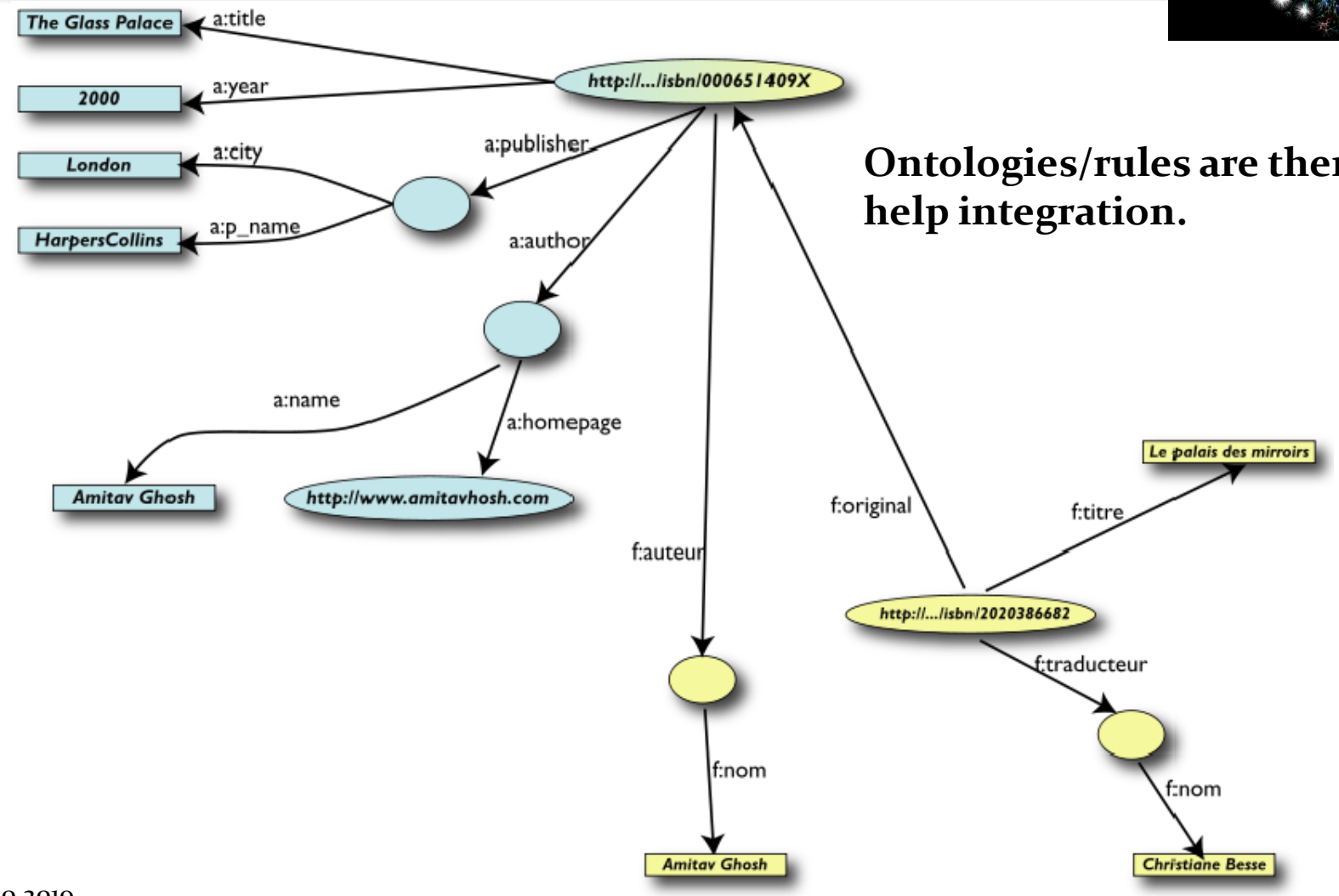
Note the role of URIs (IRIs)!



