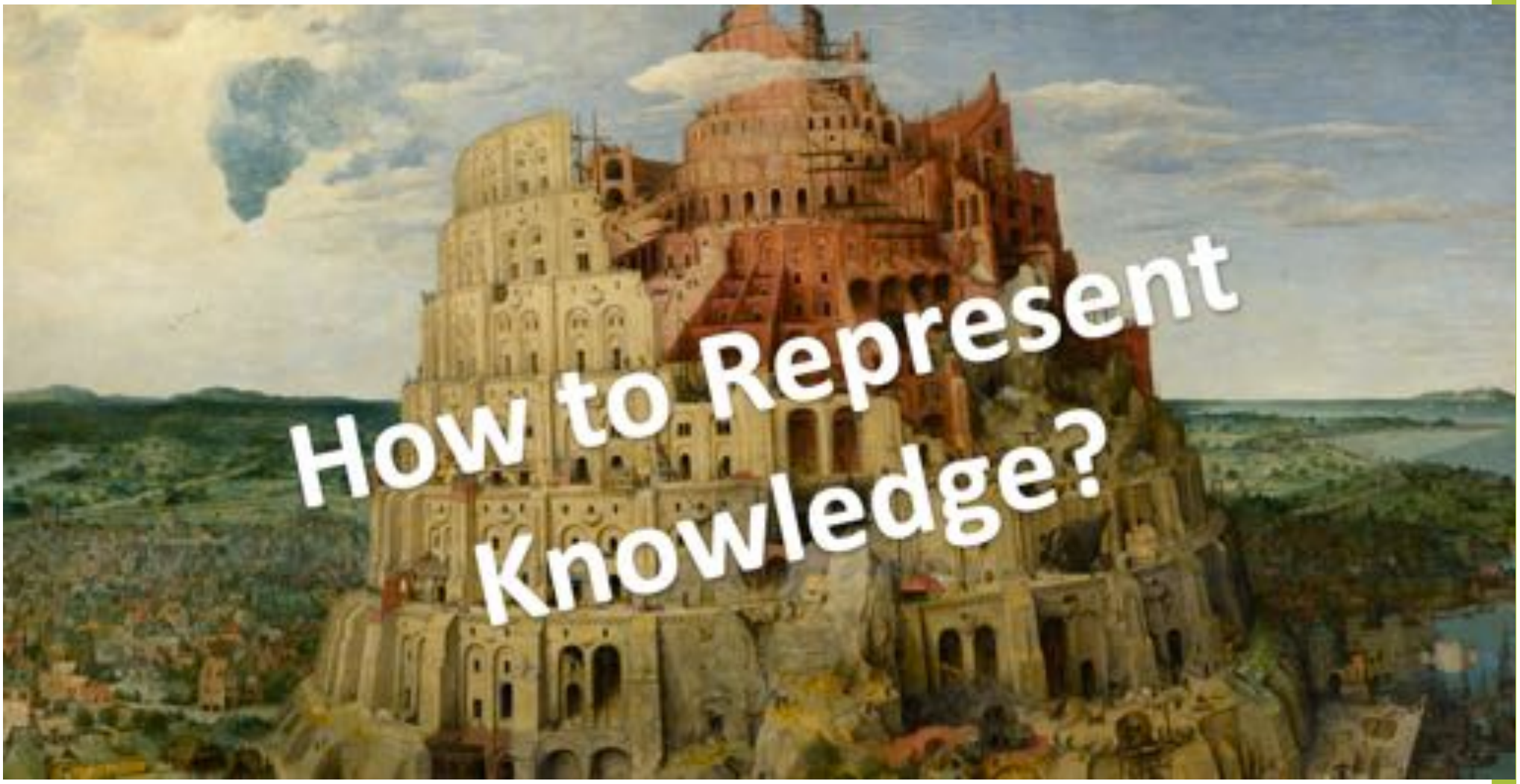
Knowledge Representation & Reasoning & Introduction To Knowledge Graphs

Week 3 | Fall 2022

Dr. Amna Basharat


How to Represent
Knowledge?

Formal Knowledge Representation

- **Formal Knowledge Representation**
  - is a field of **artificial intelligence (AI),**
  - which (unambiguously) captures the **semantics (meaning)** of **concepts, properties, relationships,** and **entities**
  - of specific **knowledge domains,** i.e., fields of interest or areas of concern,
  - as **structured data.**
- **Machines (computers)** must be able to **understand** formal knowledge representations.
- To "**understand**" a knowledge representation, the machine must be able to **interpret it correctly**.

Towards Universal Knowledge Representation

Climate Change is the **Everest** of all problems...


Climate Change is
the Everest of all
problems...

Knowledge & Understanding

Text: "Everest"

Entity Mapping
Entity Disambiguation

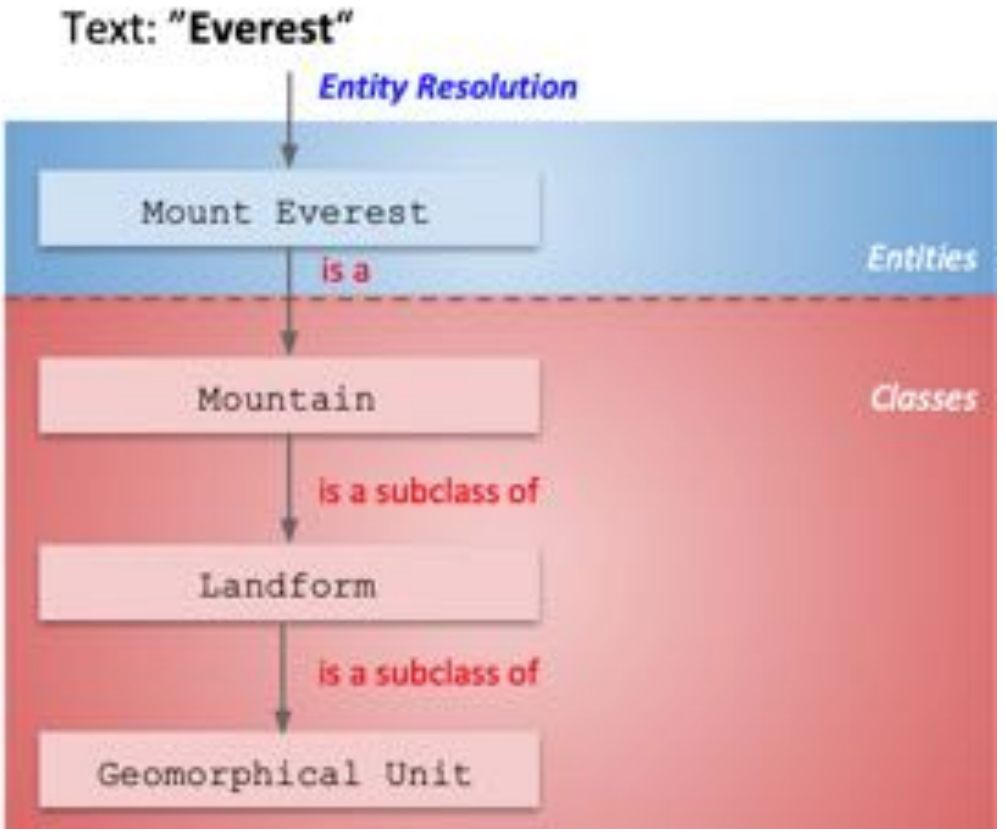
Everest, Kansas	a small village
Everest, Gasfield	a gas field near Scotland
George Everest	a Surveyor General of India
Jack Everest	an Irish football player
...	
Mount Everest	a mountain

Disambiguation

- solution of linguistic ambiguities

Knowledge & Understanding

- The **Meaning (Semantics)** of entities and classes must be defined explicitly.



Knowledge & Understanding

$MountEverest \in Mountain$

$Mountain \subseteq Landform$

$Landform \subseteq GeomorphicalUnit$

$GeomorphicalUnit \subseteq NaturalGeographicObject$

$GeorgeEverest \in Person$

$Person \cap Mountain = \varnothing$

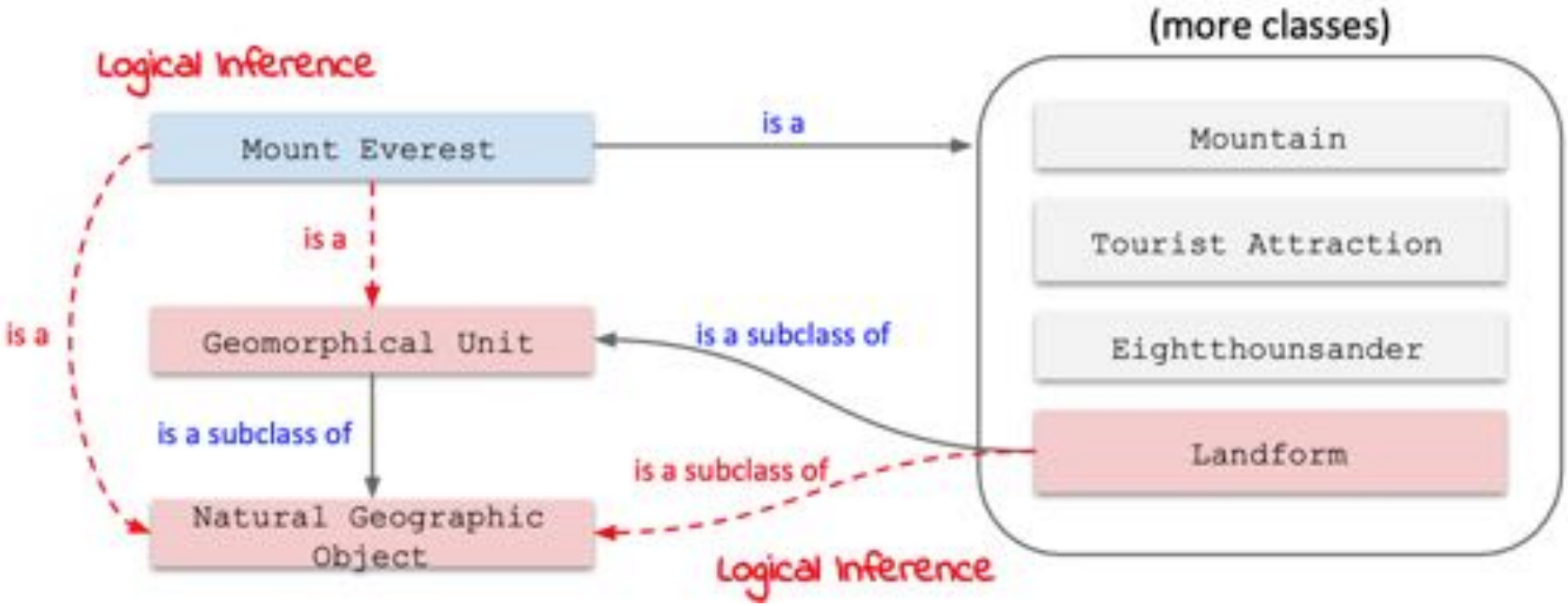
Logical Inference

$MountEverest \notin Person$

$GeorgeEverest \notin Mountain$

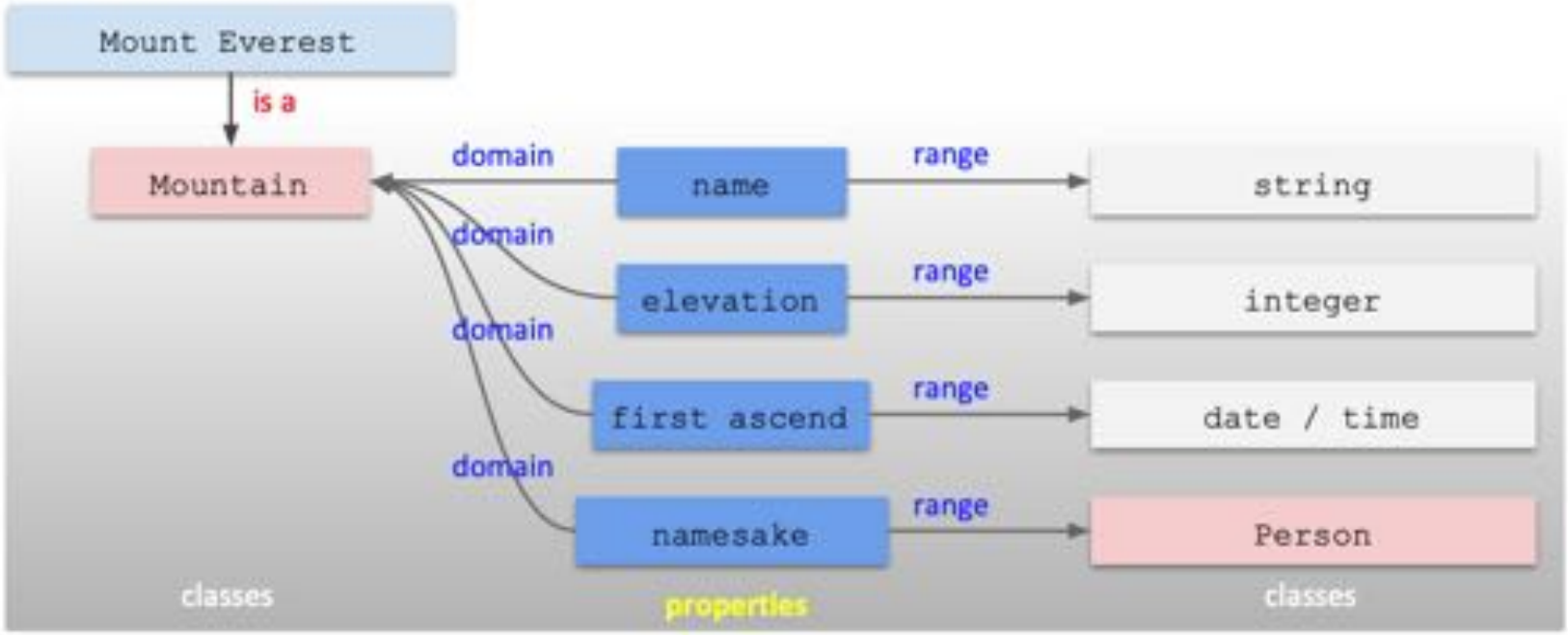
Knowledge and Understanding

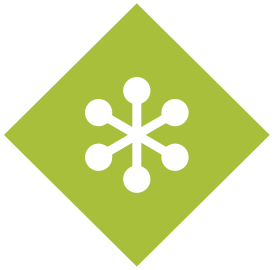
- The Meaning (Semantics) of information is expressed with the help of knowledge representations (**Ontologies**)



