

# *YAGO – A Core of Semantic Knowledge*

## *Yet Another Great Ontology*

Fabian M. Suchanek, Gjergji Kasneci, Gerhard Weikum

(Max-Planck Institute for Computer Science Saarbrücken/Germany)

Presented by

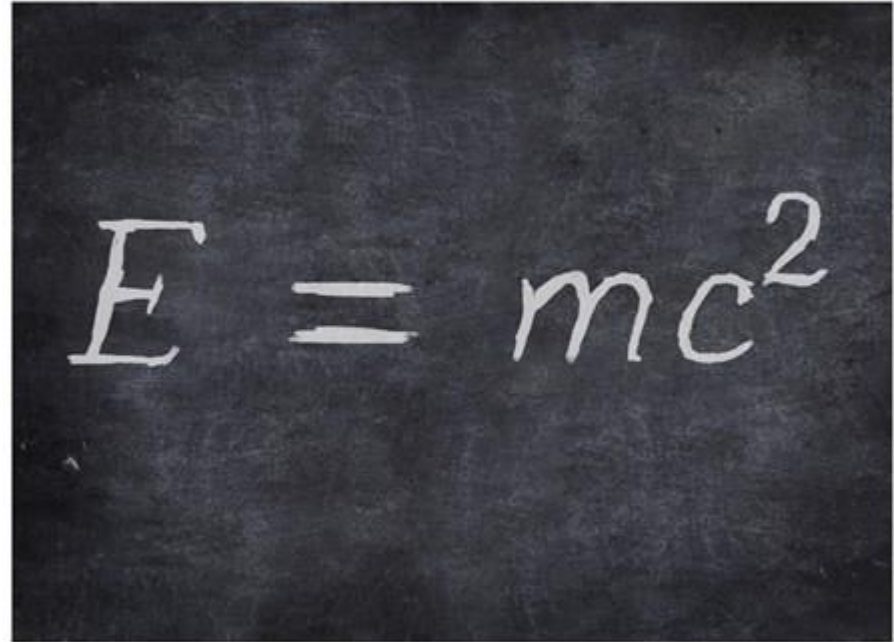
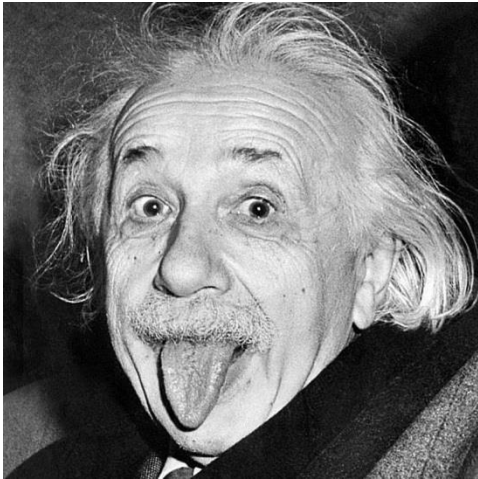
Tandra Chakraborty

Std# 20546668

# Overview

- **Motivation for Ontology**
- **The YAGO Model**
  - ✓ **Structure**
  - ✓ **Semantics**
  - ✓ **Source**
- **Conclusion**

# Motivation for Ontology



What else come on your mind?

- Nobel Prize
- Germany

About 35,900,000 results (0.25 seconds)

### Albert Einstein - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Albert\_Einstein  
Albert Einstein was a German-born theoretical physicist and philosopher of science. He developed the general theory of relativity, one of the two pillars of ...  
Hans Albert Einstein - Eduard Einstein - Elsa Einstein - Einstein family

### Albert Einstein - Biography.com

www.biography.com/people/albert-einstein-9285408  
Biography.com offers a glimpse into the life of Albert Einstein, the most influential physicist of the 20th century, who developed the theory of relativity.

