



# YAGO: Yet Another Great Ontology

Fabian M. Suchanek

(joint work with Gjergji Kasneci, Mauro Sozio and Gerhard Weikum)

(Max-Planck-Institute for Informatics, Saarbrücken/Germany)



# Before we start the talk...

---

Thank you,

Prof. Etzioni,

Prof. Weld,

Mr. Allen

for inviting me so expeditiously!





# About myself

---

Fabian M. Suchanek

2003 **BSc. in Cognitive Science**

from U Osnabrück/Germany

2005 **MSc. in Computer Science**

from U Saarbrücken/Germany

2008 **PhD in Computer Science**

at the Max-Planck Institute for Informatics

in Saarbrücken/Germany,

working on Ontologies and Information Extraction



MAX-PLANCK-GESELLSCHAFT

# The Search for Excellent Scientists



Max-Planck Institute



University of Washington





# The Search for Excellent Scientists

MAX-PLANCK-GESELLSCHAFT

Google™

*scientist musician*

## [Invisible Gorilla steals the Nobel Prize](#)

...The gorilla, plus dropped food and country **music**, were honored...

[news\*\*scientist\*\*.org/article/invisibleGorilla.htm](http://news<b>scientist</b>.org/article/invisibleGorilla.htm)   [Cached](#)   [Similar pages](#)



MAX-PLANCK-GESELLSCHAFT

# The Search for Excellent Scientists

---

*scientist who are musicians and won a prize*

## [Invisible Gorilla steals the Nobel Prize](#)

...The gorilla, plus dropped food and country **music**, were honored...

[news\*\*scientist\*\*.org/article/invisibleGorilla.htm](http://news<b>scientist</b>.org/article/invisibleGorilla.htm) [Cached](#) [Similar pages](#)



# The Search for Excellent Scientists

MAX-PLANCK-GESellschaft



*Please give me IMMEDIATELY the scientists who are...*

## [Invisible Gorilla steals the Nobel Prize](#)

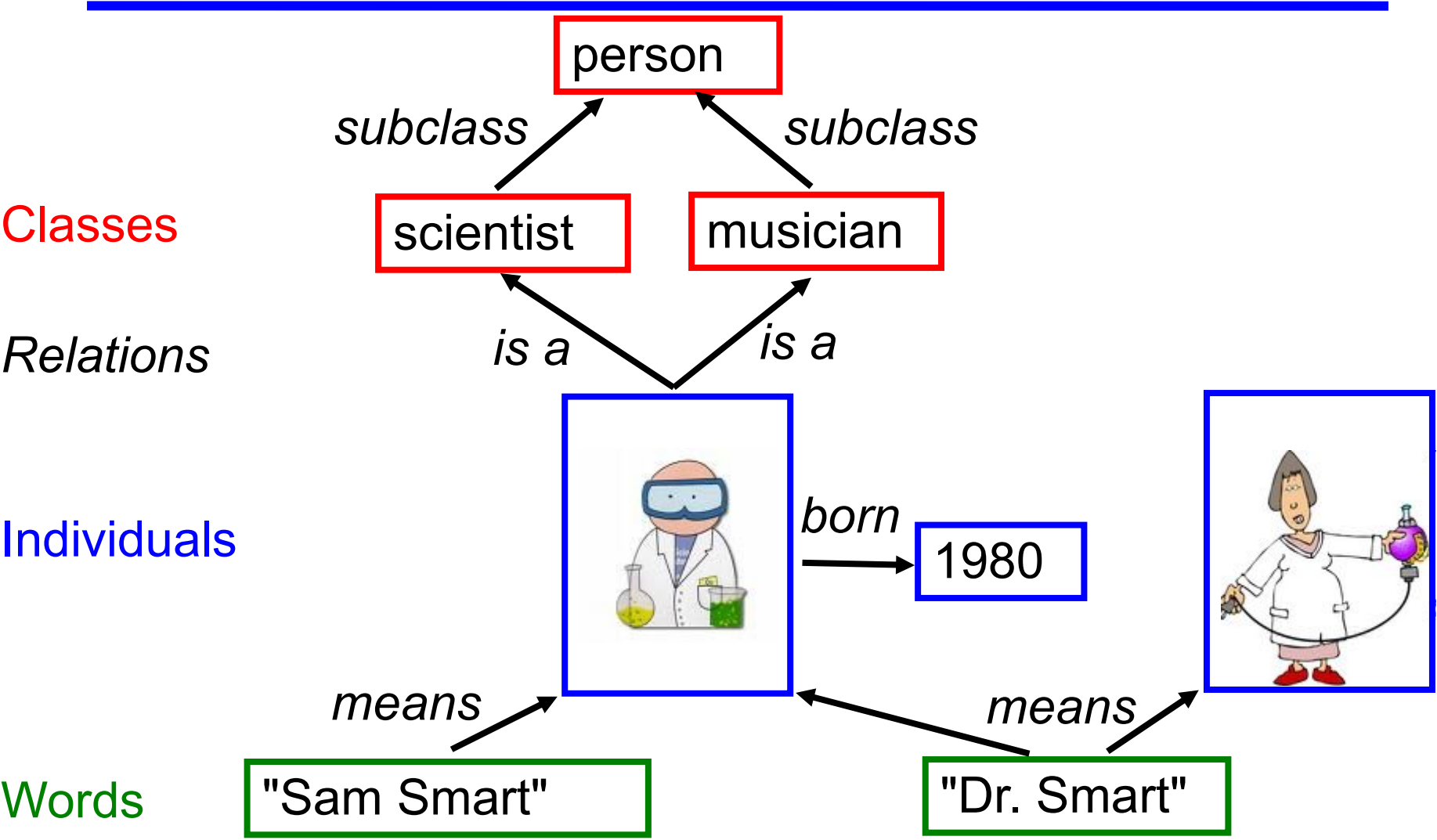
...The gorilla, plus dropped food and country **music**, were honored...

[news scientist.org/article/invisibleGorilla.htm](http://news scientist.org/article/invisibleGorilla.htm)   [Cached](#)   [Similar pages](#)



# Solution: An Ontology

MAX-PLANCK-GESELLSCHAFT



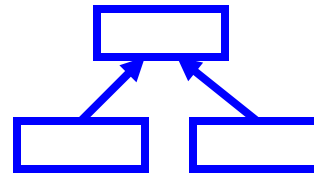




Gathering the knowledge of this world  
in a structured ontology.

- ⌋ Semantic Search
- ⌋ Question answering
- ⌋ Machine Translation
- ⌋ Document classification
- ⌋ ...

The world I'd  
like to say, even  
though some  
may contradict,  
is not as it  
seems. It rather  
seems as if the  
world seems not  
what it seems





# Where do we get the ontology from?

---

- ┆ Assemble the ontology manually

(WordNet, SUMO, Cyc, GeneOntology)

Question: Can we increase coverage? (MPI is in none of these)

- ┆ Use community work (Semantic Wikipedia, Freebase)

Question: Will it take off?

- ┆ Extract the ontology from a semi-structured corpus (Wikipedia)

(KYLIN/KOG, DBpedia)

Question: Can we go beyond Wikipedia?

- ┆ Extract information from corpora (e.g. the Web)

(Espresso, Snowball, LEILA, TextRunner, KnowItAll)

Question: Can we increase accuracy? Can we achieve canonicity?




# Our Approach (= this talk)

---

- ➔ Extract the ontology from a semi-structured corpus (Wikipedia)  
as a core ontology: YAGO
- ↳ Extract information from corpora (e.g. the Web)  
to extend the core ontology: SOFIE



# YAGO Construction: Infoboxes



Smart, S

blah blah blub Elvis (don't read this! Better listen to the talk!) laber fasel suez. Insbesondere, blub, texte zu, und so weiter blah blah blub Elvis laber fasel suez. Blub, aber blah! Insbesondere, blub, texte zu, und so weiter blah blah blub Elvis laber fasel suez. Insbesondere, blub, texte zu, und so weiter

Name: Sam Smart

Born in: Berlin

...



*bornIn* → Berlin

Exploit infoboxes



# YAGO Construction: Categories

MAX-PLANCK-GESELLSCHAFT



Smart, S

blah blah blub Elvis (don't read this! Better listen to the talk!) laber fasel suelz. Insbesondere, blub, texte zu, und so weiter blah blah blub Elvis laber fasel suelz. Blub, aber blah! Insbesondere, blub, texte zu, und so weiter blah blah blub Elvis laber fasel suelz. Insbesondere, blub, texte zu, und so weiter

Categories:

1980\_births



Exploit infoboxes

Exploit relational categories



# YAGO Construction: Categories

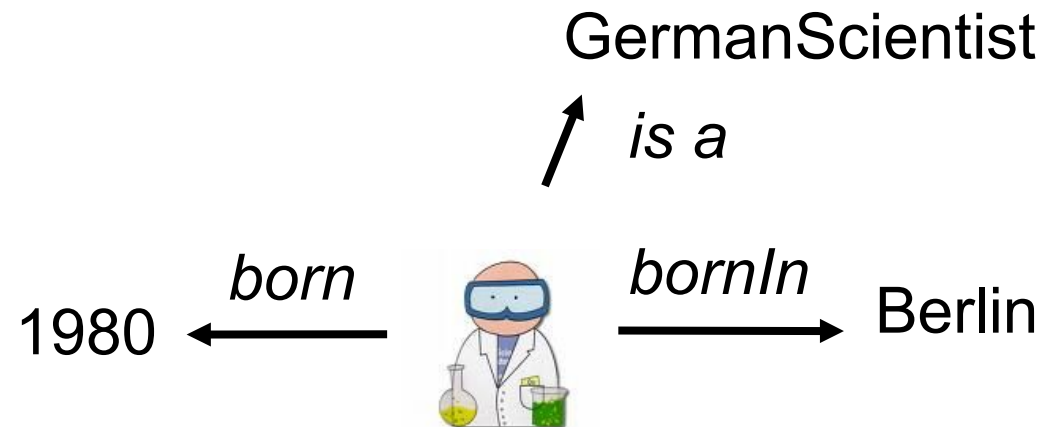
MAX-PLANCK-GESELLSCHAFT



Smart, S

blah blah blub Elvis (don't read this! Better listen to the talk!) laber fasel suezl. Insbesondere, blub, texte zu, und so weiter blah blah blub Elvis laber fasel suezl. Blub, aber blah! Insbesondere, blub, texte zu, und so weiter blah blah blub Elvis laber fasel suezl. Insbesondere, blub, texte zu, und so weiter

