

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

Fabian Suchanek¹, Johannes Hoffart², Joanna Biega²,
Thomas Rebele¹, Erdal Kuzey², Gerhard Weikum²



¹ Télécom ParisTech ² Max Planck Institute for Informatics

2017-06-23

Plan

What is a knowledge base?

YAGO

- Combining WordNet with Wikipedia

- Multilingual: facts and entities from 10 Wikipedias

- Time and space

- Accuracy

Applications

- DBpedia

- IBM Watson

- AIDA

- Semantic Culturomics

AMIE

- Partial completeness assumption

- Rule mining in knowledge bases

- Canonicalization of knowledge bases

- Wikilinks semantification

- Completeness

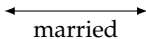
Conclusion

References

What is a knowledge base?



Mileva Marić



Albert Einstein

Image source: <https://wikipedia.de>

What is a knowledge base?

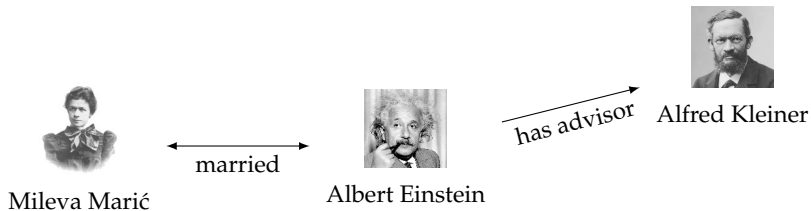


Image source: <https://wikipedia.de>

What is a knowledge base?

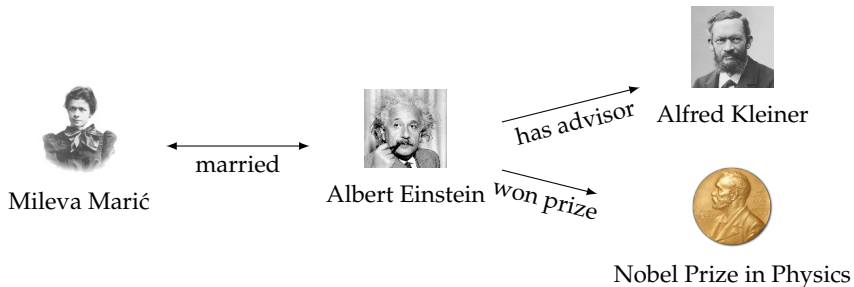


Image source: <https://wikipedia.de>

Where do knowledge bases help?

Applications of knowledge bases

- ▶ question answering
- ▶ semantic search
- ▶ intelligent personal assistants
- ▶ investigative journalism
- ▶ text analysis
- ▶ machine translation

Hence, everybody does it


- ▶ Google: Knowledge Graph
- ▶ Amazon: Evi
- ▶ Microsoft: Satori

What is YAGO?

- ▶ knowledge base with 10 million entities and >210 million facts
- ▶ multilingual facts from 10 languages
- ▶ focus on precision
- ▶ developed by Max-Planck Institute for Informatics and Télécom ParisTech



What is YAGO?



WIKIPEDIA
The Free Encyclopedia

- [Main page](#)
- [Contents](#)
- [Featured content](#)
- [Current events](#)
- [Random article](#)
- [Donate to Wikipedia](#)
- [Wikipedia store](#)

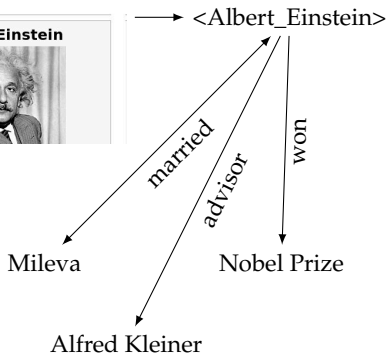
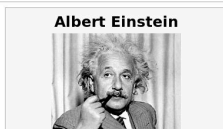
Article [Talk](#) Read [View source](#) [View history](#)

Albert Einstein

From Wikipedia, the free encyclopedia

Albert Einstein

(/ˈaɪnstaɪn/^[4] German: [ˈalbɛçt ˈaɪnʃtaɪn] • listen);
14 March 1879 – 18 April
1955) was a German-born
theoretical physicist.