- The URIs made the merge possible
- URIs ground RDF into the Web
- URIs make this the Semantic Web

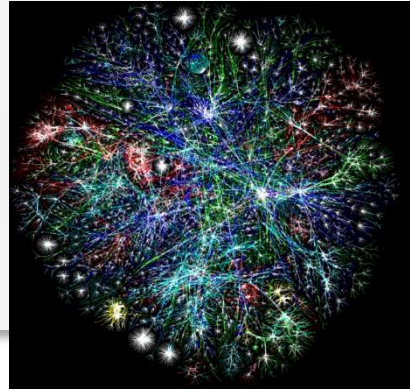
So what is then the role of ontologies and/or rules?





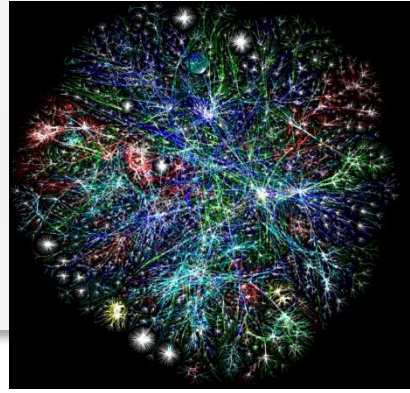
Ontologies/rules are there to help integration.

