

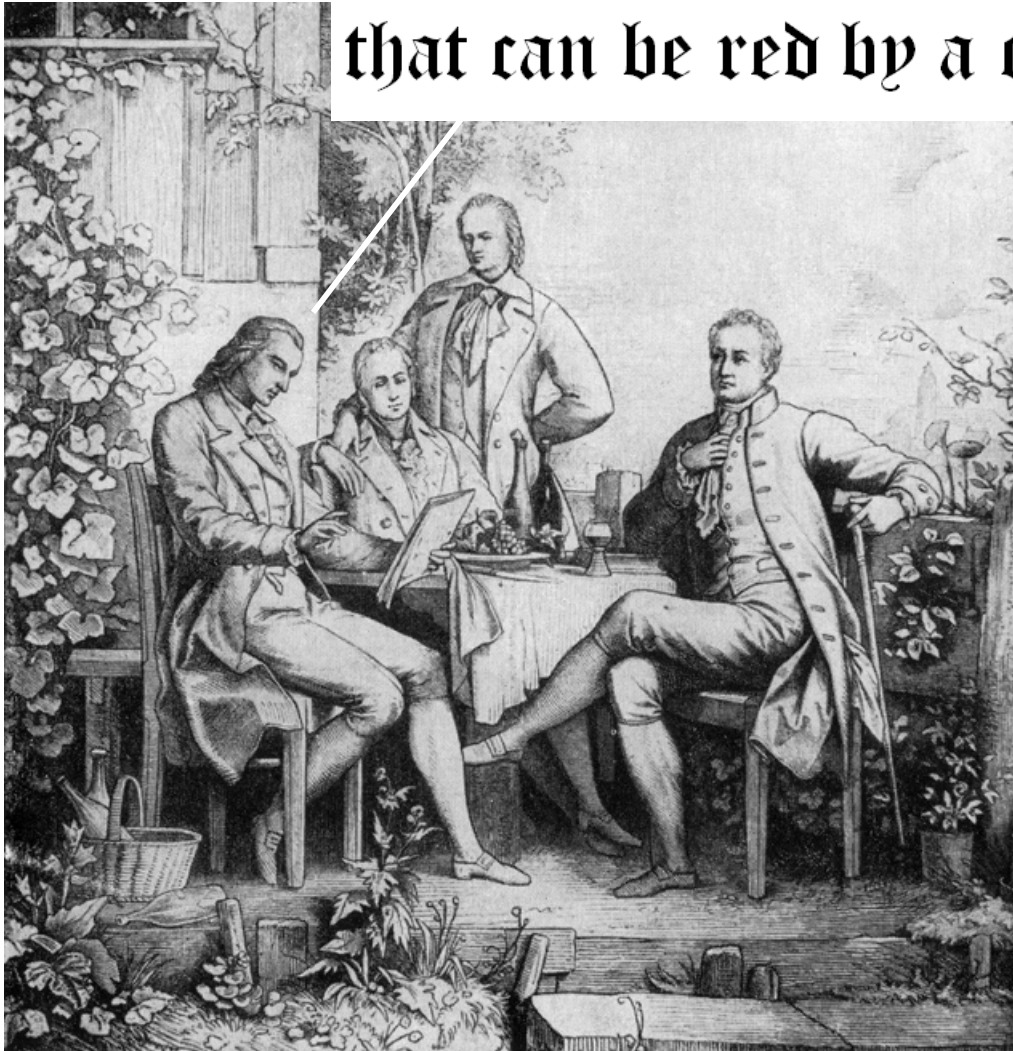
The YAGO Knowledge Base

Fabian M. Suchanek
Télécom ParisTech University
Paris, France

See the [original SVG slides](#)

A bit of history...

There shalleth be a base of wysdom
that can be red by a computing devis.



Unknown painter

A.D. 2007

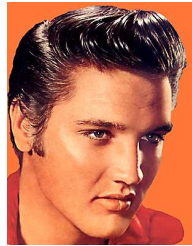
YAGO: First large KB from Wikipedia

Elvis Presley



WIKIPEDIA
The Free Encyclopedia

Elvis Presley was one of the best blah blah blub blah don't read this, listen to the speaker! blah blah blah blubl blah you are still reading this! blah blah blah blah blabbel blah