### Albert Einstein - Biographical - Nobelprize.org

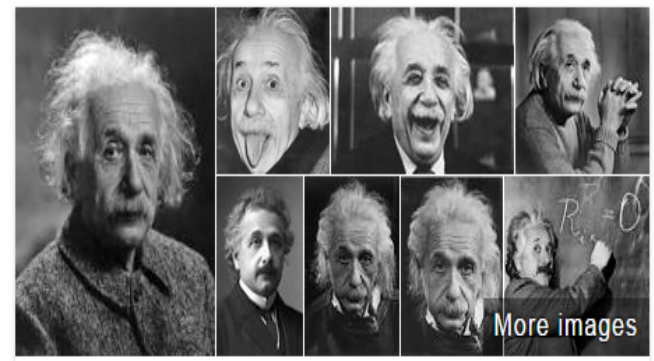
www.nobelprize.org/nobel\_prizes/physics/laureates/.../einstein-bio.html  
Albert Einstein - Biographical. Questions and Answers on Albert Einstein. Albert Einstein was born at Ulm, in Württemberg, Germany, on March 14, 1879.

### Albert Einstein Quotes - BrainyQuote

www.brainyquote.com/quotes/authors/a/albert\_einstein.html  
Enjoy the best Albert Einstein Quotes at BrainyQuote. Quotations by Albert Einstein, German Physicist, Born March 14, 1879. Share with your friends.

### Albert Einstein | Albert Einstein Official Site

einstein.biz  
Albert Einstein official Web Site and Fan Club, featuring biography, photos, trivia, rights representation, licensing, contact and more.



## Albert Einstein

Theoretical Physicist

Albert Einstein was a German-born theoretical physicist and philosopher of science. He developed the general theory of relativity, one of the two pillars of modern physics. Wikipedia

- Born: March 14, 1879, Ulm, Germany
- Died: April 18, 1955, Princeton, New Jersey, United States
- Children: Eduard Einstein, Hans Albert Einstein, Lieserl Einstein
- Spouse: Elsa Einstein (m. 1919–1936), Mileva Marić (m. 1903–1919)
- Education: University of Zurich (1905), more
- Awards: Nobel Prize in Physics, more

Wednesday, October 15, 2014  
YAGO - A Core of Semantic Knowledge CS 743@Tandra



when albert einstein became famous

Web Images News Videos More Search tools

About 1,170,000 results (0.39 seconds)

Albert Einstein became famous for the theory of relativity, which laid the basis for the release of atomic energy. In **1905** Albert Einstein formulates Special Theory of Relativity.

**Albert Einstein - Tesla Memorial Society of New York**  
[www.teslasociety.com/einstein.htm](http://www.teslasociety.com/einstein.htm)

Feedback

### Albert Einstein - Tesla Memorial Society of New York

[www.teslasociety.com/einstein.htm](http://www.teslasociety.com/einstein.htm)

Each of them became the basis of a new branch of physics. Einstein treated matter and energy as exchangeable. **Albert Einstein** became famous for the theory ...

### Kids - Albert Einstein Biography

[www.einstein-website.de/z\\_kids/biographykids.html](http://www.einstein-website.de/z_kids/biographykids.html)

In 1909 he became professor for theoretical physics at the University of Zurich. ...  
When the 12-year-old Eduard asked his father why he was so famous he got ...

### Albert Einstein (1879-1955) - Planet Science

[www.planet-science.com](http://www.planet-science.com) > Over 11s > Science celebrities

Instead, Einstein went to work in a patent office. Luckily ... This is where Einstein developed his most famous ideas. .... Einstein became a clerk in a patent office.



albert einstein nobel prize



**Web** Images Videos News Maps More ▾ Search tools

About 689,000 results (0.30 seconds)

### The Nobel Prize in Physics 1921 - Nobelprize.org

[www.nobelprize.org/nobel\\_prizes/physics/laureates/1921/](http://www.nobelprize.org/nobel_prizes/physics/laureates/1921/) ▾

Albert Einstein received his Nobel Prize one year later, in 1922. During the selection process in 1921, the Nobel Committee for Physics decided that none of the ...

