deepschema.org: An Ontology for Typing Entities in the Web of Data

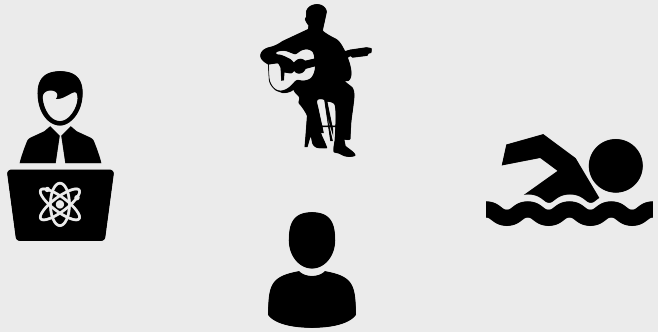
Panayiotis Smeros, Amit Gupta, Michele Catasta and Karl Aberer

LDOW '17. 3 April, 2017. Perth, WA, Australia.

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

What is deepschema.org?

- Class Hierarchy, Taxonomy
- Describes all the **possible environments** than an entity exists in



- Comprises information from **Wikidata** and **schema.org**
- Main features:
  - generic, cross-domain
  - rich, evolving as fast as the Web
  - traversable
  - accurate

Related Work

YAGO Taxonomy

✓ rich, evolving as fast as the Web

× traversable

✓ accurate

<wikicat_People_murdered_in_British_Columbia>
rdfs:subClassOf <wordnet_person_100007846> .

Dbpedia Ontology

× rich, evolving as fast as the Web

✓ traversable

✓ accurate

Manually constructed;
Only 685 classes

Wikidata Class Hierarchy

✓ rich, evolving as fast as the Web

? traversable

✓ accurate

Crowdsourced schema;
No tree structure

schema.org

? rich, evolving as fast as the Web

✓ traversable

✓ accurate

deepschema.org

Manually constructed;
Used by billions of web pages

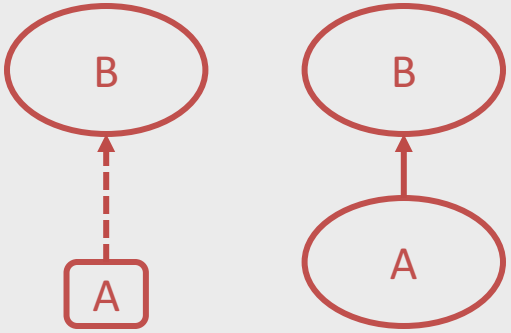
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Wikidata Class Hierarchy

Wikidata Class Hierarchy

Extraction Phase

RDFS Entailment Rules



A rdf:type B ⇒ class(B)
A rdfs:subClassOf B ⇒ class(A) ∧ class(B) ∧ subclass(A, B)

rdfs:subClassOf ≡ P279 (subclass of)
rdfs:type ≡ P31 (instance of)

