

The NoRDF Project

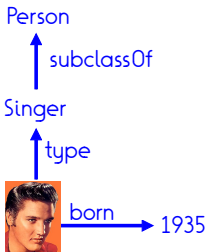
Fabian Suchanek

Amazing! This talk is free
of the Corona virus!

(about the speaker, we don't know...)



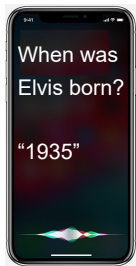
Knowledge Bases



For us, a knowledge base (KB) is a graph, where the nodes are entities and the edges are relations.

(We do not distinguish T-Box and R-Box.)

Cool knowledge-based applications



Apple Siri

How long was the
Thirty Years' War?



Amazon Echo

Discovered 6 kinesin
proteins that relate
to cancer



IBM Watson

These applications feed from **knowledge bases**.

There are plenty of knowledge bases



NELL



TextRunner



BabelNet



Plus industrial projects at

Google



Microsoft

ebay

facebook



Sponsored Message: YAGO



NELL



TextRunner

We develop YAGO, one of the largest open general purpose KBs.

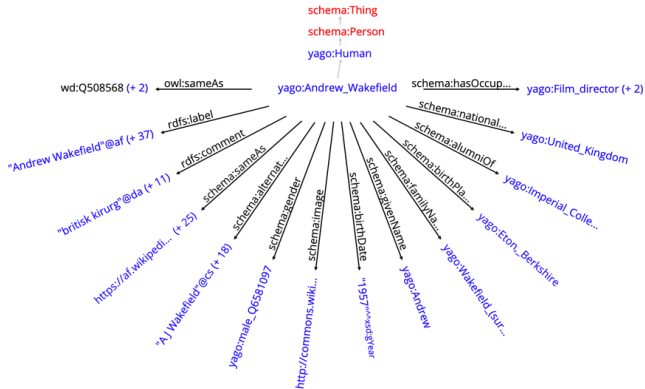
The newest version, YAGO4,

- combines Wikidata and schema.org
- contains 50 million entities and 2 billion facts
- is so clean that it allows for automated reasoning

<https://yago-knowledge.org>

What's in a knowledge base?

Essentially binary facts ("triples") in the knowledge format "RDF":



What's in the real world?

In February 1998, Andrew Wakefield published a paper in the medical journal *The Lancet*, which reported on twelve children with developmental disorders. The parents were said to have linked the start of behavioral symptoms to vaccination. The resulting controversy became the biggest science story of 2002. As a result, vaccination rates dropped sharply. In 2011, the *BMJ* detailed how Wakefield had faked some of the data behind the 1998 *Lancet* article.

Beliefs

Events

Stories

Claims

Reasons

Falsifications

...none of which is in a knowledge base!

The NoRDF Project: Go Beyond Triples

If we want tomorrow's intelligent applications to be really intelligent, we have to extend their knowledge bases by

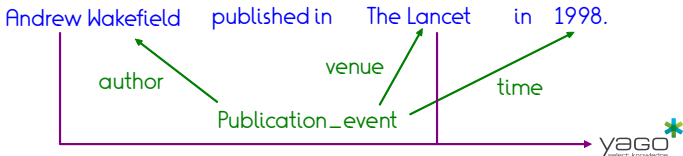
Beliefs Events Stories
Claims Reasons Falsifications

- 1) We have to be able to extract complex knowledge from text
(a process called "Information Extraction", "IE")
- 2) We have to be able to represent such knowledge and to reason on it

IE: What is possible already

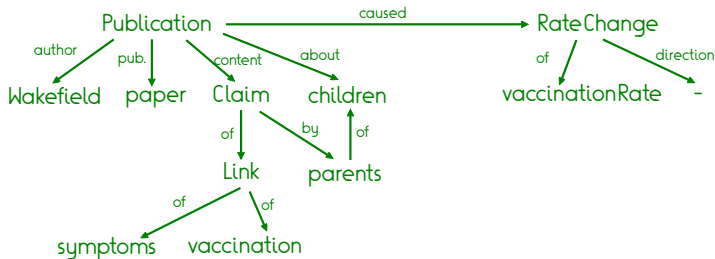
Several cool approaches can extract non-binary information:

- FRED
- K-Parser
- Document spanners
- ClausIE
- StuffIE
- OpenIE
- HighLife
- Advanced Meaning Representation (AMR)



IE: What we need

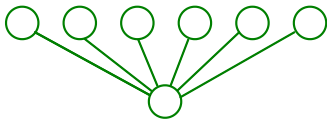
"Wakefield published a paper that reported on children. Their parents were said to have linked the start of behavioral symptoms to vaccination. The resulting controversy caused vaccination rates to fall. ..."



Cross-sentence analysis, advanced co-reference resolution, standardized types of frames, relationships between events, negation, hypothetical stances, storylines, ...