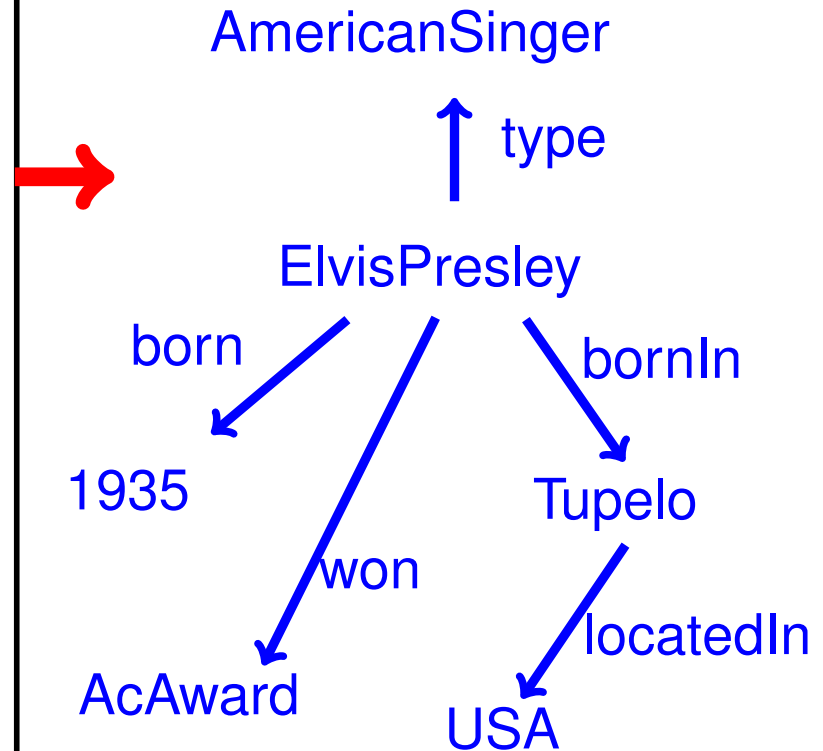
Born: 1935

In: Tupelo

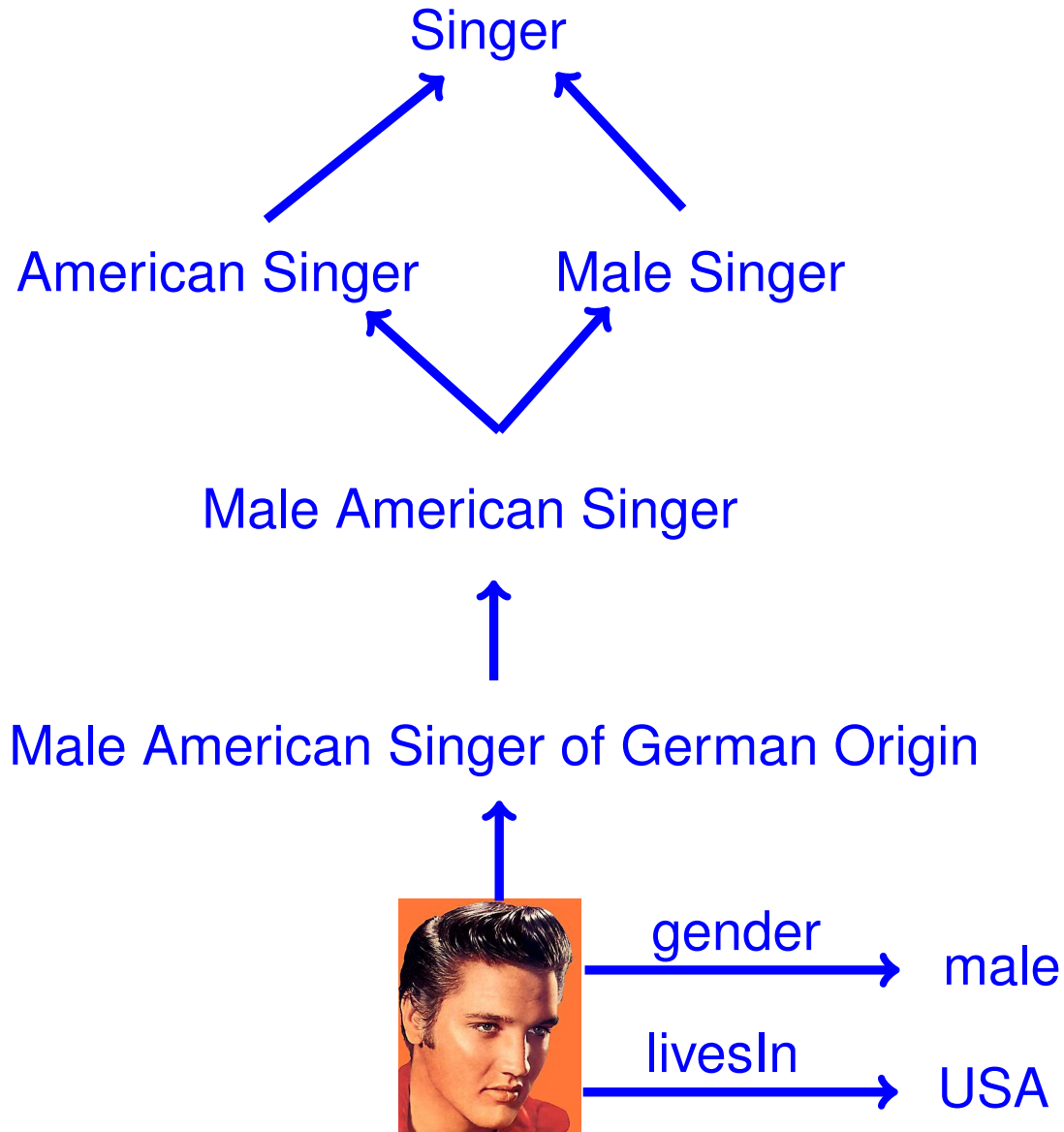
...

Categories:

Rock&Roll, American Singers, Academy Award winners...



Type Hierarchy

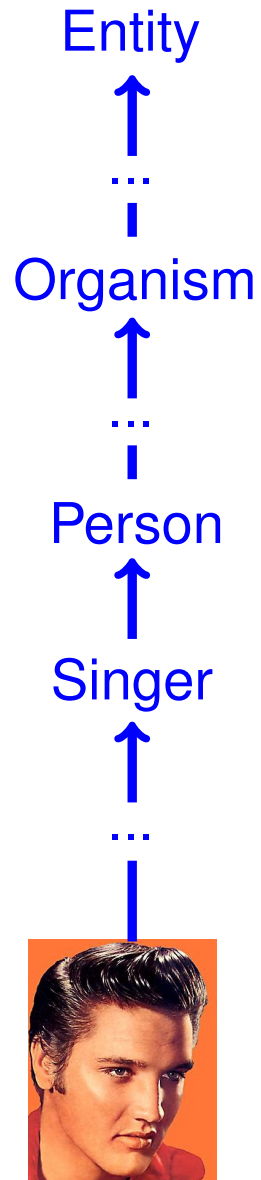


from Wikipedia
categories,
automatically
selected

Shallow noun
phrase parsing

Regular
expressions
+ demonyms

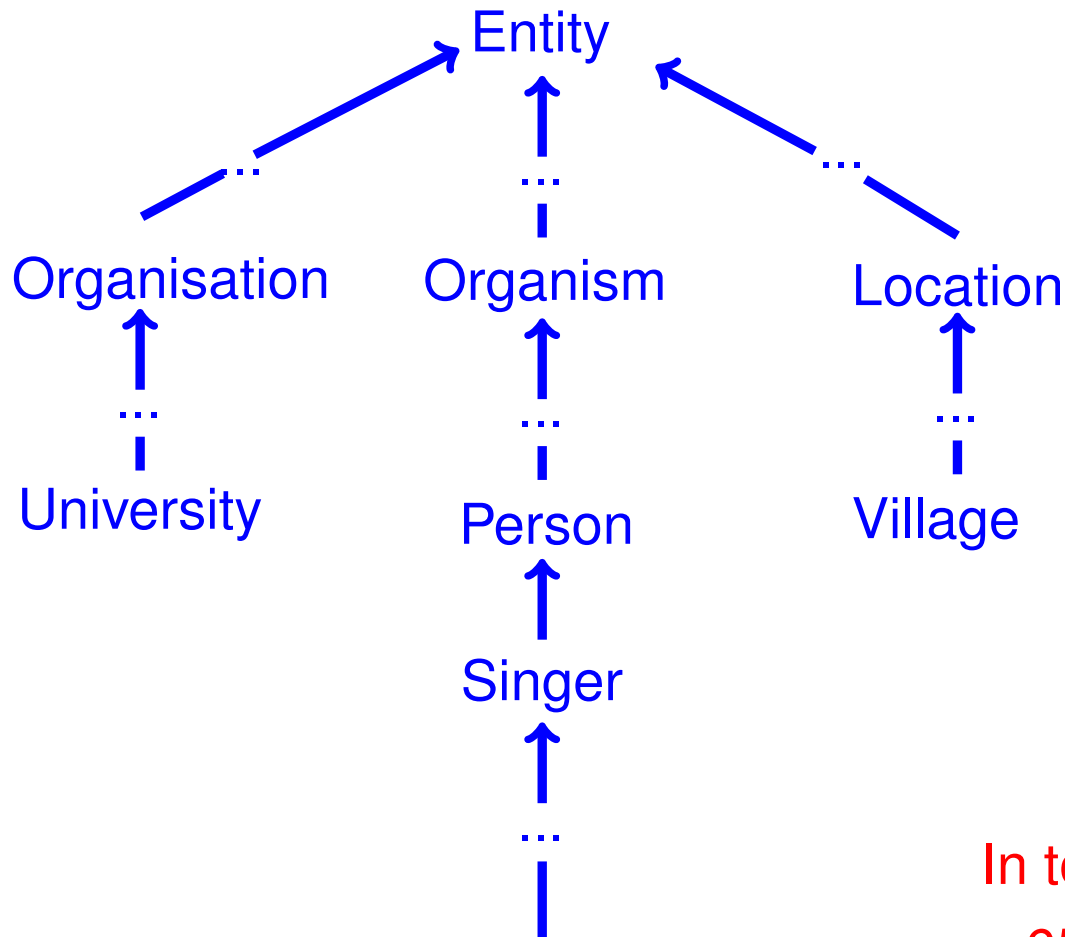
Type Hierarchy continued



from WordNet

by linking
Wikipedia &
WordNet

Type Hierarchy continued

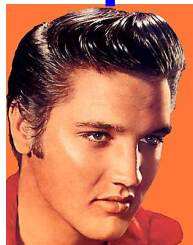


from WordNet

by linking
Wikipedia &
WordNet

In total:

- 650k classes
- every entity has ≥ 1 class
- link precision of 97%



Knowledge Representation

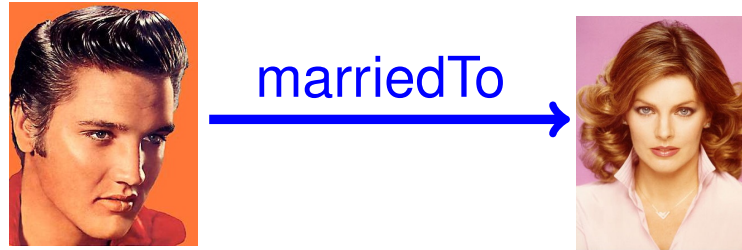


Fact Id	Subject	Predicate	Object
<fact_42>	<Elvis>	<marriedTo>	<Priscilla>
<fact_43>	<fact_42>	<since>	"1967-05-01"8sd:date
<fact_44>	<fact_42>	<until>	"1973-08-18"8sd:date
<fact_45>	<fact_42>	<extractedFrom>	<http://wikipedia...>
<fact_46>	<fact_45>	<byTechnique>	"Infobox "spouse""

(exported to RDF w/out factIds)

(Example for illustration only)

Knowledge Representation



Fact Id	Subject	Predicate	Object
<fact_42>	<Elvis>	<marriedTo>	<Priscilla>
<fact_43>	<fact_42>	<since>	"1967-05-01"8sd:date
<fact_44>	<fact_42>	<until>	"1973-08-18"8sd:date
<fact_45>	<fact_42>	<extractedFrom>	<http://wikipedia...>
<fact_46>	<fact_45>	<byTechnique>	"Infobox "spouse""

(exported to RDF w/out factIds)

(Example for illustration only)

Typechecked,
clean facts

Creating a large knowledge base

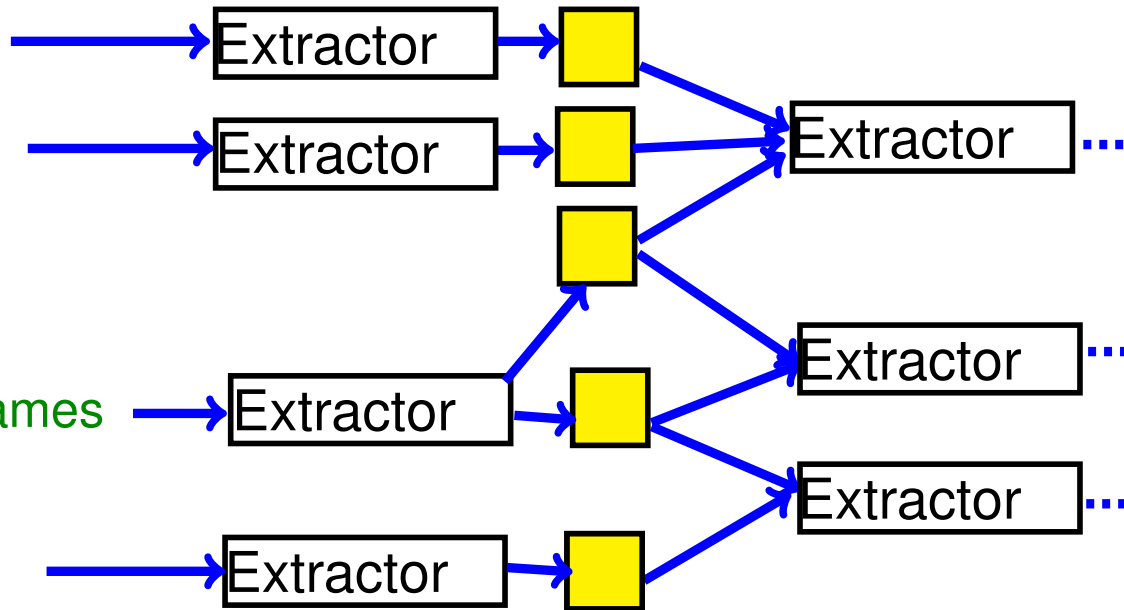


WIKIPEDIA
The Free Encyclopedia



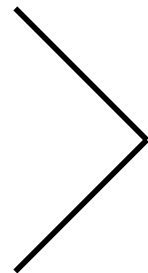
GeoNames

WordNet



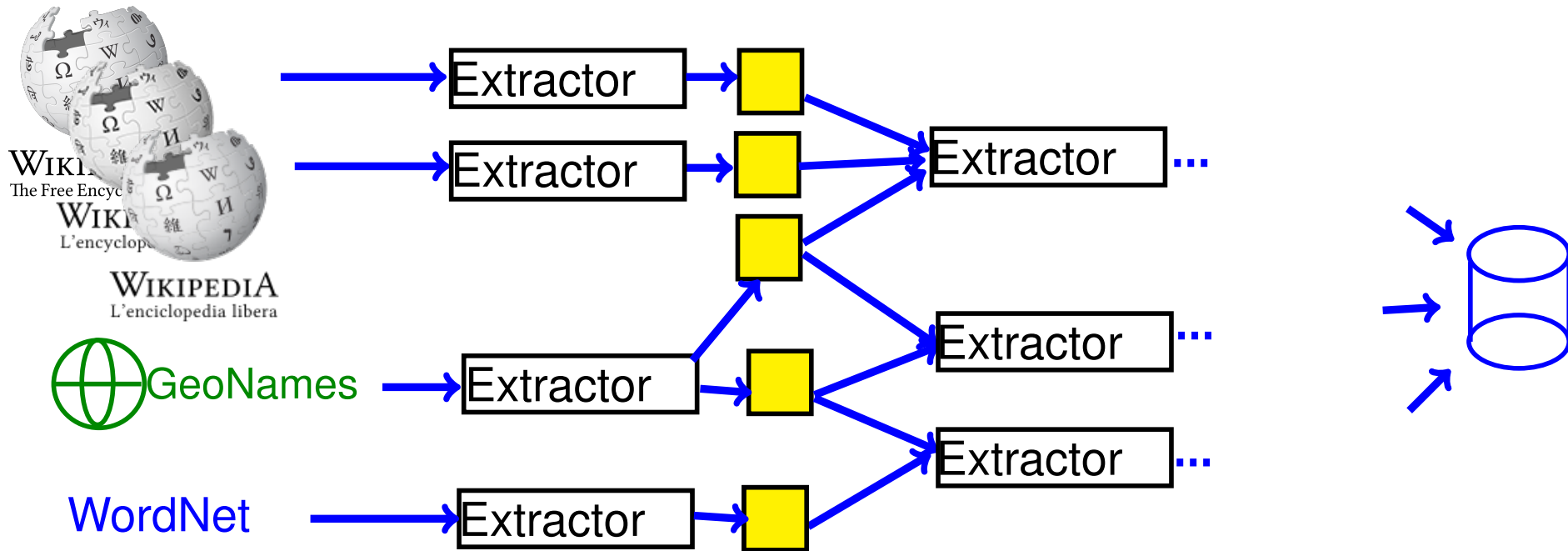
Intermediate extractors

- clean facts
- deduplicate facts and entities
- check consistency



ensuring high quality (95%)

Creating a large knowledge base

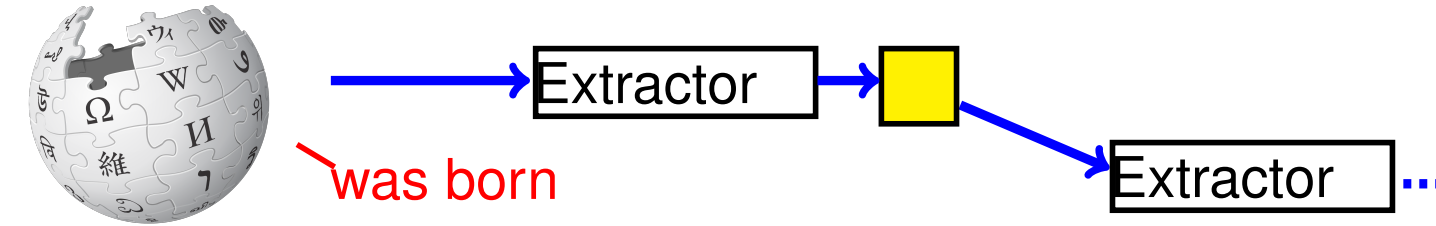


Intermediate extractors

- clean facts
- deduplicate facts and entities
- check consistency

ensuring high quality (95%)