Categories:  
German\_scientists



- Exploit infoboxes
- Exploit relational categories
- Exploit conceptual categories



# YAGO Construction: Categories

MAX-PLANCK-GESELLSCHAFT

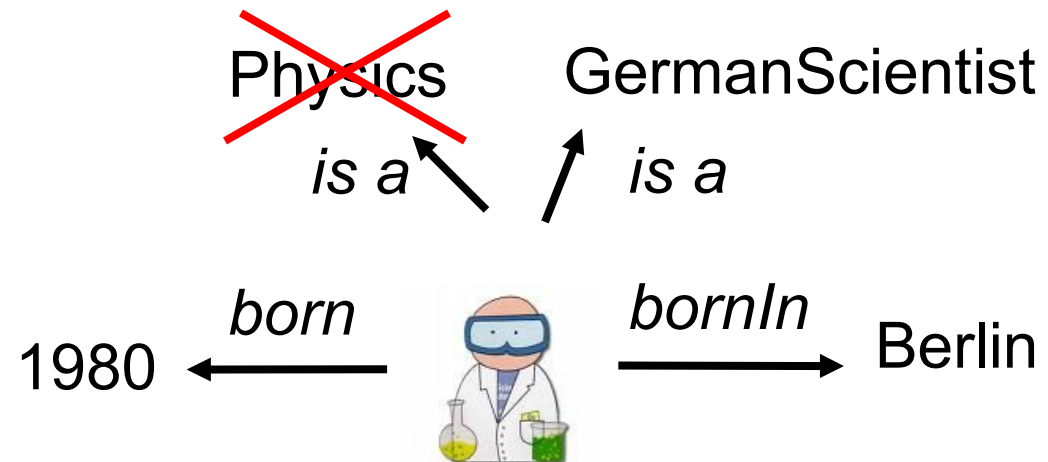


Smart, S

blah blah blub Elvis (don't read this! Better listen to the talk!) laber fasel suelz. Insbesondere, blub, texte zu, und so weiter blah blah blub Elvis laber fasel suelz. Blub, aber blah! Insbesondere, blub, texte zu, und so weiter blah blah blub Elvis laber fasel suelz. Insbesondere, blub, texte zu, und so weiter

Categories:

Physics



- Exploit infoboxes
- Exploit relational categories
- Exploit conceptual categories
- Avoid thematic categories



# YAGO Construction: Thematic vs Conceptual

MAX-PLANCK-GESELLSCHAFT

😊 **conceptual:**

American scientists **s** of German origin

☹️ **thematic:**

Biomedical science in Germany

Shallow linguistic  
noun phrase parsing:

*Premodifier*

*Head*

*Postmodifier*

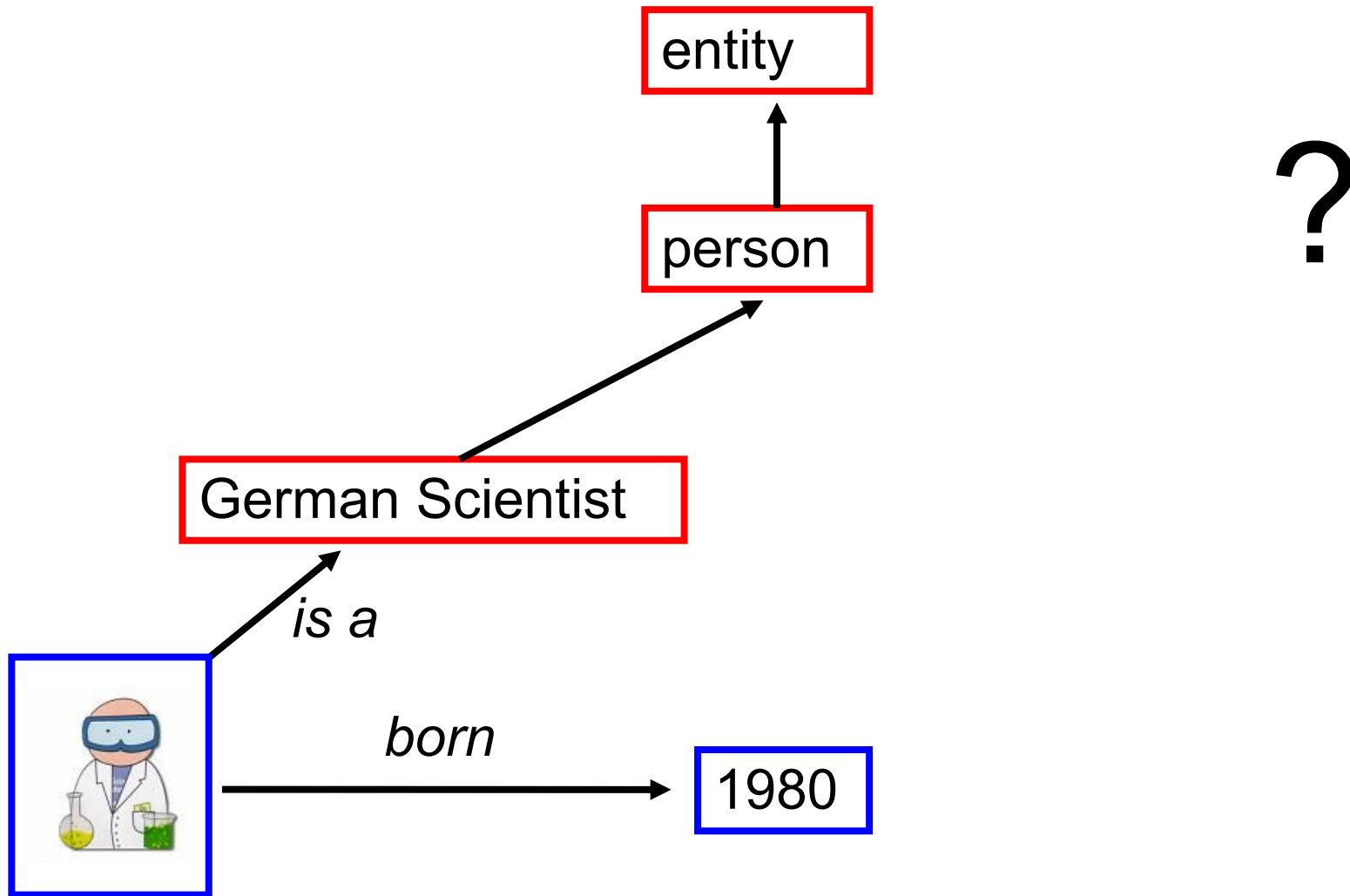
Heuristics: If the head is a plural word, the category is conceptual