Filtering Phase

Ontologies from domain-specific KBs

Classes with no English label

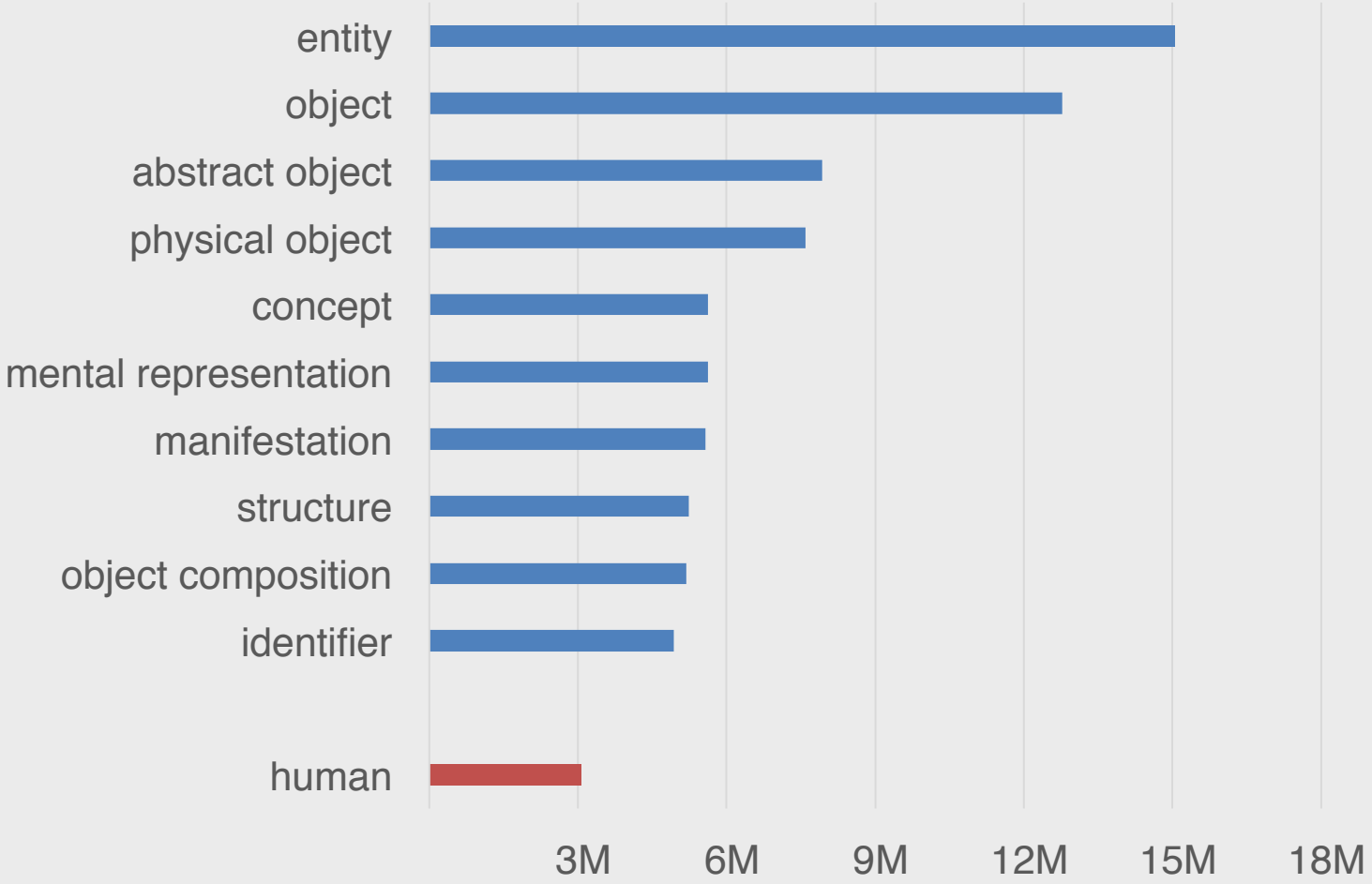
Keep deepschema.org generic

Structure of the Class Hierarchy

Classes	**123,033**
Subclass Relations	**126,688**
Disconnected Subgraphs	**4,263**
Root Classes	14,084
Leaf Classes	102,434
Avg num of Subclasses per Class	1.03
Avg depth of Hierarchy	**7.93**