Knowledge and Understanding

- The Meaning (Semantics) is expressed with the help of knowledge representations (**Ontologies**)





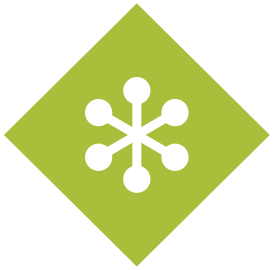
Knowledge Representation vs Data Structures

- **What's the difference to traditional data structures?**
  - **Mathematical Logic** provides a framework to **formally express the semantics** of knowledge representations.
  - **Semantics** of knowledge representations can be defined **explicitly.**
  - **Mathematical Logic** enables **logical inferences and reasoning** for knowledge representations.

Whats All the Hype About KGs?

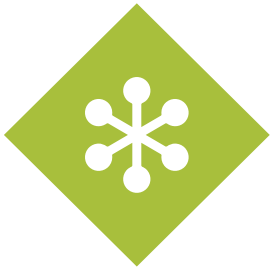
Knowledge Graphs Constitute the Backbone of Today's State-of-the-Art Information Systems & Intelligent Agents

From improving search results over question answering and recommender systems up to explainable AI systems, the applications of knowledge graphs are manyfold.



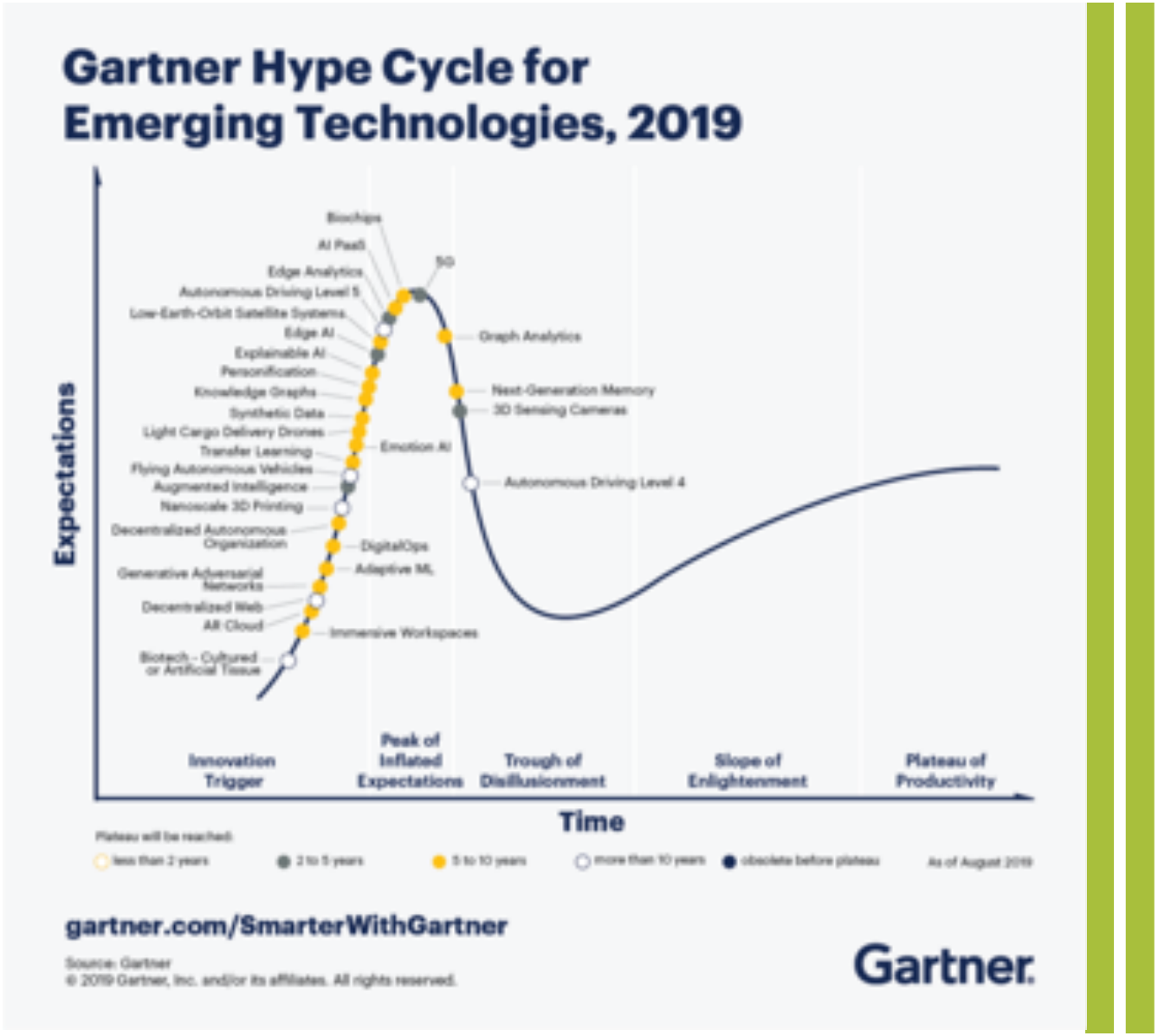
Knowledge Graphs - the Hype

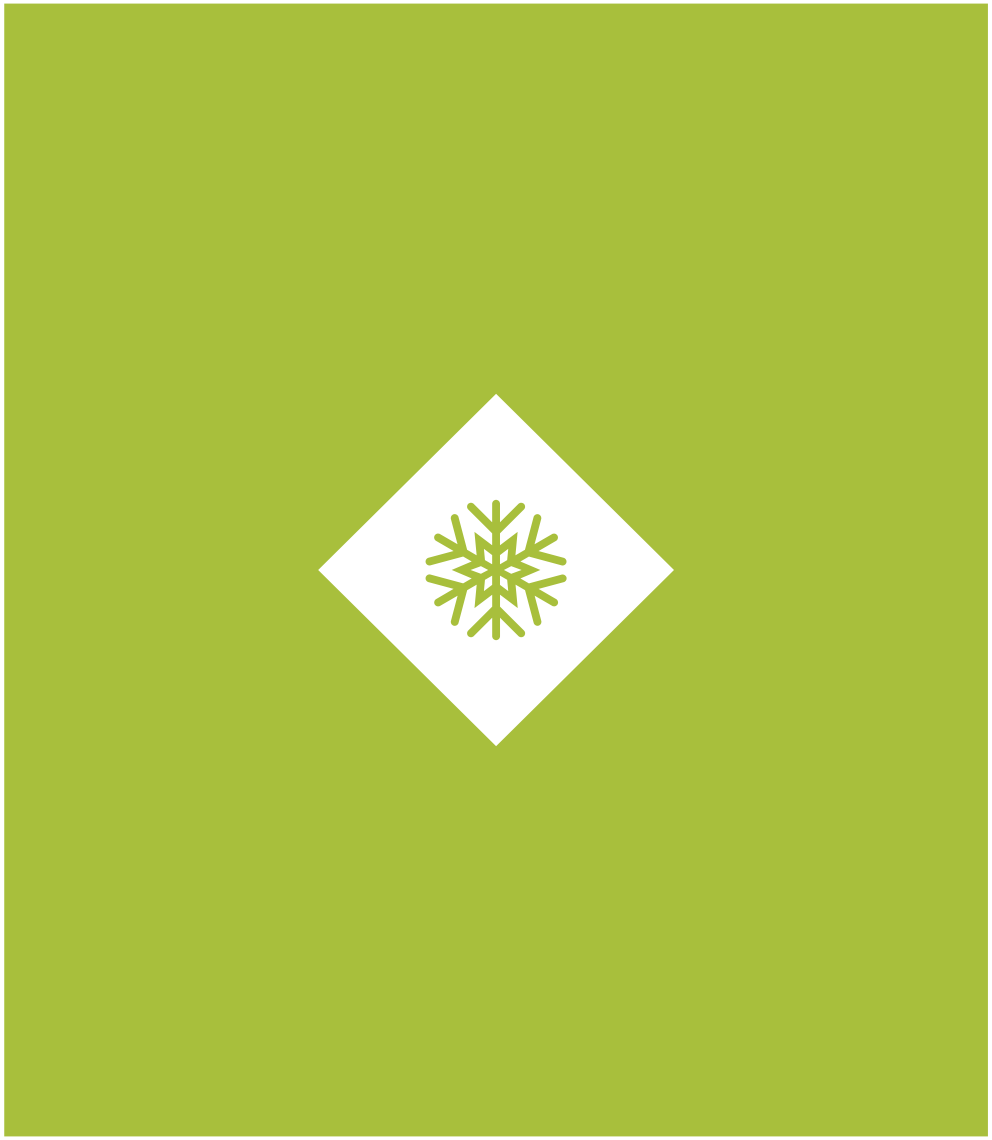
- Knowledge graphs have emerged as a compelling abstraction for organizing world's structured knowledge over the internet
  - capturing relationships among key entities of interest to enterprises
  - a way to integrate information extracted from multiple data sources
- Knowledge graphs have also started to play a central role in machine learning and natural language processing
  - as a method to incorporate world knowledge
  - as a target knowledge representation for extracted knowledge
  - and for explaining what is being learned



Knowledge Graphs - Emerging on the Hype Cycle

- Getting the Semantic Web and Knowledge Graph researchers excited!





More About Knowledge Graphs

THE KNOWLEDGE GRAPH COOKBOOK

RECIPES THAT WORK

ANDREAS BLUMAUER

AND **HELMUT NAGY**

1st edition, 2020

WHY KNOWLEDGE GRAPHS?

How do you "cook" a knowledge graph? Before we discuss specific variants of recipes and dishes, examine the individual ingredients, tools and methods or classify recipes, I would like to explain the main reasons why you should learn how to cook knowledge graphs. This chapter will outline the excellent results you can achieve. Here is a brief preview:

- Knowledge graphs (KGs) solve well-known data and content management problems.
- KGs are the ultimate linking engine for enterprise data management.
- KGs automatically generate unified views of heterogeneous and initially unconnected data sources, such as Customer 360.
- KGs provide reusable data sets to be used in analytics platforms or to train machine learning algorithms.
- KGs help with the dismantling of data silos. A semantic data fabric is the basis for more detailed analyses.

A BRIEF HISTORY OF KNOWLEDGE GRAPHS

> Cooking is culture, and culture is based on history. History is not only what has happened, but also what has been piled up—the ground upon which we stand and build. Therefore, we should also have an understanding of where knowledge graphs come from if we want to become a maestro KG chef. Understanding the historical context is always paramount to understanding the possible paths one can take in the future.

FAST FORWARD

- In 1736, graph theory was born: Leonhard Euler formulated the 'Königsberg Bridge Problem.'
- In 1976, John F. Sowa published his first paper on Conceptual Graphs.[4]
- In 1982, Knowledge Graphs were invented in the Netherlands. The theory of Knowledge Graphs was initiated by C. Hoede, a mathematician at the University of Twente, and F.N. Stokman, a mathematical sociologist at the University of Groningen.
- In 1999, Resource Description Framework (RDF) Model was published as a W3C Recommendation to lay a foundation for a Semantic Web.
- In 2001, Tim Berners-Lee, Jim Hendler and Ora Lassila published their ground-breaking article 'The Semantic Web'[5] in the Scientific American Magazine.
- In 2006, the DBpedia[6] project created a seed for the emergence of the Linked Open Data cloud by transforming Wikipedia content into linked data.
- In 2012, Google introduced their Knowledge Graph, and since then a lot of companies have started to build their own projects using knowledge graphs in various flavours.
- In 2018, The GQL Manifesto[7] was published to agree on a standard for a property graph query language.

