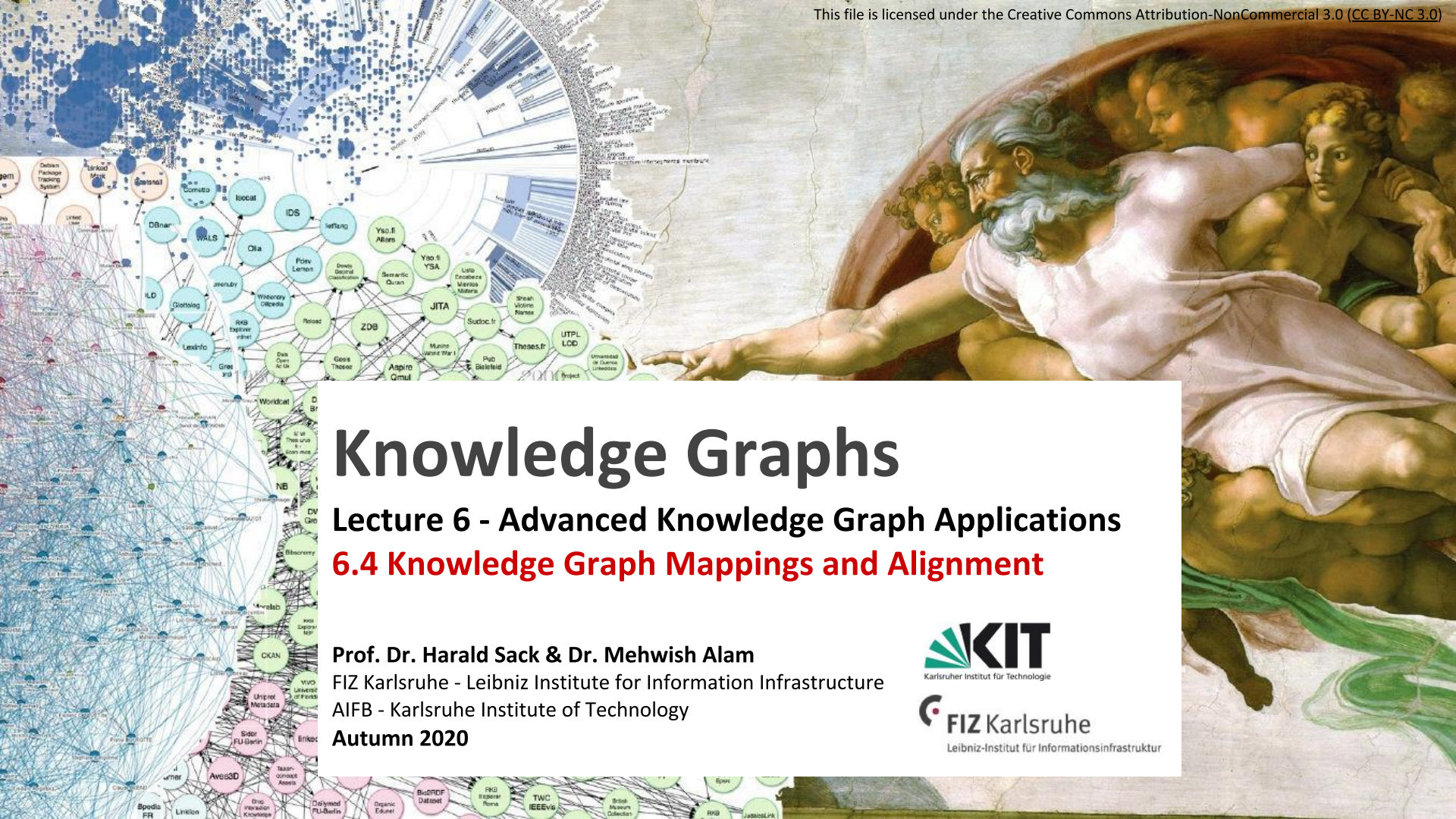


Knowledge Graphs

Lecture 6 - Advanced Knowledge Graph Applications

6.4 Knowledge Graph Mappings and Alignment

Prof. Dr. Harald Sack & Dr. Mehwish Alam
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure
AIFB - Karlsruhe Institute of Technology
Autumn 2020