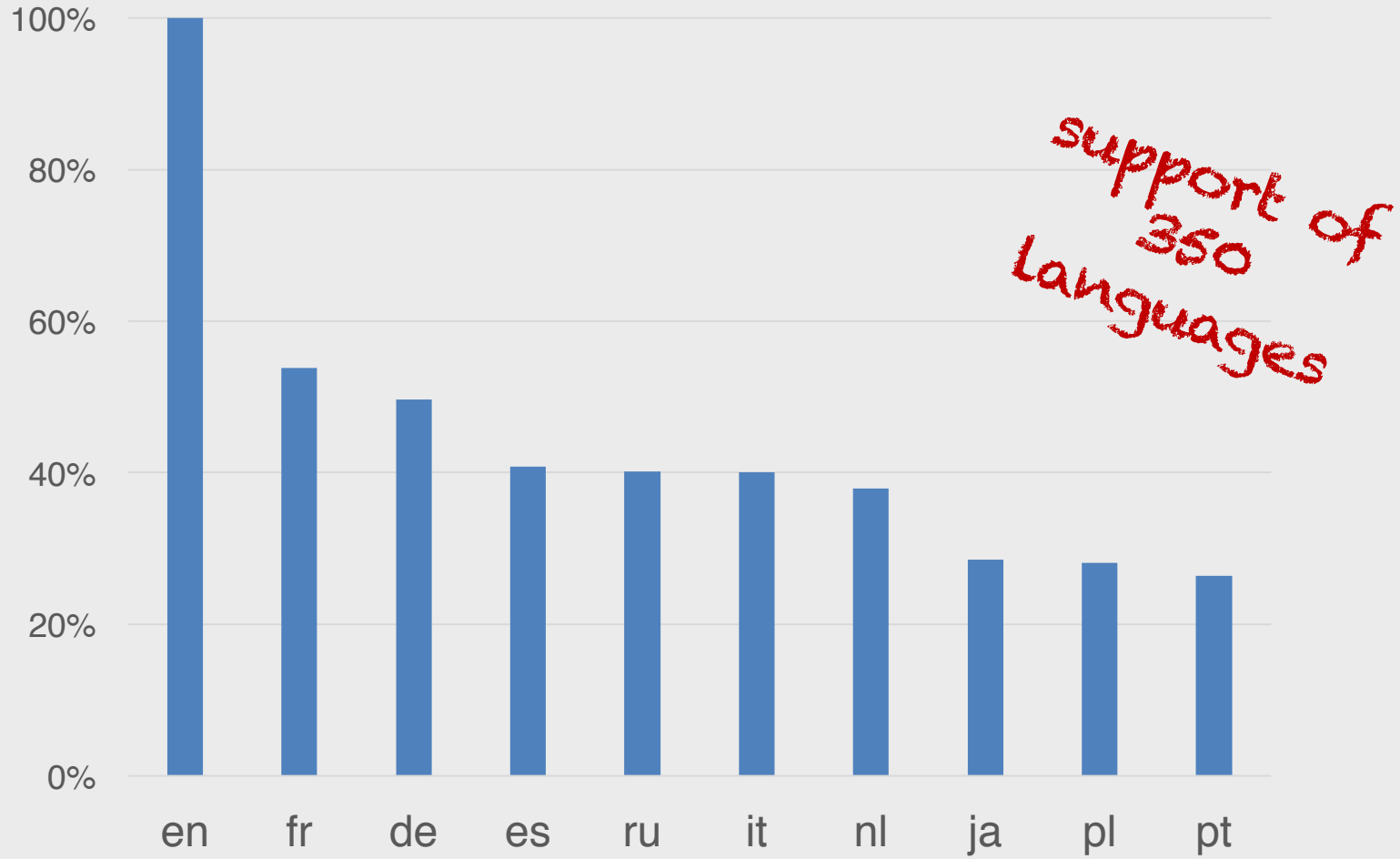
One subgraph contains **96%** of the total classes and **97%** of the total subclass relations