What does YAGO know?

theme name	number of facts
taxonomy-related facts	95 m
simplified taxonomy	17 m
main facts	55 m
GeoNames facts	39 m
meta-facts	203 m
multilingual class labels	787 k
maps to other knowledge bases	4 m
raw information Wikipedia	296 m
redirect links and anchor texts	471 m

Themes of YAGO

What does YAGO know? - Sources

- Wikipedia



WIKIPEDIA
The Free Encyclopedia

- WordNet

WordNet
A lexical database for English

- GeoNames

 **GeoNames**

YAGO

- 2006 idea: combine entities from Wikipedia with WordNet taxonomy
- 2007 YAGO: first version [SKW07]
- 2008 extract facts from infoboxes [SKW08]
- 2011 YAGO2: temporal and geographical meta facts [Hof+11a]; [Hof+13]
- 2013 YAGO2s: modular architecture [Suc+13]
- 2015 YAGO3: entities and facts from 10 languages [MBS15]



Fabian Suchanek



Johannes Hoffart



Joanna Biega



Gerhard Weikum



Erdal Kuzey



F. Mahdisoltani

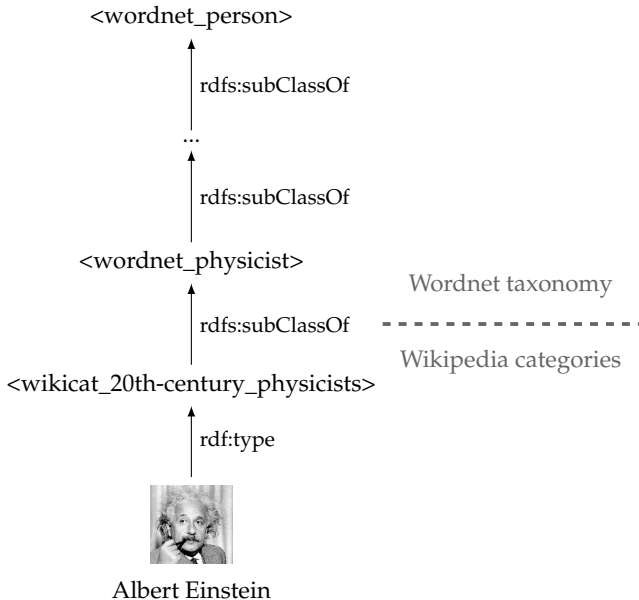


Gjergji Kasneci

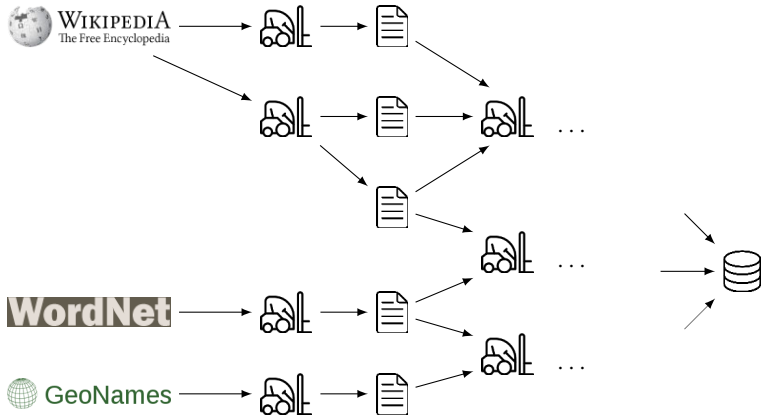


Thomas Rebele

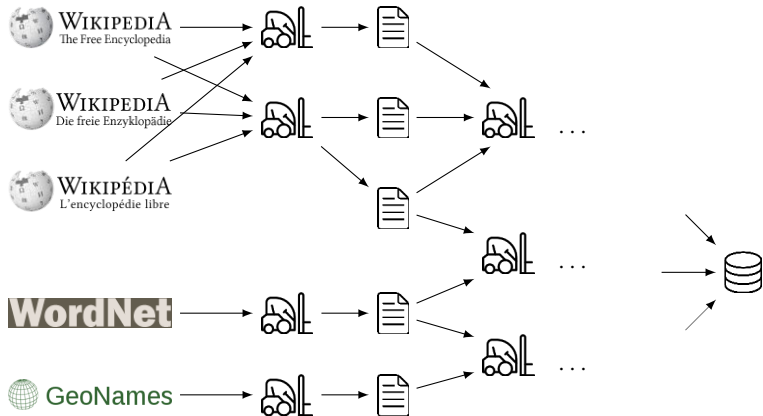
YAGO - Combining WordNet with Wikipedia



YAGO - Multilingual: facts and entities from 10 Wikipedias



YAGO - Multilingual: facts and entities from 10 Wikipedias



intermediate extractors

- ▶ clean facts
- ▶ deduplicate entities and facts
- ▶ check consistency

YAGO - Multilingual: facts and entities from 10 Wikipedias



↓
birth_place

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



↓
GEBURTSORT

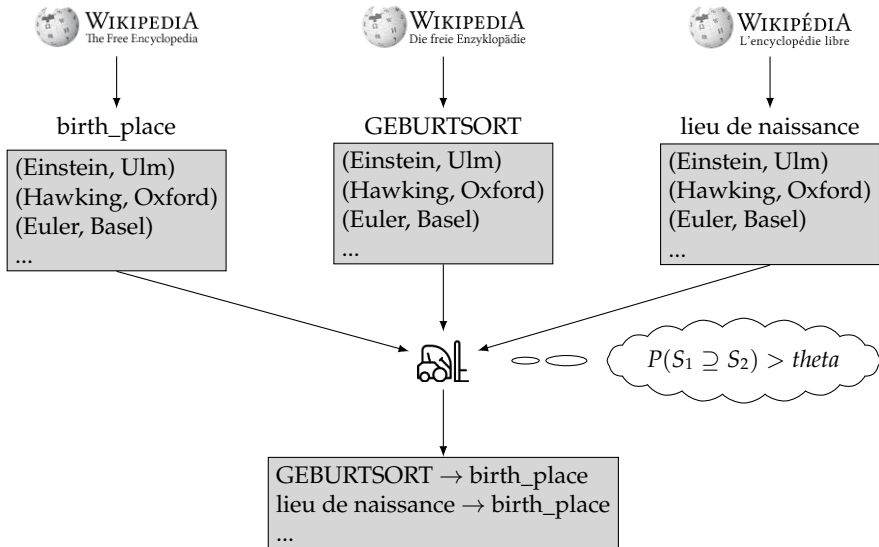
(Einstein, Ulm)
(Hawking, Oxford)
(Euler, Basel)
...