Karlsruher Institut für Technologie

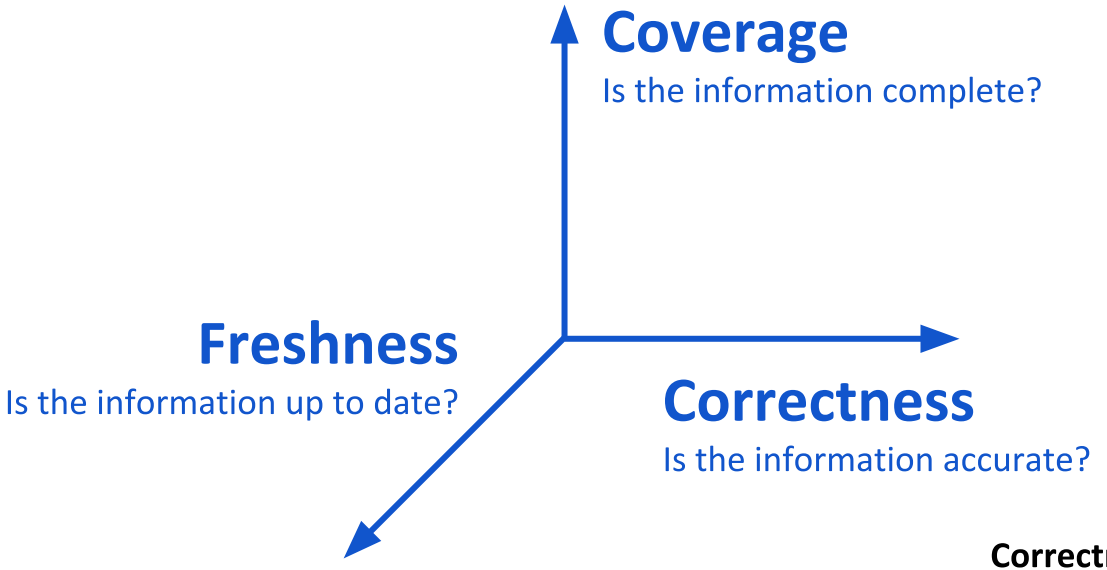
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Knowledge Graphs

Lecture 6: Advanced Knowledge Graph Applications

Knowledge Graph Challenges

- Building a small KG is easy but building a vast system like Google Knowledge Graph is a huge challenge



Increase **Freshness** & **Coverage**

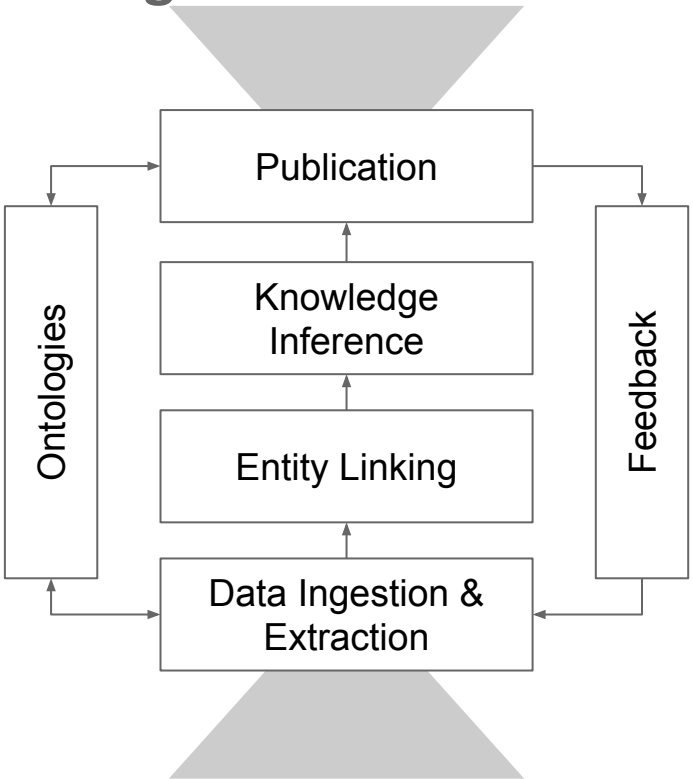
Harder to ensure **Correctness**

Increase **Correctness**

Harder to ensure **Freshness** & **Coverage**

Correctness is always hard – what is true and correct?

Towards Automated Knowledge Graph Management



- Unsupervised knowledge extraction from unstructured data in open domain
- Semantic embedding via Ontologies
- Ultra-scale knowledge representations
- Large scale entity linking and disambiguation
- Autonomous knowledge inference & verification
- Knowledge Graph versioning and archiving
- Knowledge Precision vs Comprehensiveness

How to Automate Knowledge Graph Construction?