Number of Instances per Wikidata Class



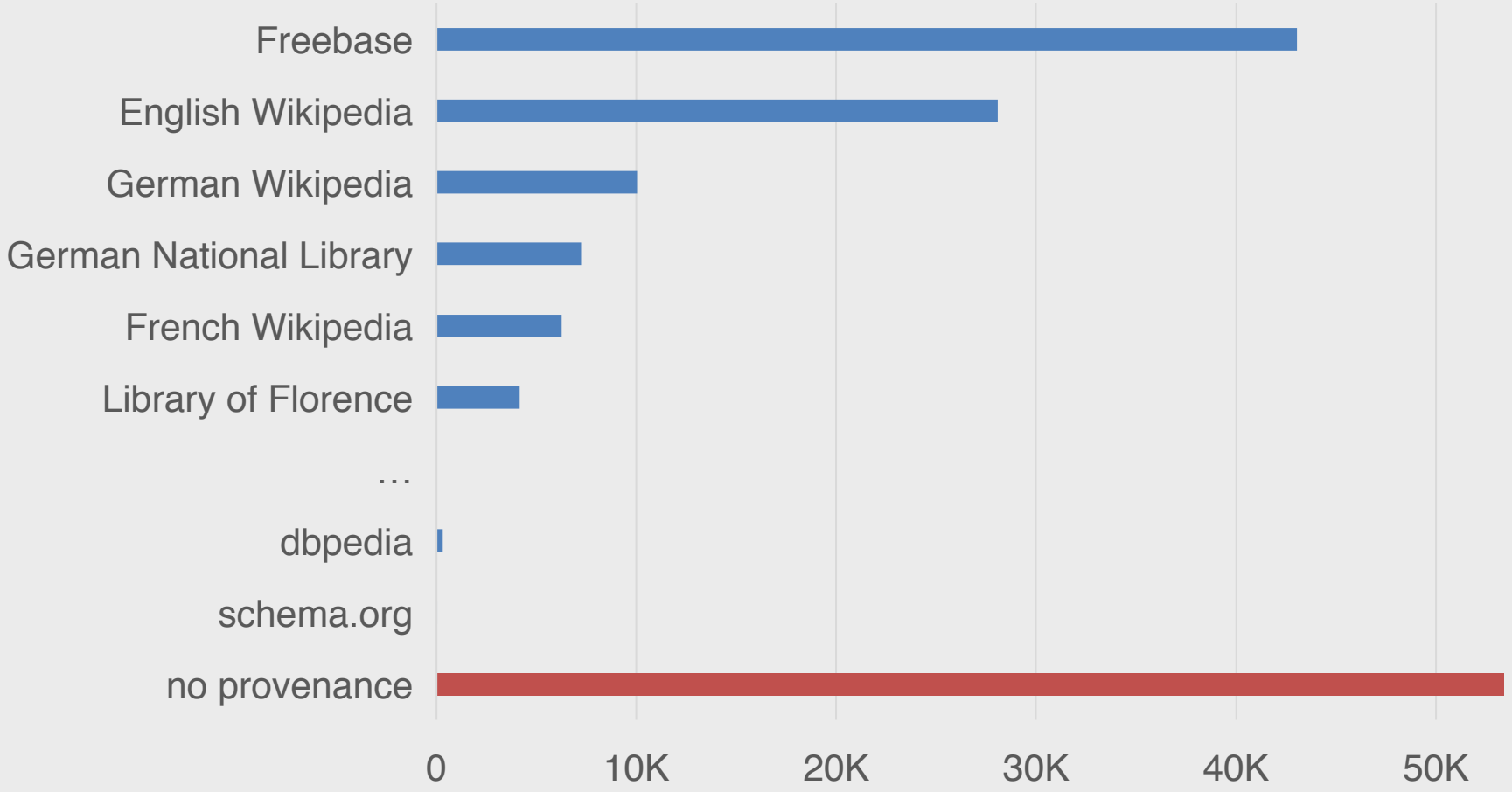
The first class with more **direct** than **inherited** instances is ***human***

Language Coverage of Wikidata Classes



For a ***non-English*** class hierarchy we get ***at least ~50%*** loss of information