### Albert Einstein - Facts - Nobelprize.org

[www.nobelprize.org/nobel\\_prizes/physics/laureates/.../einstein-facts.html](http://www.nobelprize.org/nobel_prizes/physics/laureates/.../einstein-facts.html) ▾

... and especially for his discovery of the law of the photoelectric effect". Field: theoretical physics. Albert Einstein received his Nobel Prize one year later, in 1922.

### Albert Einstein - Biographical - Nobelprize.org

[www.nobelprize.org/nobel\\_prizes/physics/laureates/.../einstein-bio.html](http://www.nobelprize.org/nobel_prizes/physics/laureates/.../einstein-bio.html) ▾

Questions and Answers on Albert Einstein. Albert Einstein was born at Ulm, in Württemberg, Germany, on March 14, 1879. Six weeks later the family moved to ...

### News for albert einstein nobel prize

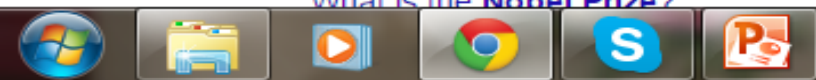


#### The Tragic Nobel Peace Prize Story You've Probably Never ...

TIME - 2 days ago

Albert Einstein and French author Romain Rolland were among the period's ... For nearly a year the Nobel Peace Prize Committee has been ...

What is the Nobel Prize?





Give me name of Nobel Prize winners when Albert Einstein win Nobel Prize

Web News Images Videos More Search tools

About 2,050,000 results (0.69 seconds)

News for Give me name of Nobel Prize winners ...



Light bulb moment: Low-energy LED wins Nobel prize

Reuters - 5 days ago

The award is a notable example of a practical discovery winning the prize, ... join some of the biggest names in science such as Albert Einstein, ...

A bright idea indeed: Inventors of low-energy LED light win ...

GMA News - 5 days ago

More news for Give me name of Nobel Prize winners when Albert Einstein win Nobel Prize

The Nobel Prize in Physics 1921 - Nobelprize.org

www.nobelprize.org/nobel\_prizes/physics/laureates/1921/

Nobel Prizes and Laureates. Physics Prizes, Chemistry Prizes ... Albert Einstein received his Nobel Prize one year later, in 1922. During the selection process in ...

All Nobel Prizes in Physics - Nobelprize.org

www.nobelprize.org/nobel\_prizes/physics/laureates/

Collage: Erwin Schrödinger, Marie Curie, Albert Einstein, Nils Bohr and Henri Becquerel ... John Bardeen is the only Nobel Laureate who has been awarded the Nobel Prize in Physics ... Click on the links to get more information. .... of gamma radiation and his





Give me the name of scientists who won Nobel prize at the same time as Alb

Web Images News Videos More Search tools

About 6,550,000 results (0.84 seconds)

### Albert Einstein - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Albert\_Einstein

He received the 1921 Nobel Prize in Physics "for his services to theoretical physics, and ... Einstein published more than 300 scientific papers along with over 150 .... It was during his time in Italy that he wrote a short essay with the title "On the ... also enrolled at the Polytechnic that same year, the only woman among the six ...

### Images for Give me the name of scientists who won ...

Report images



More images for Give me the name of scientists who won Nobel prize at the same time as Albert Einstein

### Albert Einstein - Tesla Memorial Society of New York

www.teslasociety.com/einstein.htm

He was one of the greatest scientists of all time. In 1905 ... Einstein received the