# YAGO Construction: Upper Model

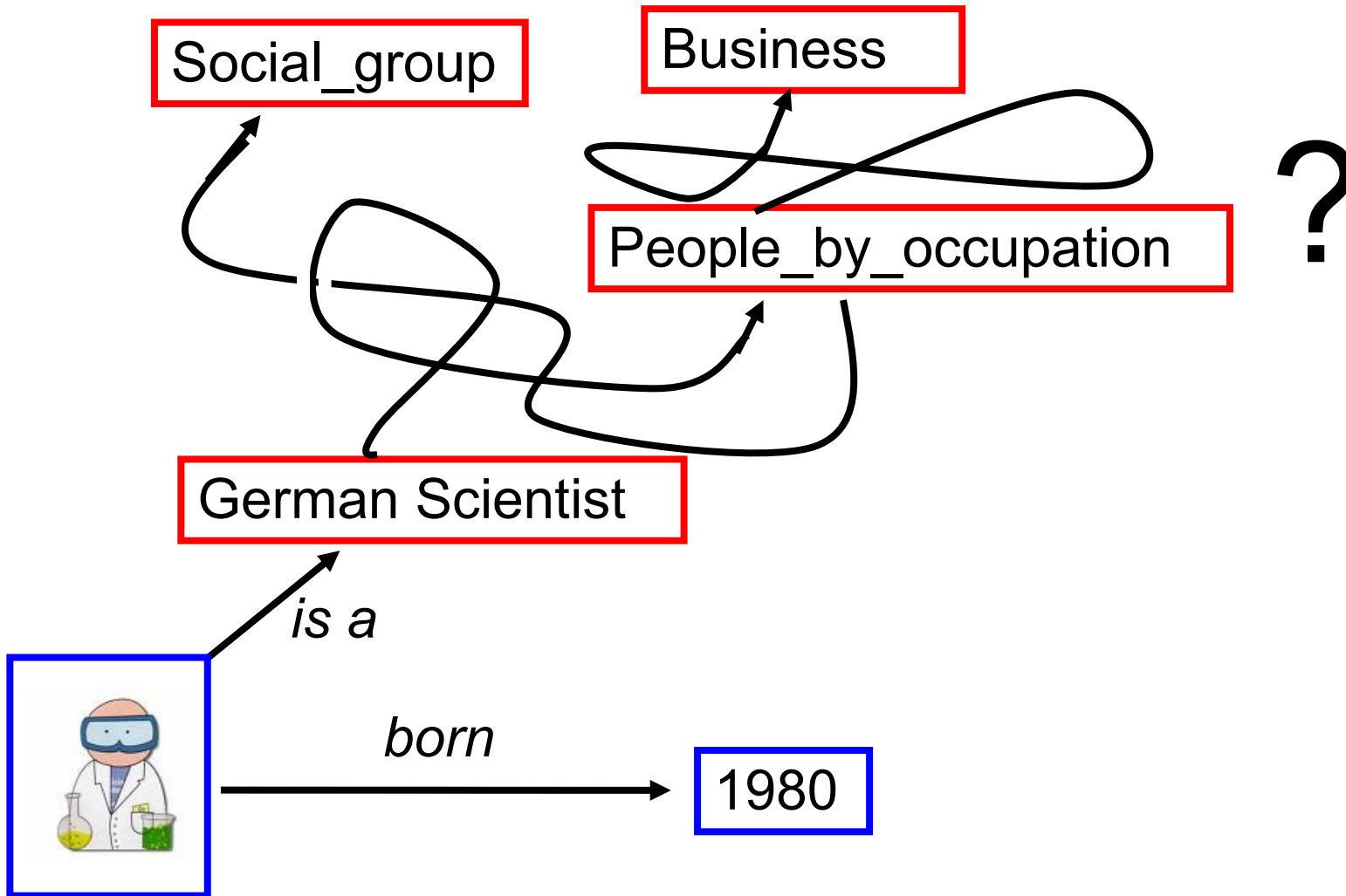
MAX-PLANCK-GESELLSCHAFT





# YAGO Construction: Upper Model

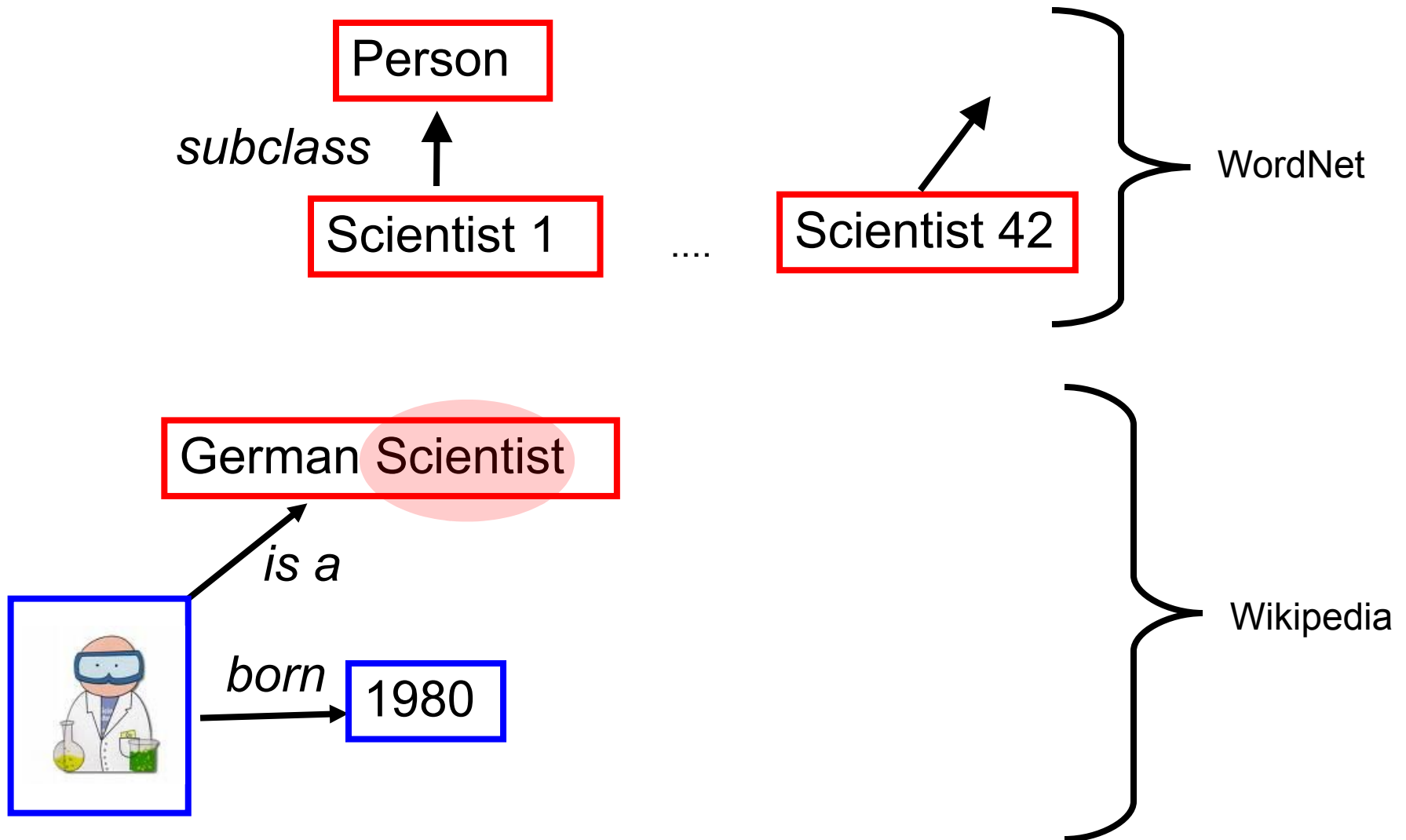
MAX-PLANCK-GESELLSCHAFT





# YAGO Construction: Upper Model

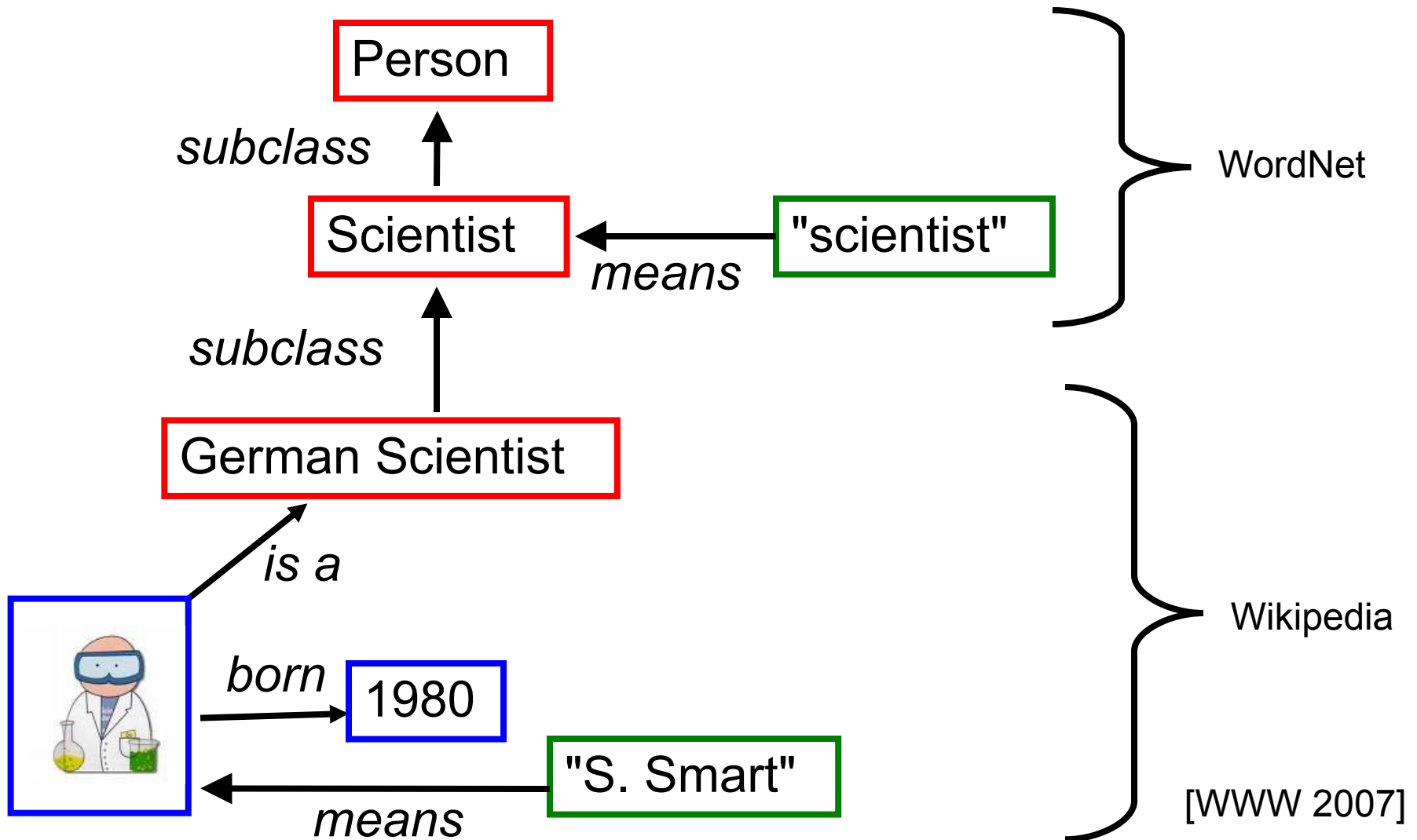
MAX-PLANCK-GESELLSCHAFT





# YAGO Construction: Upper Model

MAX-PLANCK-GESELLSCHAFT





# YAGO: Relations

MAX-PLANCK-GESELLSCHAFT

---

<b>is a</b>	<b>establishedOnDate</b>
<b>familyName</b>	<b>isMarriedTo</b>
<b>givenName</b>	<b>hasPopulation</b>
<b>bornOnDate</b>	<b>hasHeight</b>
<b>diedOnDate</b>	<b>hasWeight</b>
<b>bornIn</b>	<b>hasInflation</b>
<b>diedIn</b>	<b>actedIn</b>
<b>locatedIn</b>	<b>...</b>