Number of Classes per Wikidata External Contributor



50% of the classes has ***no provenance*** information

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Integration with schema.org

Integration Heuristics

Exact Match

(language)^{w} → (Language)^{s}

Lemma Match

(languages)^{w} → (Language)^{s}

Head Match

(Kalapuyan languages)^{w} → (Language)^{s}

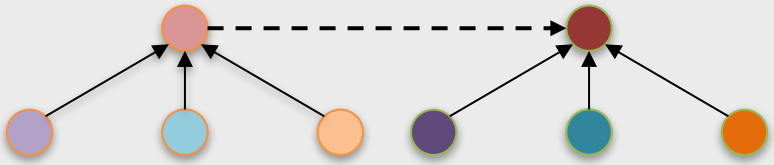
Vector Cosine Similarity

(warehouse)^{w} → (Store)^{s}

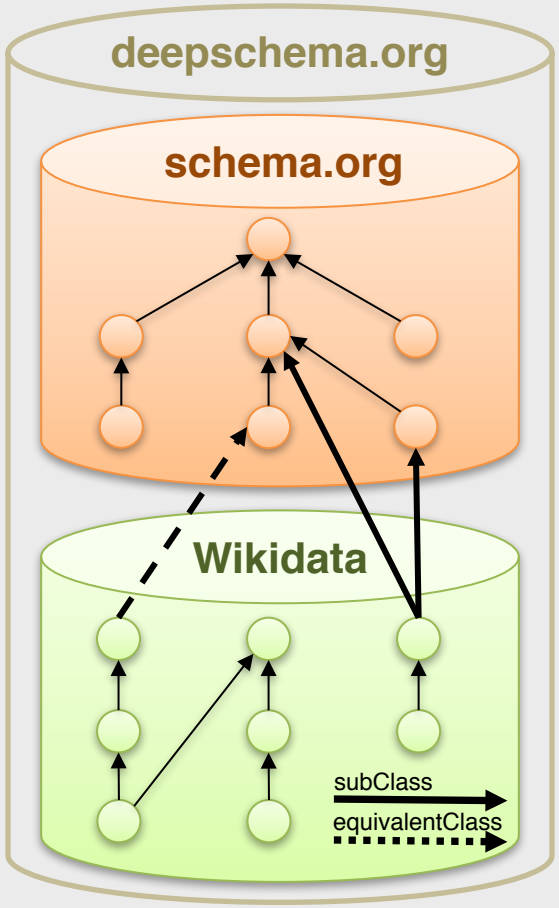
Head Similarity

(Survey motor boat)^{w} → (Vessel)^{s}

Subclass/Instance Similarity



Resulting pairs



Cosine Similarity Threshold	# of pairs
0.5	15,112
0.6	8,494
0.7	7,329
0.8	6,120
0.9	5,586

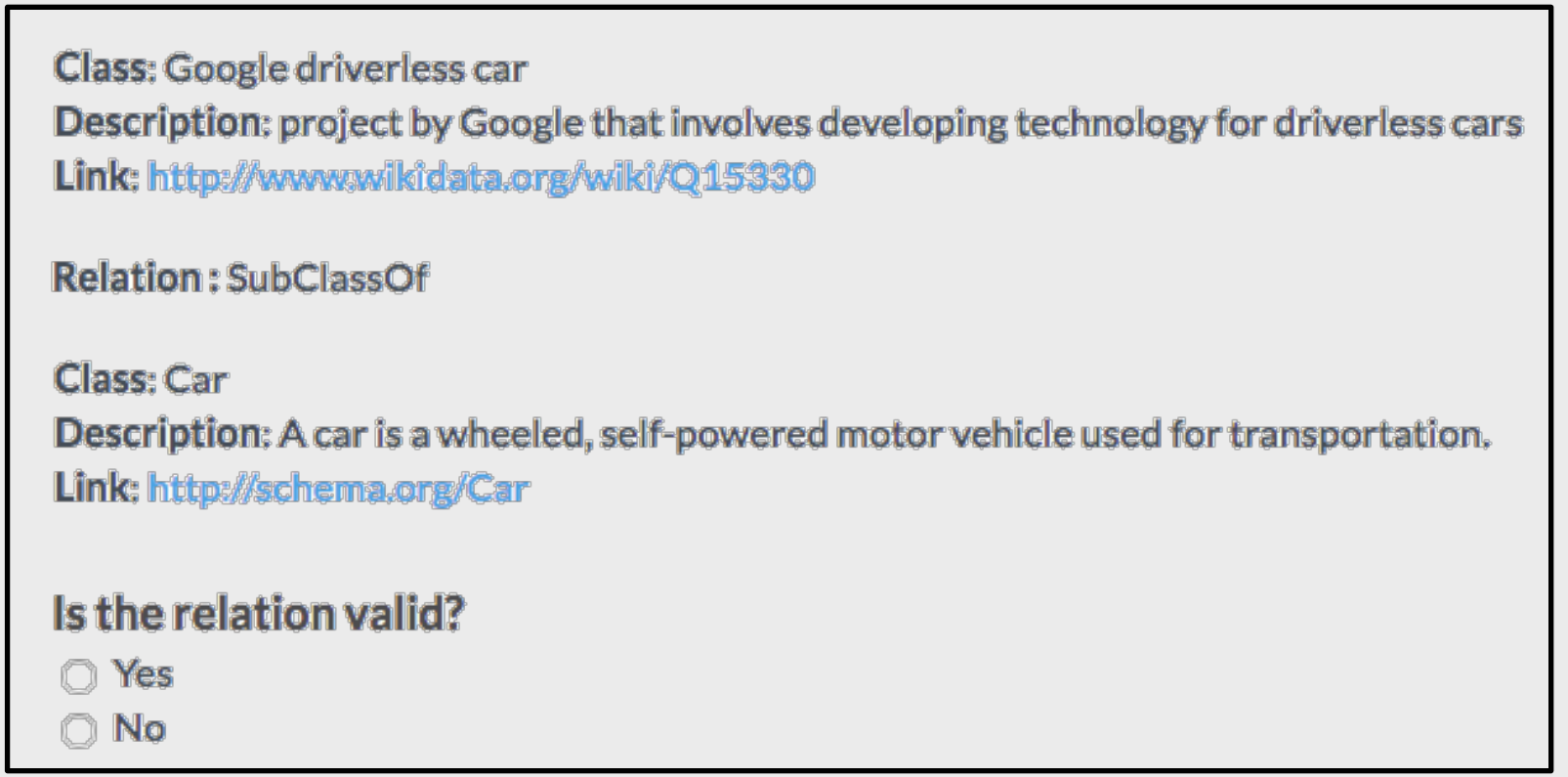
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Evaluation