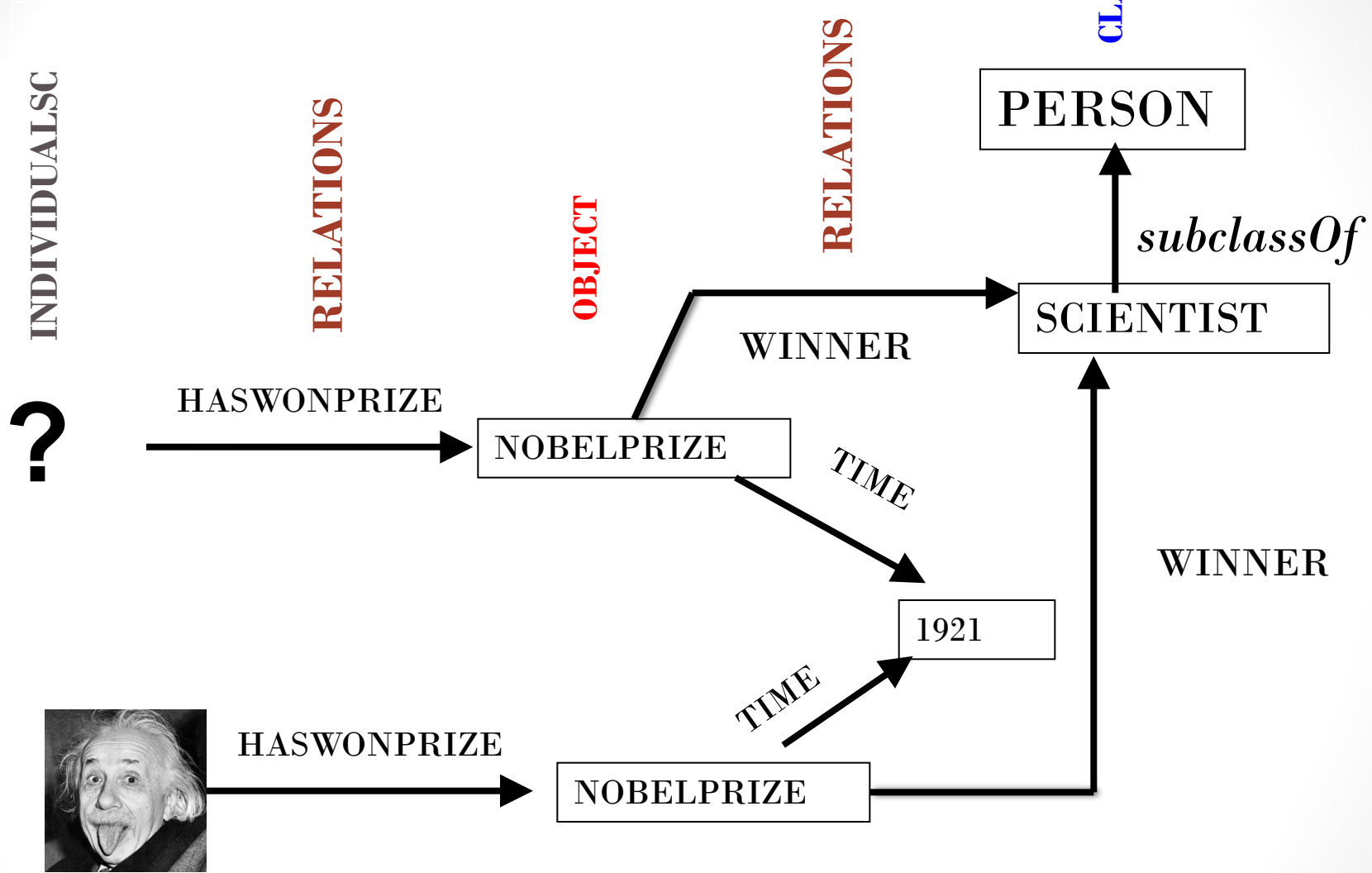
Google is searching for Webpage not Knowledge

Wednesday, October 15, 2014

YAGO - A Core of Semantic Knowledge CS 743@Tandra



# Solution: An ontology



An **ontology** is a formal framework for representing knowledge [Wikipedia]

# Application of Ontology

- Machine Translation
- Word Sense Disambiguation
- Document Classification
- Question Answering
- Entity and fact-oriented Web Search
- Data cleaning
- Record Linkage

# Idea behind YAGO

Manual categorization used for extraction :