- Sound **Knowledge Graph Construction** relies on **Ontologies**
- Ontologies don't come for free, i.e. Ontology Design is very expensive wrt. time and resources
- Ontologies can be „learned" automatically

- **Ontology Learning** defines a set of methods and techniques
  - for **fundamental development** of new ontologies
  - for **extension or adaption** of already existing ontologies
- in a (partly) automated way from various resources.

Fundamental Types of Ontology Learning

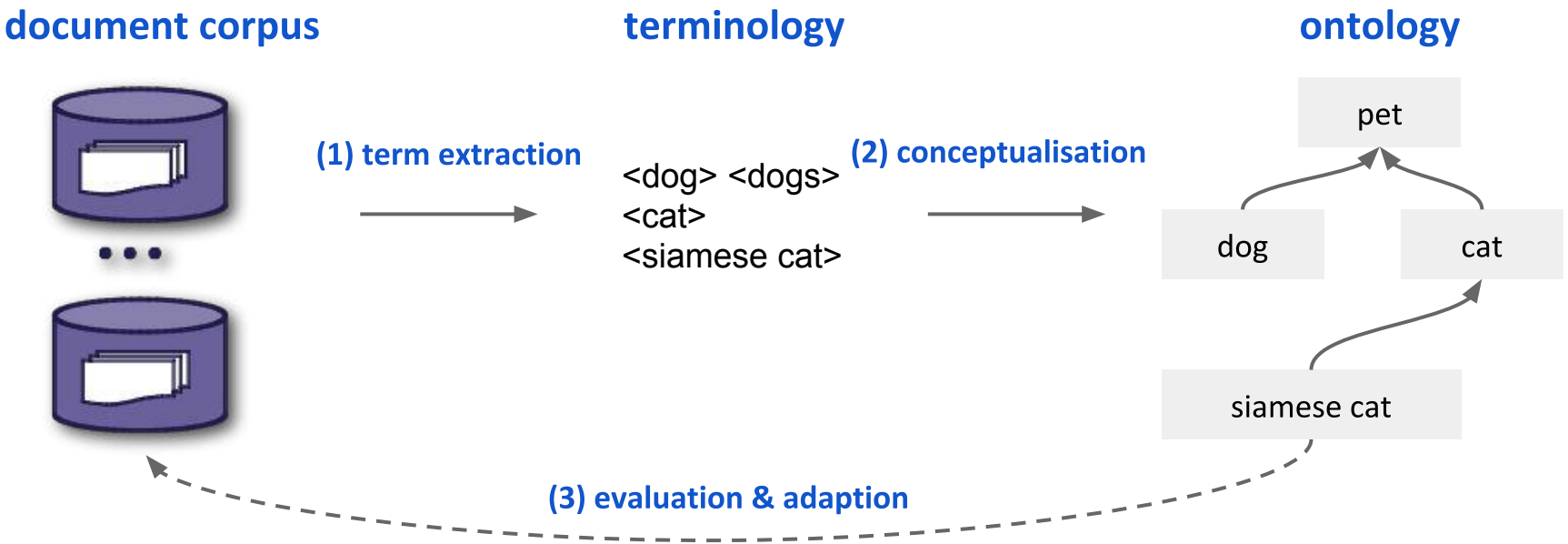
- Ontology Learning from Text
  - automatic or semi-automatic generation of lightweight ontologies by means of text mining and information extraction
- Linked Data Mining
  - detecting meaningful patterns in RDF graphs via statistical schema induction or statistical relational learning
- Concept Learning in Description Logics and OWL
  - learning schema axioms from existing ontologies and instance data mostly based on Inductive Logic Programming
- Crowdsourcing Ontologies
  - combines the speed of computers with the accuracy of humans, as e.g. taxonomy construction via Amazon Turk or games with a purpose

Ontology Learning from Text

- **Ontology Learning from text** is the process of identifying terms, concepts, relations, and optionally axioms from textual information and using them to construct and maintain an ontology.

