

Hierarchical Target Type Identification for Entity-oriented Queries

Krisztian Balog and Robert Neumayer

Norwegian University of Science and Technology, Trondheim, Norway

Resources are available at <http://bit.ly/SdIpbZn>

Hierarchical target type identification

Given an entity-oriented input query, find the single most specific type from an ontology that is general enough to cover all entities that are relevant to the query.

Queries

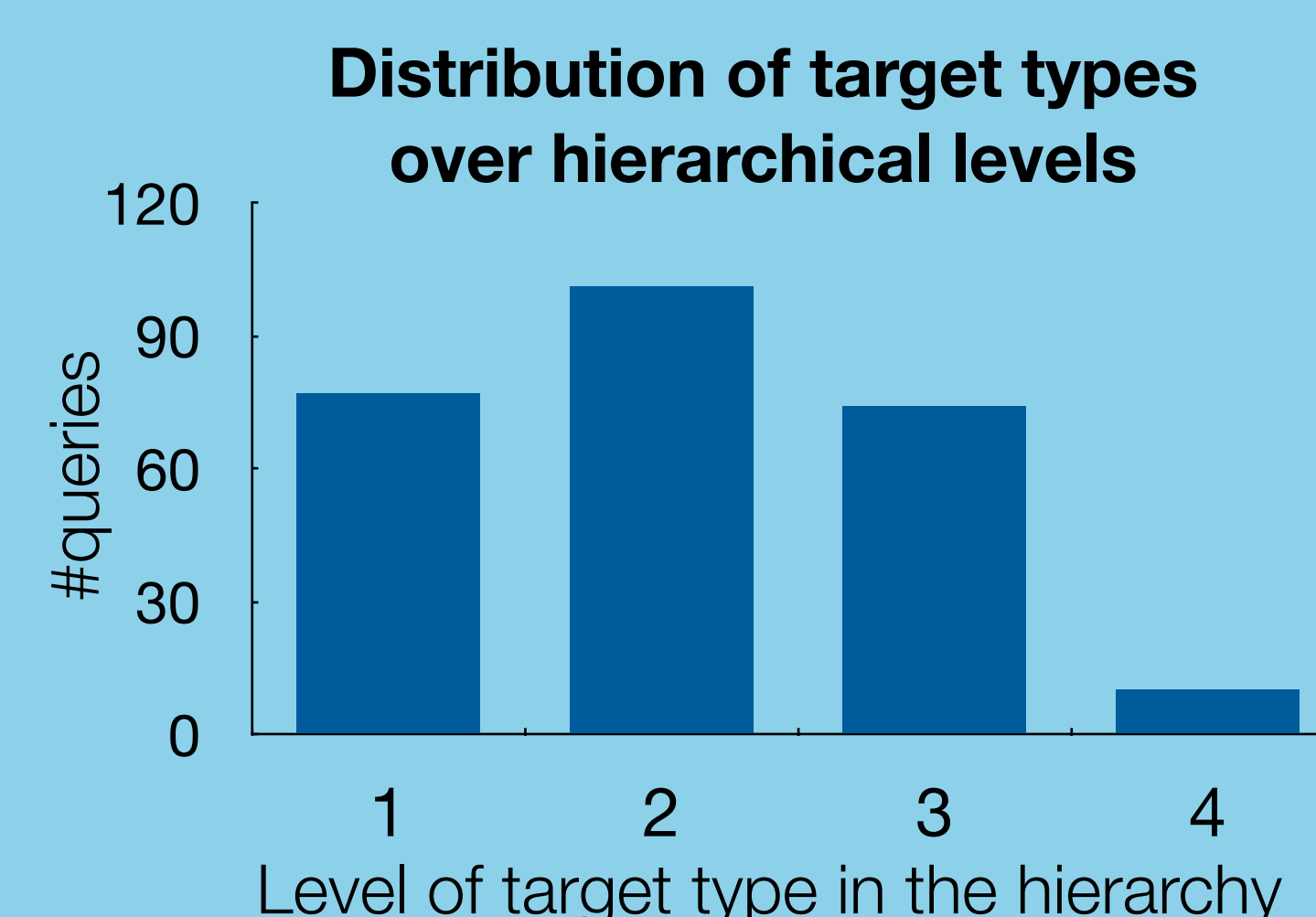
Collected from a number of recent benchmarking evaluation efforts:

Evaluation effort	Example
TREC Entity track 2009-11 (120) focus on specific relationships between entities	Airlines that currently use Boeing 747 planes
INEX Entity Ranking track 2009 (55) seek a list of entities	US presidents since 1960
SemSearch Challenge 2010-11, ES task (142) refer to a particular entity	Ben Franklin
SemSearch Challenge 2011, LS task (50) target a group of entities	Axis powers of World War II

Annotation

Queries are labeled with types from the DBpedia Ontology.

Query type	Count
Single target type	262
Multiple (top-level) target types	30
Target type missing from ontology	46
Query not interpretable	29



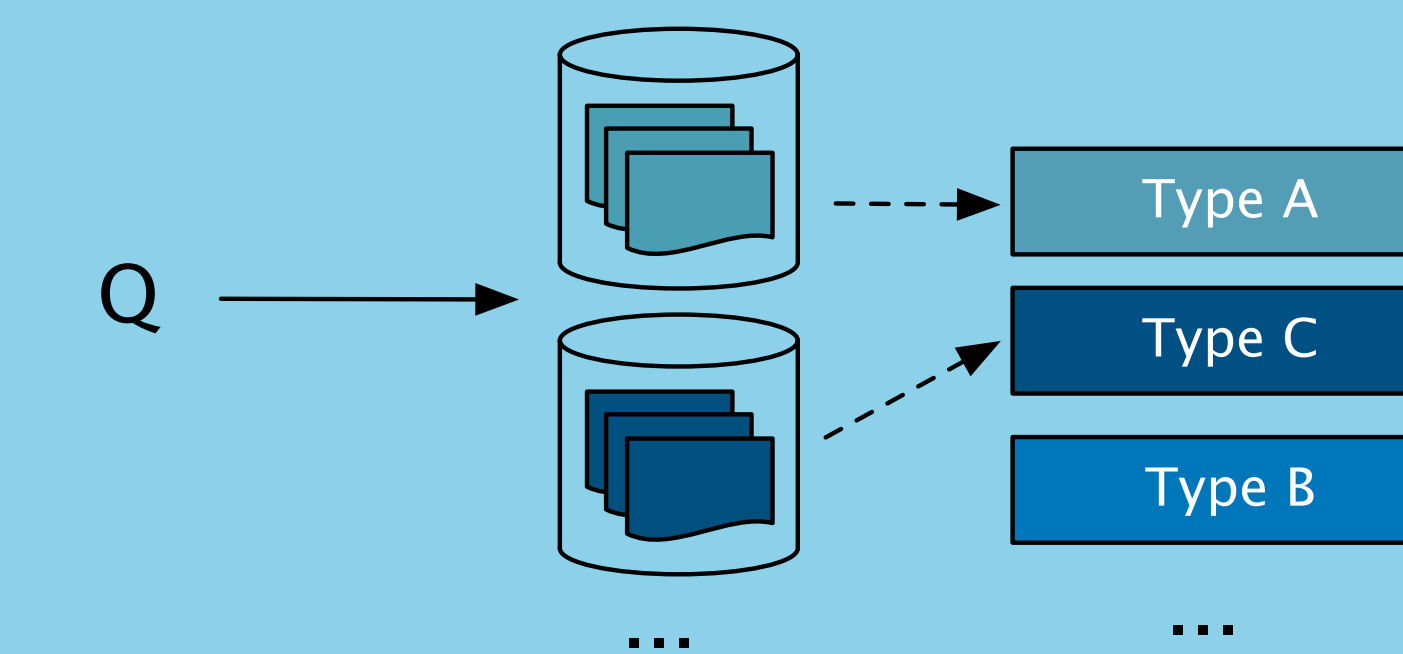
Evaluation metrics

- *Strict*: binary judgments, only the correct answer is credited
- *Lenient*: graded relevance, rewarding near-misses too
 - Two variants: linear and exponential decay

