

# YAGO: a Multilingual Knowledge Base from Wikipedia, Wordnet, and Geonames

Thomas Rebele<sup>1</sup>, Fabian Suchanek<sup>1</sup>, Johannes Hoffart<sup>2</sup>,  
Joanna Biega<sup>2</sup>, Erdal Kuzey<sup>2</sup>, Gerhard Weikum<sup>2</sup>



<sup>1</sup> Télécom ParisTech    <sup>2</sup> Max Planck Institute for Informatics

2016-10-17

# Outline

What is a knowledge base?

YAGO

Applications

Conclusion

# What is a knowledge base?

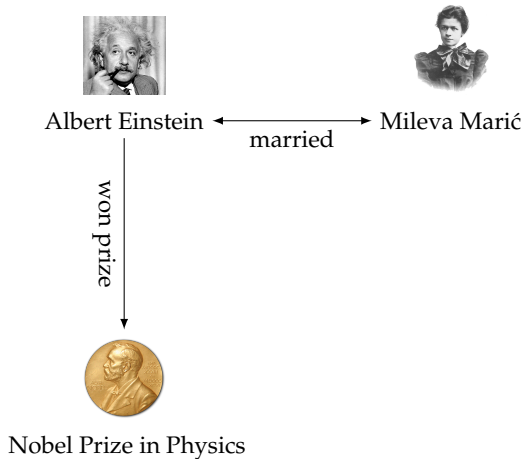
- ▶ What is a knowledge base?

YAGO

Applications

Conclusion

# What is a knowledge base?



# YAGO

What is a knowledge base?

► YAGO

Applications

Conclusion



- ▶ >100m facts and >100 relations
- ▶ combines Wordnet and Wikipedia
- ▶ merges facts from 10 Wikipedias
- ▶ includes space and time metafacts
- ▶ manually evaluated

## Contributors:



Fabian Suchanek



Johannes Hoffart



Joanna Biega



Gerhard Weikum



Erdal Kuzey



Farzaneh  
Mahdisoltani



Gjergji Kasneci



Thomas Rebele

# YAGO - Taxonomy: WordNet and Wikipedia

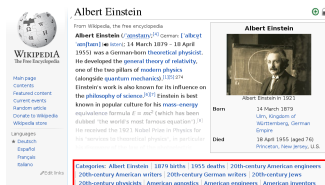
What is a knowledge base?

## YAGO

- Taxonomy: WordNet and Wikipedia
- Multilingual: 10 Wikipedias
- Metafacts: Time and Space
- Accuracy: Manual Evaluation

Applications

Conclusion



**Albert Einstein**

From Wikipedia, the free encyclopedia

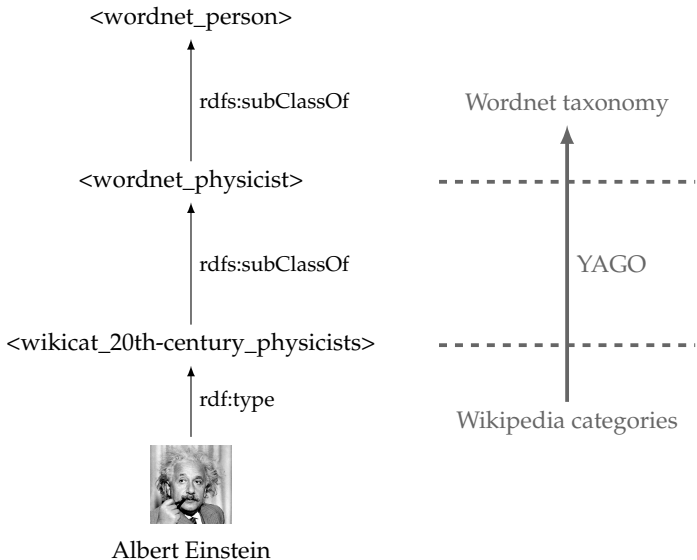
**Albert Einstein** (listen<sup>ⓘ</sup>)<sup>[de]</sup> German (listen<sup>ⓘ</sup>)<sup>[en]</sup> listen<sup>ⓘ</sup> listen<sup>ⓘ</sup> **14 March 1879** – **18 April 1955**) was a German-born theoretical physicist. He developed the general theory of relativity, one of the two pillars of modern physics (alongside quantum mechanics).<sup>[E]</sup><sup>[2]</sup><sup>[3]</sup> Einstein's work is also known for its influence on the philosophy of science.<sup>[4]</sup><sup>[5]</sup> Einstein is best known in popular culture for his mass-energy equivalence formula  $E=mc^2$  (which has been dubbed "the world's most famous equation").<sup>[6]</sup> He received the 1921 Nobel Prize in Physics for his "services to theoretical physics", in particular his development of the photoelectric effect.