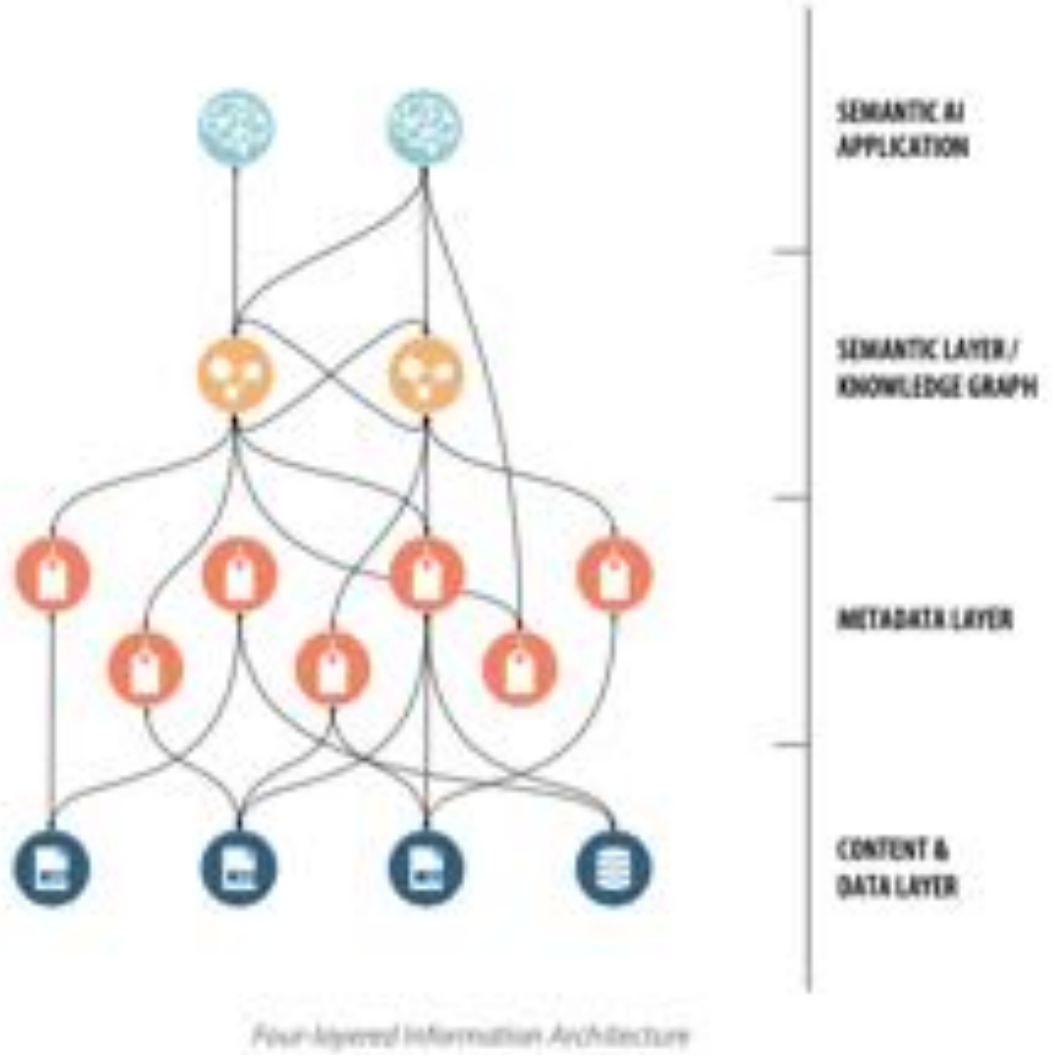
- By the end of 2019 knowledge graphs had become mainstream. For example, Gartner states that "... a semantic knowledge graph can be used to power other data management tasks such as data integration in helping automate a lot of redundant and recurring activities."[1]
- After decades of developing KGs, the discipline has also been influenced by a lot of other knowledge domains including mathematical logic, graph theory, information retrieval, computer linguistics, knowledge representation and reasoning, and most recently, the Semantic Web and machine learning.

SEMANTIC WEB

In 2001, when the WWW was still in its infancy, its founder Tim Berners-Lee was already talking about the next big step: "The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users."

20 years later, we all know that things have developed more slowly and somehow in a different direction than expected; nevertheless, the W3C has laid the groundwork for a Semantic Web by publishing several important recommendations:

- 1999: Resource Description Framework (RDF) Model and Syntax Specification as a foundation for processing metadata to provide interoperability between applications that exchange machine-understandable information on the Web.
- 2004: Resource Description Framework (RDF) and RDF Vocabulary Description Language 1.0: RDF Schema (RDFS) as a standard for representing information about resources in the WWW. As a major update, RDF 1.1 was published in 2014.
- 2004: OWL Web Ontology Language as a language for defining and instantiating Web ontologies.
- 2008: SPARQL Protocol and RDF Query Language (SPARQL) to retrieve and manipulate data stored in RDF via so-called SPARQL endpoints. As a major update, SPARQL 1.1 was published in 2013.
- 2009: Simple Knowledge Organization System (SKOS) for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary.



Four-layered Information Architecture

Four Layered Information Architecture

Powered by
Knowledge Graphs

Better Knowledge Management

DIKW Pyramid: From Data to Wisdom

Towards Explainable AI

MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE: MAKE IT EXPLAINABLE

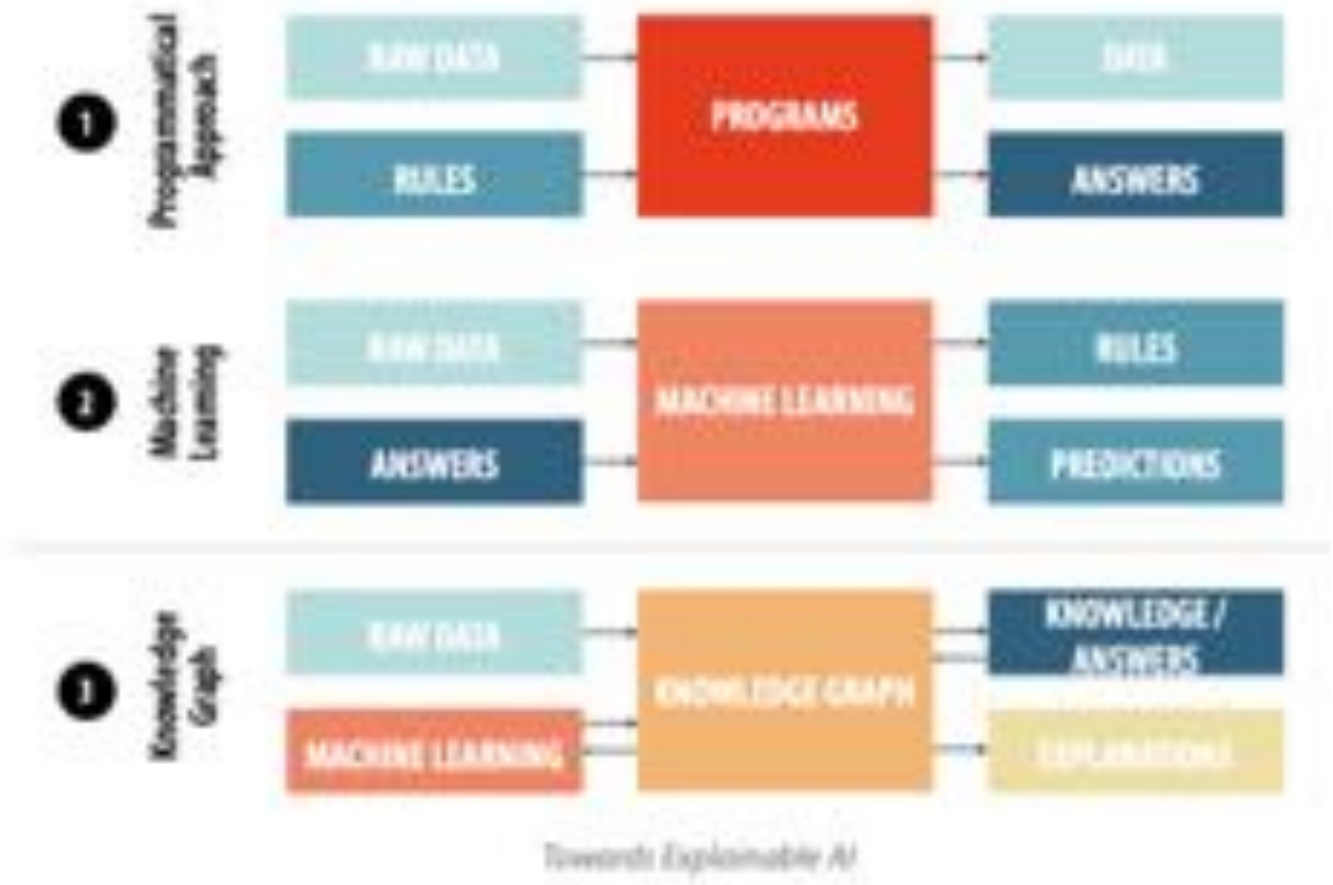
"MACHINE LEARNING ALGORITHMS LEARN FROM HISTORICAL DATA, BUT THEY CANNOT DERIVE NEW INSIGHTS FROM IT"

While AI is becoming a part of our daily lives, many people are still skeptical. Their main concern is that many AI solutions work like black boxes and seem to magically generate insights without explanation.

In addition to the benefits they can bring to the area of enterprise data management, knowledge graphs are increasingly being identified as building blocks of an AI strategy that enables explainable AI following the Human-in-the-Loop (HITL) design principle.

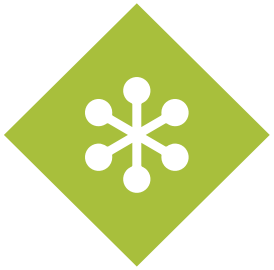
Why does artificial intelligence often work like a black box?

Towards Explainable AI



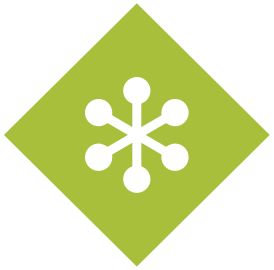
Towards Explainable AI

Review Questions



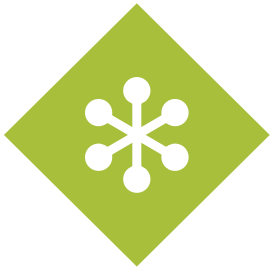
Review

- What is the difference between the Internet and the Web?
- Distinguish Web 1.0 vs 2.0 vs. 3.0
- Distinguish semantic vs syntactic search
- What 5 things correct meaning and interpretation of knowledge depends upon?
- What is the limitation of the traditional web?



Review - II

- What is the semantic web?
- Give another name for semantic web.
- How is meaning expressed on the semantic web?
- Give three inherent benefits of the semantic web.
- What are the key elements in first three layers of the semantic web stack?



Review - III

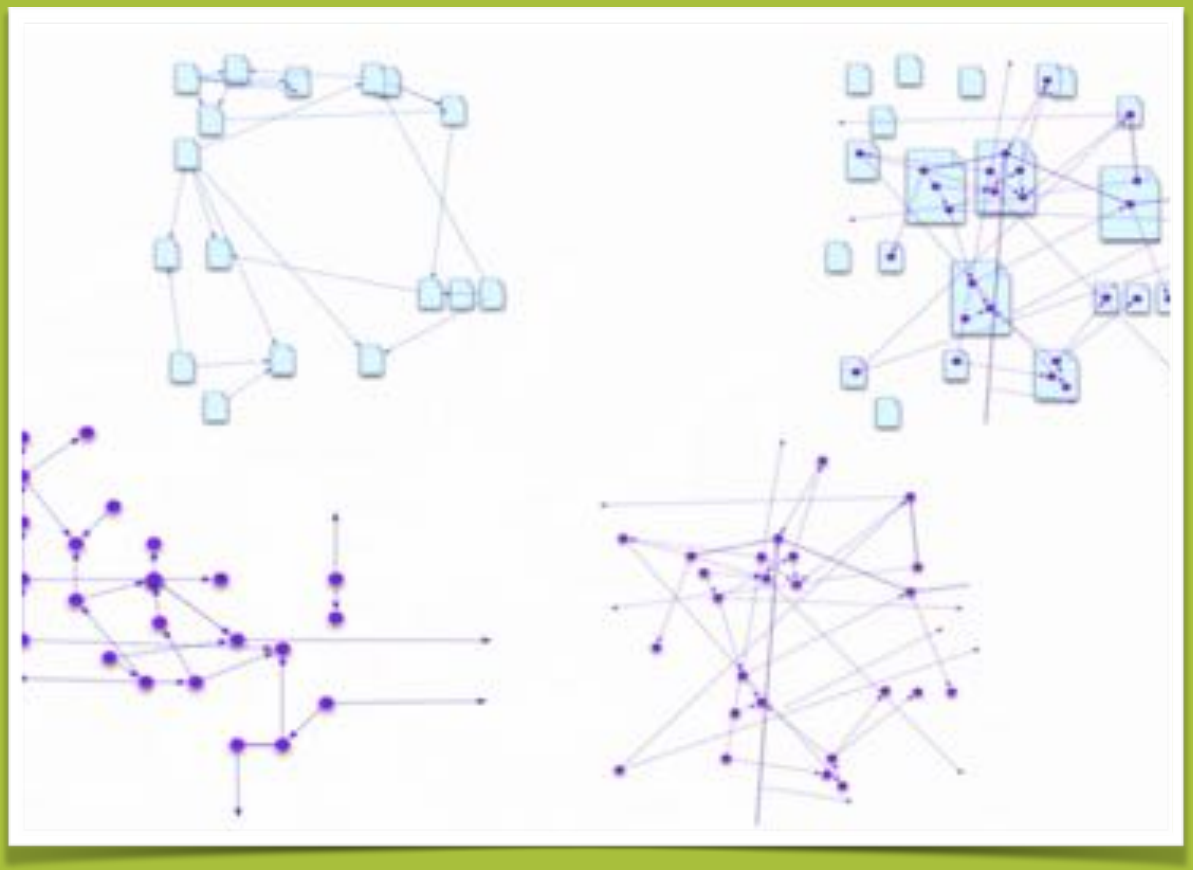
- What is the linked data cloud?
- What is considered the nucleus of linked data cloud?
- What is dbpedia?