- WordNet, SUMO, GeneOntology
- Problem: Usually low coverage

Extracting from Web:

- KnowItAll, Espresso, Snowball, LEILA
- Problem: Usually low accuracy (many false positive result)

# What's new in YAGO?

YAGO approach:

- Combine Wikipedia and Wordnet
- Arranged concepts in a taxonomy (Wordnet) ( $\Rightarrow$  good coverage)
- Use the category system of Wikipedia ( $\Rightarrow$  good accuracy)
- Extension of RDF(Resource Description Framework )
- It supports built in transitive relation (not in RDF)
- Simple and Decidable

# Structure

- Represents knowledge as
  - Entities
  - Relations
  - Facts

Example:

**AlbertEinstein** *HASWONPRIZE* **NobelPrize**

**AlbertEinstein** *BORNINYEAR* **1879**

# Structure

- All objects are entity(cities, people, URL)
- Number, dates, strings are entity
- Words are also entity
- Similar entities are grouped in classes

*AlbertEinstein* **TYPE** *physicist*

- Classes are also entities.

*physicist* **SUBCLASSOF** *scientist*

- Relations are entity as well

**subclassOf** **TYPE** *transitiveRelation*

The triple of entity, a relation and an entity is called a fact

# Structure

- Entities are *argument* of fact
- Each fact is given a fact identifier

#1 (AlbertEinstein HASWONPRIZE NobelPrize)

#2 (#1 TIME 1921)

#3 (#1 FPUNDIN <http://www.wikipedia.org/Einstein> )

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

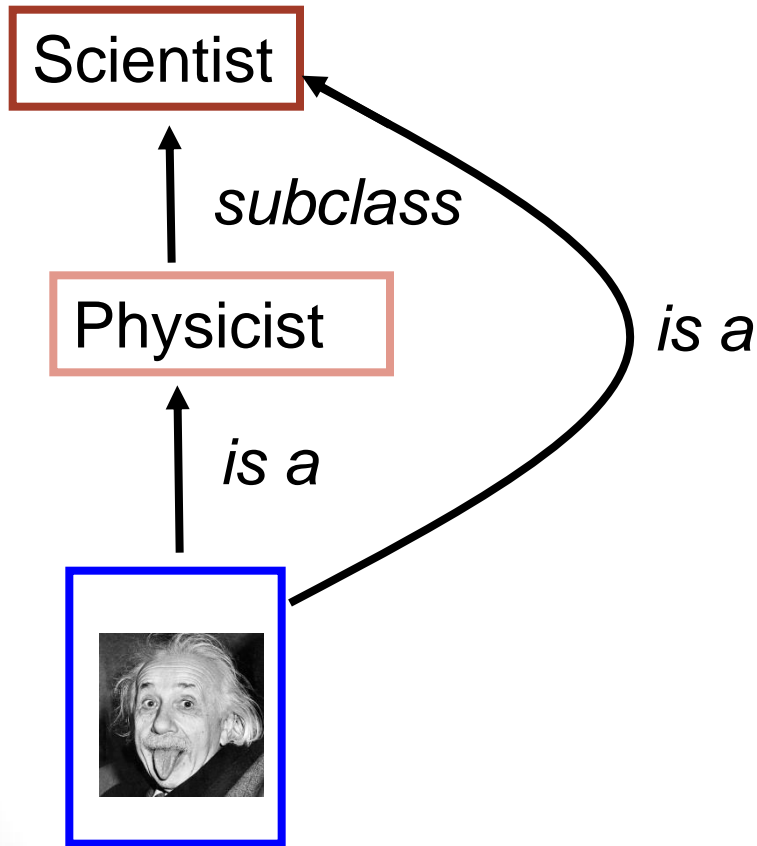
# Structure

*A structure for a YAGO ontology  $y$  is a triple of*

- *a set  $U$ (the universe)*
- *a function  $D: I \cup C \cup R \rightarrow U$ (the denotation)*
- *a function  $\varepsilon: D(R) \rightarrow U \times U$ (the extension function)*



# Transitivity



Axioms:

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

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

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

...