IE: Why Deep Learning is not enough

"Wakefield published a paper that reported on children. Their parents were said to have linked the start of behavioral symptoms to vaccination. The resulting controversy caused vaccination rates to fall. ..."



Did Wakefile publish a paper? ✓

Who published a paper? ✓

Were vaccination rates higher before the publication? ?

What caused the controversy? ✗

Does vaccination cause autism? ✗

What nationality is the person who caused the vaccine controversy? ✗

Reasoning: What we have

As knowledge representation:

- Frames, JSON
- complex objects
- object-relational databases
- Fact identifiers
- RDF*
- Reification

For reasoning:

- RDFS, OWL DL, SHACL
- Description Logic
- Context logics
- Modal logics
- Epistemic logics
- Formal argumentation
- Belief revision
- Provenance and annotated logics

Cannot represent

- "All clients believe that the company delivers a good service"
 - "the loss of value on the stock market happened because the public learned of a fraudulent activity by the company"
 - "Mary believes everything Paul says, Paul says $X \Rightarrow$ Mary believes X "
- ... or if they can, they are undecidable

Reasoning: What we need

1) a very simple logic inside a context

First-order logic without \exists ?

OWL EL?

Datalog?

$\forall x: \text{scientist}(x) \Rightarrow \text{person}(x)$ (?)

2) a very simple logic about contexts

Horn Rules?

Datalog?

$\forall \phi: \text{reads}(\text{Mary}, \phi)$

$\Rightarrow \text{believes}(\text{Mary}, \phi)$ (?)

\Rightarrow a moderately simple logic
in combination

You have a great idea? Let me know!

Vagueness, fuzziness, and probability: orthogonal topics

Applications

- Analysis of fake news / fact checking:
understand an article about a controversial topic, allow reasoning
(who said what when and why, what is the evidence, ...)
- Analysis of the e-reputation of a company:
extract controversy or beliefs with reasons and supporters,
for companies or their products
- Modeling of controversies:
detect a controversial topic on the Web (in blogs, forums, Twitter),
extract opinions, and model different views

Understanding the arguments of the other side
is a prerequisite for refuting them.

Our project "NoRDF"

Our project "NoRDF" aims to extract and model complex information from natural language text. The project runs for 4 years, supported by:



MINISTÈRE DES ARMÉES

Your company name here?

Our project "NoRDF": Who's there?



Fabian Suchanek

Professor at Télécom Paris, DIG team
Knowledge Bases, Reasoning, NLP



Chloé Clavel

Professor at Télécom Paris, S²A team
Affective Computing, Sentiment Analysis

We hired

- Pierre-Henri Paris (CNAM) as a postdoc
- Chadi Helwe (American Univ. of Beirut) as PhD student
- Cyril Chhun (Polytechnique) as a PhD student
- Julie Dessaint (Polytechnique) as master's student

Summary: The NoRDF Project

Input:

“Wakefield published a paper that reported on children. [...]
The resulting controversy caused vaccination rates to fall. ...”

Output:



And we're always open to

- new sponsors
- new students and postdocs

Join our team! <https://nordf.telecom-paris.fr/>

Backup Slides

Reasoning: What we have

As knowledge representation:

- Frames, JSON
- complex objects
- object-relational databases



Reasoning: What we have

As knowledge representation:

- Frames, JSON
 - complex objects
 - object-relational databases
- great, but do not allow for reasoning



- "If X caused Y and Y caused Z, then X caused Z" ✗
- "If X did not publish a paper, X is not a scientist" ✗
- "If Mary believes what Paul says & Paul says X, then Mary believes X" ✗

Reasoning: What we have

For reasoning:

- RDFS, OWL DL, SHACL
- Description Logic



Reasoning: What we have

For reasoning:

- RDFS, OWL DL, SHACL
 - Description Logic
- great, but do not allow for statements about statements



- "The paper says that vaccines cause autism" ✗
- "Fact A caused Fact B" ✗

Reasoning: What we have

Annotated Knowledge Representations:

- Fact identifiers
- RDF*
- Reification



Reasoning: What we have

Annotated Knowledge Representations:

- Fact identifiers
 - RDF*
 - Reification
- cannot deal with hypothetical statements
- cannot do reasoning



- "Mary believes that vaccines cause autism" ✗

Reasoning: What we have

Big logic machinery:

- Context logics
- Modal logics
- Epistemic logics

Reasoning: What we have

Big logic machinery:

- Context logics
- Modal logics
- Epistemic logics

cannot quantify over contexts
(or if they can, they are propositional logics or undecidable)

- "All clients believe that the company delivers a good service" ✗
- "the loss of value on the stock market happened because the public learned of a fraudulent activity by the company" ✗

Formal argumentation has monolithic propositions.

Belief revision has monolithic agents.

Provenance and annotated logics cannot make claims about annotations.

Vagueness, fuzziness, and probability are orthogonal topics.