- Automatisation requires help from
  - Natural Language Processing (NLP)
  - Data Mining
  - Machine Learning techniques (ML)
  - Information Retrieval (IR)

Ontology Learning from Text - Basic Approach

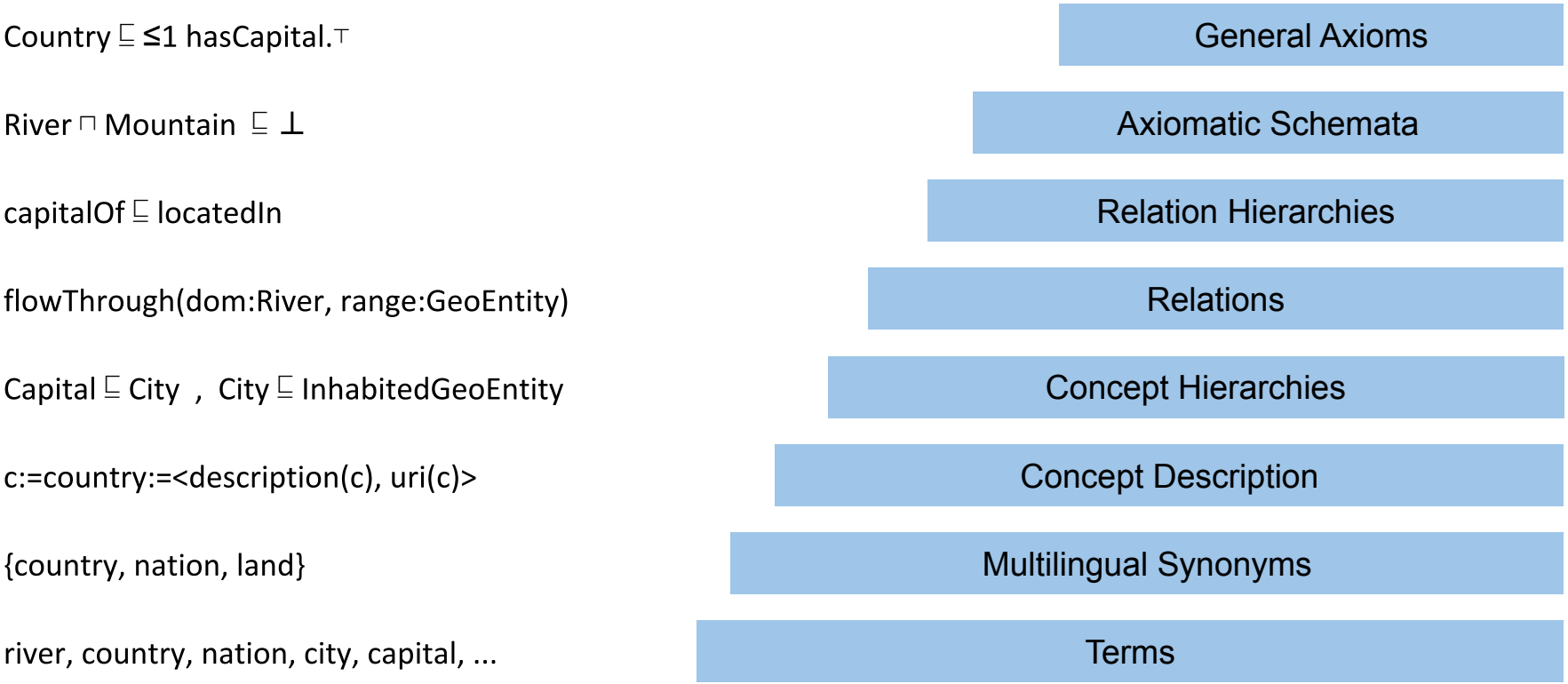


term extractions requires linguistic processing (NLP) to identify important noun phrases and their internal semantic structure

terms: linguistic realisations of domain specific concepts

Concepts: clusters of semantically related terms

The Ontology Learning Layer Cake



Ontologies are NOT the Reality

- Ontologies are a **context-dependent projection (model) of the Reality**
- **Different ontologies** might represent the **same (or similar) knowledge**, as e.g. ontologies might
  - reflect different tasks and requirements for applications
  - follow different conventions or restrictions