# Enriching YAGO

Want to add  $(x,r,y)$  to existing fact

- Map  $x$  and  $y$  to existing entities in YAGO ontology
- Add as new entity if they don't exist
- Next  $r$  has to be mapped to a relation in the YAGO ontology
- If  $(x,r,y)$  exist in ontology then add a new witness for the fact  $\longrightarrow (f, \text{FOUNDIN}, w)$
- Calculate the confidence of fact
- If  $(x,r,y)$  does not exist add this fact with a new fact identifier

# Knowledgebase For YAGO

- Wordnet
  - Semantic Lexicon for the English Language
- Wikipedia
  - Multilingual, web-based encyclopedia
- “Categories” from Wikipedia
- “Class hierarchy” from Wordnet

Example:

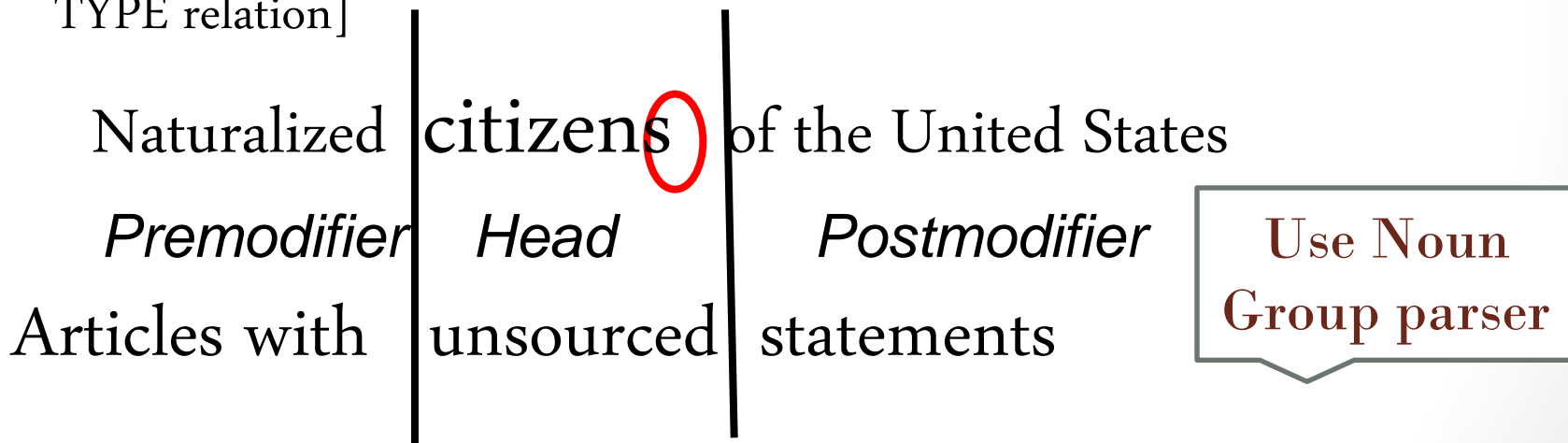
According to Wiki Zidane is in the super-category named “*Football in France*”, but *Zidane is a football player and not a football.*

Here WordNet, in contrast, provides a clean and carefully assembled hierarchy of thousands of concepts.

# Knowledge Extraction

Albert Einstein is in the category Naturalized citizens of the United States  
Albert Einstein is also in the category Articles with unsourced statements,  
Relational information (like 1879 births) and  
Thematic vicinity (Physicist).