Integrating multilingual Wikis



WIKIPEDIA
The Free Encyclopedia

est né

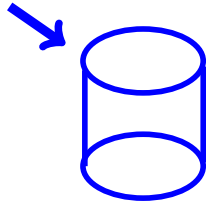


WIKIPÉDIA
L'encyclopédie libre

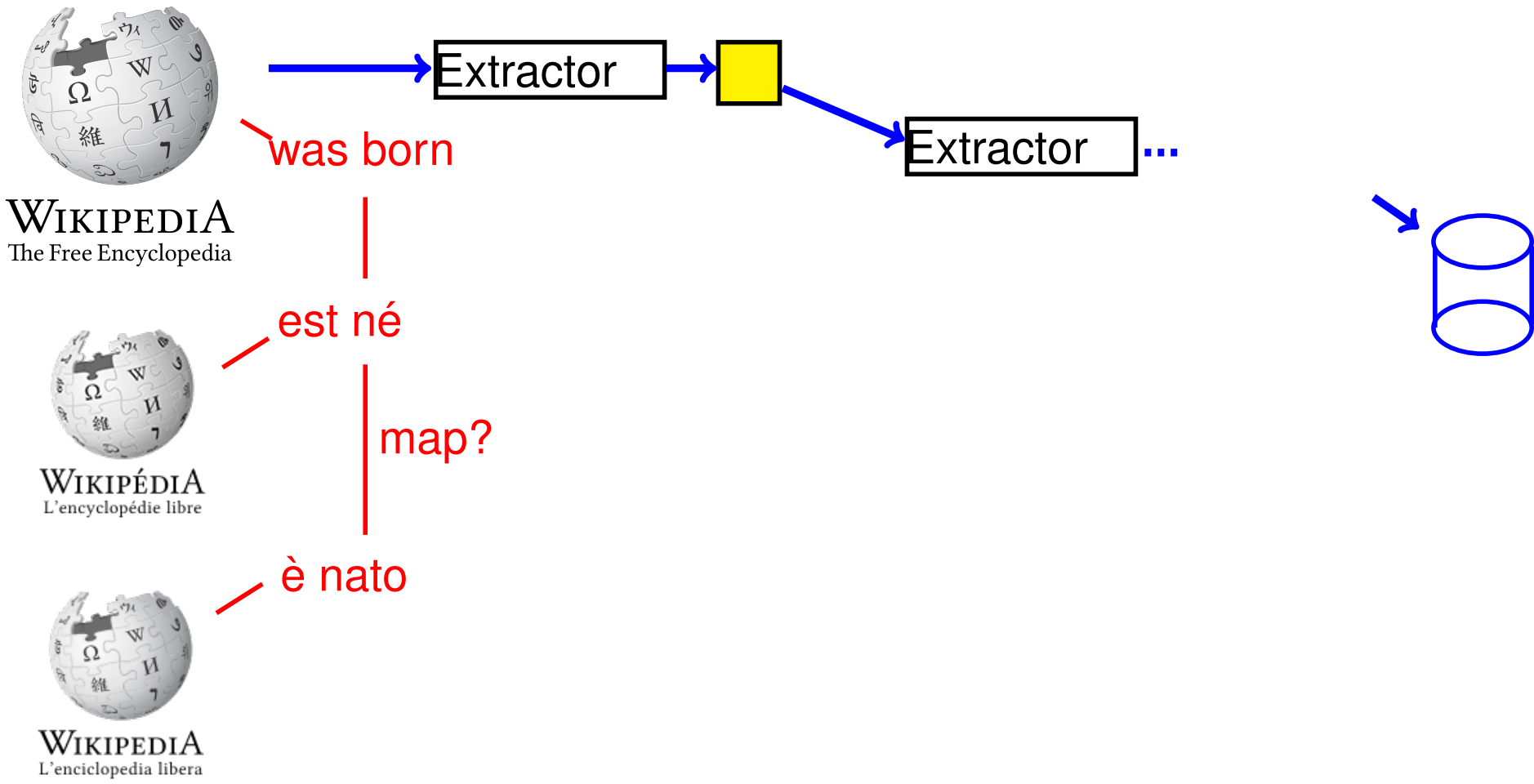
è nato



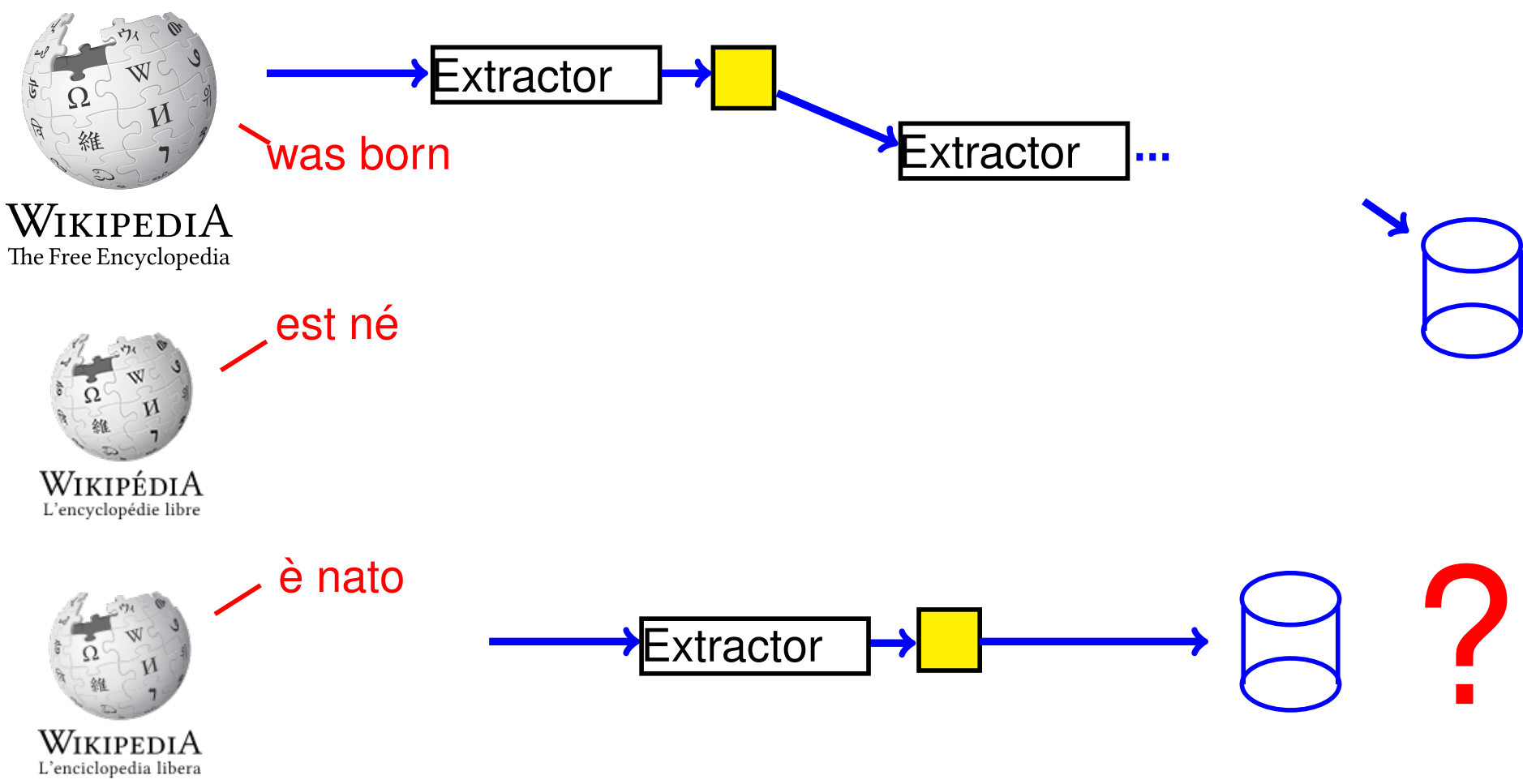
WIKIPEDIA
L'enciclopedia libera



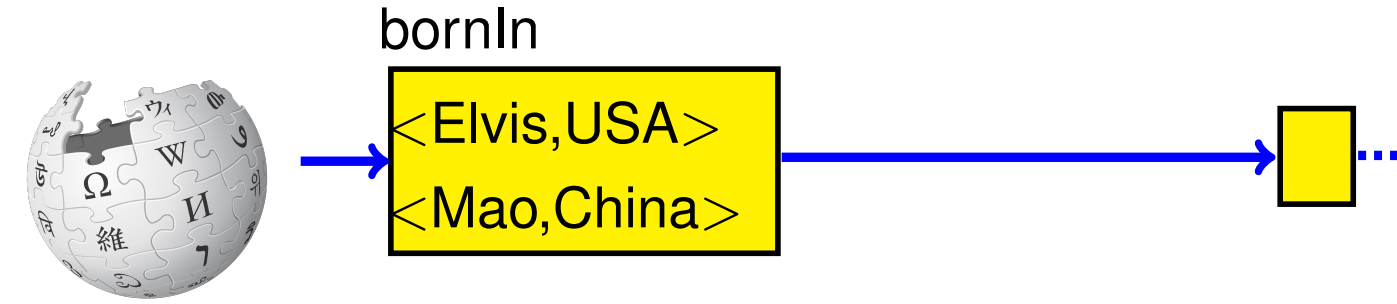