(Accuracy, Traversability and Genericity)

Accuracy: Crowdsourcing Evaluation

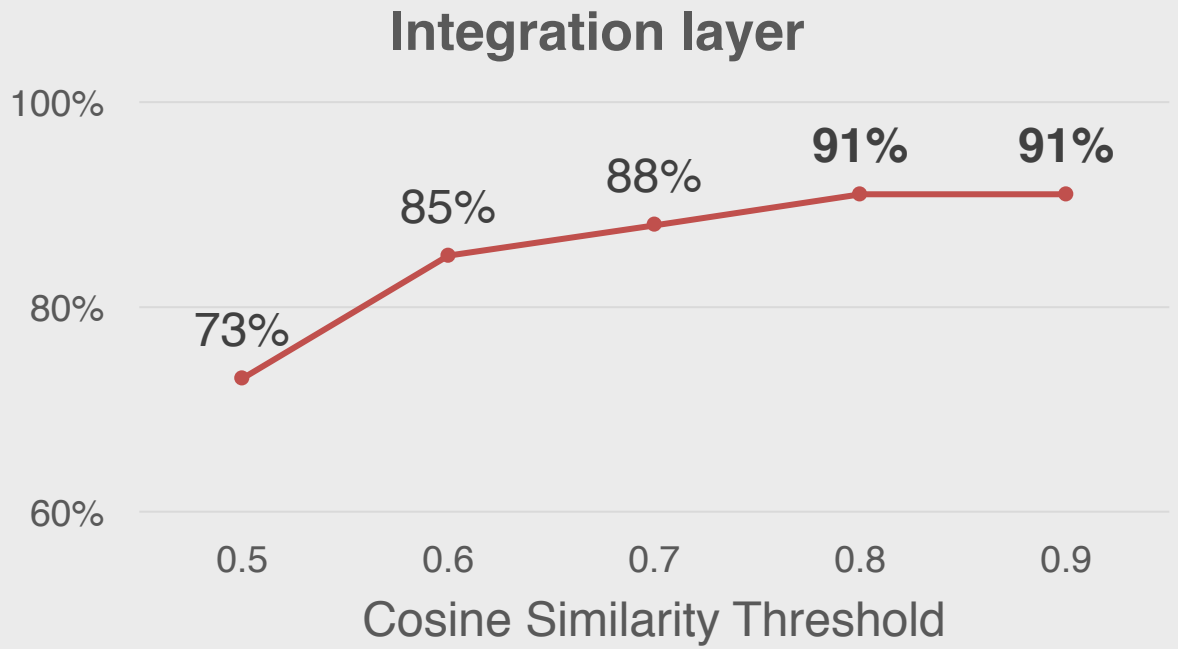
- CrowdFlower Platform
- ~100 workers
- Majority Voting (2 out of 3)