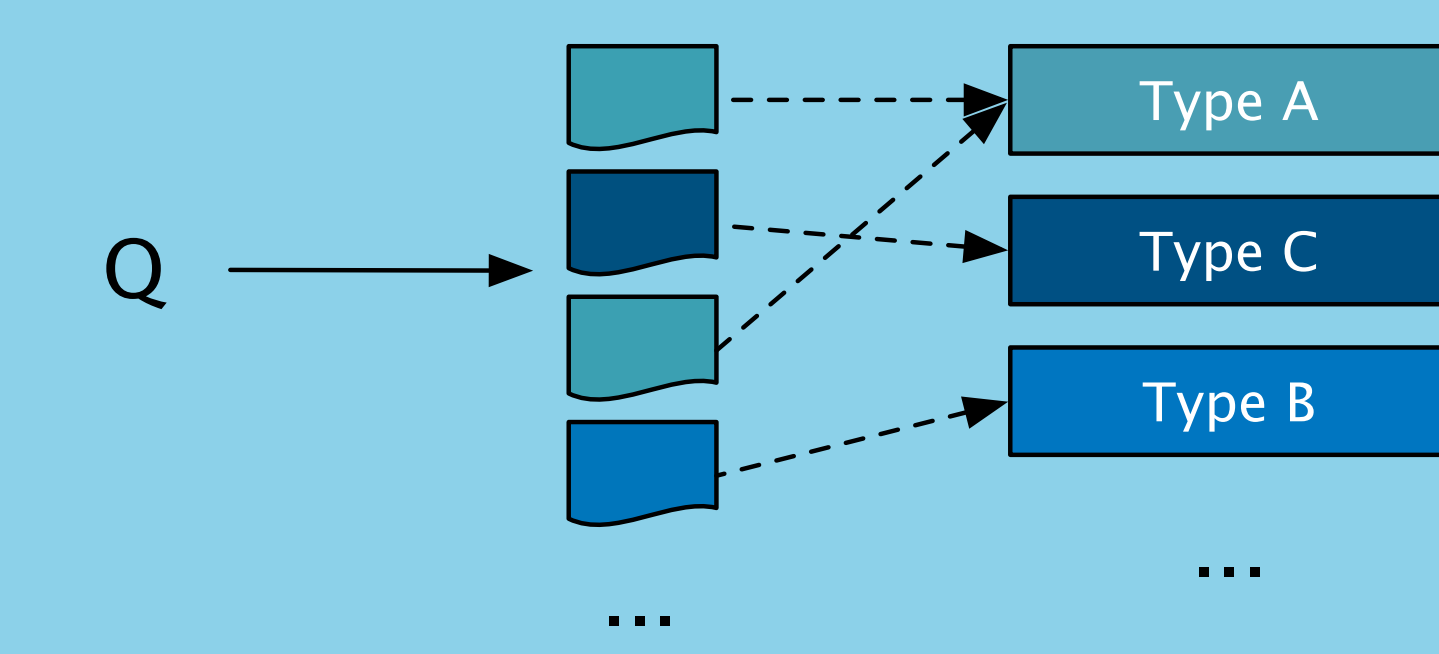
Baseline methods

- Relying on the typical labeling of entities in a knowledge base
- Adapting methods from federated search

Type-centric model



Entity-centric model



Query type		Top-lv.		Hierarchical			Linear decay		Exp. decay	
		MRR	MRR	nDCG@1	nDCG@5		nDCG@1	nDCG@5	nDCG@1	nDCG@5
Type-centric	Strict	0.528	0.299	0.192	0.309	Lenient	0.327	0.344	0.261	0.329
Entity-centric		0.695	0.351	0.163	0.397		0.414	0.454	0.294	0.417

Findings

- Even simple baselines can find the top-level type with high accuracy
- Finding the right granularity in the hierarchical case is difficult
 - Type-centric model tends to return more specific categories
 - Entity-centric model assigns more general types
- Results suggest the combination of the two strategies (future work)