Integrating multilingual Wikis



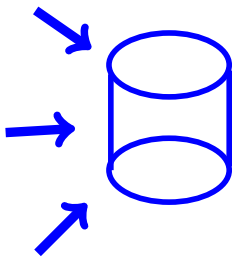
Integrating multilingual Wikis



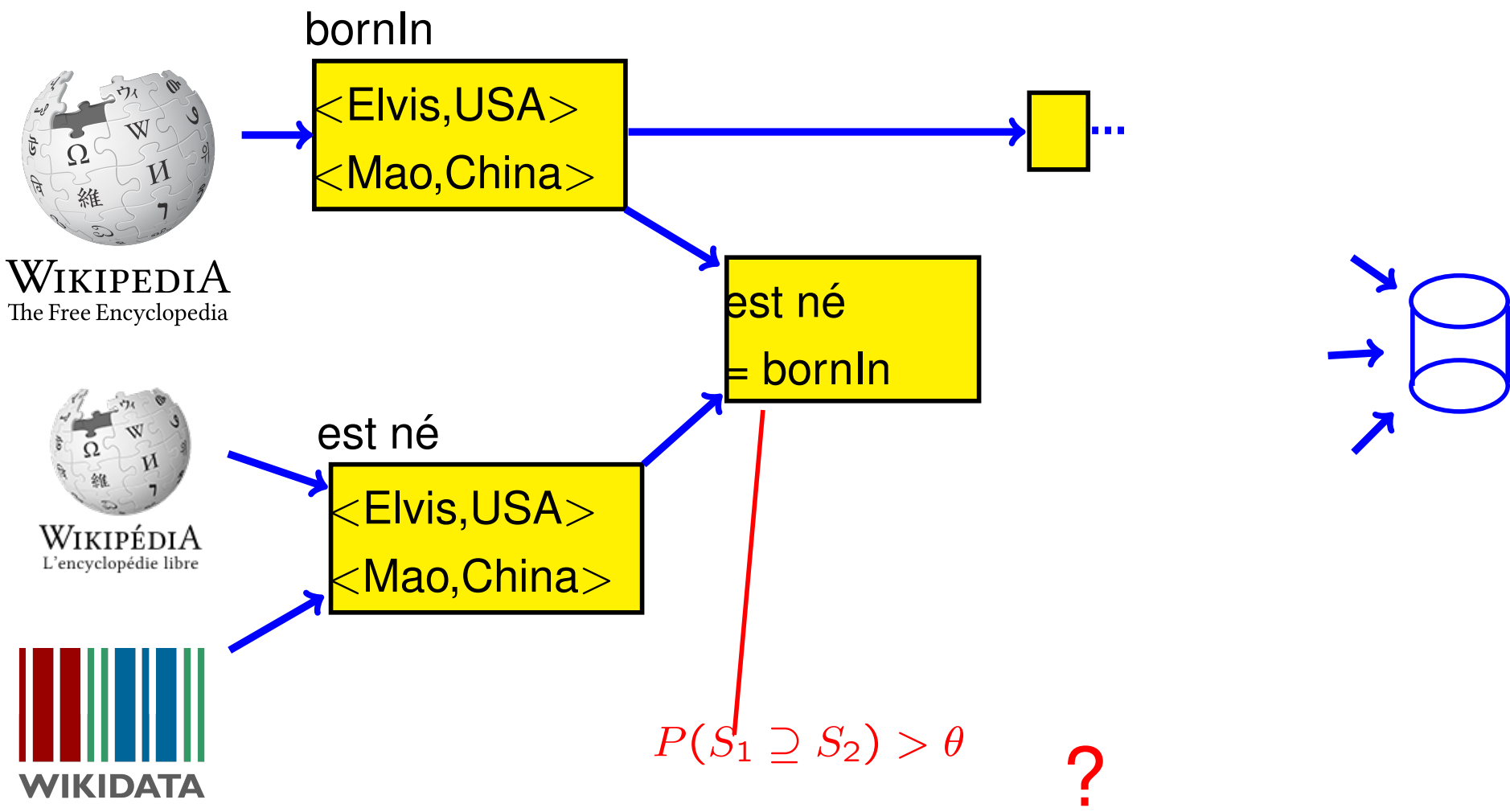
Integrating multilingual Wikis



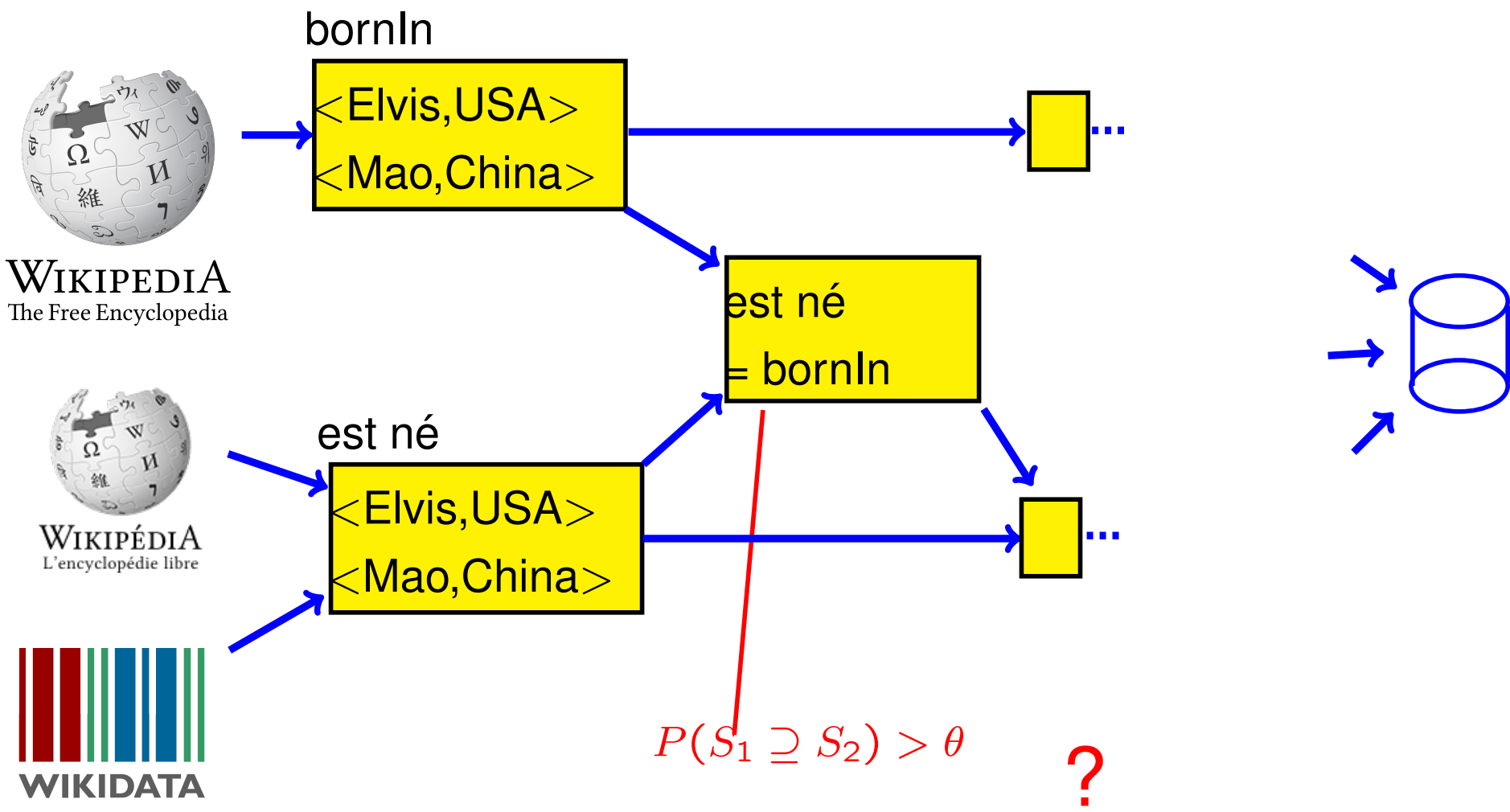
WIKIPEDIA
The Free Encyclopedia



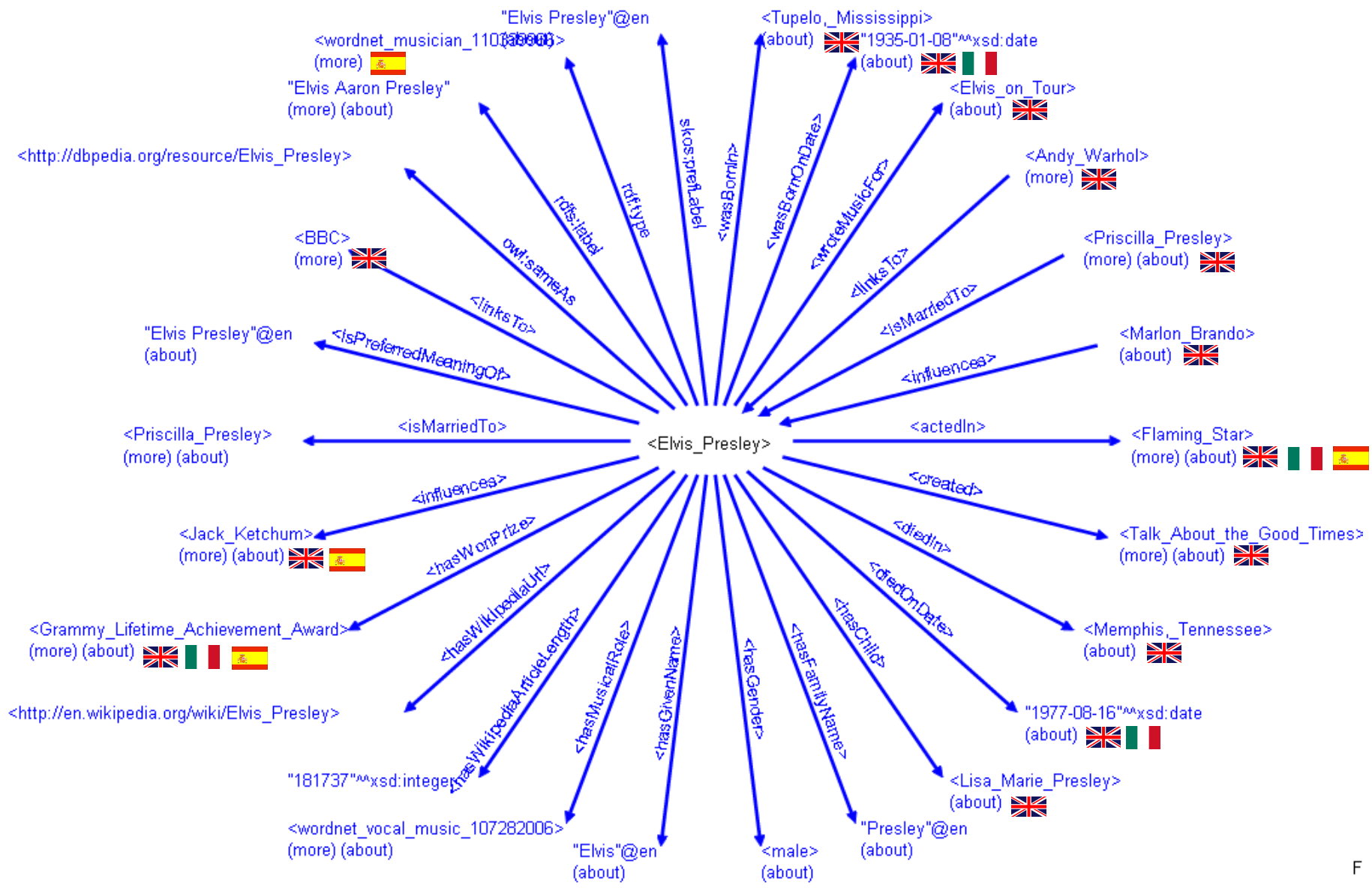
Integrating multilingual Wikis