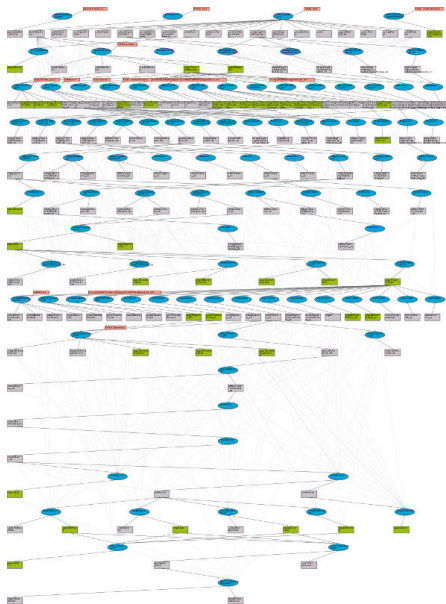
↓
lieu de naissance

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

YAGO - Multilingual: facts and entities from 10 Wikipedias



YAGO - Multilingual: facts and entities from 10 Wikipedias



YAGO architecture

YAGO - Time and space



Albert Einstein



Mileva Marić

married



Albert Einstein



Mileva Marić

married

<id_tiwmcu_16x_11ovtp>

YAGO - Time and space



Albert Einstein — married — Mileva Marić

<id_tiwmcu_16x_11ovtp>

source



The screenshot shows the top portion of the Wikipedia article for Albert Einstein. On the left is the Wikipedia logo and a sidebar with navigation links: Main page, Contents, Featured content, Current events, Random article, Donate to Wikipedia, and Wikipedia store. The main content area includes the article title "Albert Einstein", a search bar, and a sub-header "From Wikipedia, the free encyclopedia". Below this is the text "Albert Einstein (/ˈaɪnstʌɪn/; German: [ˈalbɛʁt ˈaɪnʃtaɪn] (listen); 14 March 1879 – 18 April 1955) was a German-born theoretical physicist." and a small portrait of Einstein.

<http://en.wikipedia.org/wiki/Albert_Einstein>

YAGO - Time and space



Albert Einstein — married — Mileva Marić

<id_tiwmcu_16x_11ovtp>

source

since

until

1903-##-##

1919-##-##



Article [Talk](#) [Read](#) [View source](#) [View history](#)

Albert Einstein

From Wikipedia, the free encyclopedia

Albert Einstein

/ˈaɪnʃtaɪn/^[4] German: [ˈalbɛçt ˈaɪnʃtaɪn] (ⓘ·ⓘ·ⓘ); 14 March 1879 – 18 April 1955) was a German-born theoretical physicist.

Albert Einstein



<http://en.wikipedia.org/wiki/Albert_Einstein>

YAGO 2 spotlx

Query

Id	Subject	Property	Object	Time	
?id0:	<input type="text" value="?x"/>	<input type="text" value="rdf:type"/>	<input type="text" value="scientist"/>	<input type="text" value=""/>	
?id1:	<input type="text" value="?x"/>	<input type="text" value="rdf:type"/>	<input type="text" value="politician"/>	<input type="text" value=""/>	
?id2:	<input type="text" value="?x"/>	<input type="text" value="<wasBornIn>"/>	<input type="text" value="?p"/>	<input type="text" value="before"/>	<input type="text" value="1900"/>
?id3:	<input type="text" value=""/>	<input type="text" value=""/>	<input type="text" value=""/>	<input type="text" value=""/>	
?id4:	<input type="text" value=""/>	<input type="text" value=""/>	<input type="text" value=""/>	<input type="text" value=""/>	

Results

Id	Subject	Property	Object
1	null	<Albert Gallatin>	rdf:type <wordnet scientist 1105606
	null	<Albert Gallatin>	rdf:type <wordnet politician 1104512

SPOTLX Browser ▷ w3id.org/yago/spotlx

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

Combining WordNet with Wikipedia

Multilingual: facts and entities from 10 Wikipedias

Time and space

Accuracy

Applications

DBpedia

IBM Watson

AIDA

Semantic Culturomics

AMIE

Partial completeness assumption

Rule mining in knowledge bases

Canonicalization of knowledge bases

Wikilinks semantification

Completeness

Conclusion

References



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

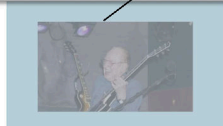


- ▶ participated in *Jeopardy* (TV quizz show)
- ▶ first place (competitors: human champions)
- ▶ used YAGO's type hierarchy

Applications - AIDA



Page played the hit Kashmir on his uniquely tuned Les Paul.