Key Terms

Key Terms

- Knowledge
- Syntax vs. Semantics
- Ontology
- Web 1.0/Web 2.0/Web 3.0
- Semantic Web
- Web of Data
- Linked Data

Understanding Web 1.0/Web 2.0/Web 3.0



Understanding the Web of Data

https://www.youtube.com/watch?v=V6BR9DrmUQA

First Generation: the Internet

Computer Centered Processing

- How does the user get the information?

1. Open Terminal
2. Connect to remote system
3. Retrieve file system data from remote system
4. Download file from remote to local system
5. Read file on local system

Problems

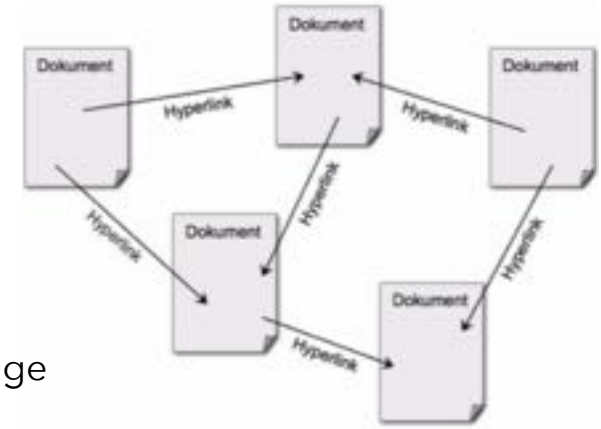
- Information access requires expert knowledge and is expensive
- Information retrieval is even more expensive

Second Generation: the Web

Document Centered Processing

- How does the user get the information?

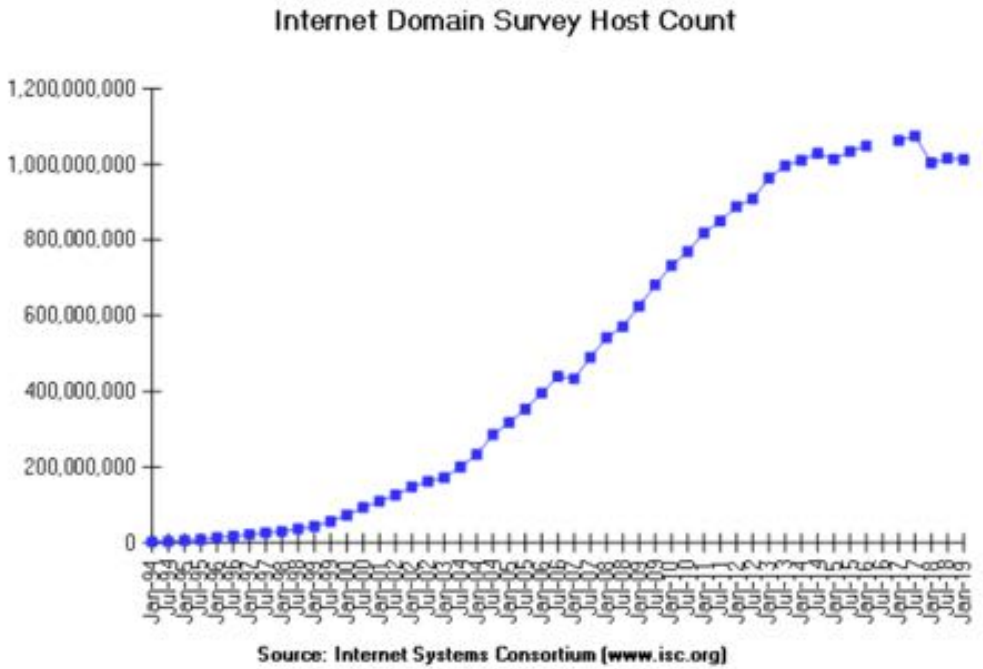
1. Open browser
2. Load Document
3. Click on the next hyperlink
4. ….



Advantages

- Information access does not require expert knowledge
- Information retrieval via search engine

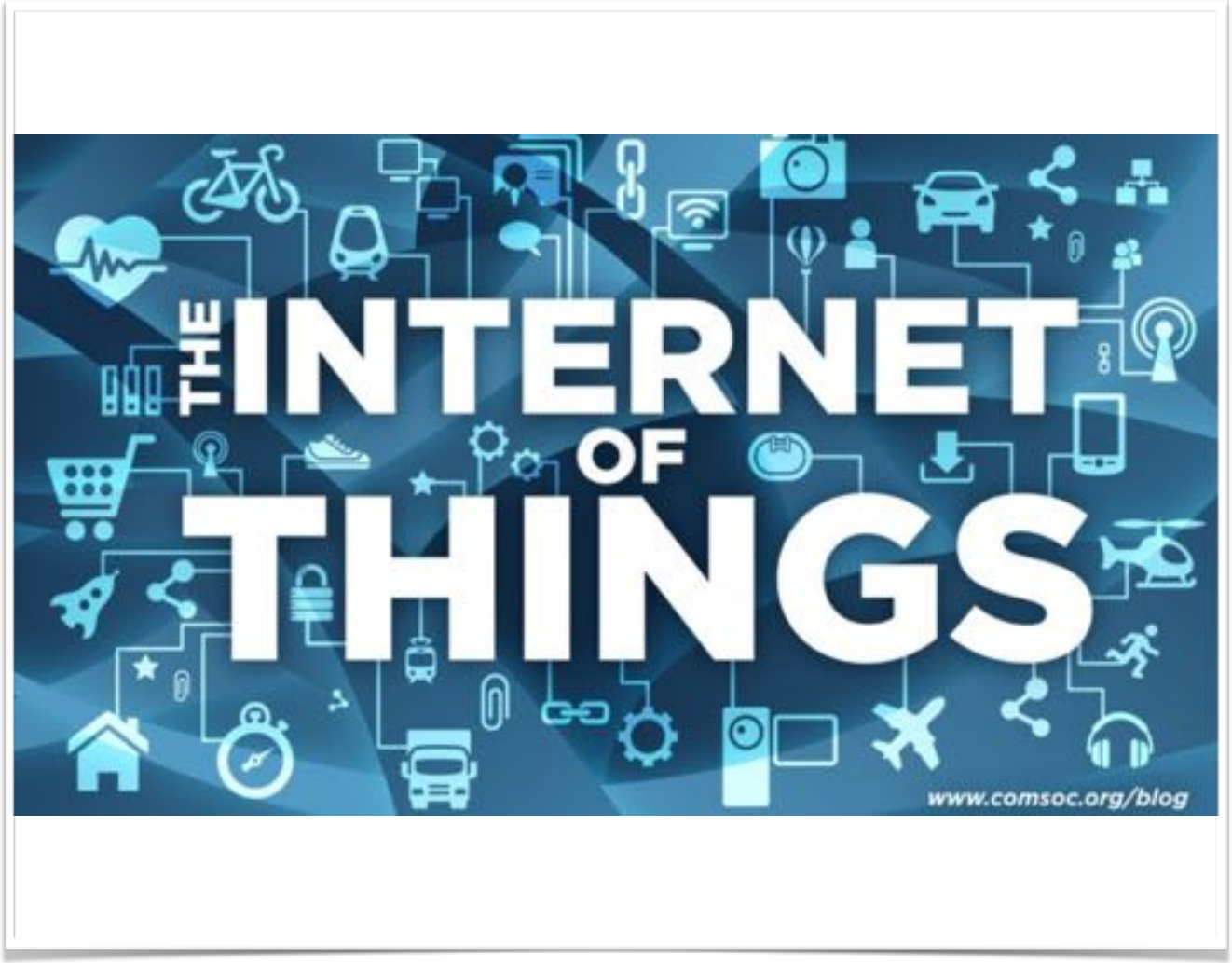
No Limits of Growth



"The Web is Big! Really Big. You just won't believe how vastly, hugely, mind-bogglingly big it is."

-According to Douglas Adams

The Next Revolution!



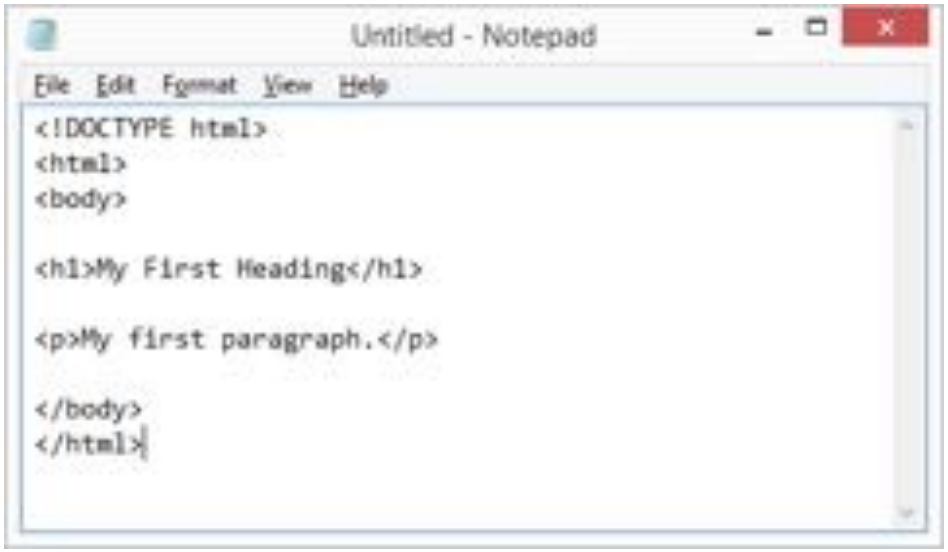
The Need for Semantics
AKA ‘Meaning'

Understanding the Limitations

Humans vs. Machines

Humans have contextual knowledge, world knowledge and experience to solve the problems

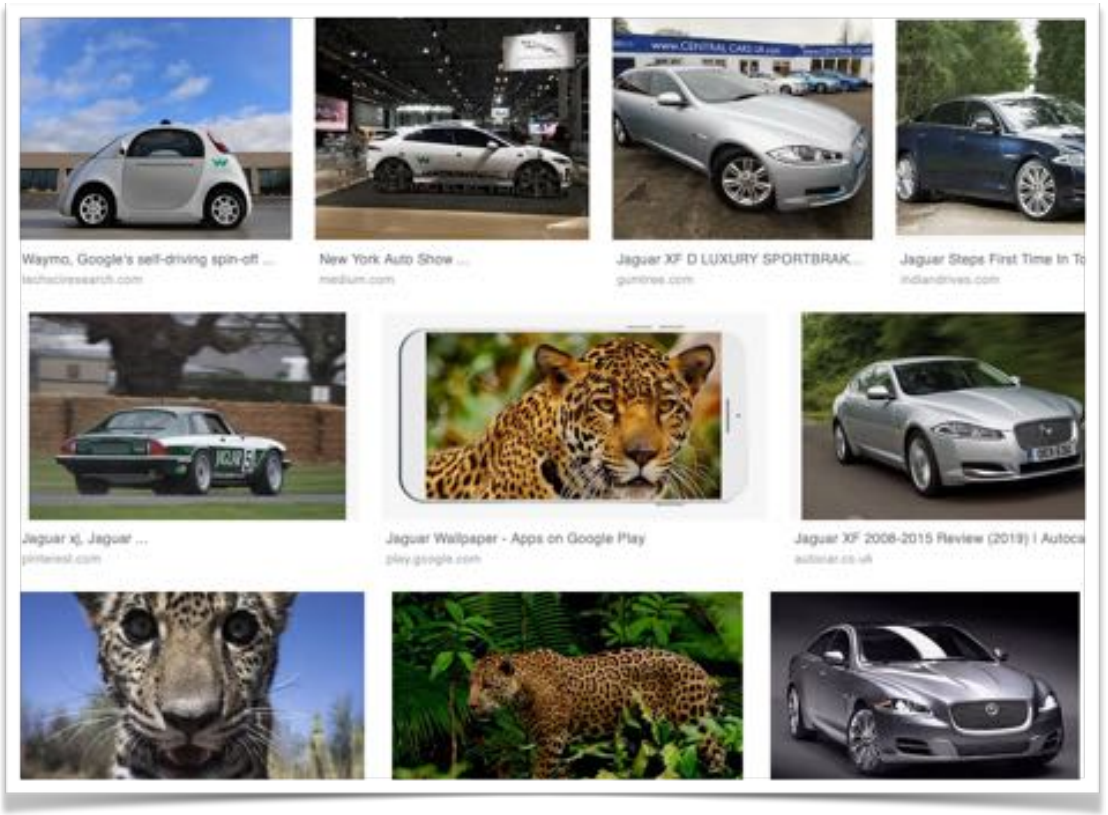
The (Document) Web Is for Humans



- The Web is based on the markup language HTML
- HTML describes
  - how information is presented
  - how information is linked
  - but NOT, what the information means