Integrating multilingual Wikis



Example: YAGO about Elvis



YAGO: a large knowledge base



<http://yago-knowledge.org>

now open-source!

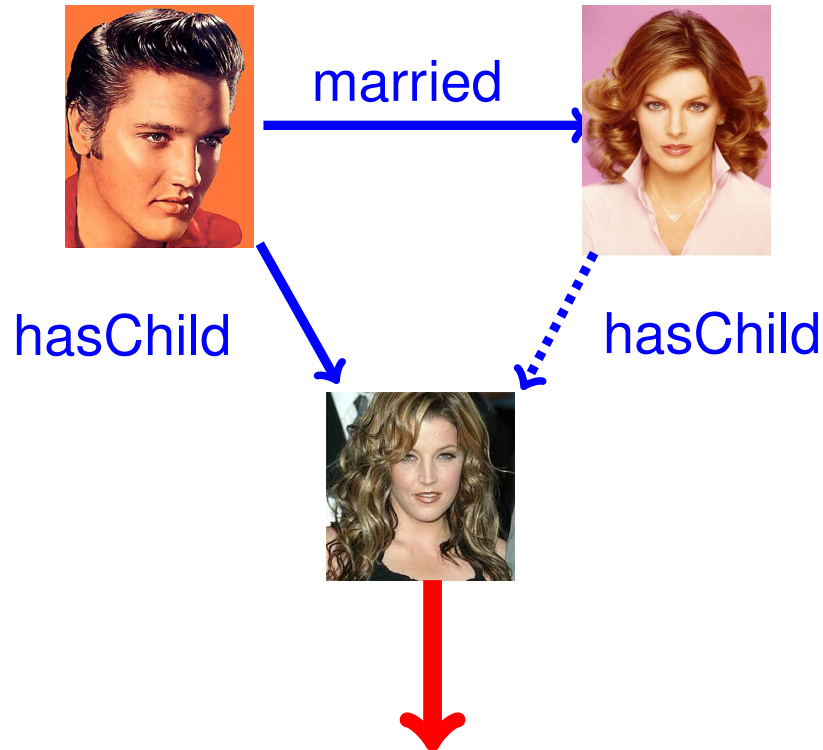


Wikipedia + WordNet
time and space
10 languages
100 relations
100m facts
10m entities
95% accuracy
used by DBpedia
and IBM Watson