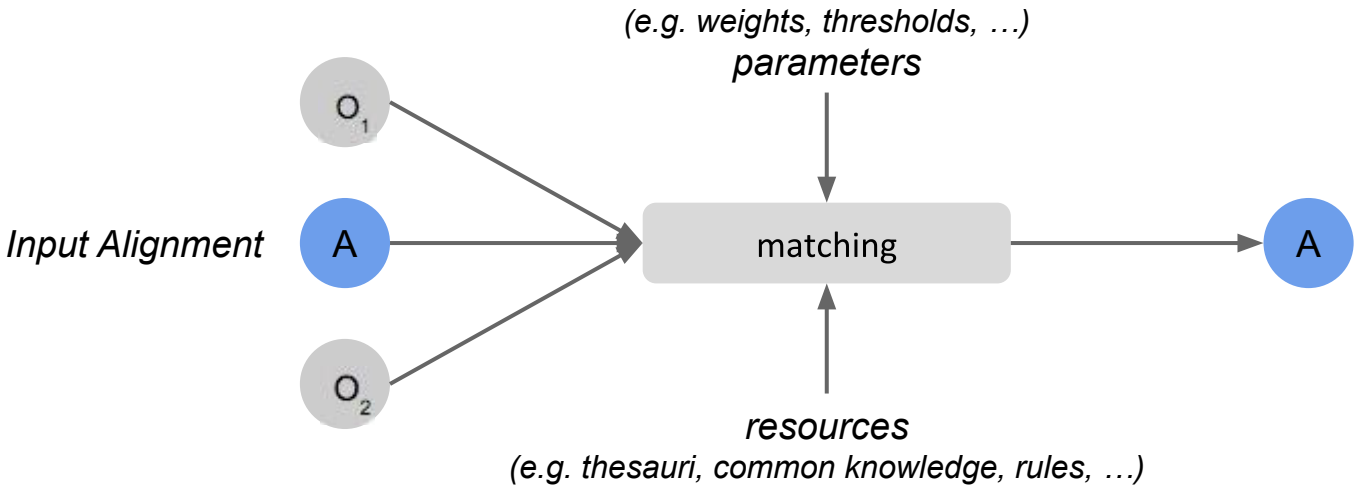
[1]

How Ontologies can Differ

- The **same term describes different concepts**
  - e.g. `Author` - *writer of a book vs. creator of a document*
- **Different terms describe the same concept**
  - e.g. `Author` vs. `Writer`
- **Different modeling conventions and paradigms**
  - e.g. `intervals` vs. `points` - *to describe temporal aspects*
- **Different level of granularity**
  - e.g. `Fiction` vs. `PoliticalFiction`, `ScienceFiction`, `RomanticFiction`, *etc. as literary Genres*
- **Different coverage or different point of view,** etc.

Ontology Alignment

- **Ontology Alignment** or **Ontology Matching** is the process of determining *correspondences* between ontological concepts



Correspondence or Mapping

- Given the ontologies O_1 and O_2, a **correspondence** or **mapping** among the entities e_1 and e_2 from O_1 and O_2 respectively, is defined as

$$\langle id, e_1, e_2, r, n \rangle$$

- with
  - **id** ... a unique **identifier** of the correspondence
  - **r** ... a **relation**, as e.g. equivalence (=), more general ($\sqsupseteq$,≥), less general ($\sqsubseteq$,≤), disjointness($\perp$), part-of, etc...
  - **n** ... a **confidence measure** (typically in the range of [0,1]) holding for the correspondence between e_1 and e_2
- the correspondence $\langle id,e_1,e_2,r,n \rangle$ asserts that the relation r holds between the entities e_1 and e_2 with confidence n

Complexity of Correspondences

- Examples of **simple correspondences**:
  - http://dbpedia.org/resource/Joseph_Fourier = https://www.wikidata.org/wiki/Q8772
  - Author = Writer
  - Gas $\geq_{1.0}$ GreenhouseGas
  - rdfs:label $\geq_{0.9}$ dc:title

Complexity of Correspondences

- Examples of **more complex correspondences**:
  - $speed = velocity \times 2.237$

    $0.477 \times speed = velocity$
  - $Book(x) \wedge author(x,y) \wedge Writer(y) \Rightarrow_{.85} writtenBy(x,concat(y.firstname, y.lastname))$