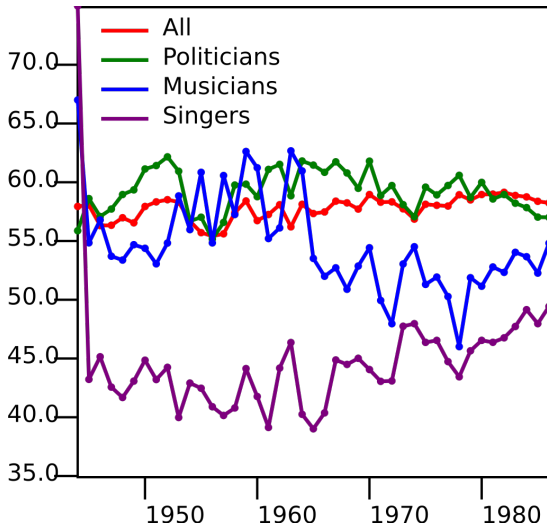


Named Entity Disambiguation with ADIA,
[Hof+11b]

▷ github.com/yago-naga/aida

Applications - Semantic Culturomics

in cooperation with Le Monde



Average age of people by occupation, [HBS13]

▷ w3id.org/yago/mpi/le-monde

What is a knowledge base?

YAGO

Combining WordNet with Wikipedia

Multilingual: facts and entities from 10 Wikipedias

Time and space

Accuracy

Applications

DBpedia

IBM Watson

AIDA

Semantic Culturomics

AMIE

Partial completeness assumption

Rule mining in knowledge bases

Canonicalization of knowledge bases

Wikilinks semantification

Completeness

Conclusion

References



Reference: [Gal+13] [Gal+15]



Luis Galárraga



Christina Teflioudi



Katja Hose



Fabian Suchanek

AMIE - Partial completeness assumption

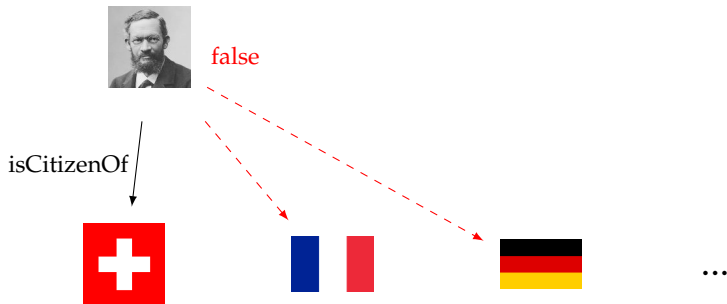


isCitizenOf

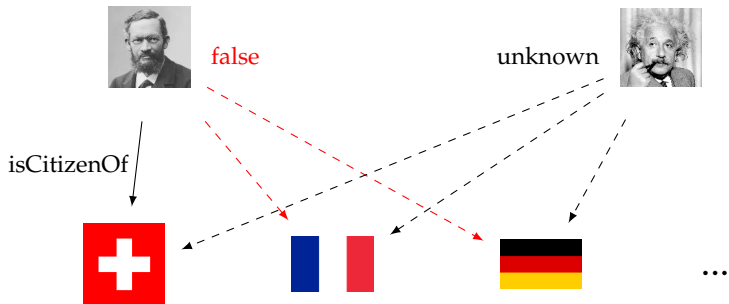


...

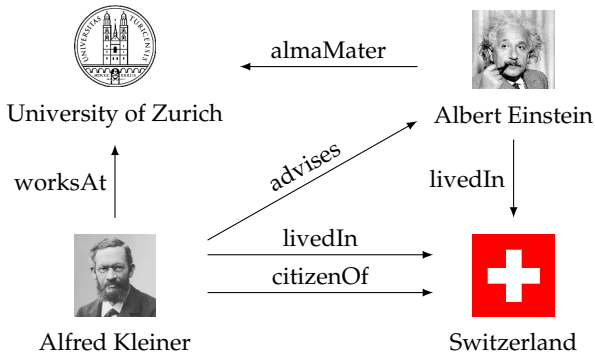
AMIE - Partial completeness assumption



AMIE - Partial completeness assumption



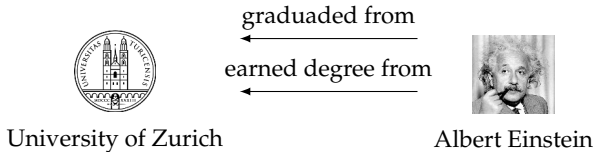
AMIE - Rule mining in knowledge bases



AMIE

$\text{almaMater}(z,y), \text{advises}(x,z) \implies \text{worksAt}(x,y)$
 $\text{livedIn}(x,y) \implies \text{isCitizenOf}(x,y)$