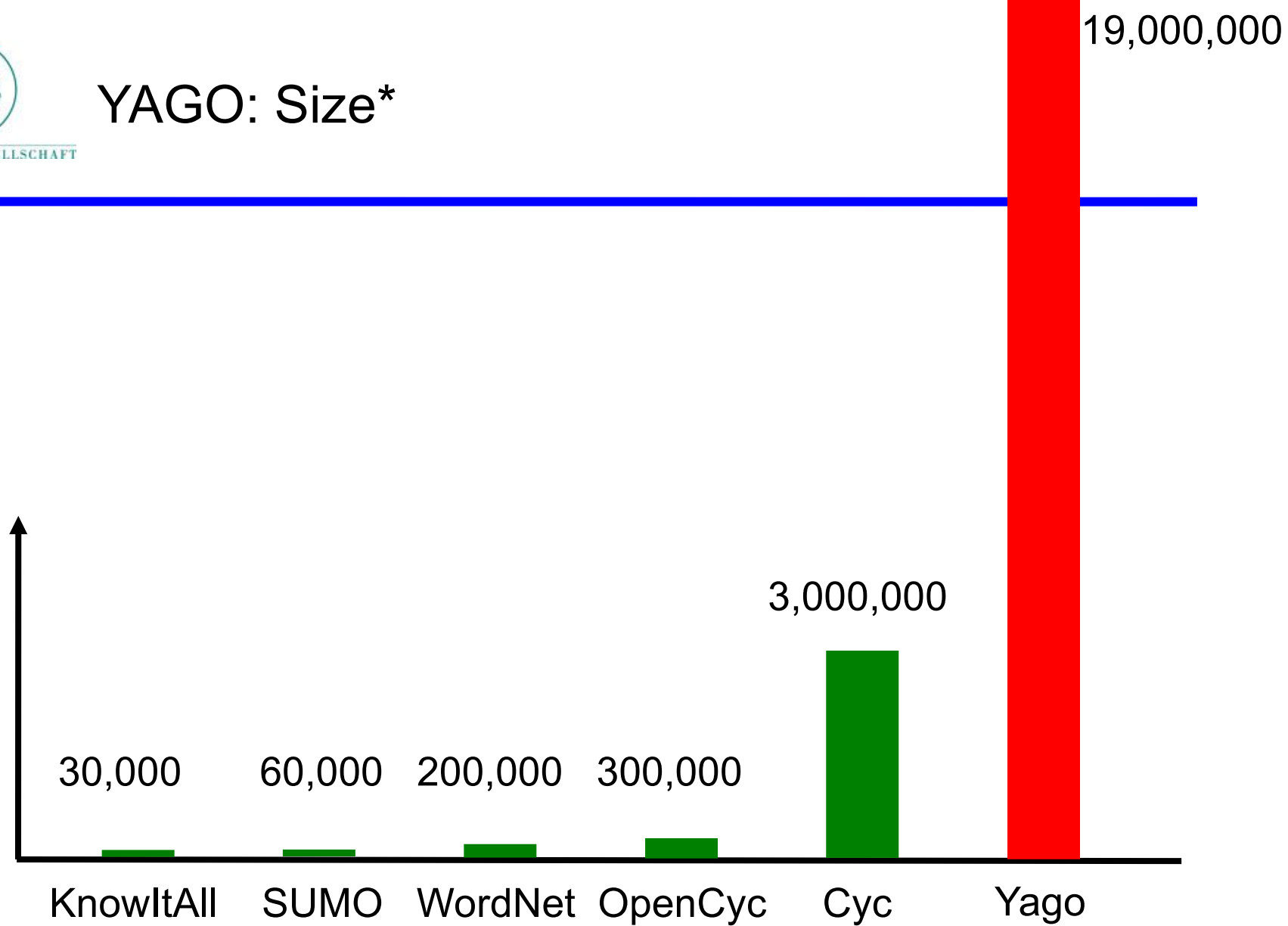
Manual evaluation:  
95% correct

---

90 relations



# YAGO: Size\*

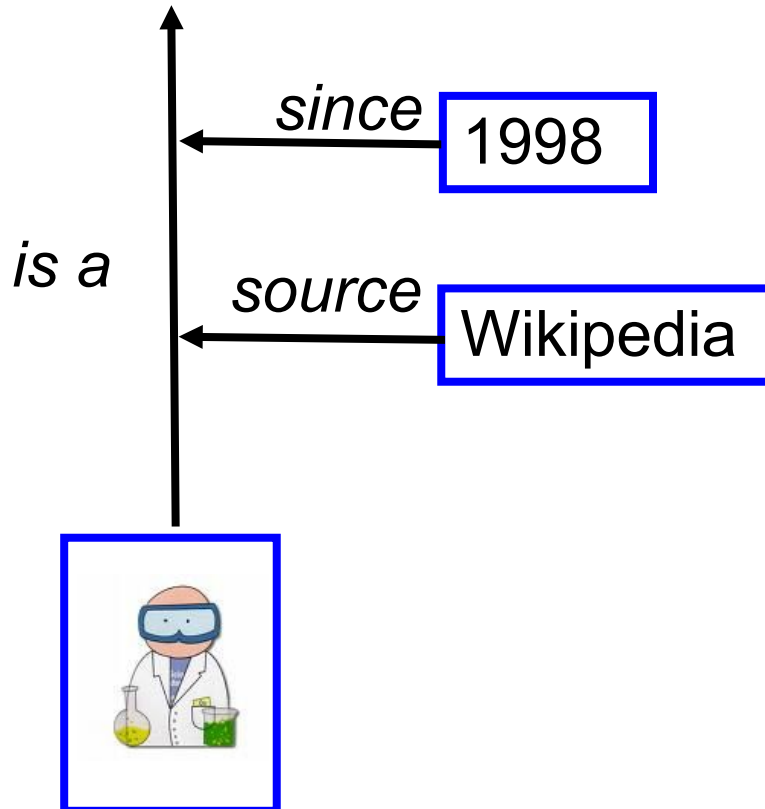


\* Publicly available ontologies with a quality guarantee. Size is not correlated with usefulness.



# YAGO Model: Why binary is not enough

scientist



- #1 (Sam, is\_a, scientist)
- #2 (#1, since, 1998)
- #3 (#1, source, Wikipedia)



# YAGO Model: Formal view

A YAGO ontology over

↳ a set of relations  $R$

↳ a set of common entities  $C$

↳ a set of fact identifiers  $I$

is a function

#1 (**Sam, is\_a, scientist**)

#2 (**#1, since, 1998**)

#3 (**#1, source, Wikipedia**)

$$I \rightarrow (R \cup C \cup I) \times R \times (R \cup I \cup C)$$

We can talk about

↳ facts (**#1, source, Wikipedia**)

↳ additional arguments (**#1, since, 1998**)

↳ relations (**time, hasRange, time\_interval**)

Still: Decidable  
Consistency

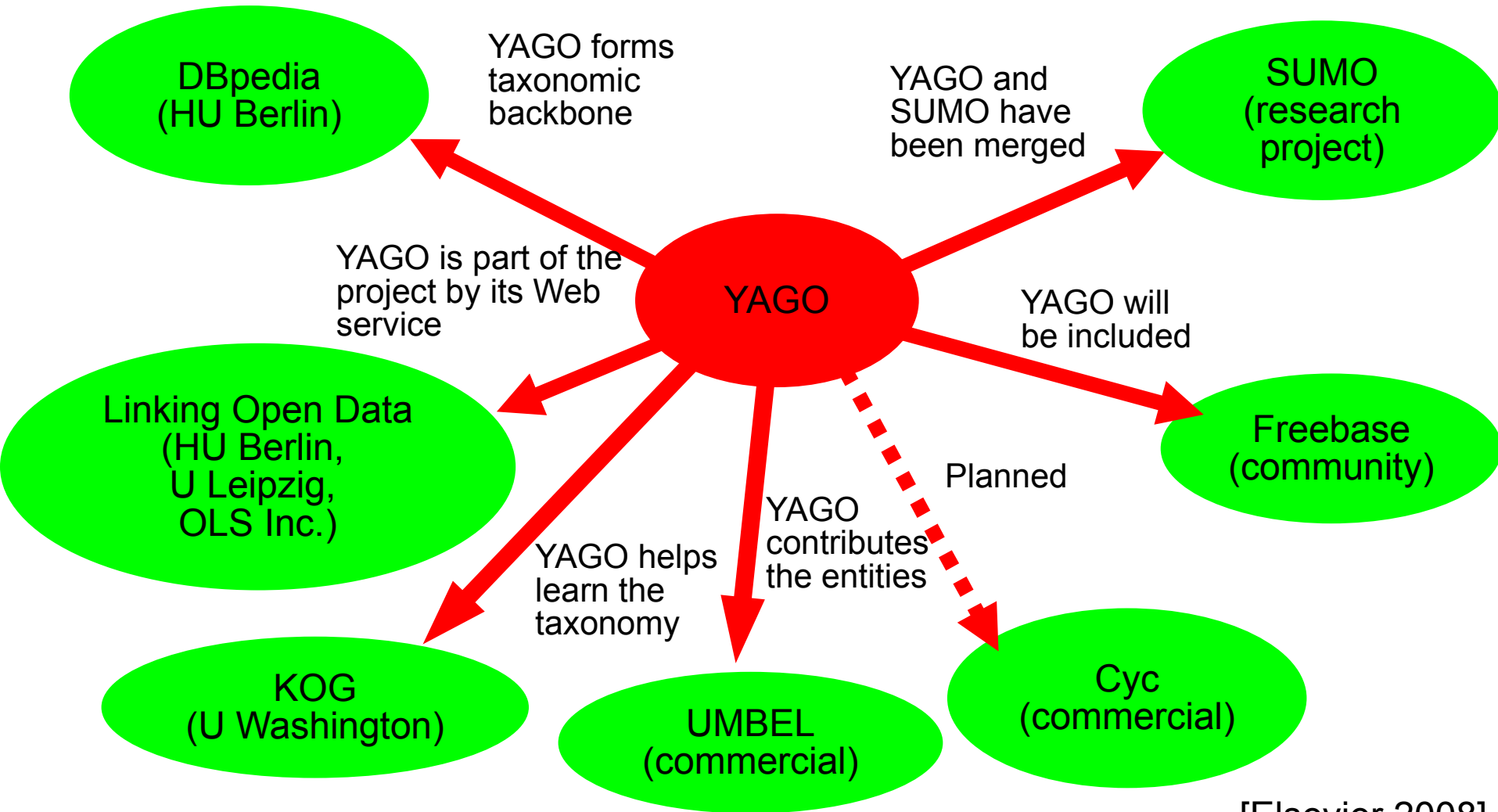
[WWW 2007]