[WWW'07, JWS'08, WWW'11 demo, AIJ'13, WWW'13 demo, CIDR'15, ISWC'16]

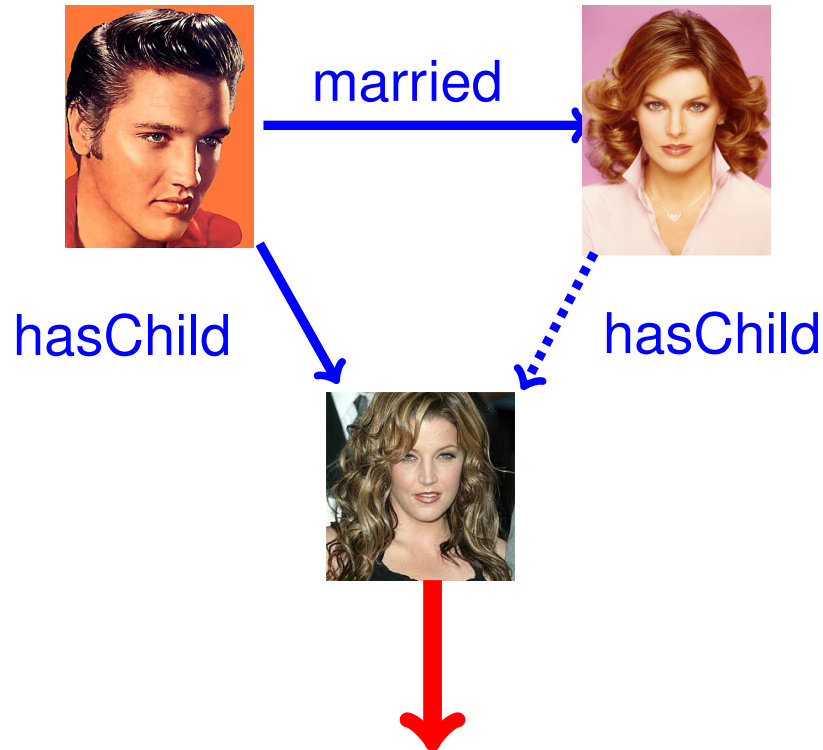


Rule Mining finds patterns



$$\text{married}(x, y) \wedge \text{hasChild}(x, z) \Rightarrow \text{hasChild}(y, z)$$

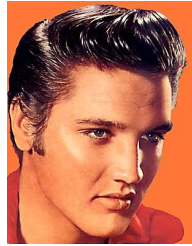
Rule Mining finds patterns



$$\text{married}(x, y) \wedge \text{hasChild}(x, z) \Rightarrow \text{hasChild}(y, z)$$

But: Rule mining needs counter examples
and RDF ontologies are positive only

Partial Completeness Assumption



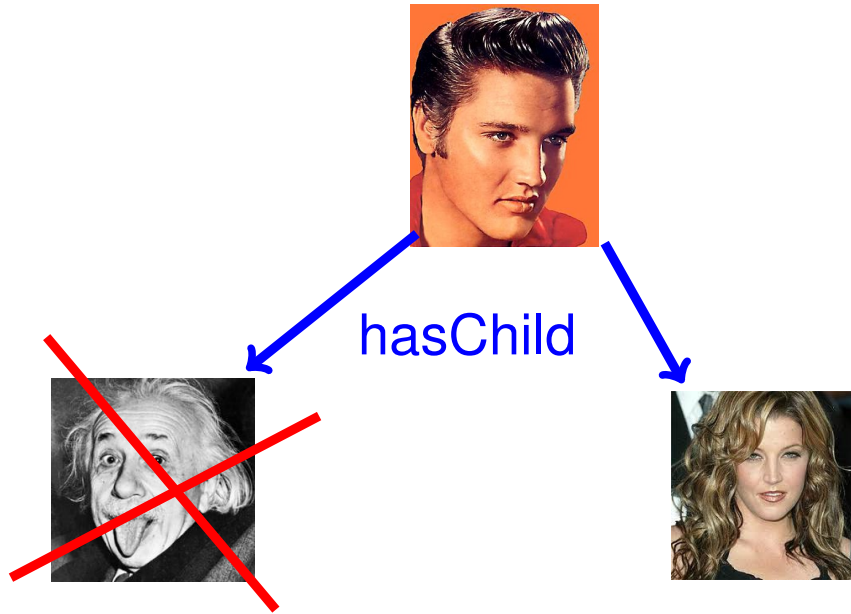
hasChild



Assumption:

If we know $r(x,y_1), \dots, r(x,y_n)$,
then all other $r(x,z)$ are false.

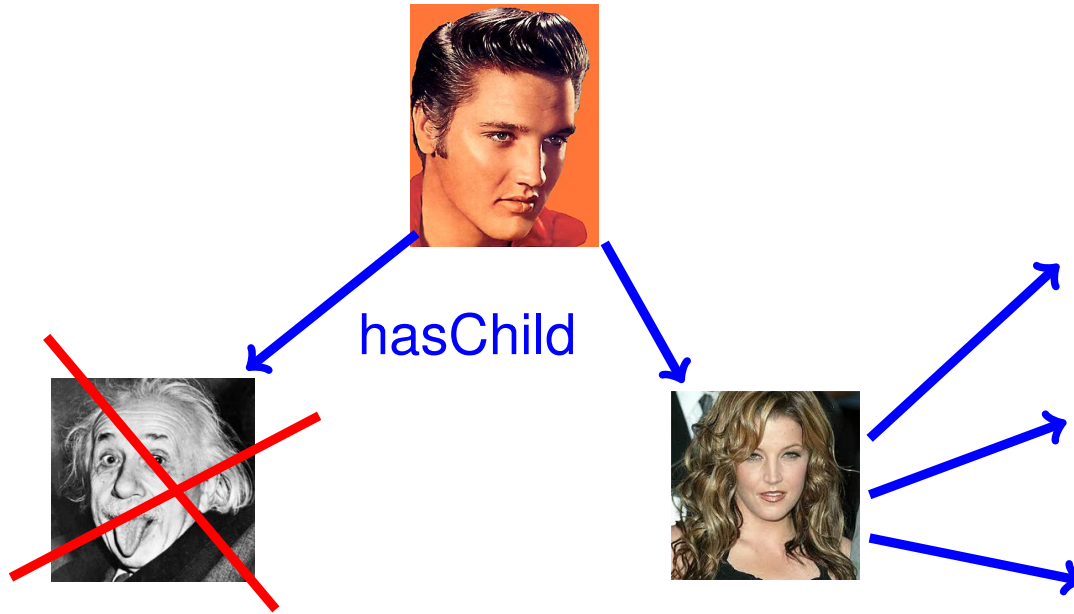
Partial Completeness Assumption



Assumption:

If we know $r(x, y_1), \dots, r(x, y_n)$,
then all other $r(x, z)$ are false.

Partial Completeness Assumption



Assumption:

If we know $r(x, y_1), \dots, r(x, y_n)$,
then all other $r(x, z)$ are false.