**Born** 14 March 1879  
Ulm, Kingdom of Württemberg, German Empire

**Died** 18 April 1955 (aged 76)  
Princeton, New Jersey, U.S.

**Categories:** Albert Einstein | 1879 births | 1955 deaths | 20th-century American engineers | 20th-century American writers | 20th-century German writers | 20th-century Jews | 20th-century physicists | American Zionists | American engineers | American inventors

# YAGO - Taxonomy: WordNet and Wikipedia







# YAGO - Multilingual: 10 Wikipedias

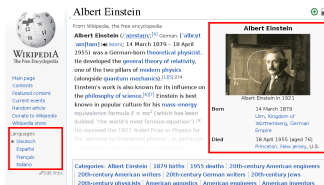
What is a knowledge base?

## YAGO

- Taxonomy: WordNet and Wikipedia
- Multilingual: 10 Wikipedias
- Metafacts: Time and Space
- Accuracy: Manual Evaluation

Applications

Conclusion



The image shows a screenshot of the English Wikipedia article for Albert Einstein. On the left side, there is a navigation menu with several items: 'Main page', 'Contents', 'Featured content', 'Current events', 'Random article', 'Donate to Wikipedia', and 'Wikipedia store'. A red box highlights the bottom four items: 'Language', 'Deutsch', 'Español', 'Français', and 'Italiano'. On the right side, there is an infobox for Albert Einstein, which includes a portrait photo, his birth and death dates, and a brief biography. A red box highlights the entire infobox area.

**Albert Einstein**

From Wikipedia, the free encyclopedia

**Albert Einstein** (listen<sup>ⓘ</sup>)<sup>[de]</sup> German (listen<sup>ⓘ</sup>)<sup>[en]</sup> 14 March 1879 – 18 April 1955) was a German-born theoretical physicist. He developed the general theory of relativity, one of the two pillars of modern physics (alongside quantum mechanics).<sup>[E]</sup><sup>[I]</sup><sup>[2]</sup><sup>[3]</sup> Einstein's work is also known for its influence on the philosophy of science.<sup>[E]</sup><sup>[I]</sup> Einstein is best known in popular culture for his mass–energy equivalence formula *E* = *mc*<sup>2</sup> (which has been dubbed "the world's most famous equation").<sup>[E]</sup><sup>[I]</sup> He received the 1921 Nobel Prize in Physics for his services to theoretical physics, in particular for his discovery of the photoelectric effect.

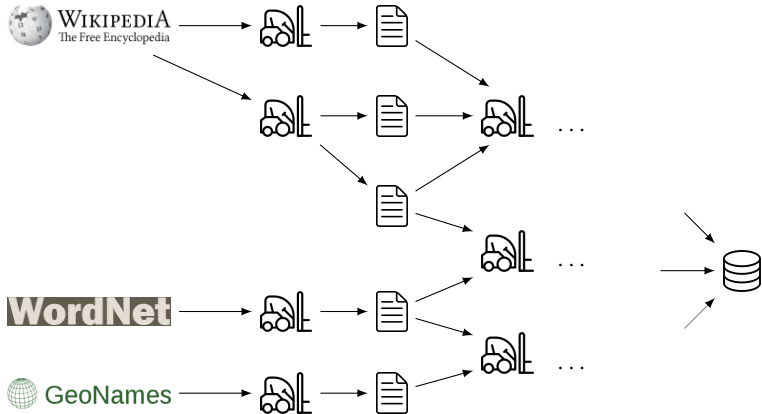
**Born** 14 March 1879  
Ulm, Kingdom of Württemberg, German Empire