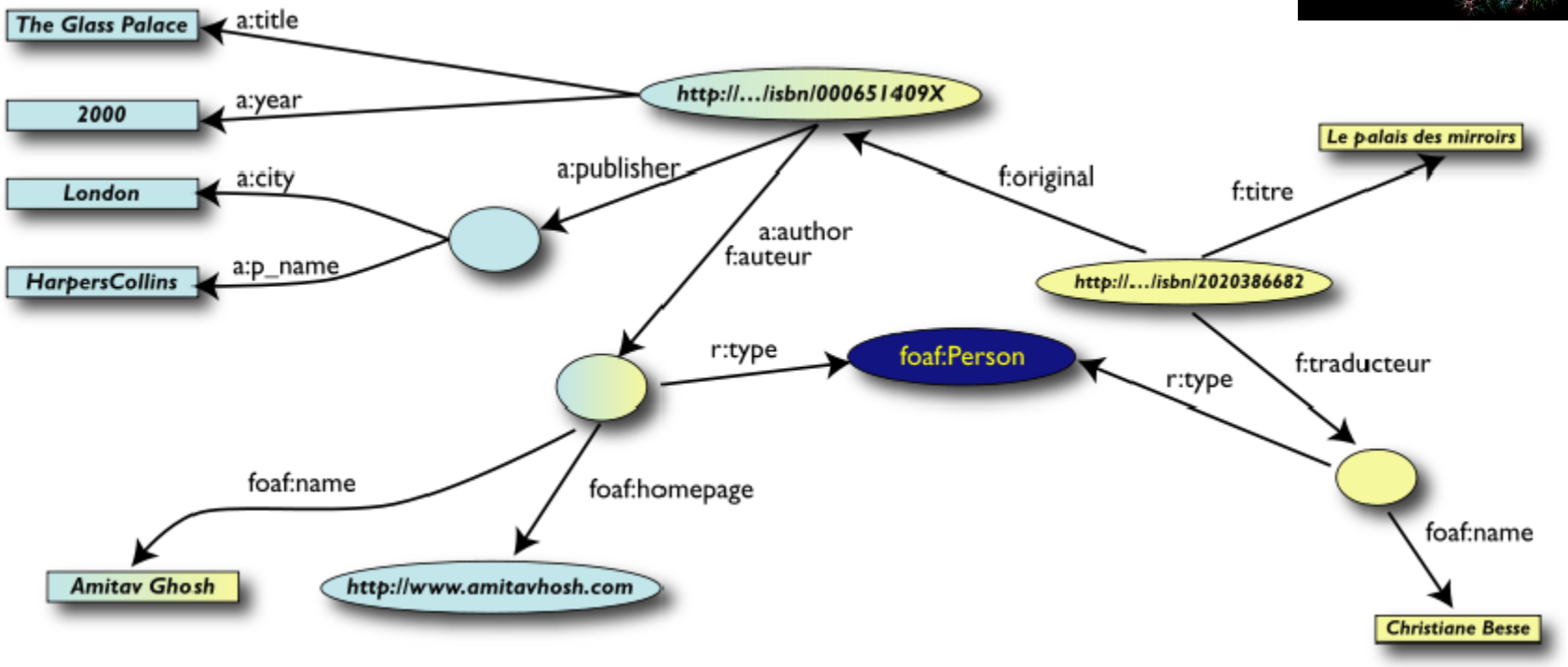
Our merge is not complete yet…



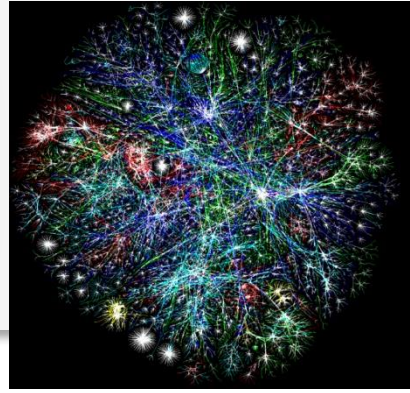
- We “feel” that a:author and f:auteur should be the same. But an automatic merge does not know that!
- Let us add some extra information to the merged data:
  - `a:author` same as `a:author`
  - both identify a “Person”:
    - A term that a community has already defined (part of the “FOAF” terminology).
    - A “Person” is uniquely identified by his/her name and, say, homepage.
    - It can be used as a “category” for certain type of resources
    - we can also identify, say, `a:name` with `foaf:name`.
- These statements can be described in an ontology (or, alternati
- vely, with rules).
- The ontology/rule serves as some sort of a “glue”.

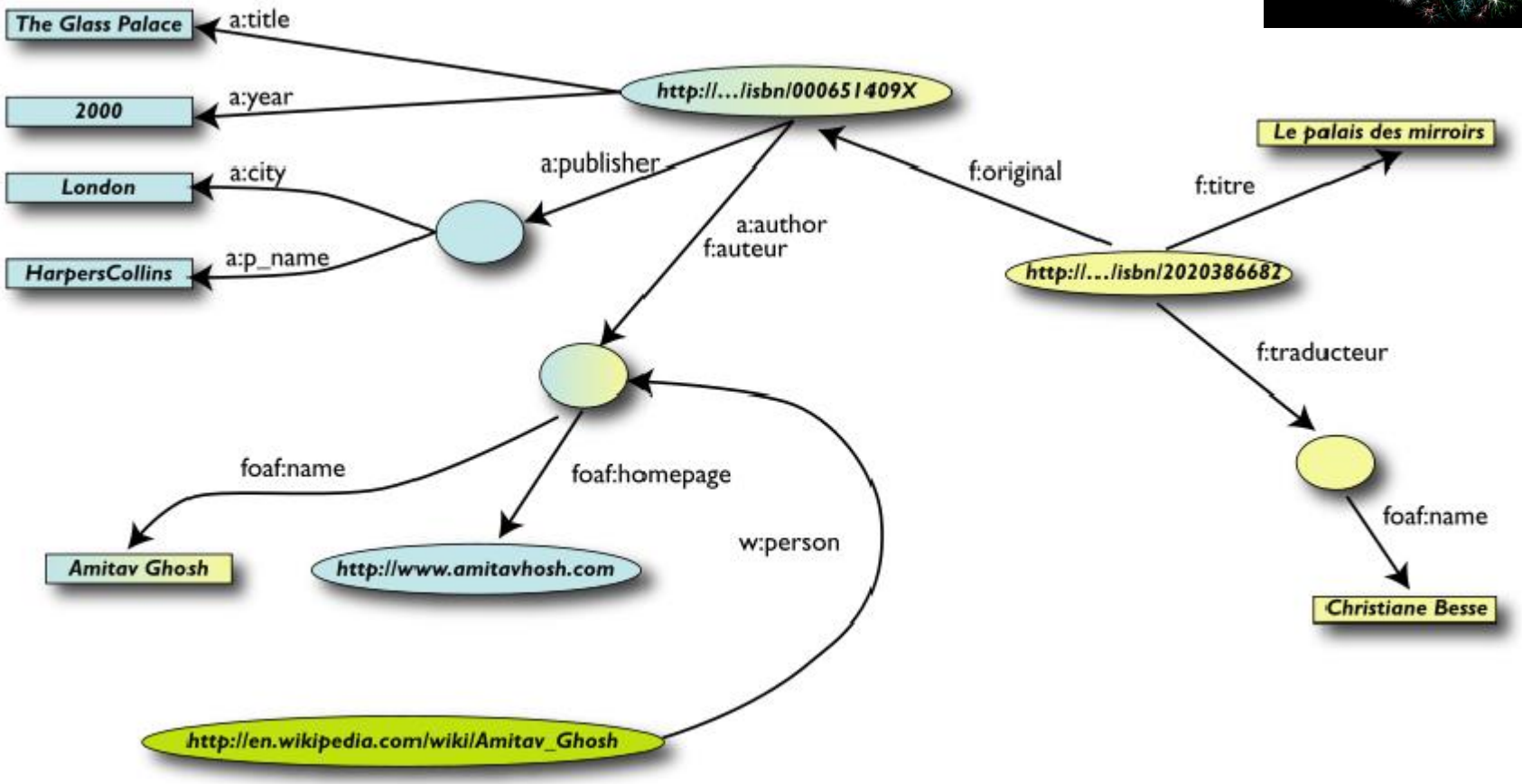
Better merge: richer queries are possible!