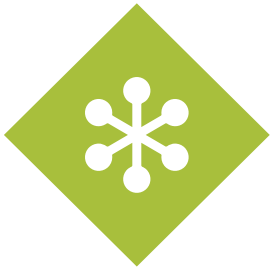
The Information Retrieval Dilemma

- **Syntactic vs. Semantic Search**
- Ambiguity of natural language (polysemy)
- Different words/expression for the same concept (synonyms/metaphors)



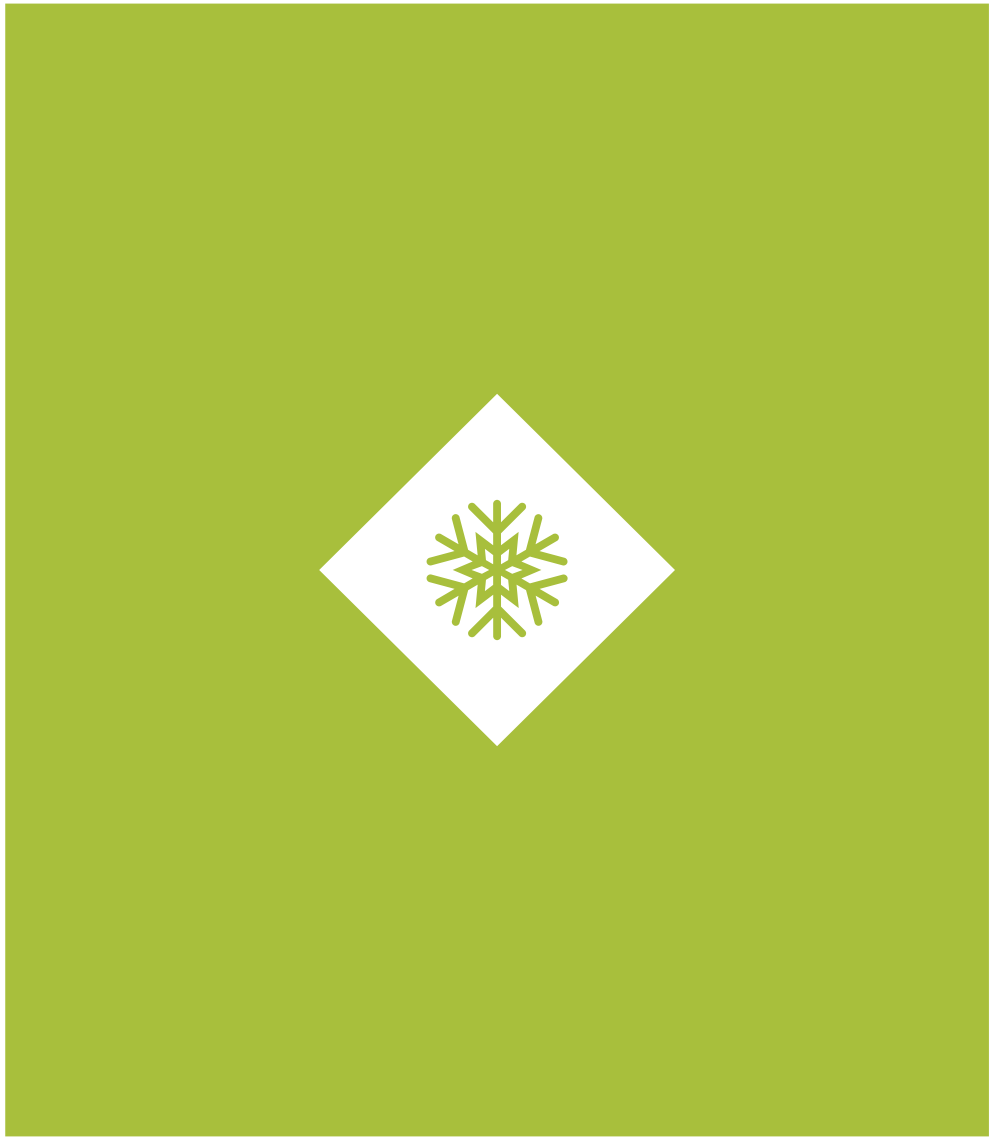
Implicit Knowledge

Information does not have to be specified explicitly, but may be derived via logical deductions from the available information



Meaning and Comprehension

- **Understanding** is the ability to grasp the **meaning of information**
- Information is conveyed in a **message** using a **specific language**
- Information is understood by the receiver of the message, if the receiver **interprets** the information **correctly**



From Syntactic to Semantic Search

No (Explicit) Semantics!

From WWW To Web of Data

- Precondition:
  - Content can be **read and interpreted correctly (= "understood") by machines**

Natural Language Processing

- Technologies of **traditional Information Retrieval**
- Statistical models & machine learning

Semantic Web Technologies

- Natural language web content is **explicitly annotated with semantic metadata**
- Semantic metadata encode the **meaning** of the content and can be **read and interpreted correctly by machines**


THE
SEMANTIC
WEB
A new form of Web content
that is meaningful to computers
will unleash a revolution of new possibilities
by
TIM BERNERS-LEE,
JAMES HENDLER and
ORA LASSILA

Web 3.0: the Semantic Web - A Web of Data

- The meaning of information (Semantics) is made explicit by **formal (structured) and standardized knowledge representations (Ontologies).**
- Thus it will be possible,
  - To process the meaning of information automatically
  - To relate and integrate heterogeneous data
  - To deduce implicit (not evident) information from existing (evident) information in an automated way.
- The Semantic Web is kind of a **global database** that contains a **universal network of semantic propositions.**

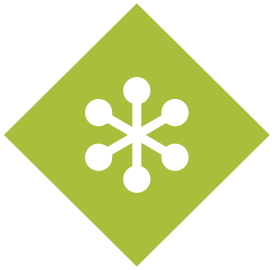
A Case Study

To Understand basics of Web 3.0, KGs, LD

Adaptation
Introduction to the Semantic Web
(tutorial)

2009 Semantic Technology Conference
San Jose, California, USA
June 15, 2009

Ivan Herman, W3C



The Rough Structure of Data Integration

- Map the various data onto an abstract data representation
  - make the data independent of its internal representation…
- Merge the resulting representations
- Start making queries on the whole!
  - queries that could not have been done on the individual data sets

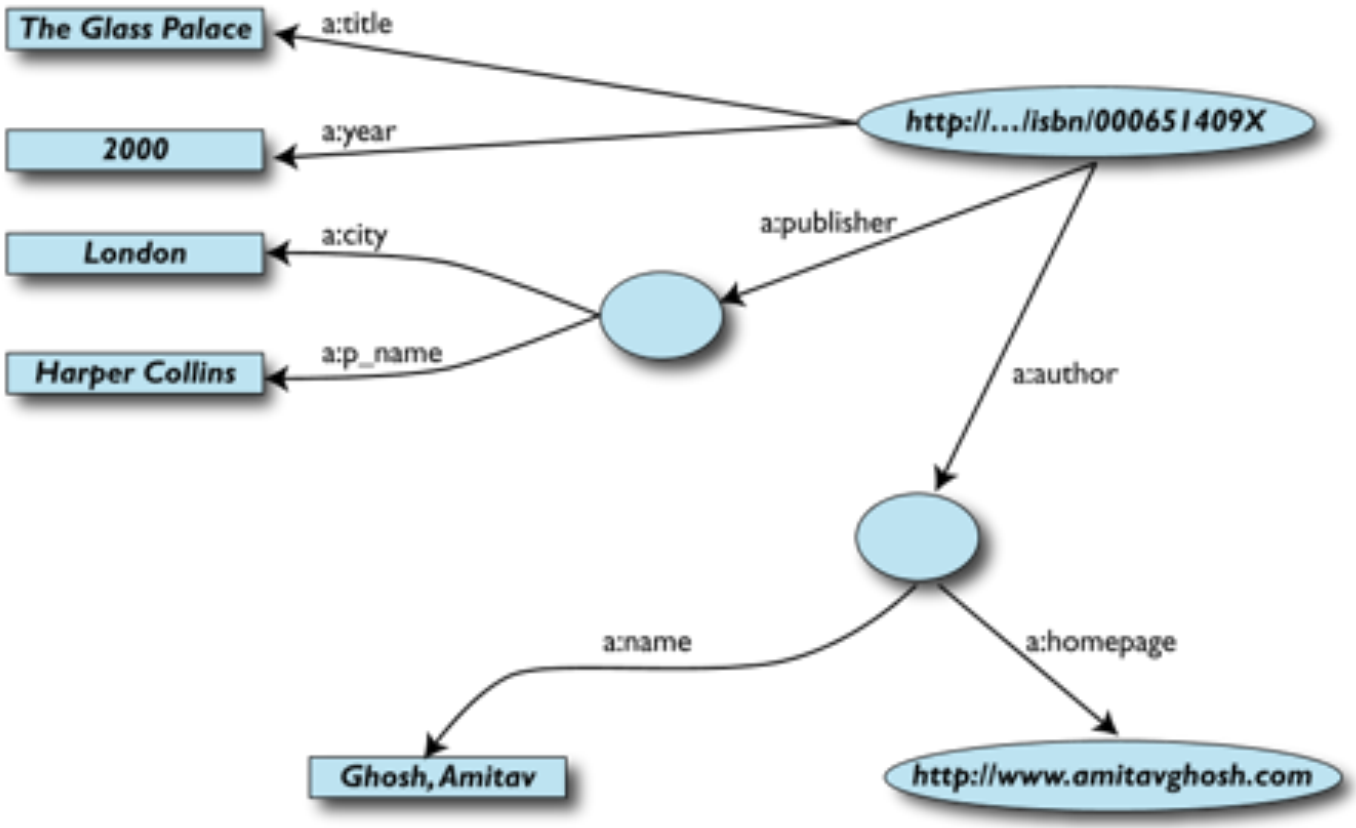
A Simplified Bookstore Data (Dataset “a”)

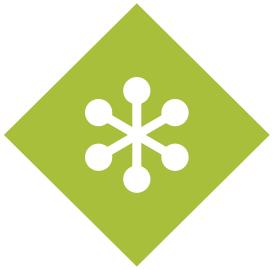
ID	Author	Title	Publisher	Year
ISBN0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

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

ID	Publ. Name	City
id_qpr	Harper Collins	London

1st: Export Your Data as a Set of Relations





Some Notes on the Exporting the Data

- Relations form a graph
  - the nodes refer to the “real” data or contain some literal
  - how the graph is represented in machine is immaterial for now
- Data export does not necessarily mean physical conversion of the data
  - relations can be generated on-the-fly at query time
    - via SQL “bridges”
    - scraping HTML pages
    - extracting data from Excel sheets
    - etc.
- One can export part of the data

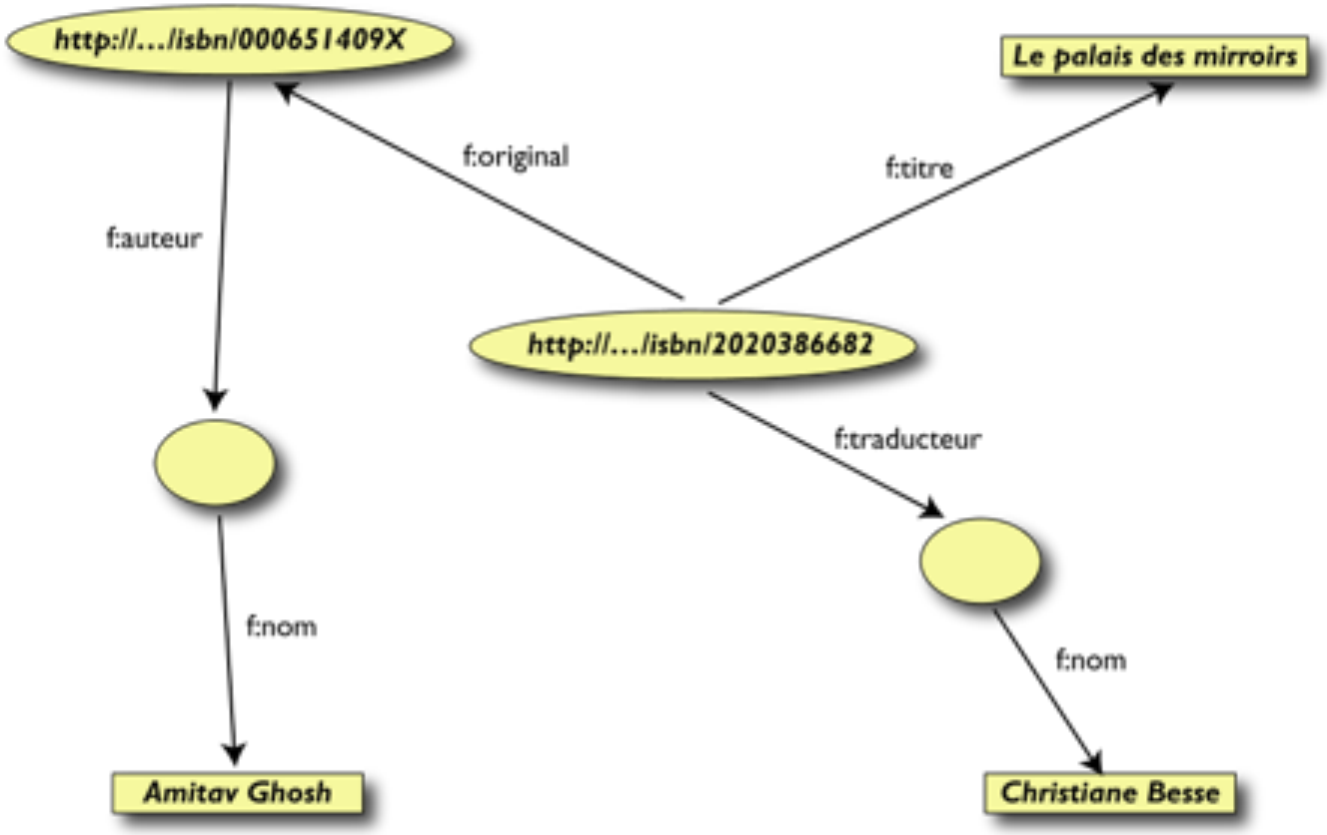
Another Bookstore Data (Dataset “F”)