**Died** 18 April 1955 (aged 76)  
Princeton, New Jersey, U.S.

Categories: Albert Einstein | 1879 births | 1955 deaths | 20th-century American engineers | 20th-century American writers | 20th-century German writers | 20th-century Jews | 20th-century physicists | American agnostics | American engineers | American inventors

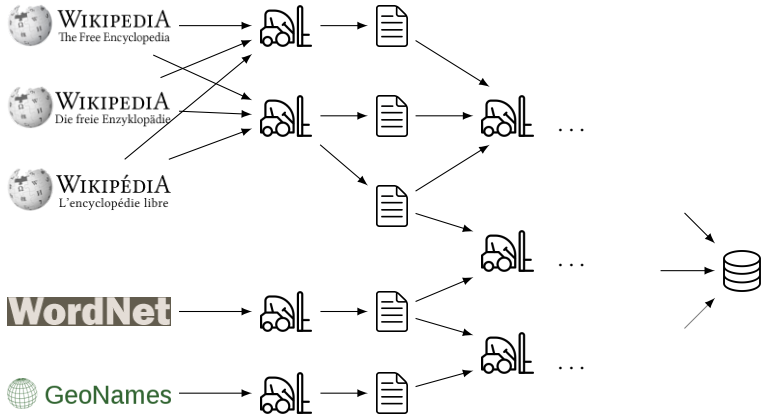
# YAGO - Multilingual: 10 Wikipedias

Icons designed by  
Madebyoliver from Flaticon



# YAGO - Multilingual: 10 Wikipedias

Icons designed by  
Madebyoliver from Flaticon



# YAGO - Multilingual: 10 Wikipedias



↓  
birth\_place

(Einstein, Ulm)  
(Hawking, Oxford)  
(Euler, Basel)  
...



↓  
geburtsort

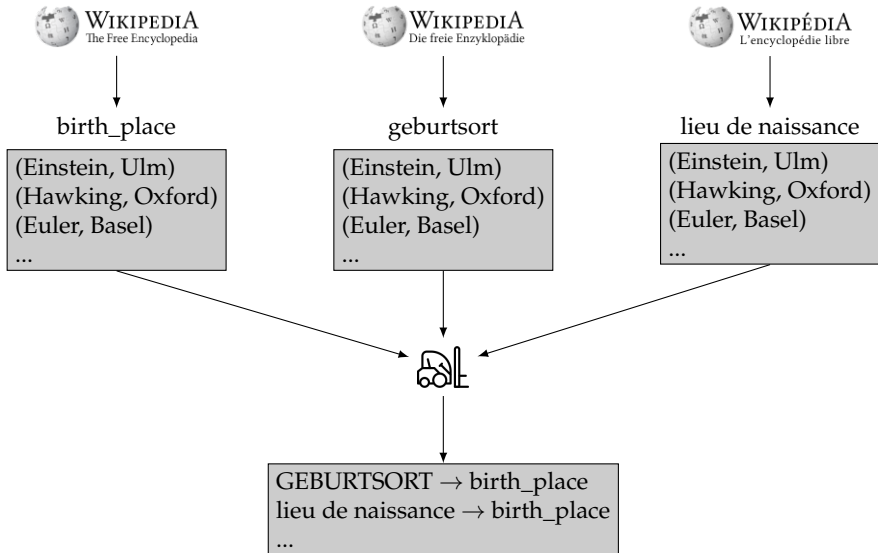
(Einstein, Ulm)  
(Hawking, Oxford)  
(Euler, Basel)  
...



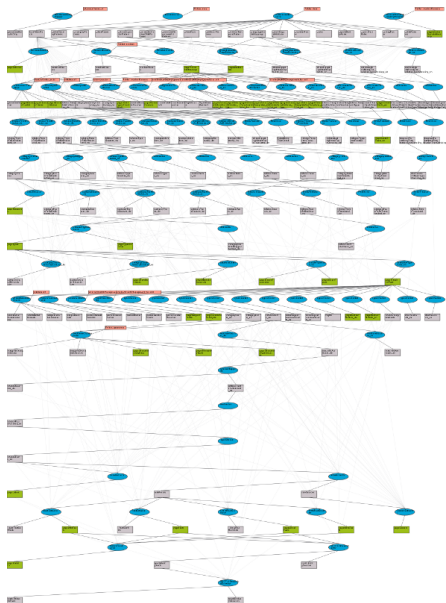
↓  
lieu de naissance

(Einstein, Ulm)  
(Hawking, Oxford)  
(Euler, Basel)  
...