# A Hitchhiker's Guide to Ontology

MAX-PLANCK-GESELLSCHAFT




[Elsevier 2008]



# This talk

MAX-PLANCK-GESELLSCHAFT

- Extract the ontology from a semi-structured corpus (Wikipedia)  
as a core ontology: YAGO ✓

 Extract information from corpora (e.g. the Web)  
to extend the core ontology: SOFIE



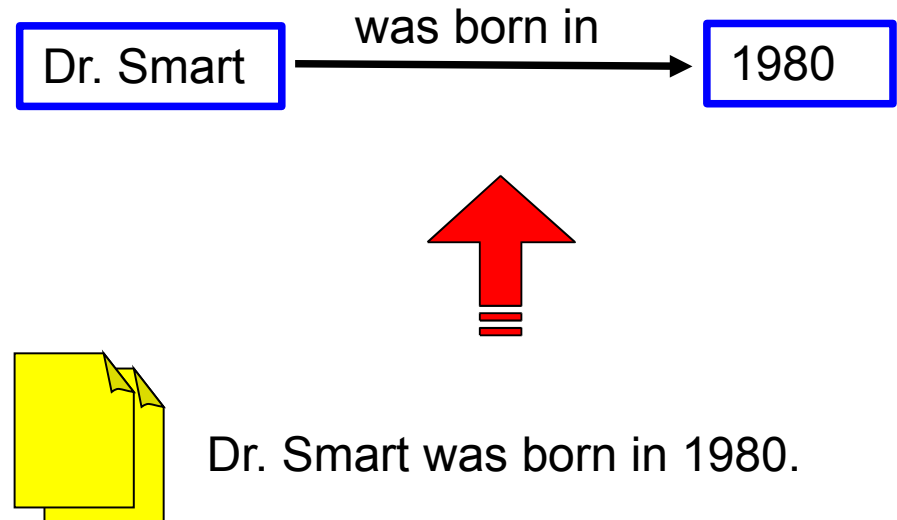
# Extending the Ontology

---

Our first approach:

LEILA - Combining Linguistic and Statistical Analysis [SIGKDD 2006]

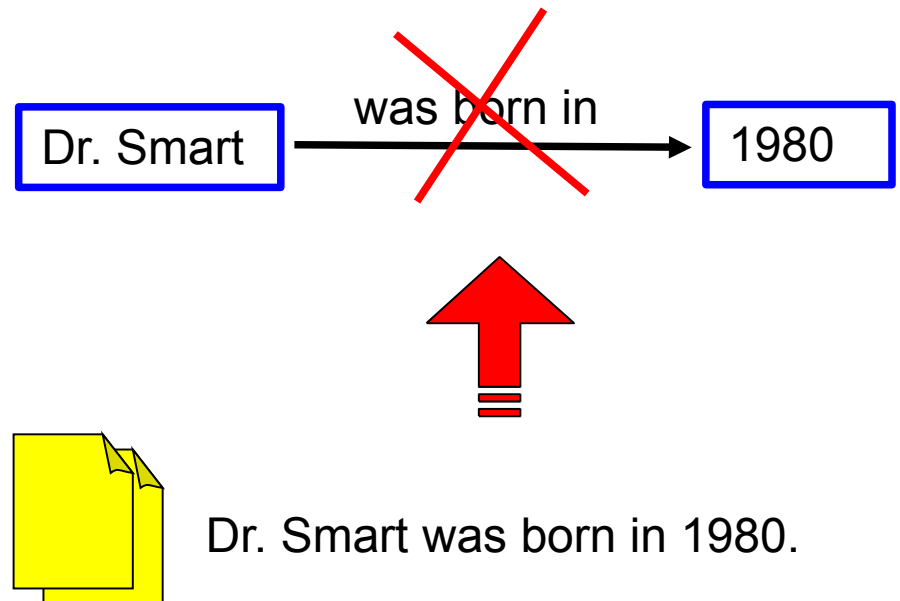
Worked well, but was slow.





# Extending the Ontology

bornInYear(Person, Year)

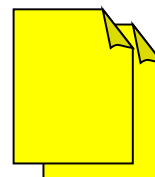
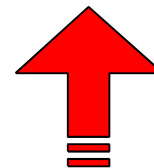
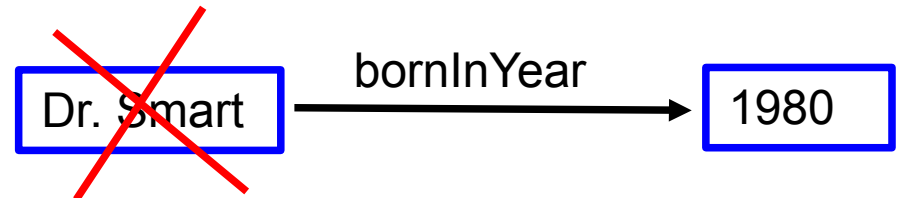




# Extending the Ontology

MAX-PLANCK-GESELLSCHAFT

## 1. Mapping patterns to relations



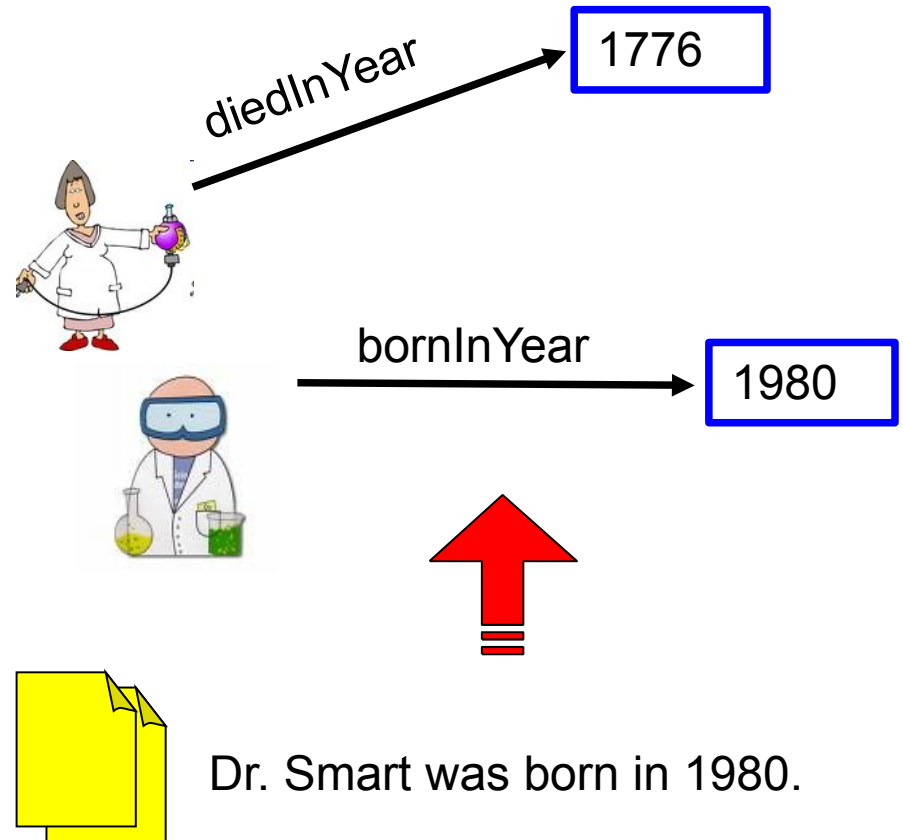
Dr. Smart was born in 1980.



# Extending the Ontology

MAX-PLANCK-GESellschaft

1. Mapping patterns to relations
2. Disambiguating entity names





# Extending the Ontology

MAX-PLANCK-GESELLSCHAFT

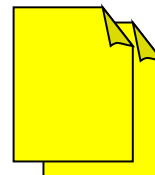
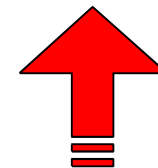
1. Mapping patterns to relations
2. Disambiguating entity names
3. Performing logical reasoning



bornInYear



1980

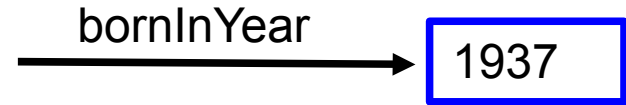


Dr. Smart was born in 1980.