	A	B	D	E
1	**ID**	**Titre**	**Traducteur**	**Original**
2	ISBN0 2020386682	Le Palais des miroirs	A13	ISBN-0-00-651409-X
3				

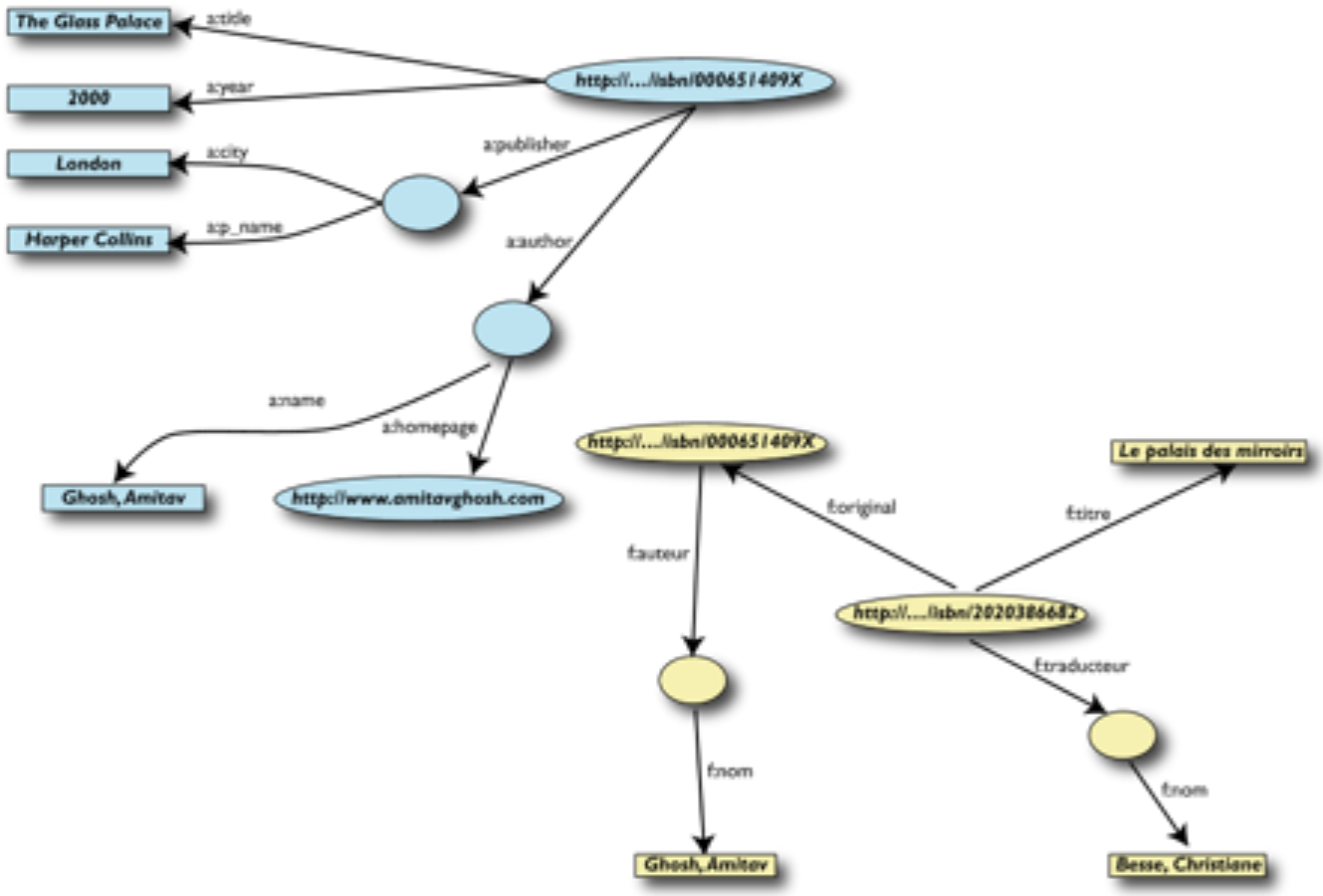
6	**ID**	**Auteur**
7	ISBN-0-00-651409-X	A12

11	**Nom**
12	Ghosh, Amitav
13	Besse, Christianne

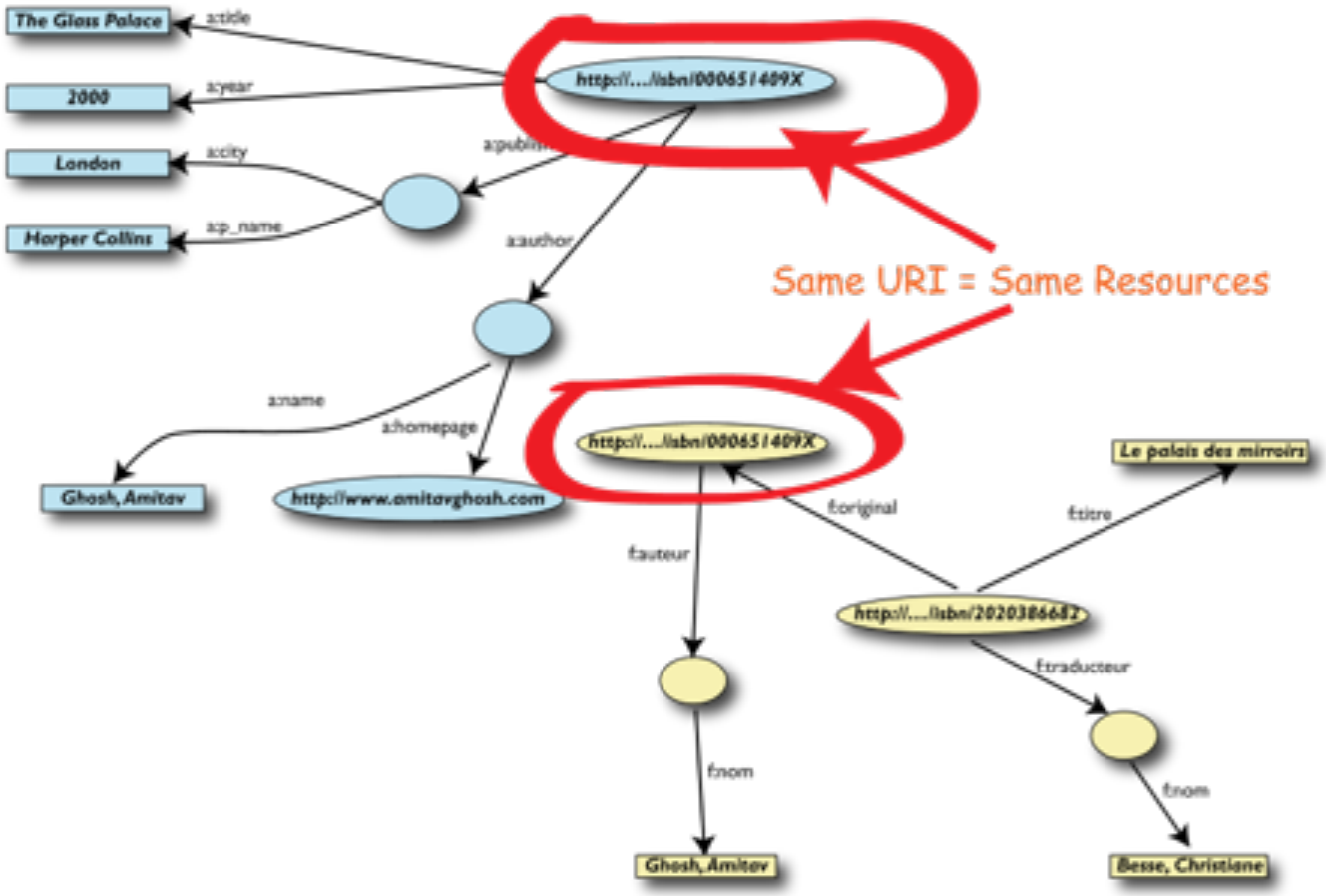
2nd: Export Your Second Set of Data



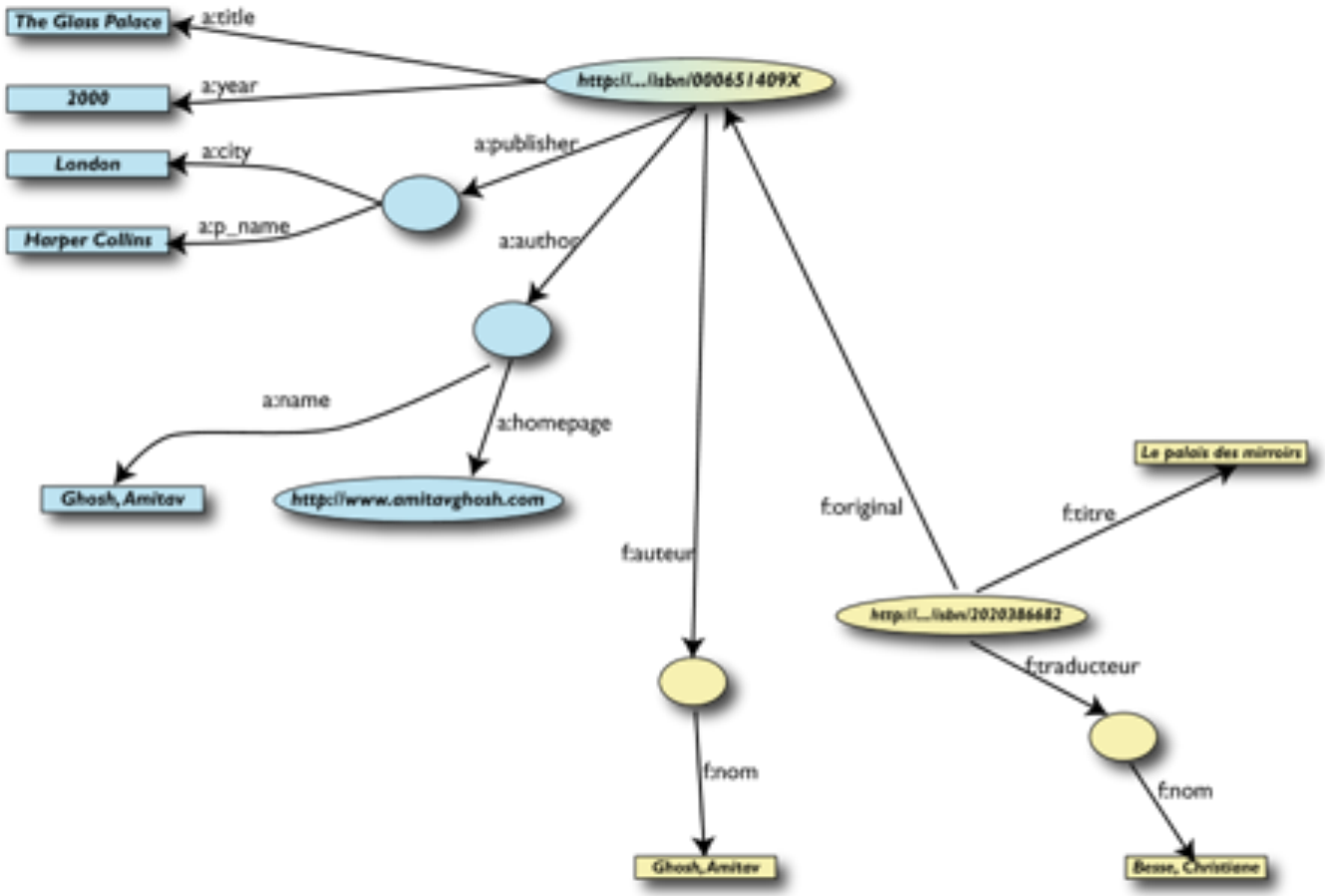
3rd: Start Merging Your Data



3rd: Start Merging Your Data (Cont.)