AMIE finds rules in ontologies



AMIE
(5min)

$$r(x, y) \wedge r'(z, y) \Rightarrow r''(x, z)$$

AMIE is based on an efficient in-memory database implementation.

Caveat: rules cannot predict the unknown with high precision

AMIE finds rules in ontologies



AMIE
(5min)

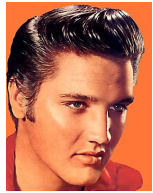
$type(x, pope) \Rightarrow$
 $diedIn(x, Rome)$



[WWW 2013, VLDB journal 2015]



Incompleteness

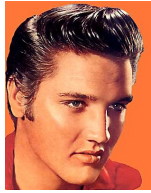


marriedTo →



the quality of YAGO w
ls a precision of 95%, as
uks to our brilliant algori

Incompleteness



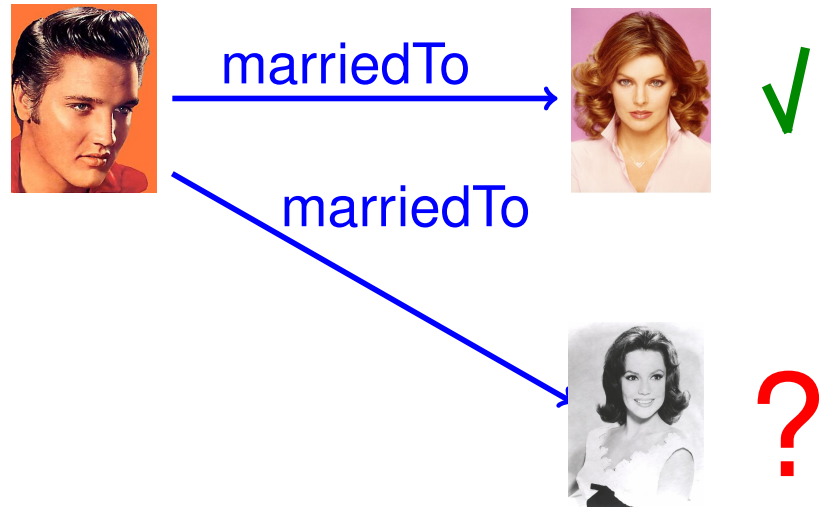
marriedTo



the quality of YAGO w
ls a precision of 95%, as
uks to our brilliant algorith



Incompleteness

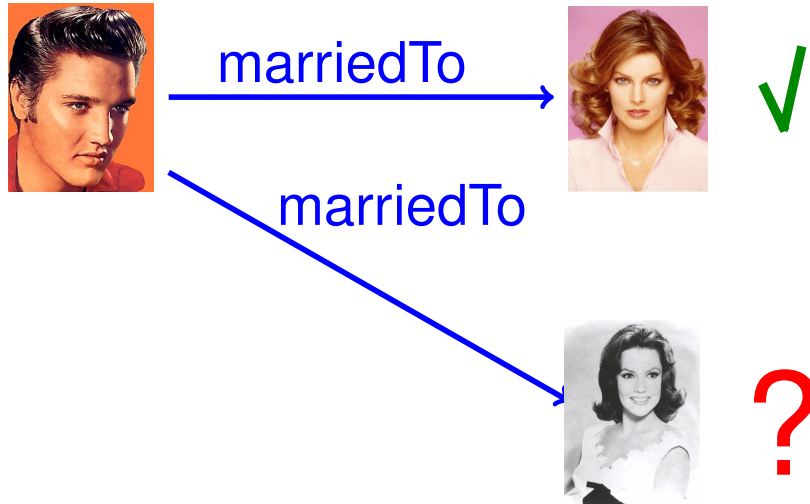


the quality of YAGO is a precision of 95%, as thanks to our brilliant algorithm



Given a subject s and a relation r , do we know all o with $r(s, o)$?

Signals for Incompleteness



Closed World Assumption
Partial Completeness Assumption
Popularity oracle

No-change oracle
Star-pattern oracle
Class-oracle

AMIE oracle: Learn rules such as

$moreThan_1(x, hasParent) \Rightarrow complete(x, hasParent)$



Signals for Incompleteness (F1)

Relation	CWA	PCA	card ₂	Popularity	No change	Star	Class	AMIE
diedIn	60%	22%	—	4%	15%	50%	99%	96%
directed	40%	96%	19%	7%	71%	0%	0%	100%
graduatedFrom	89%	4%	2%	2%	10%	89%	92%	87%
hasChild	71%	1%	1%	2%	13%	40%	78%	78%
hasGender	78%	100%	—	2%	—	86%	95%	100%
hasParent*	1%	54%	100%	—	—	0%	0%	100%
isCitizenOf*	4%	98%	11%	1%	4%	10%	5%	100%
isConnectedTo	87%	34%	19%	—	—	68%	88%	89%
isMarriedTo*	55%	7%	0%	3%	12%	37%	57%	46%
wasBornIn	28%	100%	—	5%	8%	0%	0%	100%

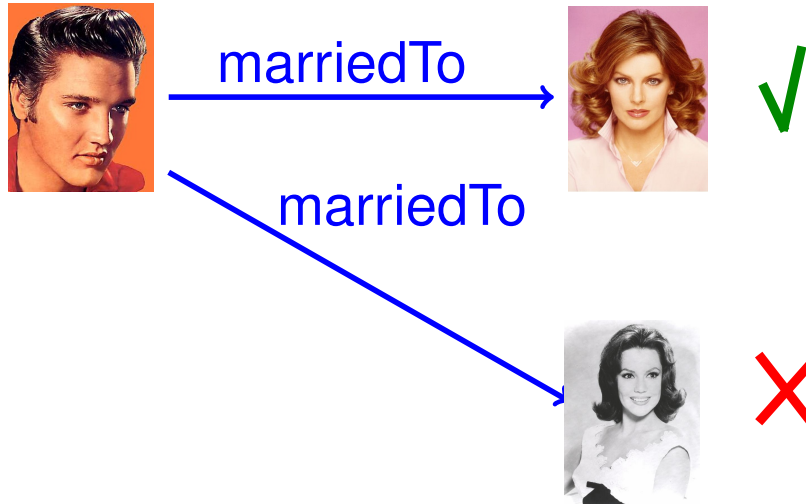


Relation	CWA	PCA	card ₂	Popularity	Star	Class	AMIE
alma_mater	90%	14%	5%	1%	87%	87%	87%
brother	93%	1%	—	1%	94%	96%	96%
child	70%	1%	—	1%	79%	72%	73%
country_of_citizenship*	42%	97%	10%	3%	0%	0%	98%
director	81%	100%	—	3%	94%	89%	100%
father*	5%	100%	6%	9%	89%	8%	100%
mother*	3%	100%	3%	10%	67%*	5%	100%
place_of_birth	53%	100%	7%	5%	55%	0%	100%
place_of_death	89%	35%	1%	2%	81%	81%	96%
sex_or_gender	81%	100%	6%	3%	92%	91%	100%
spouse*	57%	7%	—	1%	54%	54%	55%



* = biased training sample

Signals for Incompleteness



AMIE can predict incompleteness

- bornIn: 100% F1-measure
- diedIn: 96%
- directed: 100%
- graduatedFrom: 87%
- hasChild: 78%
- isMarriedTo: 46%
- ... and more.



[WSDM 2017]

We are happy to find ways of collaboration

- can WikiData make use of YAGO's type hierarchy?
- can WikiData import facts from YAGO?
- can YAGO be used to consolidate facts in WikiData?
- can the source code of YAGO be of use?
- add your idea here: _____





Happy Birthday!!!



We are happy to find ways of collaboration

- can WikiData make use of YAGO's type hierarchy?
- can WikiData import facts from YAGO?
- can YAGO be used to consolidate facts in WikiData?
- can the source code of YAGO be of use?
- add your idea here: _____