# SOFIE: A Unifying Framework

1. Mapping patterns to relations
2. Disambiguating entity names
3. Performing logical reasoning



+

„Elvis was born in 1937.“

=

„X was born in Y“  
is a good pattern for  
bornInYear





# SOFIE: A Unifying Framework

1. Mapping patterns to relations
2. Disambiguating entity names
3. Performing logical reasoning

„X was born in Y“  
is a good pattern for  
bornInYear

+

„Dr. Smart was born in 1980.“

=



bornInYear



1980



# SOFIE: A Unifying Framework

1. Mapping patterns to relations
2. Disambiguating entity names
3. Performing logical reasoning

$r(x,y) \wedge \text{occurs}(p,x,y) \Rightarrow \text{isGoodPattern}(p,r)$

$\text{isGoodPattern}(p,r) \wedge \text{occurs}(p,x',y') \Rightarrow r(x',y')$

$\text{disambiguate}(\text{„Dr. Smart“}, \text{Sam\_Smart})[0.8]$

$\text{disambiguate}(\text{„Dr. Smart“}, \text{Lisa\_Smart})[0.2]$

„...The world as such, I would like to say – even though some will contradict – is not as it seems. As **Dr. Smart** pointed out, in his ground-breaking paper „The world according to Smart“, the world rather seems not what it seems...“

0.8



0.2





# SOFIE: A Unifying Framework



1. Mapping patterns to relations
2. Disambiguating entity names
3. Performing logical reasoning

$r(x,y) \wedge \text{occurs}(p,x,y) \Rightarrow \text{isGoodPattern}(p,r)$

$\text{isGoodPattern}(p,r) \wedge \text{occurs}(p,x',y') \Rightarrow r(x',y')$

$\text{disambiguate}(\text{„Dr. Smart“}, \text{Sam\_Smart})[0.8]$

$\text{bornInYear}(x,b) \wedge \text{diedInYear}(x,d) \Rightarrow b < d$

It's all just logical formulae with weights

Find truth values for the literals so that a maximal number of formulae is happy!

[WWW 2009]



# SOEIF: A Unifying Frame

MAX-PLANCK-GESELLSCHAFT



YAG

$r(x,y) \wedge \text{occu}$   
 $\text{isGoodPatter}$

$\text{bornInYear}(\text{Lvis}, 1957)$



te  
>  
)



dis:  
bor



$\wedge \text{diedInYear}(x,d) \Rightarrow b$



It's all j  
Find tru  
maxima



e with  
terals  
ilae is





# SOFIE: Information Extraction as MAX SAT

---

We have a Weighted MAX SAT Problem

$$r(x,y) \wedge s(x,z) \Rightarrow t(x,z) [w]$$

...

Problem:

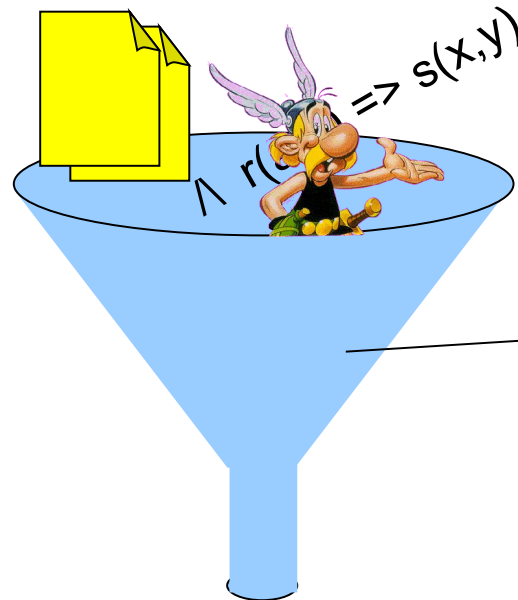
- ⌋ The Weighted MAX SAT Problem is NP-hard
- ⌋ Our instance contains YAGO (19 million facts) and textual facts (e.g. some dozens of thousands)
- ⌋ The best-known approximation algorithm cannot deal well with our specific instance



# SOFIE: A Unifying Framework

MAX-PLANCK-GESELLSCHAFT

## Weighted MAX SAT Problem



**Algorithm**  
**Functional MAX SAT**

```
FOR i=1 TO 42  
...  
NEXT i
```

Polynomial  
time

Approximation  
Guarantee

Sam\_Smart

1980



# SOFIE: Relation to Markov Logic

MAX-PLANCK-GESELLSCHAFT

$r(x,y) \wedge s(x,z) \Rightarrow t(x,z) [w]$   
...

Number of satisfied instances of the  $i^{\text{th}}$  formula

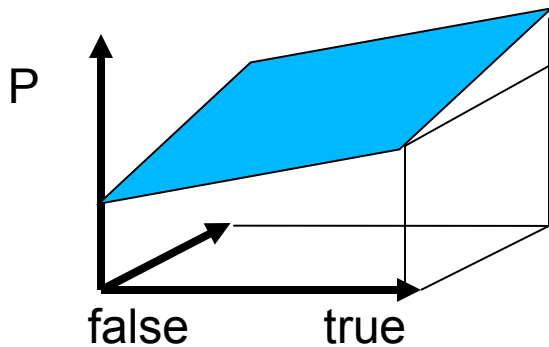
Weight of the  $i^{\text{th}}$  formula

$$P(X) \sim \prod e^{\text{sat}(i,X) w_i}$$

$$\max_x \prod e^{\text{sat}(i,X) w_i}$$

$$\max_x \log( \prod e^{\text{sat}(i,X) w_i} )$$

$$\max_x \sum \text{sat}(i,X) w_i$$



bornIn(SamSmart, 1980)

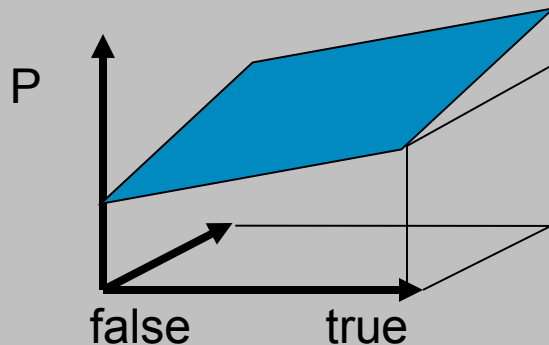


# SOFIE: Relation to Markov Logic

MAX-PLANCK-GESELLSCHAFT

$r(x,y) \wedge s(x,z) \Rightarrow t(x,z) [w]$

...



bornIn(Sam, 1980)

Number of satisfied instances of the  $i^{\text{th}}$  formula

Weight of the  $i^{\text{th}}$  formula

$$P(X) \sim \prod e^{\text{sat}(i,X) w_i}$$

$$\max_x \prod e^{\text{sat}(i,X) w_i}$$

$$\max_x \log( \prod e^{\text{sat}(i,X) w_i} )$$

$$\max_x \sum \text{sat}(i,X) w_i$$

Weighted MAX SAT problem





# SOFIE: Experiments

MAX-PLANCK-GESellschaft

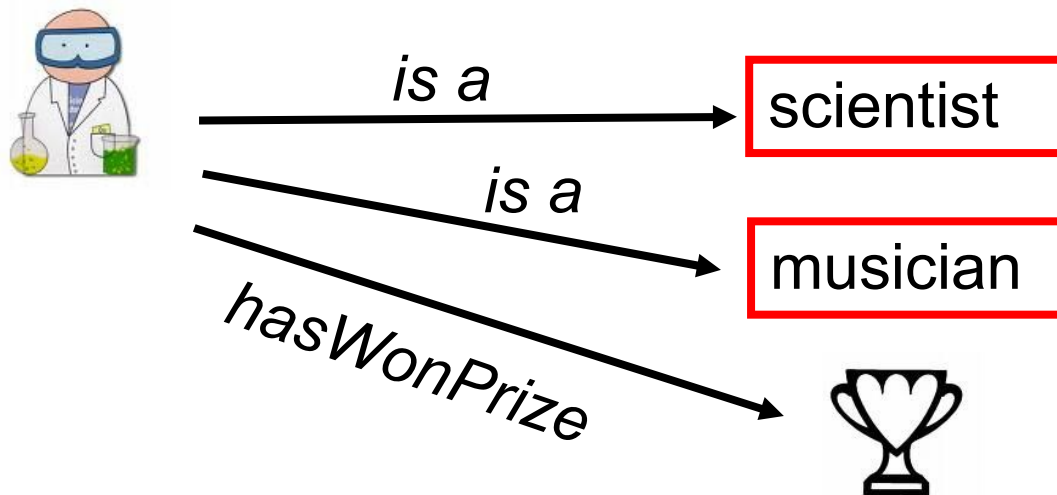
Corpus	Type	# Docs	Relations	Time	Precision
Wikipedia toy corpus	structured	100	3	8min	100%
Wikipedia subcorpus	semi-structured	2000	15	15h	94%
News article toy corpus	unstructured	150	1	24min	91%
Biographies from Web	unstructured	3440	5	15h	90%



# The Excellent Scientist

Back to the original question:

*Which scientist was also a musician and has won a prize?*





# The Excellent Scientist


MAX-PLANCK-GESELLSCHAFT


Back to the original question:


*Which scientist was also a musician and has won a prize?*


X	isa	scientist
X	isa	musician
X	hasWonPrize	Y

**Slides**


37  Slide 37

38  Slide 38

39  Slide 39

40  Slide 40

Normal Outline Notes Handout Slide Sorter



# The Excellent Scientist

---

Back to the original question:

*Which scientist was also a musician and has won a prize?*

X	isa	scientist
X	isa	musician
X	hasWonPrize	Y

Fabian M. Suchanek YAGO - Yet Another Great Ontology 38

**Slides**


37 Slide 37

38 Slide 38

39 Slide 39

40



Normal Outline Notes Handout Slide Sorter



# The Excellent Scientist

---

Back to the original question:

Which   *and has won a prize?*

isa scientist  
 isa musician  
 hasWonPrize Y

Fabian M. Suchanek YAGO - Yet Another Great Ontology 38



## Query Form

YAGO-query:

?id0:	<input type="text" value="?x"/>	isA	<input type="text" value="scientist"/>
?id1:	<input type="text" value="?x"/>	isA	<input type="text" value="musician"/>
?id2:	<input type="text"/>		<input type="text"/>

Daten absenden

- Home
- Use YAGO
- Query YAGO
- References
- Related Projects
- Acknowledgements



## Query Form

YAGO-query:

?id0:	<input type="text" value="?x"/>	isA	<input type="text" value="scientist"/>
?id1:	<input type="text" value="?x"/>	isA	<input type="text" value="musician"/>
?id2:	<input type="text"/>		<input type="text"/>

Daten absenden

Solving query, please wait...