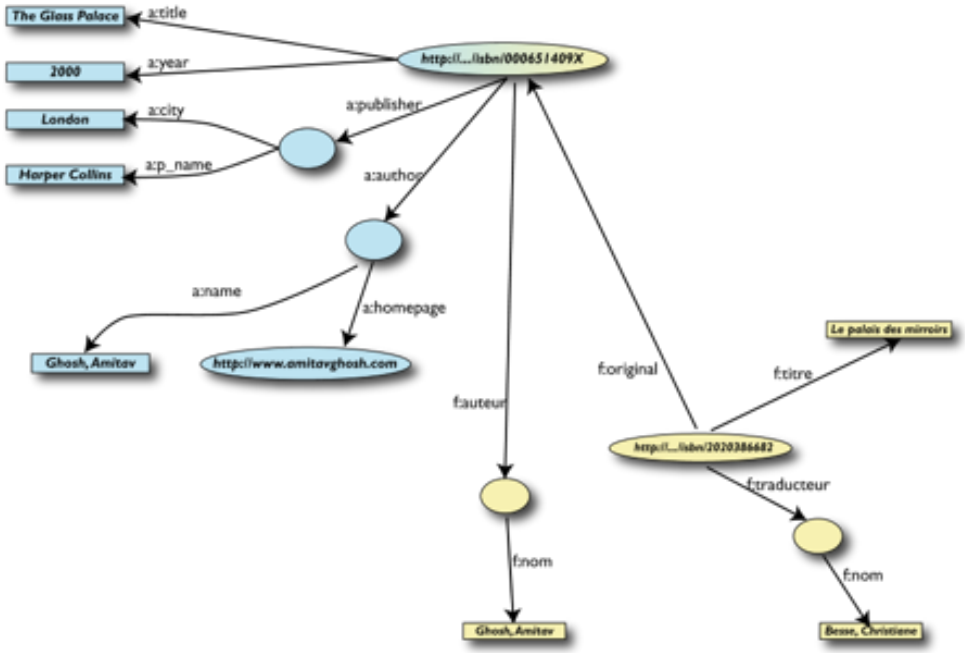


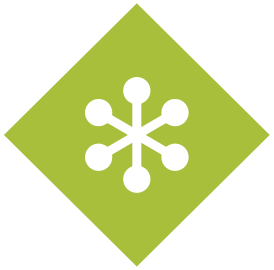
3rd: Merge Identical Resources



Start Making Queries…

- User of data “F” can now ask queries like:
  - “give me the title of the original”
    - well, … « donnes-moi le titre de l’original »
- This information is not in the dataset “F”…
- …but can be retrieved by merging with dataset “A”!

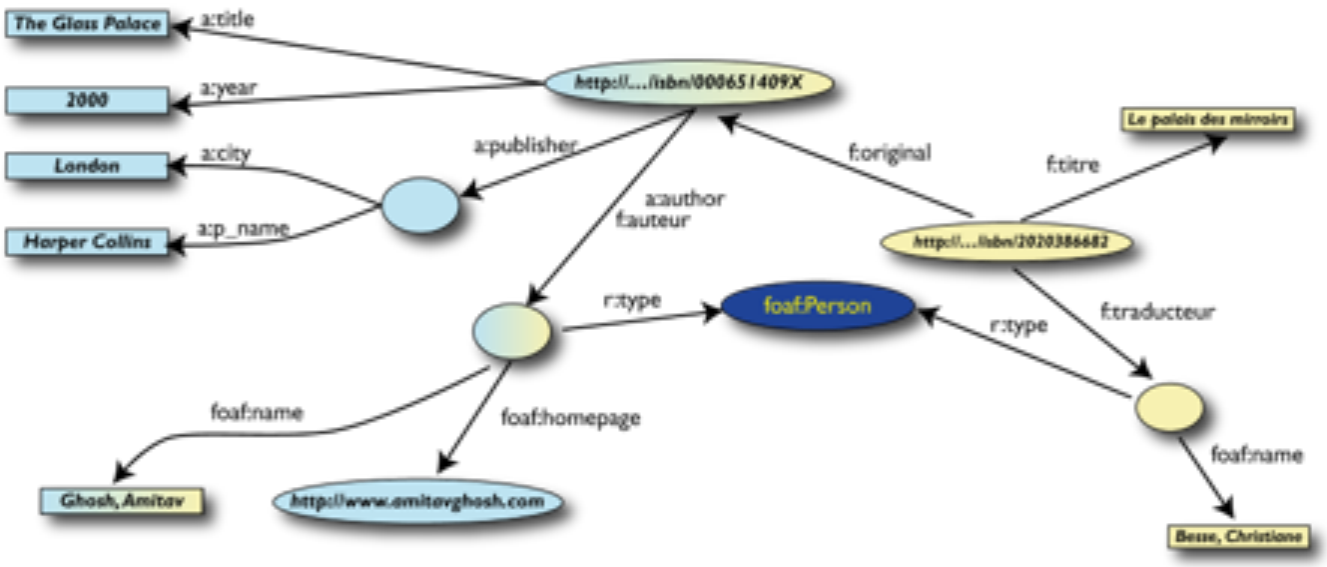


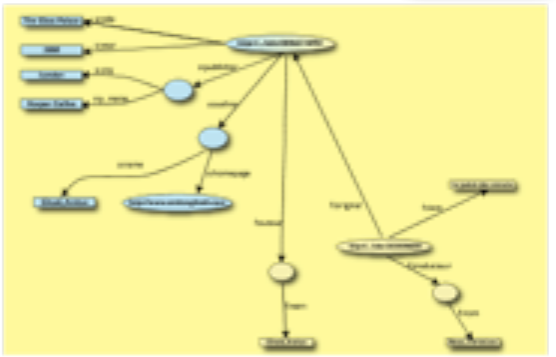


However, More Can Be Achieved...

- We “feel” that a:author and f:auteur should be the same
- But an automatic merge doest not know that!
- Let us add some extra information to the merged data:
  - a:author same as f:auteur
  - both identify a “Person”
  - a term that a community may have already defined:
    - a “Person” is uniquely identified by his/her name and, say, homepage
    - it can be used as a “category” for certain type of resources

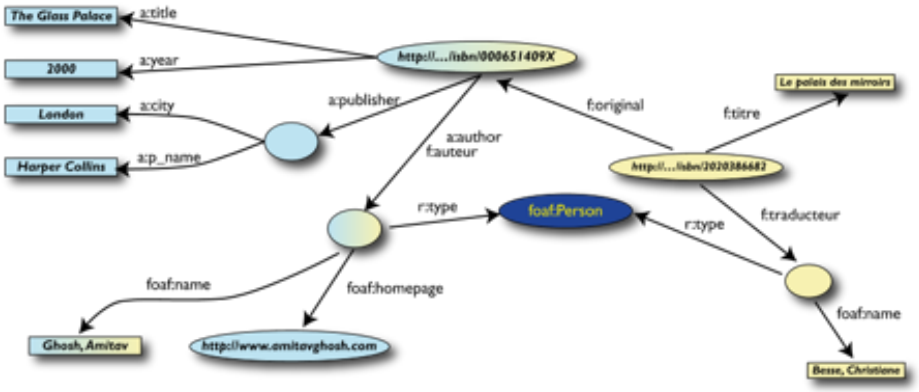
3rd Revisited: Use the Extra Knowledge

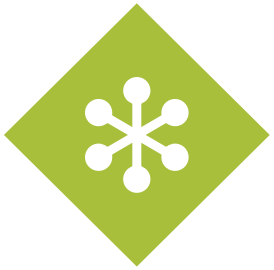




Start Making Richer Queries!

- User of dataset “F” can now query:
  - “donnes-moi la page d’accueil de l’auteur de l’originale”
    - well… “give me the home page of the original’s ‘auteur’”
- The information is not in datasets “F” or “A”…
- …but was made available by:
  - merging datasets “A” and datasets “F”
  - adding three simple extra statements as an extra “glue”