# YAGO - Multilingual: 10 Wikipedias



# YAGO - Multilingual: 10 Wikipedias



YAGO architecture

What is a knowledge base?

## YAGO

Taxonomy: WordNet and Wikipedia

Multilingual: 10 Wikipedias

► Metafacts: Time and Space

Accuracy: Manual Evaluation

Applications

Conclusion



# YAGO - Metafacts: Time and Space



Albert Einstein



married



Mileva Marić

# YAGO - Metafacts: Time and Space



Albert Einstein



Mileva Marić

married

<id\_tiwmcu\_16x\_11ovtp>

# YAGO - Metafacts: Time and Space



Albert Einstein — married — Mileva Marić

<id\_tiwmcu\_16x\_11ovtp>

since

1903-##-##

until

1919-##-##

# YAGO - Metafacts: Time and Space



Albert Einstein

married

Mileva Marić

<id\_tiwmcu\_16x\_11ovtp>

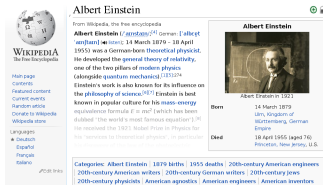
source

since

until

1903-##-##

1919-##-##



The screenshot shows the Wikipedia article for Albert Einstein. It includes the title "Albert Einstein", a brief biography, a photo of Einstein, and a table of his birth and death information. The birth date is listed as 14 March 1879 in Ulm, Kingdom of Württemberg, German Empire, and the death date is 18 April 1955 (aged 76) in Princeton, New Jersey, U.S. The page also features a sidebar with navigation links and a list of categories at the bottom.

<[http://en.wikipedia.org/wiki/Albert\\_Einstein](http://en.wikipedia.org/wiki/Albert_Einstein)>

What is a knowledge base?

## YAGO

Taxonomy: WordNet and Wikipedia

Multilingual: 10 Wikipedias

Metafacts: Time and Space

- ▶ Accuracy: Manual Evaluation

Applications

Conclusion

# YAGO - Accuracy: Manual Evaluation

## YAGO3 Evaluation - Judge The Fact

There is an airline connection from **Dalaman Airport** to **Leeds Bradford International Airport**

byTransport

byTransport

Submit

**Jet2.com**

**Monarch Airlines**

right wrong ignore [help](#)



Evaluating YAGO3 Facts as  
Thomas R.

Extraction technique:  
"FlightExtractor"

### Dalaman Airport

The screenshot shows the Wikipedia article for Dalaman Airport. The page title is "Dalaman Airport". The introductory text states: "From Wikipedia, the free encyclopedia **Dalaman Airport** (IATA: **DLM**, ICAO: **LTBS**) is an international airport and one of three serving south-west Turkey, the others being Milas-Bodrum Airport and Antalya Airport. It has two terminals. The old terminal is used for domestic flights and the new terminal is for international flights. The airport serves the surrounding tourist areas and environs of Dalaman. Flights are available to and from over 120". A featured image shows the Dalaman International Airport terminal with the caption "International Terminal" and "IATA: DLM - ICAO: LTBS". The left sidebar contains navigation links such as "Main page", "Contents", "Featured content", "Random article", "Donate to Wikipedia", "Wikipedia store", "Interaction", "Help", "About Wikipedia", and "Community portal". The top navigation bar includes "Article", "Talk", "Read", "Edit", "View history", and a search box.