Alignment Example

Book =1.0 Volume

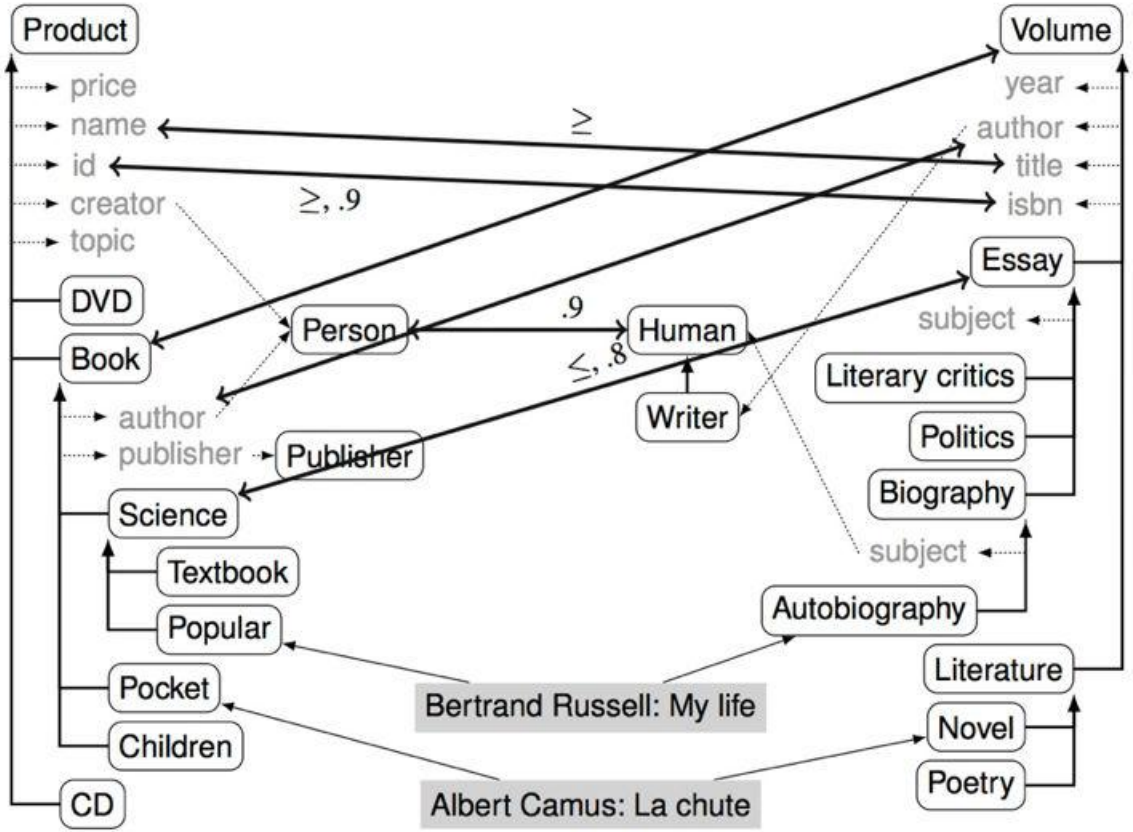
id ≥0.9 isbd

Person =0.9 Human

name ≥1.0 title

author =1.0 author

Science ≤0.9 Essay



Jérôme Euzenat, Pavel Shvaiko: Ontology Matching, Springer, 2007, p.48

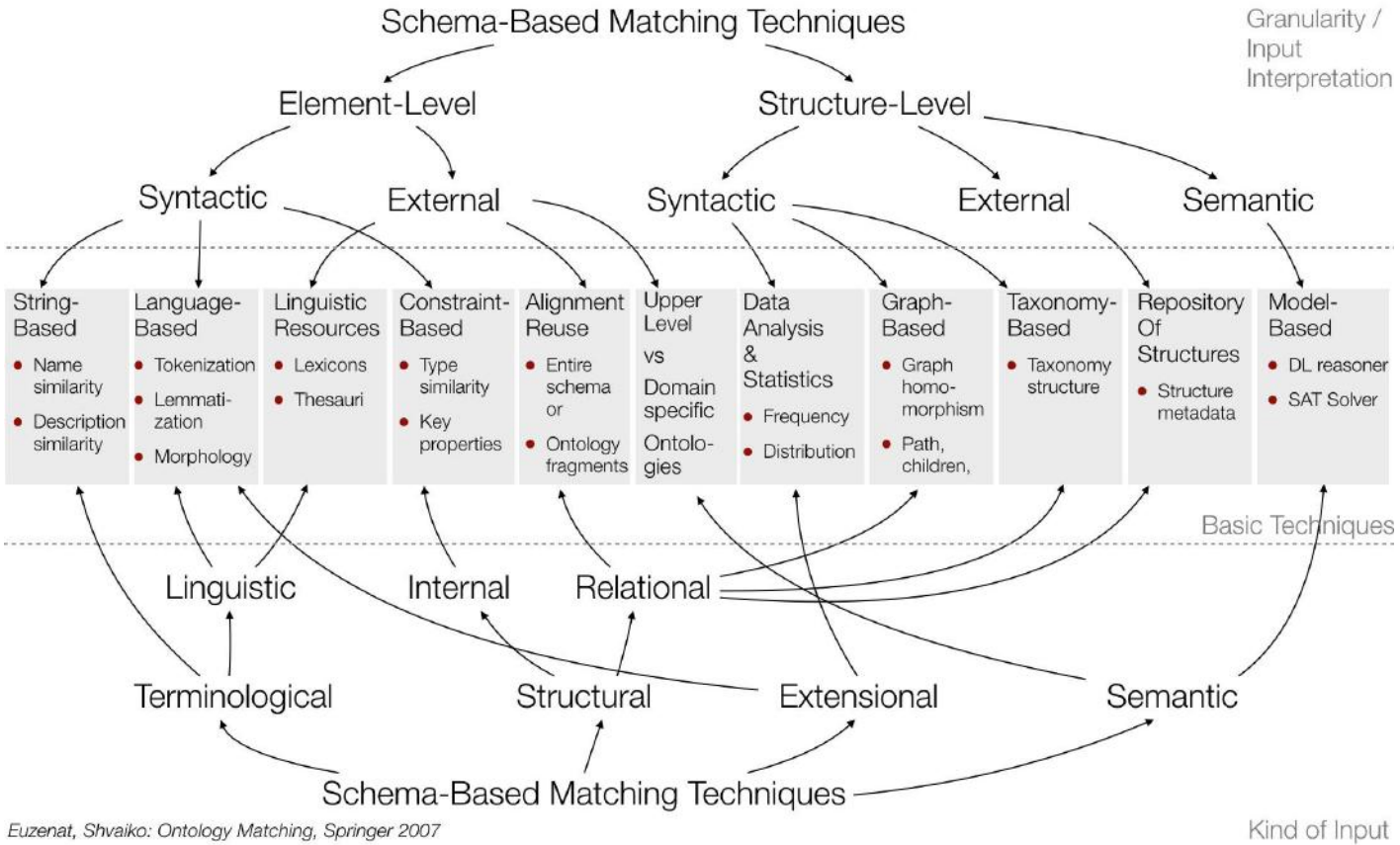
Ontology Matching Techniques

- **Element-level Ontology Matching Techniques** consider ontology entities or their instances in isolation from their relations with other entities or their instances
  - **String-Based** - *matching names or descriptions of entities*
  - **Linguistic-Based** - *use NLP, lexicons, or domain specific thesauri to match words based on linguistic relations (homonymy, synonymy, partonomy, etc.), or exploiting morphological properties*
  - **Constrained-Based** - *take into account internal constraints applied to the definitions of entities, as e.g. types, cardinality of properties, etc.*
  - **Extensional-Based** - *use individual representation of classes, i.e. classes are considered similar if they share many instances*

Ontology Matching Techniques

- **Structure-level Ontology Matching Techniques** consider ontology entities or their instances to compare their relations with other entities or their instances
  - **Graph-Based** - *consider ontologies as labeled graphs, assumption: if nodes are similar, then also their neighbors must be similar*
  - **Taxonomy-Based** - *like graph-based algorithms, but consider only specialization/generalization relation*
  - **Method-Based** - *take into account semantic interpretation of the ontologies, assumption: if two entities are the same, then they share the same interpretation*
  - **Data Analysis and Statistics** - *take a large sample, try to find regularities, discrepancies, allows grouping or determining distance metrics, ...*

Ontology Alignment



Euzenat, Shvaiko: Ontology Matching, Springer 2007

Next Lecture...

[2]

Picture References:

- [1] Francisco Goya: The Sleep of Reason Produces Monsters, 1797-98, [Public Domain]
  https://commons.wikimedia.org/wiki/File:Goya-Capricho-43.jpg
- [2] "Indiae Orientalis, Insularumque Adiacientium Typus," from the Theatrum Orbis Terrarum of Abraham Ortelius, 1603 Latin edition
  https://commons.wikimedia.org/wiki/File:Two_fearsome_sea_monsters_attack_a_disabled_ship_off_the_shore_of_America,_which_is_shown_in_close_proximity_to_Japan.jpg