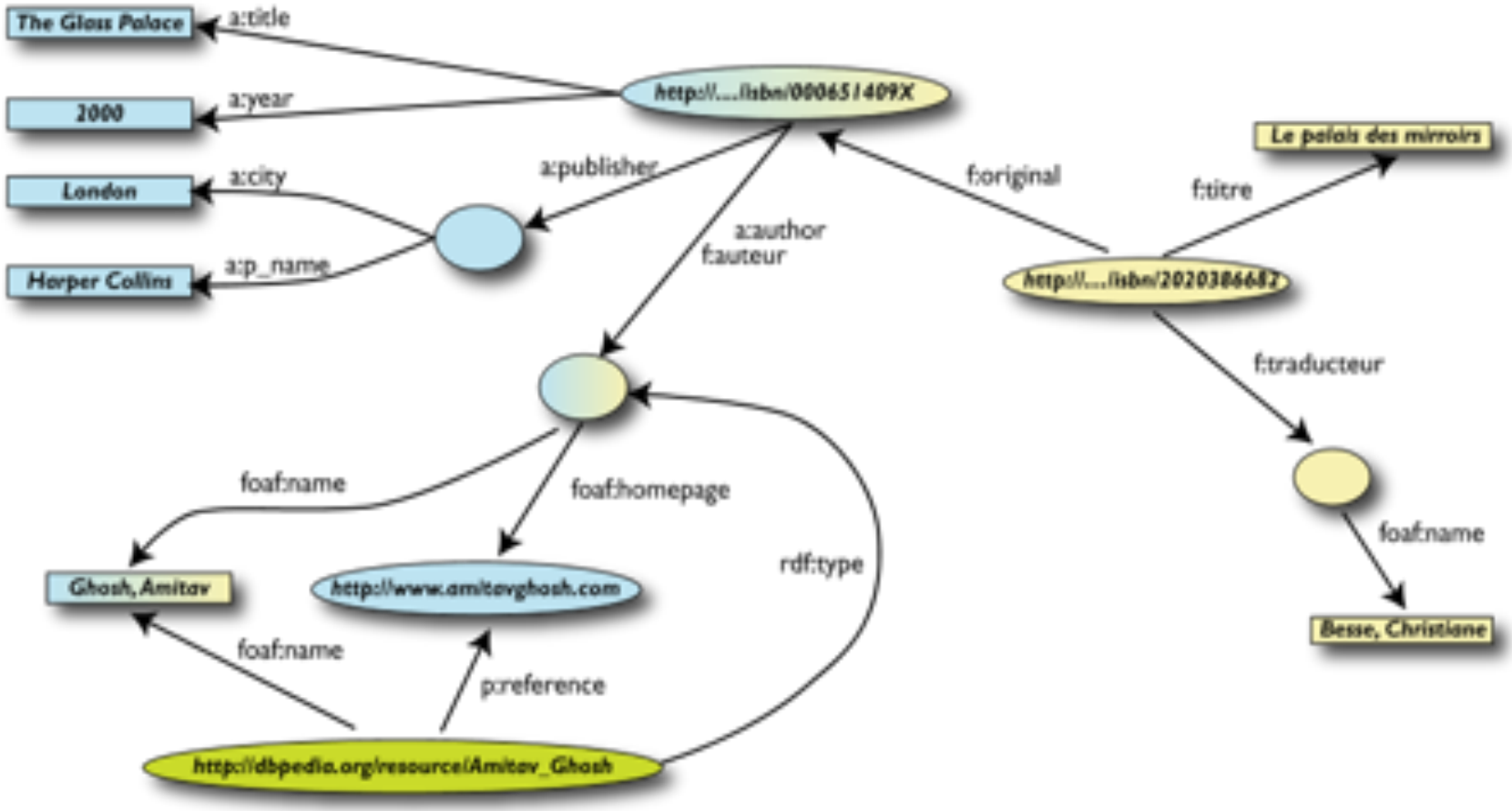




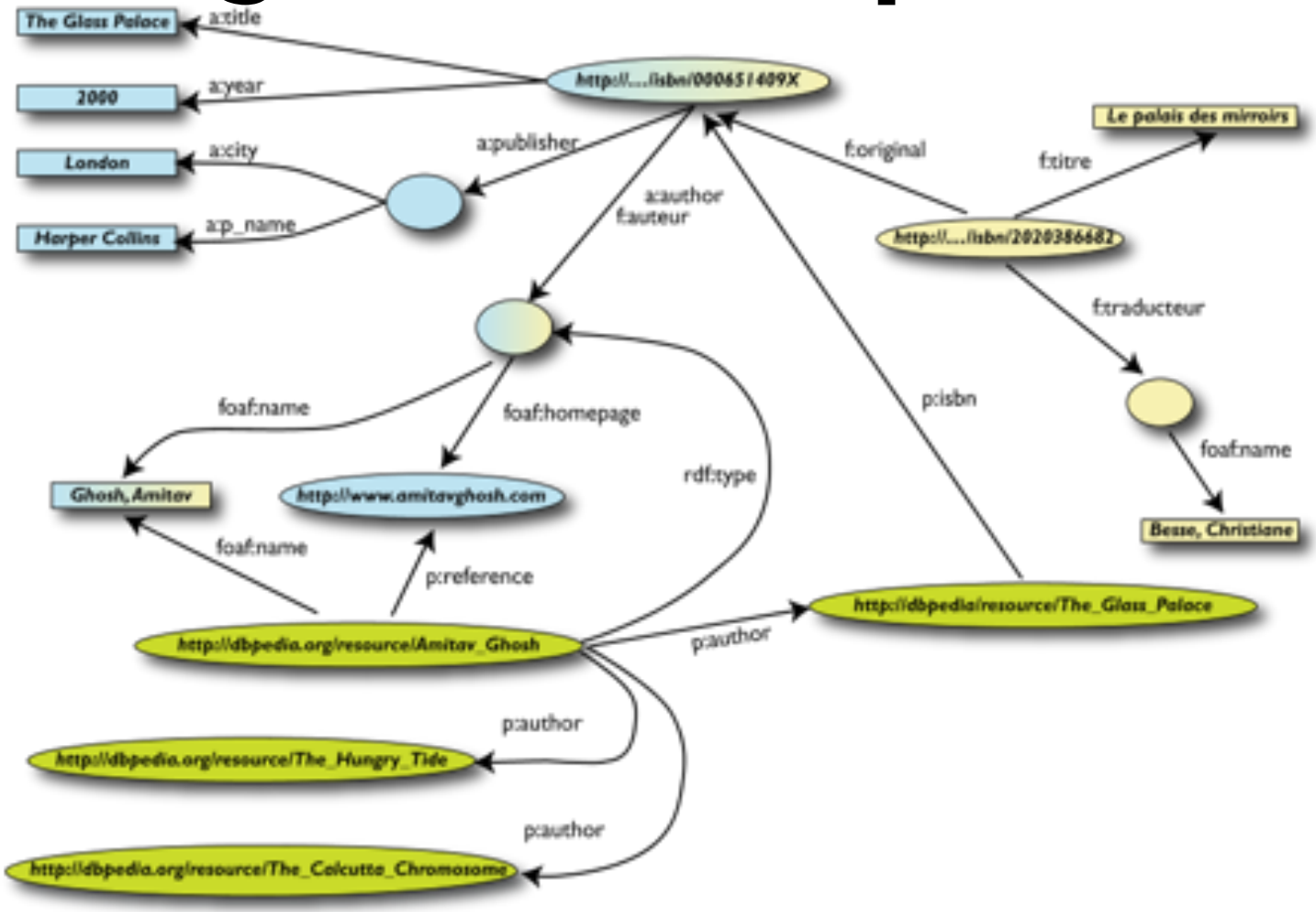
Combine With Different Datasets

- Using, e.g., the “Person”, the dataset can be combined with other sources
- For example, data in Wikipedia can be extracted using dedicated tools
  - e.g., the “dbpedia” project can extract the “infobox” information from Wikipedia already…

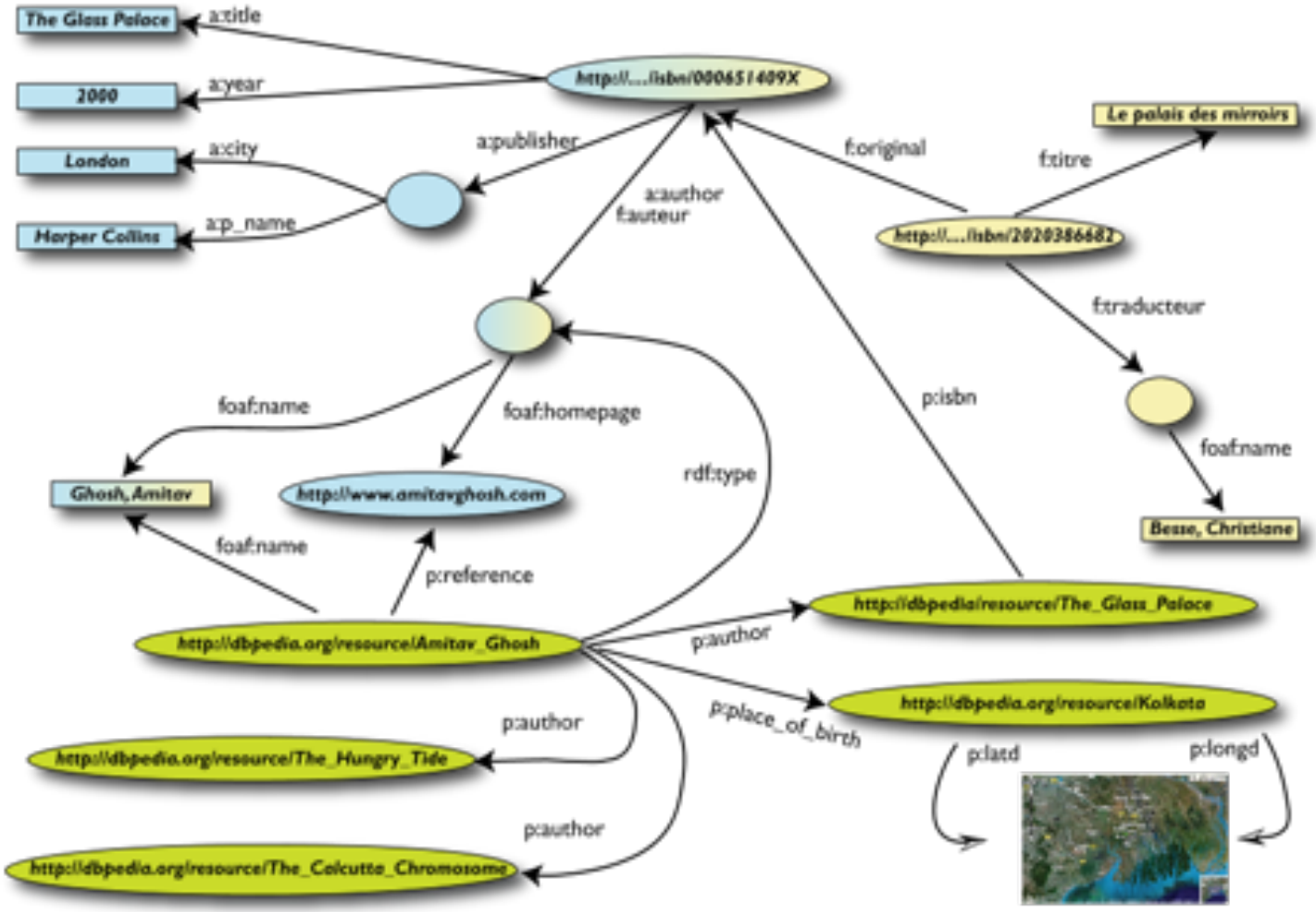
Merge With Wikipedia Data

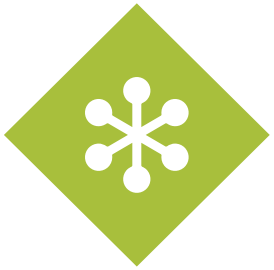


Merge With Wikipedia Data



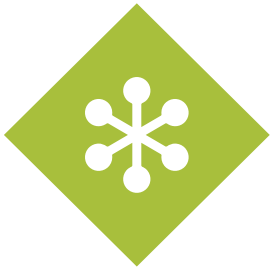
Merge With Wikipedia Data





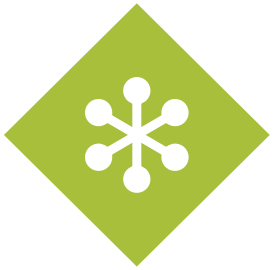
Is That Surprising?

- It may look like it but, in fact, it should not be…
- What happened via automatic means is done every day by Web users!
- The difference: a bit of extra rigour so that machines could do this, too



What Did We Do?

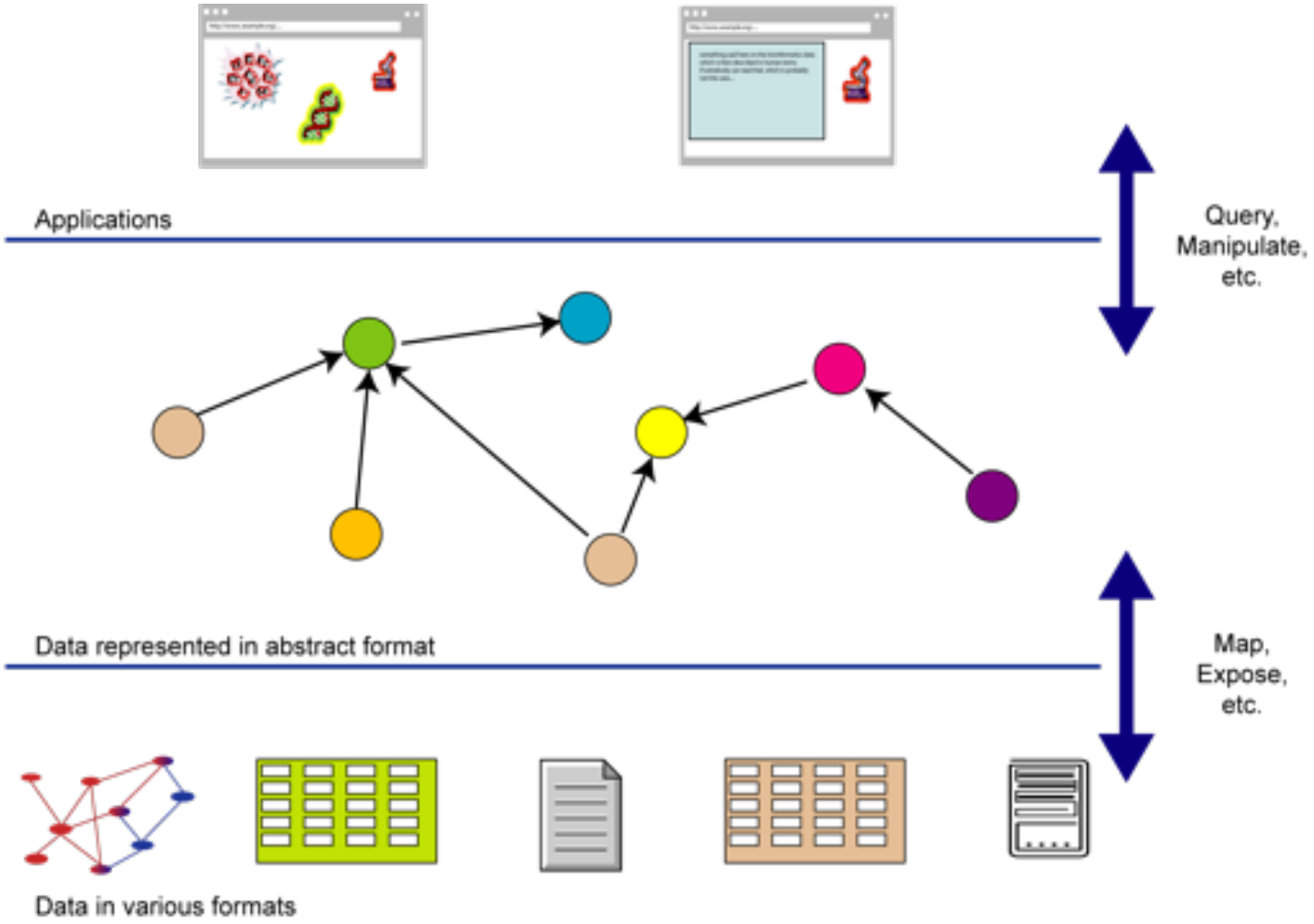
- We combined different datasets that
  - are somewhere on the web
  - are of different formats (mysql, excel sheet, XHTML, etc)
  - have different names for relations
- We could combine the data because some URI-s were identical (the ISBN-s in this case)
- We could add some simple additional information (the “glue”), possibly using common terminologies that a community has produced
- As a result, new relations could be found and retrieved



It Could Become Even More Powerful

- We could add extra knowledge to the merged datasets
  - e.g., a full classification of various types of library data
  - geographical information
  - etc.
- This is where ontologies, extra rules, etc, come in
  - ontologies/rule sets can be relatively simple and small, or huge, or anything in between…
- Even more powerful queries can be asked as a result

What Did We Do? (Cont)



The Semantic Web - A Web of Data

„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, James Hendler, Ora Lassila: The Semantic Web, Scientific American, 284(5), pp. 34-43(2001)

The Semantic Web - A Web of Data

- The meaning of information (Semantics) is made explicit by **formal (structured) and standardized knowledge representations (Ontologies).**
- Thus it will be possible,
  - To process the meaning of information automatically
  - To relate and integrate heterogeneous data
  - To deduce implicit (not evident) information from existing (evident) information in an automated way.
- The Semantic Web is kind of a **global database** that contains a **universal network of semantic propositions.**

The Take Away

Simple Take Away

- If we use standard languages and semantic markup for our data on the web
  - we allow for easier discovery and reuse of data, even automatically by machines/ agents
  - we can build "smarter applications"
- Semantic Web technologies are not ONLY for the open Web