Accuracy: Crowdsourcing Evaluation

Wikidata: **92%**

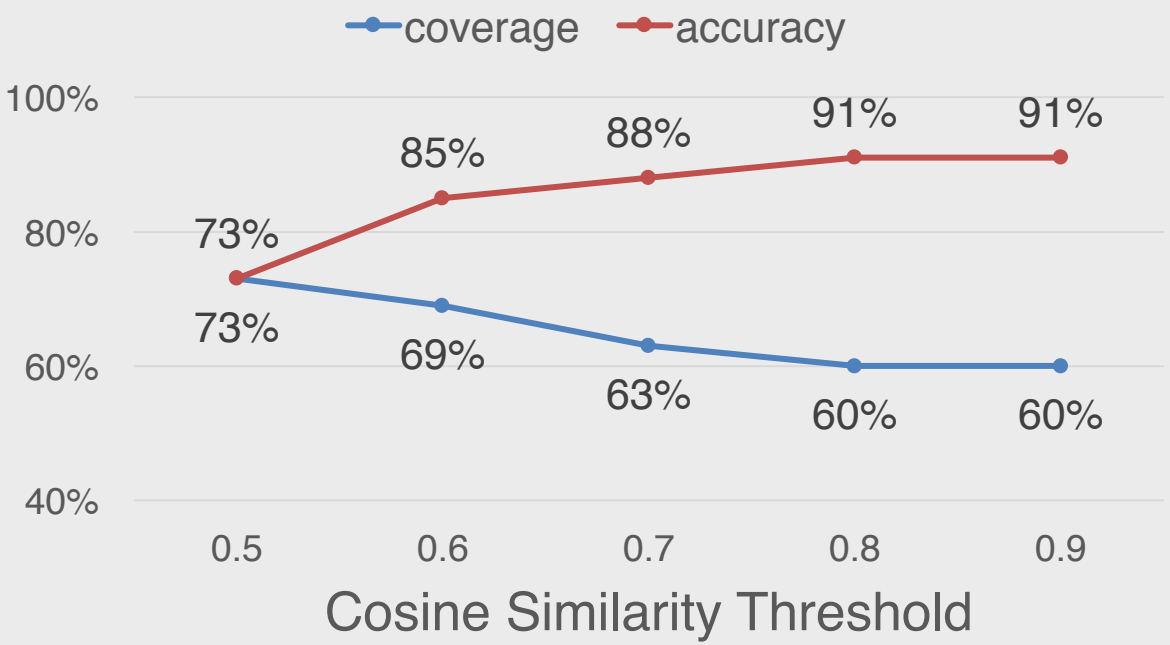
schema.org: **100%** (assumed)



- *YAGO (Wikipedia Categories - WordNet):* ***95%***
- *PARIS (YAGO - DBpedia):* ***90%***

Comparable with integration techniques for similar data sources

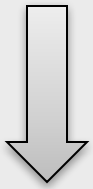
Traversability



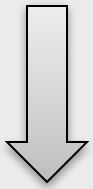
- False class definition ☹
  - **partOf** relations interpreted into **subclassOf**
- Topic not covered by schema.org ☺
  - Class **Child Abuse**

Genericity

Oxford English Dictionary



3000 most frequent English words



Nouns and Noun Phrases

81% coverage

Conclusions & Future Work

deepschema.org

- ✓ generic, cross-domain
- ✓ rich, evolving as fast as the Web
- ✓ traversable
- ✓ accurate

- Integrate more data sources
  - Facebook's Open Graph
- Include customizable filters
  - Control the granularity of the ontology
  - Choose specific topics of interest
- Employ deepschema.org in use-cases
  - Discover the appropriate type (class) of an entity, given a context

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Thanks for your attention!
Questions?