How to identify Conceptual category? [We need it for defining  
TYPE relation]



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

# Knowledge Extraction

- The *subclassOf* is reflected by thematic category of Wiki
- Only with Wiki we don't get total accuracy
- So for taxonomy purpose leaf categories are taken from Wikipedia
- Wordnet is used on those leaf categories to establish hierarchy of classes
- Each synset(A set of words share one sense) becomes a class for YAGO.
- Extract *MEANS* relation from redirect pages of Wiki

**Einstein,Albert** *MEANS* **Albert Einstein**

# Knowledge Extraction

Other relations from Category:

- **BornInYear**
- **DiedInYear**
- **EstablishedIn**
- **LocatedIn**
- **WrittenInYear**
- **PoliticianOf**
- **HasWonPrize**

# Knowledge Extraction

## Meta Relations

- Describes (individual and URL of the corresponding Wikipedia page)
- Witness (The page from where knowledge extracted)
- ExtractedBy (Technique of Extraction)
- FoundIn (Relation between Fact and URL)
- Context(Albert Einstein, Relativity Theory)

# The YAGO ontology: Accuracy

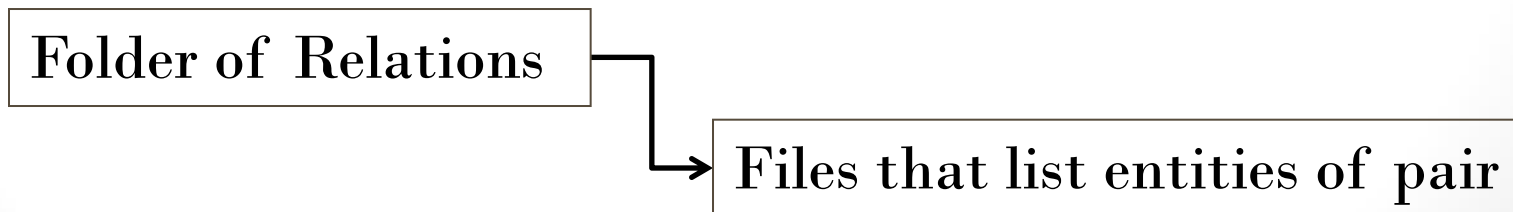
Relation	Accuracy
<b>subclass</b>	97.70% +/- 1.59%
<b>is a</b>	94.54% +/- 2.36%
<b>familyName</b>	97.81% +/- 1.75%
<b>givenName</b>	97.62% +/- 2.08%
<b>establishedIn</b>	90.84% +/- 4.28%
<b>bornInYear</b>	93.14% +/- 3.71%
<b>diedInYear</b>	98.70% +/- 1.30%
<b>locatedIn</b>	98.41% +/- 1.52%
<b>politicianOf</b>	92.43% +/- 3.93%
<b>writtenInYear</b>	94.35% +/- 3.33%
<b>hasWonPrize</b>	98.47% +/- 1.53%

Ref: <http://suchanek.name/work/publications/www2007.ppt>



# YAGO Storage

- When the paper was written YAGO has 1 million entities with 5 million facts
- In the next iteration YAGO-2 has 2 million of entities with 20 million of facts
- Simple text files as an internal format
- Store only facts that are unique, not derivable from other facts [Canoncialization]



# Size of YAGO

Table 2: Size of YAGO (facts)