AMIE - Canonicalization of knowledge bases



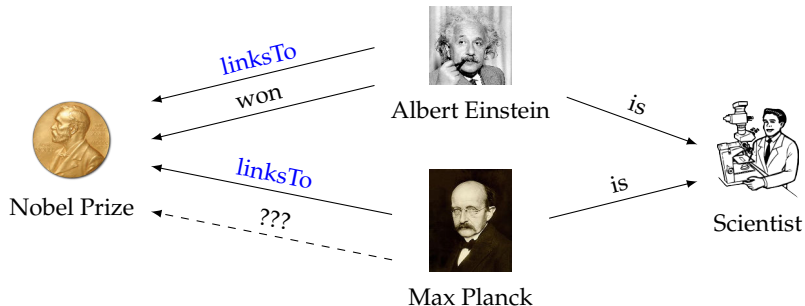
AMIE

graduated from \iff earned degree from

Reference: [Gal+14]

Diagram by Luis Galárraga
Image source: <https://wikipedia.de>

AMIE - Wikilinks semantification



AMIE

$\text{is}(x, \text{Scientist}), \text{linksTo}(x, \text{Nobel Prize}) \implies \text{won}(x, \text{Nobel Prize})$

Reference: [GSM15]

AMIE - Completeness



Albert Einstein

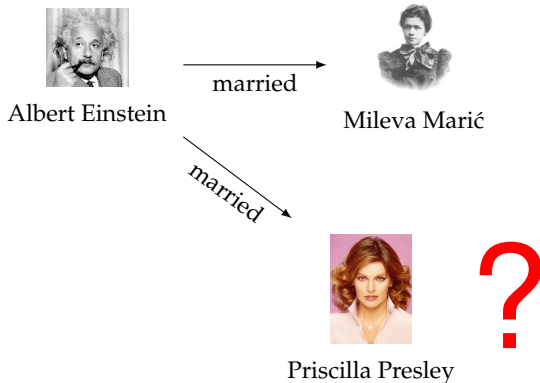
→
married



Mileva Marić

Image source: <https://wikipedia.de>

AMIE - Completeness



AMIE can mine completeness rules

AMIE - Completeness

- ▶ closed world assumption
- ▶ partial completeness assumption
- ▶ popularity oracle
- ▶ no-change oracle
- ▶ star-pattern oracle
- ▶ class oracle
- ▶ AMIE oracle

Reference: [Gal+17]



Luis Galárraga



Simon Razniewski



Antoine Amarilli



Fabian Suchanek

AMIE - Completeness

- ▶ closed world assumption
- ▶ partial completeness assumption
- ▶ popularity oracle
- ▶ no-change oracle
- ▶ star-pattern oracle
- ▶ class oracle
- ▶ AMIE oracle

$$\text{moreThan}_1(x, \text{hasParent}) \implies \text{complete}(x, \text{hasParent})$$

Reference: [Gal+17]



Luis Galárraga



Simon Razniewski



Antoine Amarilli



Fabian Suchanek

AMIE can predict incompleteness

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

Conclusion

What is a knowledge base?

YAGO

- Combining WordNet with Wikipedia

- Multilingual: facts and entities from 10 Wikipedias

- Time and space

- Accuracy

Applications

- DBpedia

- IBM Watson

- AIDA

- Semantic Culturomics

AMIE

- Partial completeness assumption

- Rule mining in knowledge bases

- Canonicalization of knowledge bases

- Wikilinks semantification

- Completeness

Conclusion

References



- ▶ knowledge base with 10 million entities and >210 million facts
- ▶ combines Wikipedia with Wordnet
- ▶ multilingual facts from 10 languages
- ▶ has facts about time and space
- ▶ contains >95% correct facts
- ▶ available at
<http://yago-knowledge.org>



- ▶ knowledge base with 10 million entities and >210 million facts
- ▶ combines Wikipedia with Wordnet
- ▶ multilingual facts from 10 languages
- ▶ has facts about time and space
- ▶ contains >95% correct facts
- ▶ available at
<http://yago-knowledge.org>

Announce: in the following weeks

- ▶ new data release
- ▶ YAGO goes Open Source

References

What is a knowledge base?