- Home
- Use YAGO
- Query YAGO
- References
- Related Projects
- Acknowledgements



- Home
- Use YAGO
- Query YAGO
- References
- Related Projects
- Acknowledgements

### Query Form

YAGO-query:

?id0:	<input type="text" value="?x"/>	isA	<input type="text" value="scientist"/>
?id1:	<input type="text" value="?x"/>	isA	<input type="text" value="musician"/>
?id2:	<input type="text"/>		<input type="text"/>

Daten absenden

- ?musician = [musician](#)
- ?scientist = [scientist](#)
- ?x = [Brian May](#)
- ?y = [Ivor Novello Award](#)

---

- ?musician = [musician](#)
- ?scientist = [scientist](#)
- ?x = [Ginger Rogers](#)
- ?y = [Academy Award for Best Actor](#)

---

- ?musician = [musician](#)
- ?scientist = [scientist](#)
- ?x = [Carol Channing](#)
- ?y = [Tony Award](#)

---

- ?musician = [musician](#)
- ?scientist = [scientist](#)
- ?x = [Carol Channing](#)
- ?y = [Sarah Siddons Award](#)

I am a Screenshot

Brian May – Musician and Astrophysicist

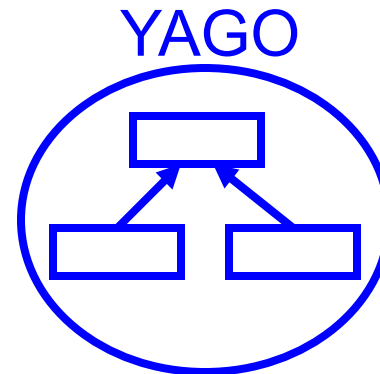
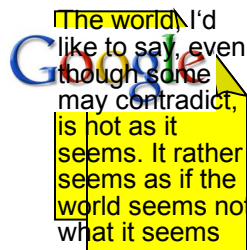




# Conclusion

MAX-PLANCK-GESELLSCHAFT

- ⌋ We made a step towards gathering the knowledge of this world in a structured ontology



- ⌋ They do exist



<http://mpii.de/yago>



# References

---

- [SIGKDD 2006] Fabian M. Suchanek, Georgiana Ifrim and Gerhard Weikum  
"Combining Linguistic and Statistical Analysis to Extract Relations from Web Documents"  
International Conference on Knowledge Discovery and Data Mining (SIGKDD 2006)
- [WWW 2007] Fabian M. Suchanek, Gjergji Kasneci and Gerhard Weikum  
"YAGO - A Core of Semantic Knowledge"  
International World Wide Web conference (WWW 2007)
- [Elsevier 2008] Fabian M. Suchanek, Gjergji Kasneci and Gerhard Weikum  
"YAGO - A Large Ontology from Wikipedia and WordNet"  
Elsevier Journal of Web Semantics 2008
- [WWW 2009] Fabian M. Suchanek, Mauro Sozio, Gerhard Weikum  
„SOFIE – A Self-Organizing Framework for Information Extraction“  
International World Wide Web conference (WWW 2009)



# Future Work at the Max-Planck Institute

---

- personalize YAGO (Shady Elbassuoni, Maya Ramanath)
- use social networks to extend YAGO (Maya, Sharat, Ashwin)
- make YAGO multilingual (Gerard de Melo)
- add Web services (Nicoleta Preda)
- make querying efficient (Gjergji Kasneci)
- store YAGO efficiently (Thomas Neumann)
- make reasoning efficient (Mauro Sozio, Martin Theobald)
- provide good visualization (Shady Elbassuoni)
- add biomedical knowledge (Alessandro Fiori)
- add multimodal support (Martin Schreiber)
- add natural language support (with Alexander Koller)



# Future Work at UW?

---

- ♪ learn also relations and classes (KOG, KYLIN)
- ♪ investigate multilingual support (PanImages)
- ♪ investigate reasoning/inferencing (Holmes, MLN)
- ♪ feed information back into Wikipedia (Intelligence in Wikipedia)
- ♪ extract information on a larger scale (KnowItAll/TextRunner)
- ♪ guide the information extraction process (Alice/LLL)

Instead of competing with these projects,  
join forces with them!



# YAGO: Knowledge Representation

**OWL Full**

**RDFS**

**YAGO**

ADTs

Reification

subClassOf

Acyclicity

Datatypes

Transitivity

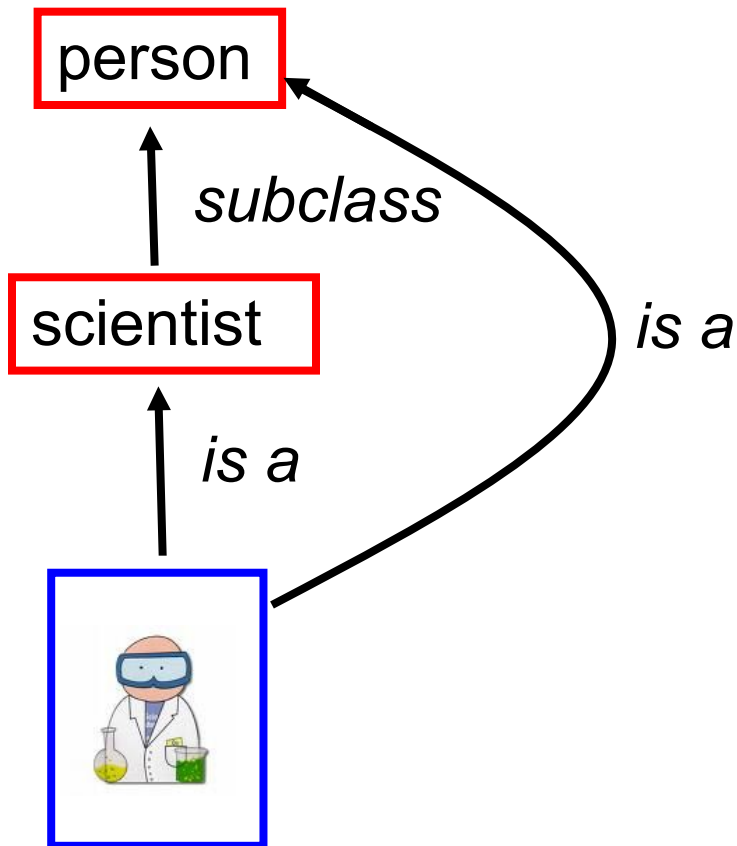
Property  
Restrictions

**OWL DL**



# YAGO Model: Logical aspects

MAX-PLANCK-GESELLSCHAFT



Axioms:

$(x, \text{is\_a}, y)$

$(y, \text{subclass}, z)$

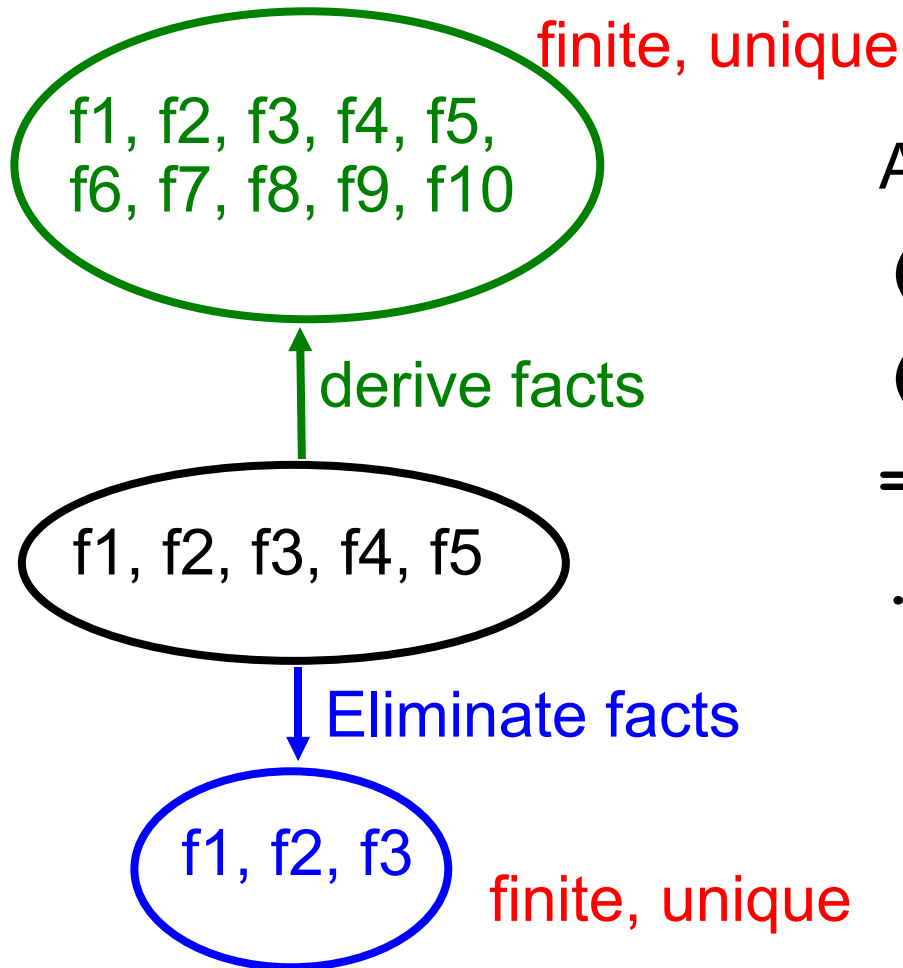
$\Rightarrow (x, \text{is\_a}, z)$

...