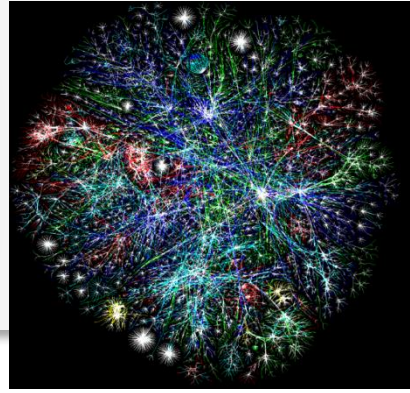


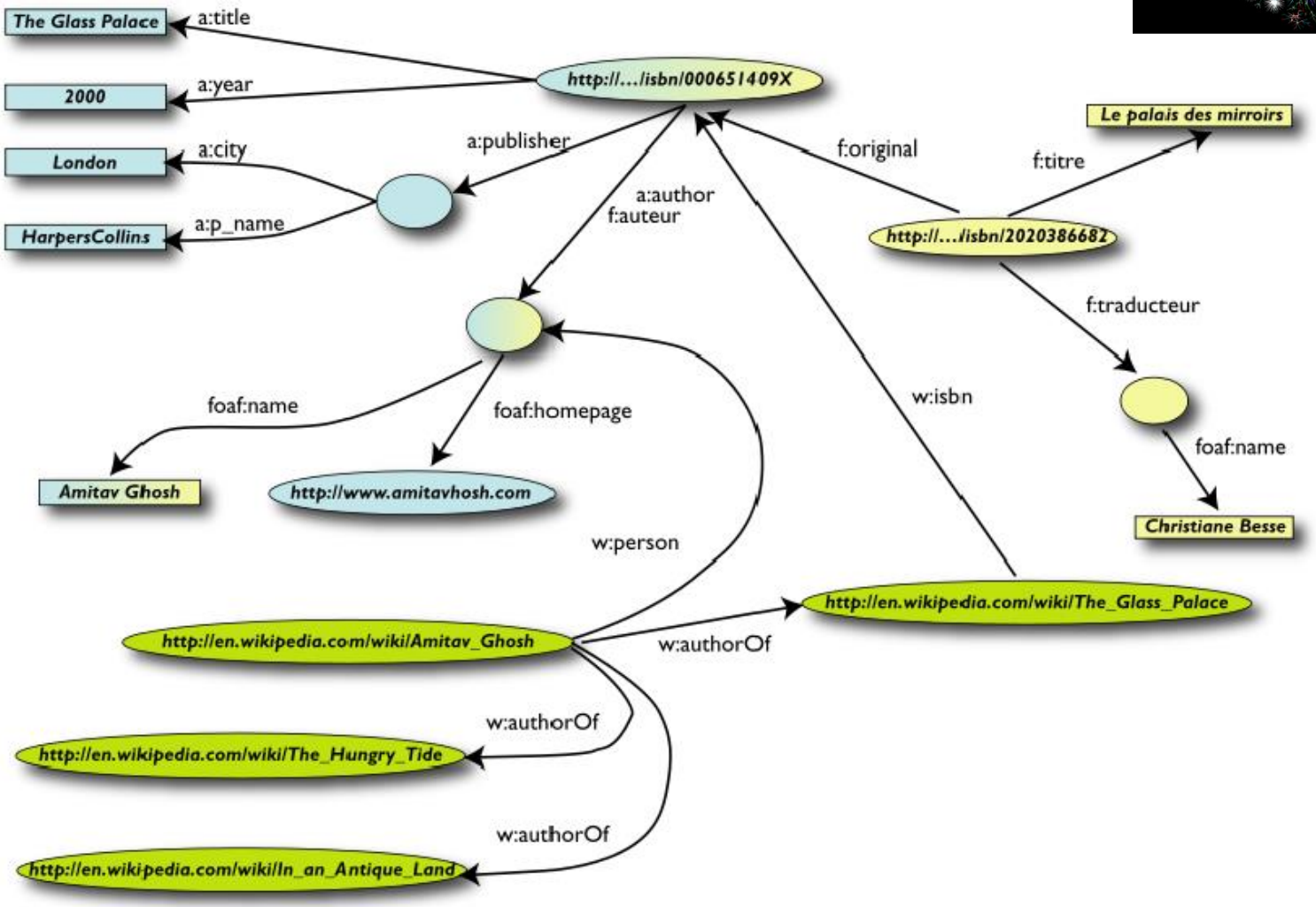
And then the merge may go on...