YAGO

- Combining WordNet with Wikipedia

- Multilingual: facts and entities from 10 Wikipedias

- Time and space

- Accuracy

Applications

- DBpedia

- IBM Watson

- AIDA

- Semantic Culturomics

AMIE

- Partial completeness assumption

- Rule mining in knowledge bases

- Canonicalization of knowledge bases

- Wikilinks semantification

- Completeness

Conclusion

References

References



Joanna Biega, Erdal Kuzey, and Fabian M. Suchanek. Inside YAGO2s: A transparent information extraction architecture. In: *Proceedings of the 22nd international conference on World Wide Web companion*. 00016. International World Wide Web Conferences Steering Committee, 2013, pp. 325–328



Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In: *WWW*. 2013



Luis Galárraga, Jeremy Heitz, Kevin Murphy, and Fabian M. Suchanek. Canonicalizing Open Knowledge Bases. In: 00008. ACM Press, 2014, pp. 1679–1688. ISBN: 978-1-4503-2598-1. DOI: 10.1145/2661829.2662073



Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. Fast Rule Mining in Ontological Knowledge Bases with AMIE+. In: *VLDBJ*. 00002
<https://suchanek.name/work/publications/vldb2015.pdf>. 2015

References



Luis Galárraga, Simon Razniewski, Antoine Amarilli, and Fabian M. Suchanek. Predicting Completeness in Knowledge Bases. In: 00000. ACM Press, 2017, pp. 375–383. ISBN: 978-1-4503-4675-7. DOI: 10.1145/3018661.3018739



Luis Galarraga, Danai Symeonidou, and Jean-Claude Moissinac. Rule Mining for Semantifying Wikilinks. In: *Linked Open Data Workshop at WWW*. 2015



Thomas Huet, Joanna Biega, and Fabian M. Suchanek. Mining history with Le Monde. In: *Proceedings of the 2013 workshop on Automated knowledge base construction*. 00010. ACM Press, 2013, pp. 49–54. ISBN: 978-1-4503-2411-3. DOI: 10.1145/2509558.2509567



Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia. Research Report. Max-Planck-Institut für Informatik, 2010

References



Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, Edwin Lewis-Kelham, Gerard De Melo, and Gerhard Weikum. YAGO2: exploring and querying world knowledge in time, space, context, and many languages. In: *Proceedings of the 20th international conference companion on World wide web*. 00187. ACM, 2011, pp. 229–232



Johannes Hoffart et al. Robust disambiguation of named entities in text. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 00288. Association for Computational Linguistics, 2011, pp. 782–792







Johannes Hoffart, Fabian M Suchanek, Klaus Berberich, and Gerhard Weikum. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. In: *Artificial Intelligence* 194 (2013). 00374, pp. 28–61



Farzaneh Mahdisoltani, Joanna Biega, and Fabian Suchanek. YAGO3: A knowledge base from multilingual Wikipedias. In: *7th Biennial Conference on Innovative Data Systems Research*. 00022. CIDR 2015, 2015

References

-  Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. YAGO: a multilingual knowledge base from Wikipedia, Wordnet, and Geonames. In: *ISWC 2016*. 2016
-  Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In: *Proceedings of the 16th international conference on World Wide Web*. 01699. ACM, 2007, pp. 697–706
-  Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: A large ontology from wikipedia and wordnet. In: *Web Semantics: Science, Services and Agents on the World Wide Web 6.3* (2008). 00575, pp. 203–217
-  Fabian M. Suchanek, Johannes Hoffart, Erdal Kuzey, and Edwin Lewis-Kelham. YAGO2s: Modular High-Quality Information Extraction with an Application to Flight Planning.. In: *BTW*. Vol. 214. 00012. 2013, pp. 515–518

Acknowledgements:

Icons designed by Madebyoliver from Flaticon