# YAGO Model: Logical aspects

MAX-PLANCK-GESELLSCHAFT



Axioms:

$(x, \text{is\_a}, y)$

$(y, \text{subclass}, z)$

$\Rightarrow (x, \text{is\_a}, z)$

...

[WWW 2007]



# SOFIE rules!

$R(X, Y)$   
 $\wedge R(X, Z)$   
 $\wedge \text{type}(R, \text{functionalRelation})$   
 $\Rightarrow Y = Z$

$\text{occurs}(P, WX, WY)$   
 $\wedge \text{refersTo}(WX, X)$   
 $\wedge \text{refersTo}(WY, Y)$   
 $\wedge R(X, Y)$   
 $\Rightarrow \text{expresses}(P, R)$

$\text{occurs}(P, WX, WY)$   
 $\wedge \text{expressed}(P, R)$   
 $\wedge \text{refersTo}(WX, X)$   
 $\wedge \text{refersTo}(WY, Y)$   
 $\wedge \text{range}(R, D1)$   
 $\wedge \text{domain}(R, D2)$   
 $\wedge \text{type}(X, D1)$   
 $\wedge \text{type}(Y, D2)$   
 $\Rightarrow R(X, Y)$

$\text{disambiguationPrior}(W, X) \Rightarrow \text{refersTo}(W, X)$

$\neg R(X, Y)$

$\text{bornInYear}(X, B) \wedge \text{diedInYear}(X, D) \Rightarrow B < D$





# SOFIE: Clause transformation

MAX-PLANCK-GESELLSCHAFT

## Rules

$$r(X,Y) \wedge s(X,Y) \Rightarrow t(X,X)$$
$$u(a)$$

## Grounded Rules

$$r(a,a) \wedge s(a,a) \Rightarrow t(a,a)$$
$$r(a,b) \wedge s(a,b) \Rightarrow t(a,a)$$
$$r(b,a) \wedge s(b,a) \Rightarrow t(b,b)$$
$$r(b,b) \wedge s(b,b) \Rightarrow t(b,b)$$
$$u(a)$$

Entities {a,b}

## Clauses

$$\neg r(a,a) \vee \neg s(a,a) \vee t(a,a)$$
$$\neg r(a,b) \vee \neg s(a,b) \vee t(a,a)$$
$$\neg r(b,a) \vee \neg s(b,a) \vee t(b,b)$$
$$\neg r(b,b) \vee \neg s(b,b) \vee t(b,b)$$
$$u(a)$$



# SOFIE: Clause transformation

## Clauses

1

 $\neg r(a,a) \vee \neg \text{s(a,a)} \vee t(a,a)$  $\neg r(a,b) \vee \neg s(a,b) \vee t(a,a)$  $\neg r(b,a) \vee \neg s(b,a) \vee t(b,b)$  $\neg r(b,b) \vee \neg s(b,b) \vee t(b,b)$  $u(a)$ 

## Textual Facts

 $r(a,a)$  [w1] $r(a,b)$  [w2] $r(b,a)$  [w3] $r(b,b)$  [w4]

## YAGO

 $s(a,a)$



# SOFIE: Clause weighting

MAX-PLANCK-GESELLSCHAFT

## Clauses

$\neg 1 \quad \vee \neg 1 \quad \vee t(a,a) [w1]$

$\neg 1 \quad \vee \neg s(a,b) \vee t(a,a) [w2]$

$\neg 1 \quad \vee \neg s(b,a) \vee t(b,b) [w3]$

$\neg 1 \quad \vee \neg s(b,b) \vee t(b,b) [w4]$

$u(a) [W]$

## Textual Facts

$r(a,a) [w1]$

$r(a,b) [w2]$

$r(b,a) [w3]$

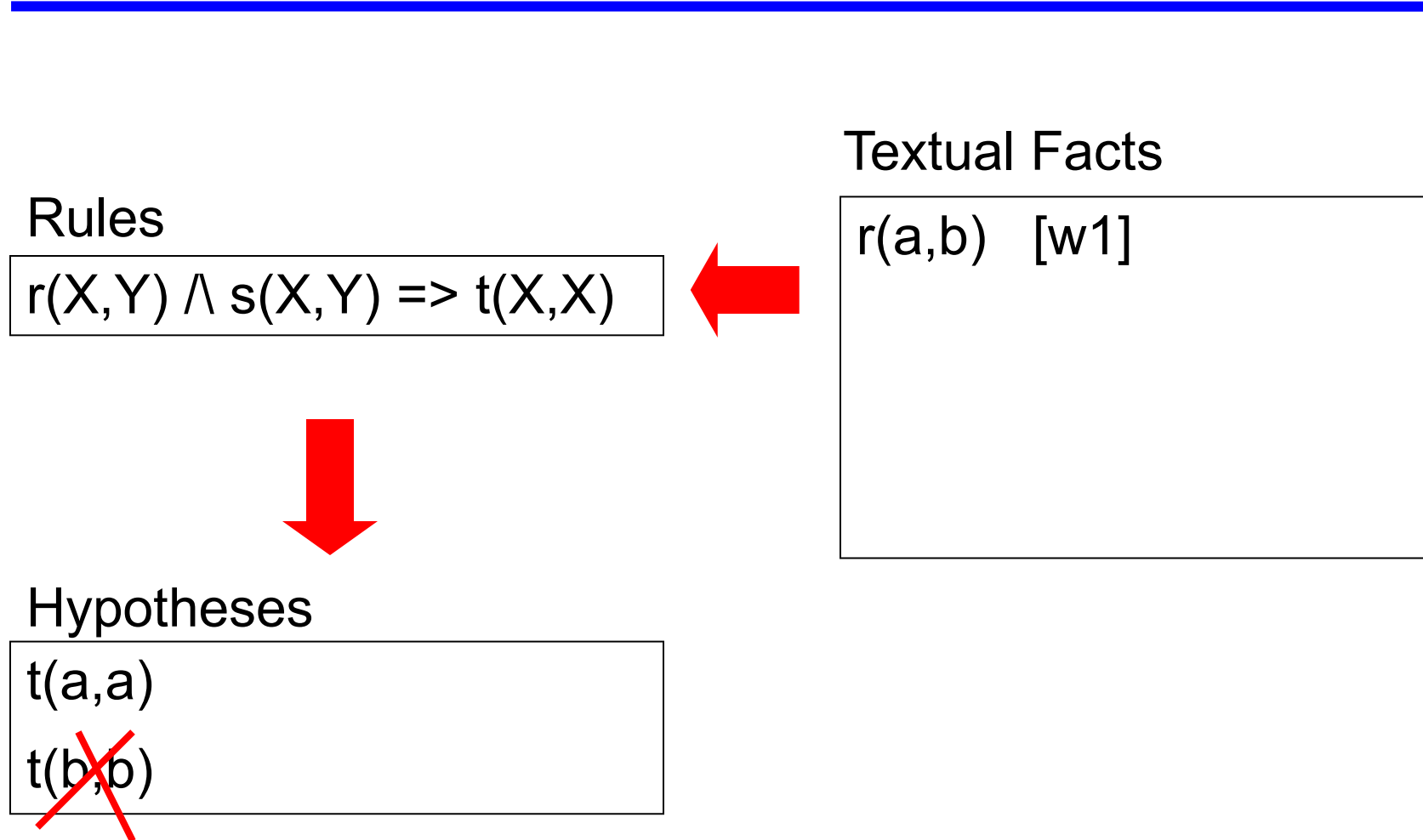
$r(b,b) [w4]$

## YAGO

$s(a,a)$



# SOFIE: Hypothesis generation



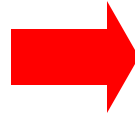


# SOFIE: Hypothesis generation

MAX-PLANCK-GESELLSCHAFT

## Rules

$r(X, Y) \wedge s(X, Y) \Rightarrow t(X, X)$



## Grounded Rules

$r(a, a) \wedge s(a, a) \Rightarrow t(a, a)$   
 $r(a, b) \wedge s(a, b) \Rightarrow t(a, a)$



## Hypotheses

$t(a, a)$



# SOFIE: Functional MAX SAT Algorithm

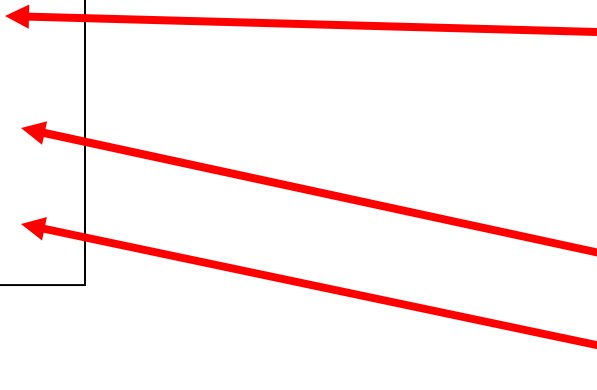
The functional MAX SAT Algorithm considers only unit clauses.

Variables

X	=0
Y	=0
Z	=1

Clauses

<del><math>\neg X \vee \neg Z</math></del>	[w1]
<del><math>\neg X \vee \neg Y</math></del>	[w1]
<del><math>\neg Y \vee \neg Z</math></del>	[w1]
<del>Z</del>	[w1]





# SOFIE: Experiments

MAX-PLANCK-GESELLSCHAFT

Corpus	Type	# Docs	# Rel	Time	# Facts	Precision	Recall
Wikipedia toy corpus	structured	100	3	8min	165	100%	98%
Wikipedia toy corpus	semi-structured 50% infoboxes removed	100	3	8min	165	100%	57%
Wikipedia subcorpus	semi-structured	2000	15	15h	505	94%	?
News article toy corpus	unstructured	150	1	24min	35, 46	91%	24%, 31%
Snowball					65	56%	31%
Biographies from Web	unstructured	3440	5	15h	744	90%	?