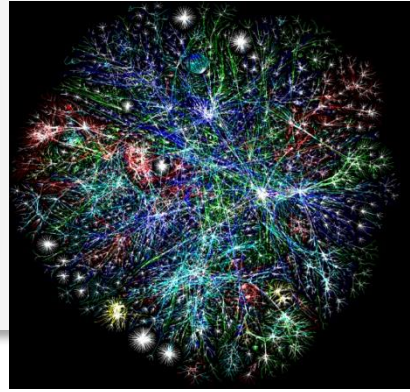


…and on…



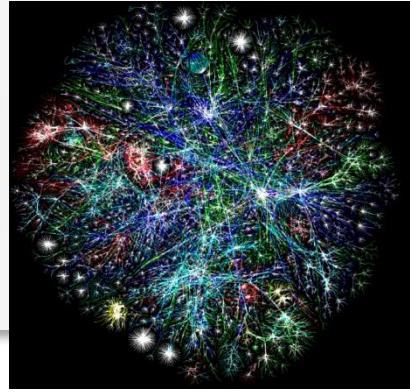

The Glass Palace
a:title
2000
a:year
London
a:city
HarpersCollins
a:p_name
http://…/isbn/000651409X
a:publisher
a:author
f:auteur
f:original
Le palais des mirroirs
f:titre
http://…/isbn/2020386682
f:traducteur
foaf:name
foaf:homepage
Amitav Ghosh
http://www.amitavhosh.com
w:isbn
foaf:name
Christiane Besse
w:person
http://en.wikipedia.com/wiki/The_Glass_Palace
http://en.wikipedia.com/wiki/Amitav_Ghosh
w:authorOf
w:authorOf
http://en.wikipedia.com/wiki/The_Hungry_Tide
w:authorOf
http://en.wikipedia.com/wiki/In_an_Antique_Land

Is that surprising?



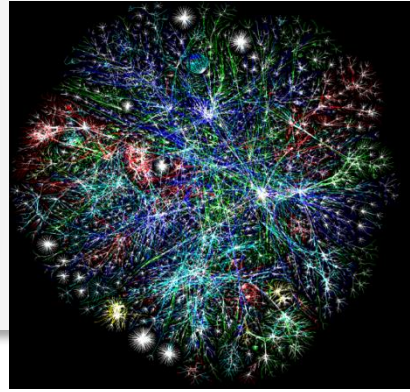
- Maybe but, in fact, no!
- What happened via automatic means is done all the time by the (human) users of the Web!
- The difference: a bit of extra rigor (eg, naming the relationships), extra information (eg, identifying relationships)
- …and machines could do this, too.

