

Problem

KBs are incomplete

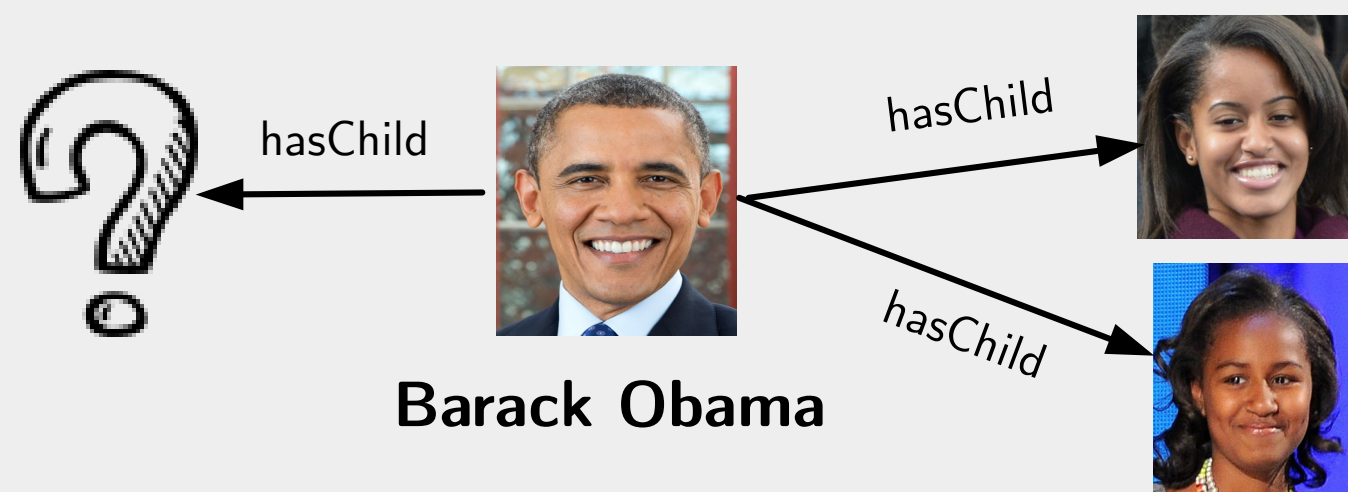
- 2% of people in YAGO have a known citizenship.
- No guarantees that queries on KBs return complete results.

KBs do not know how much they know

- A person without a spouse in the KB could be single or her spouse unknown.
- Data producers do not know where to focus information extraction efforts.

Completeness

Given the *real-world* KB K^* , a query q is complete in a KB K iff $q(K^*) \subseteq q(K)$.



We focus on queries like:

`SELECT ?x WHERE { Barack Obama hasChild ?x }`

We want to predict if K knows all the results of the query.

Completeness oracles

Simple

Closed World Assumption oracle: $cwa(s, r)$

Baseline oracle: The KB is complete.

Partial Completeness Assumption oracle: $pca(s, r)$

(s, r) is complete if the KB knows at least one object.

Popularity: $popular_{pop}(s, r)$

(s, r) is complete if s is among the top 5% entities with most entities in the KB.

No change: $nochange_{pop}(s, r)$

(s, r) is complete if the objects of (s, r) have not changed w.r.t. a previous version of the KB.

Parameterized

Star oracle: $star_{r_1, \dots, r_n}(s, r)$

(s, r) is complete if we know object values for other properties $r_1, \dots, r_n$ of s .

`producer(x, z), writer(x, w) → complete(x, director)`

Class oracle: $class_c(s, r)$

The KB is complete for entities in class C .

`Pope(x) → complete(x, hasChild)`

AMIE oracle

It uses Horn rules [1] combining all other oracles to predict completeness. In case of contradictions, the rule with higher support and confidence prevails.

`President(x), moreThan0(x, hasChild) → complete(x, hasChild)`

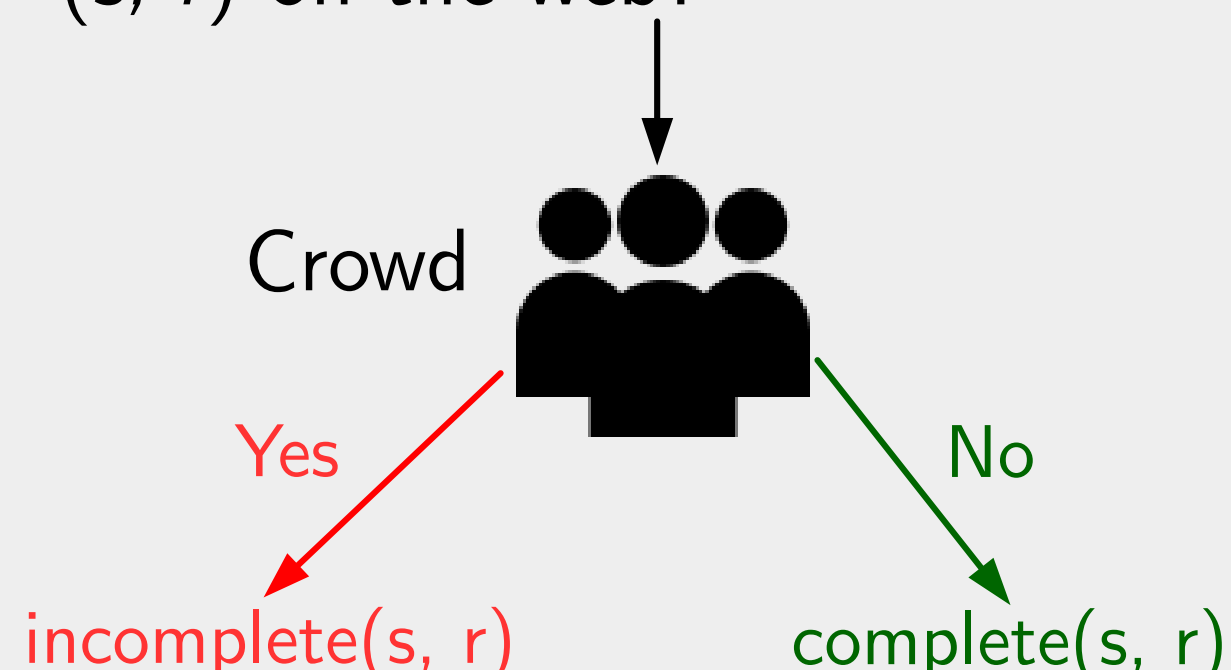
`dateOfDeath(x, y), lessThan1(x, placeOfDeath) → incomplete(x, placeOfDeath)`

Experimental evaluation

Training data



Can you find more objects for (s, r) on the web?



F1-measure on YAGO

Relation	CWA	PCA	Pop.	N. chg	Star	Class	AMIE
diedIn	60%	22%	4%	15%	50%	99%	96%
directed	40%	96%	7%	71%	0%	0%	100%
graduatedFrom	89%	4%	2%	10%	89%	92%	87%
hasChild	71%	1%	2%	13%	40%	78%	78%
hasGender	78%	100%	2%	-	86%	95%	100%
hasParent	1%	54%	-	-	0%	0%	100%
isCitizenOf	4%	98%	1%	4%	10%	5%	100%
isConnectedTo	87%	34%	-	-	68%	88%	89%
isMarriedTo	55%	7%	3%	12%	37%	57%	46%
wasBornIn	28%	100%	5%	8%	0%	0%	100%