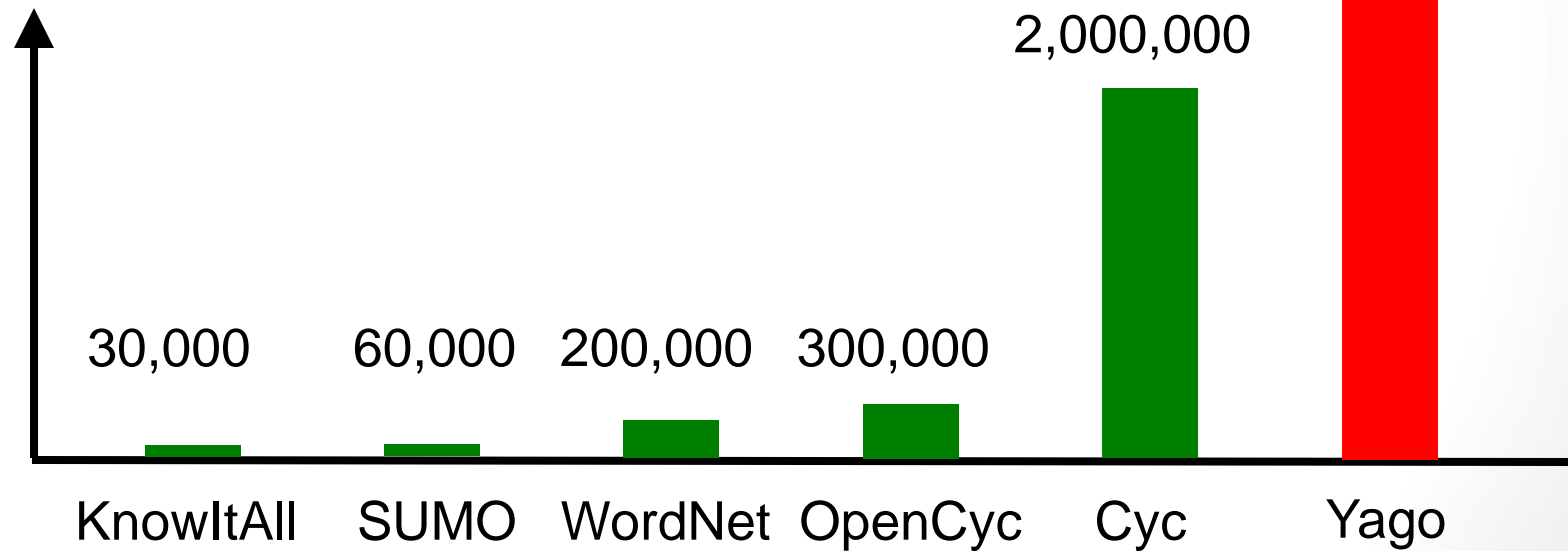
Relation	Domain	Range	# Facts
SUBCLASSOF	class	class	143,210
TYPE	entity	class	1,901,130
CONTEXT	entity	entity	~40,000,000
DESCRIBES	word	entity	986,628
BORNINYEAR	person	year	188,128
DIEDINYEAR	person	year	92,607
ESTABLISHEDIN	entity	year	13,619
LOCATEDIN	object	region	59,716
WRITTENINYEAR	book	year	9,670
POLITICIANOF	organization	person	3,599
HASWONPRIZE	person	prize	1,016
MEANS	word	entity	1,598,684
FAMILYNAMEOF	word	person	223,194
GIVENNAMEOF	word	person	217,132

Table 3: Size of YAGO (entities)

Relations	14
Classes	149,162
Individuals (without words)	907,462

# The YAGO ontology: Number of Facts

Ontologies should not be judged purely by the number of facts! This is just an informational overview.



Ref: <http://suchanek.name/work/publications/www2007.ppt>

# Compatibility

- YAGO is available as a simple XML version of the text files
- YAGO is loadable from Oracle, Postgress or MYSQL
- YAGO can be converted to a database table.
- The table has the simple schema FACTS(factId, arg1,relation,arg2, confidence).

# Sample Queries

Query	Result
When was "Mostly Harmless" written? (Mostly Harmless, writtenInYear, \$y)	\$y=1992
Which humanists were born in 1879? (\$h, type subclassOf*, humanist) (\$h, bornInYear, 1879)	\$h=Albert Einstein and 2 more
Which locations in Texas and Illinois bear the same name? (\$n, means, \$t) (\$n, means, \$k) (\$k, locatedIn, Illinois)	\$n="Farmersville" and 121 more

# Conclusion

- Briefly explained the structure, knowledge extraction technique of YAGO
- “YAGO, a large and extendable ontology of high quality” – this statement by author is already proved in the iteration YAGO-2 and YAGO-2.5
- It has 95% human accuracy
- It is importable in existing database system

Explore YOGA, utilize its application in the vision of the semantic Web

<http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/demo/>

# Thank You, Any Questions?