Important issue: “schema independence”



- The queries (ie, the application) sees the RDF data only (with references to “real”data)
- **If the structure (“schema”) of the database changes, only the mapping to RDF has to be changed**
  - this is a very local change
- Ie, the RDF layer is very robust vis-a-vis schema evolution (not only to schema differences)

Questions …

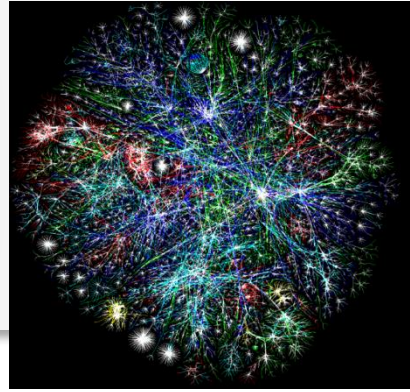


It relies on giant, centrally controlled ontologies for "meaning"

NO!

Ontologies are usually developed by communities and they are to be shared. In fact, in our example, we used an ontology called "FOAF"

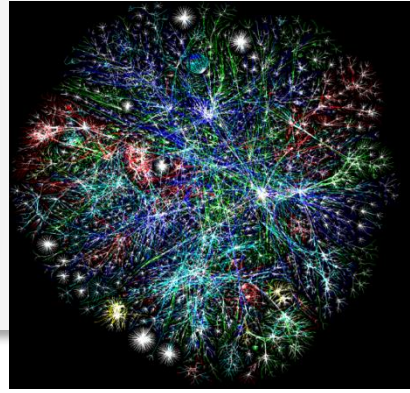
Core Vocabularies

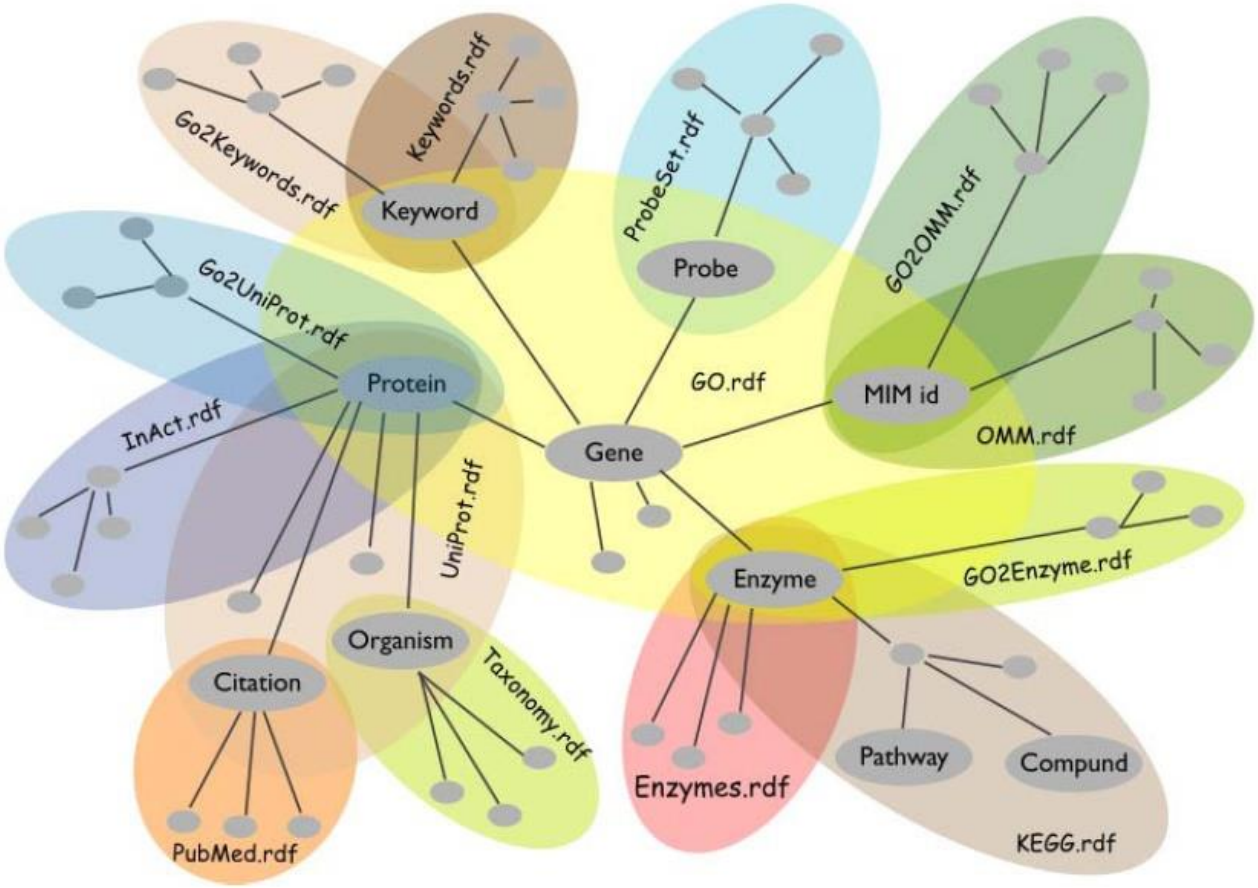


A number of public “core” vocabularies evolve to be used by applications, e.g.:

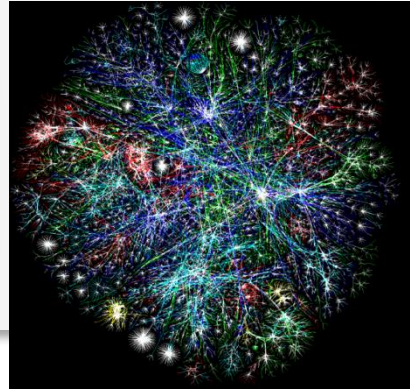
- **SKOS Core**: about knowledge systems
- **Dublin Core**: about information resources, digital libraries, with extensions for rights, permissions, digital right management
- **FOAF**: about people and their organizations
- **DOAP**: on the descriptions of software projects
- **MusicBrainz**: on the description of CDs, music tracks, …
- **SIOC**: Semantically-Interlinked Online Communities
- …

A mix of ontologies (a life science example)...





Further Questions …



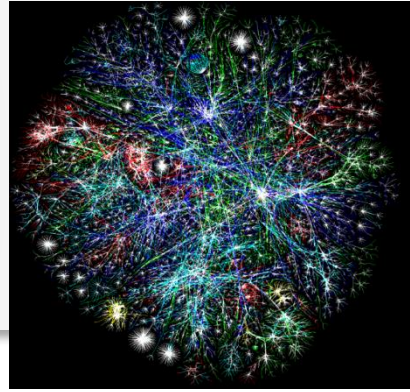