Screenshot of evaluation interface

# YAGO - Accuracy: Manual Evaluation

## YAGO3 Evaluation - Current Standings

### Overall State of the Evaluation

98.07% of 4412 evaluations were judged to be correct. This gives a weighted average Wilson center of 95.03% (4.19 % width)

### Evaluation Results for Relations

Evaluation Target	Evaluations	Correct	Ratio (%)	Wilson Center (%)	Wilson Width (%)	Progress
<happenedIn>	87	87	100	97.89	2.11	
<byTransport>	120	119	99.17	97.64	2.21	
<hasExpenses>	135	133	98.52	97.18	2.42	
<hasExport>	60	60	100	96.99	3.01	
<hasISBN>	59	59	100	96.94	3.06	
<exports>	58	58	100	96.89	3.11	
<isMarriedTo>	57	57	100	96.84	3.16	

### Screenshot of evaluation result

- ▶ 2 months evaluation, 15 participants
- ▶ evaluated 4412 facts of 76 relations (with 60m total facts)
- ▶ 98% facts of the sample were correct
- ▶ Wilson center: 95%, interval width: 4.2%

# Applications

What is a knowledge base?

YAGO

► Applications

Conclusion





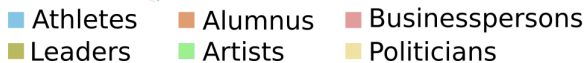
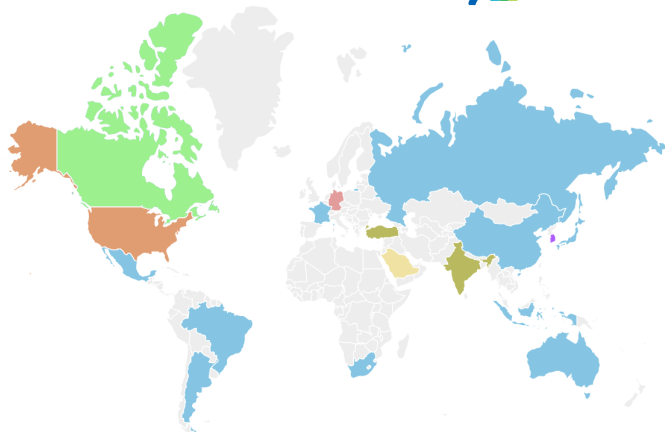
- ▶ research and community project
- ▶ knowledge base built from Wikipedia
- ▶ imported YAGO's taxonomy

# Applications - IBM Watson



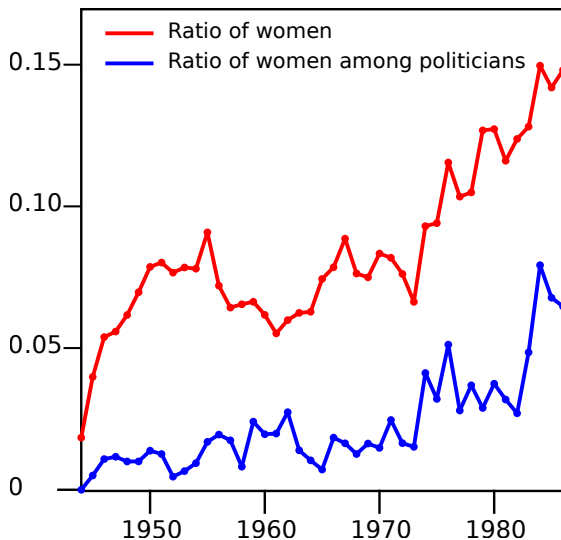
IBM Watson

- ▶ participated in *Jeopardy* (TV quizz show)
- ▶ first place against humans
- ▶ used YAGO's type hierarchy



Analysis of the panama papers [▷ github.com/yago-naga/aida](https://github.com/yago-naga/aida)

## Applications - Semantic Culturomics



Gender imbalance over time [▷ w3id.org/yago/mpi/le-monde](http://w3id.org/yago/mpi/le-monde)

# Conclusion

What is a knowledge base?

YAGO

Applications

► Conclusion



- ▶ combines Wikipedia with Wordnet
- ▶ facts from 10 Wikipedias
- ▶ time and space metafacts
- ▶ >100 relations
- ▶ >100m facts
- ▶ >95% precision

## Future work:

- ▶ improve release cycle
- ▶ textual extension
- ▶ commonsense knowledge
- ▶ intensional knowledge
- ▶ completeness