Isn't This Research Only? (or: does this have any industrial relevance whatsoever?)

- **Not any more…**
  - Lots of tools are available. Are listed on W3C's wiki:
    - RDF programming environment for 14+ languages, including C, C++ , Python, Java, Javascript, Ruby, PHP,…
    - 13+ Triple Stores, ie, database systems to store (sometimes huge!) datasets
    - a number programming environments (in Java, Prolog, …) include OWL reasoners
    - there are also stand-alone reasoners (downloadable or on the Web)
    - etc.
  - Some of the tools are Open Source, some are not; some are very mature, some are not: it is the usual picture of software tools, nothing special any more!

- **The Semantic Web has indeed a strong foundation in research results but remember…**
  - (1) the Web was born at CERN…
  - (2) …was first picked up by high energy physicists…
  - (3) …then by academia at large…
  - (4) …then by small businesses and start-ups…
  - (5) "big business" came only later!

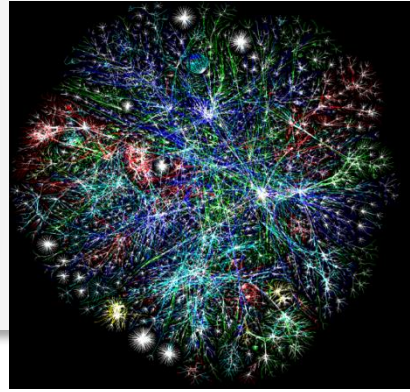
  Semantic Web is now at #4, and moving to #5!

Summary



- The Semantic Web is not as complex as people believe
- The Semantic Web does not require huge investments before seeing its value
- The Semantic Web is not only for geeks…

Further Links



- http://www.semantic-web.at/main.php
- http://www.w3.org/
- http://lov.okfn.org/dataset/lov/
- http://www.know-center.tugraz.at/