A Hitchhiker’s Guide to Ontology

Fabian M. Suchanek
Télécom ParisTech University,
Paris, France
Fabian M. Suchanek

2003

2005

2008

money  freedom  permanent
Fabian M. Suchanek
2003: BSc in Cognitive Science
   University of Osnabrück/DE
2005: MSc in Computer Science
   Saarland University/DE
2008: PhD in Computer Science
   Max Planck Institute/DE
Fabian M. Suchanek

2003: BSc in Cognitive Science
University of Osnabrück/DE

2005: MSc in Computer Science
Saarland University/DE

2008: PhD in Computer Science
Max Planck Institute/DE
Fabian M. Suchanek

2009: PostDoc at Microsoft Research
Silicon Valley/US
Fabian M. Suchanek

2009: PostDoc at Microsoft Research
Silicon Valley/US
Fabian M. Suchanek

2009: PostDoc at Microsoft Research
    Silicon Valley/US

2010: PostDoc
    INRIA Saclay/FR
Fabian M. Suchanek

2009: PostDoc at Microsoft Research
Silicon Valley/US

2010: PostDoc
INRIA Saclay/FR

money
freedom
permanent
Fabian M. Suchanek

2009: PostDoc at Microsoft Research
    Silicon Valley/US

2010: PostDoc
    INRIA Saclay/FR

2012: Research group leader
    Max Planck Institute/DE
Fabian M. Suchanek

2009: PostDoc at Microsoft Research
    Silicon Valley/US
2010: PostDoc
    INRIA Saclay/FR
2012: Research group leader
    Max Planck Institute/DE
Fabian M. Suchanek

2009: PostDoc at Microsoft Research
    Silicon Valley/US
2010: PostDoc
    INRIA Saclay/FR
2012: Research group leader
    Max Planck Institute/DE
2013: Maître de Conférences
    Télécom ParisTech/FR
Fabian M. Suchanek

2009: PostDoc at Microsoft Research
     Silicon Valley/US
2010: PostDoc
     INRIA Saclay/FR
2012: Research group leader
     Max Planck Institute/DE
2013: Maître de Conférences
     Télécom ParisTech/FR
Fabian M. Suchanek

I am an Elvis fan!

Elvis, when I need you, I can hear you!

Will there ever be someone like him?
Searching with Google

Another singer called Elvis, young
Searching with Google

Carol Connors talks about her first boyfriend - a young singer called Elvis Presley.
Searching with Google

Another singer called Elvis, young

Younger Elvis, singer
Rephrasing does not help

Google: Another singer called Elvis, young

Google: Younger Elvis, singer

Google: Google, you don’t understand! I want
We need structured knowledge

To answer the question, the computer would need structured knowledge: an “ontology”
I work on Ontologies
I work on Ontologies

Constructing ontologies

Wikipedia: The Free Encyclopedia
I work on Ontologies

Constructing ontologies

Mining ontologies

\[ A \land B \Rightarrow C \]
I work on Ontologies

Constructing ontologies

Mining ontologies

Aligning ontologies

\[ A \land B \Rightarrow C \]
I work on Ontologies

Constructing ontologies

Mining ontologies

Protecting ontologies

Aligning ontologies

\[ A \land B \Rightarrow C \]
I work on Ontologies

Le Monde

Constructing ontologies

Mining ontologies

Applying ontologies

Protecting ontologies

Aligning ontologies

Wikipedia

\[ A \land B \Rightarrow C \]
Elvis Presley

Elvis Presley was one of the best blah blah blub blah don’t read this, listen to the speaker! blah blah blah blubl blah you are still reading this! blah blah blah blah blabbel blah

Born: 1935
In: Tupelo
...

Categories:
Rock&Roll, American Singers, Academy Award winners...
Elvis Presley

Elvis Presley was one of the best blah blah blub blah don’t read this, listen to the speaker! blah blah blah blubl blah you are still reading this! blah blah blah blah blabbel blah

Born: 1935
In: Tupelo
...

Categories:
Rock&Roll, American Singers, Academy Award winners...
Elvis Presley was one of the best blah blah blub blah don’t read this, listen to the speaker! blah blah blah blubl blah you are still reading this! blah blah blah blah blabbel blah

Born: 1935
In: Tupelo
...

Categories:
Rock&Roll, American Singers, Academy Award winners...
Elvis Presley

Elvis Presley was one of the best blah blah blub blah don’t read this, listen to the speaker! blah blah blah blubl blah you are still reading this! blah blah blah blah blabbel blah

Categories:
Rock&Roll, American Singers, Academy Award winners...

Born: 1935
In: Tupelo
Elvis Presley

Elvis Presley was one of the best blah blah blub blah don’t read this, listen to the speaker! blah blah blah blubl blah you are still reading this! blah blah blah blah blabbel blah

Born: 1935
In: Tupelo

Categories:
Rock&Roll, American Singers, Academy Award winners...

AmericanSinger

ElvisPresley

1935

Tupelo

won

AcAward

type

born

bornIn
Elvis Presley was one of the best...
Elvis Presley

Elvis Presley was one of the best blah blah blub blah don’t read this, listen to the speaker! blah blah blah blubl blah you are still reading this! blah blah blah blah blah blabbel blah

Born: 1935
In: Tupelo
...

Categories:
Rock&Roll, American Singers, Academy Award winners...

Person
↓ subclass
Singer
↓ subclass
AmericanSinger
↓ type
ElvisPresley

born
1935

bornIn
Tupelo

won
AcAward

locatedIn
USA
Adding WordNet

WordNet
a lexicon of the English language developed at Princeton
We added a temporal and spatial dimension to facts and entities.
YAGO: a large knowledge base

YAGO is a knowledge base that
• combines WordNet classes and Wikipedia instances
• has time and space
• has a manually evaluated accuracy of 95%
YAGO: a large knowledge base

[Wikipedia, GeoNames, WordNet]

Try it online
Try locally

2007 — 2013

YAGO: a large knowledge base

100m facts
10m entities
95% accuracy
100 Web site visitors/day
1000 citations

linked to Freebase and DBpedia

http://yago-knowledge.org
Elvis died in 528

Elvis Presley died in 1955 and was born in 1935.
Elvis died in 528

occurs("Elvis","died in",528)
SOFIE extracts by Reasoning

Elvis died in 528

Einstein died in 1955

occurs("Elvis","died in",528)

occurs("Einstein","died in",1955)
SOFIE extracts by Reasoning

Elvis died in 528

Einstein died in 1955

occurs("Elvis","died in",528)
occurs("Einstein","died in",1955)
died(Einstein,1955), born(Elvis, 1935)
SOFIE extracts by Reasoning

Elvis died in 528

Einstein died in 1955

doctorates("Elvis","died in",528)
doctorates("Einstein","died in",1955)
died(Einstein,1955), born(Elvis, 1935)
doctorates(X’,P,Y) & means(X’,X) & R(X,Y) \rightarrow pattern(P,R)
doctorates(X’,P,Y) & means(X’,X) & pattern(P,R) \rightarrow R(X,Y)
SOFIE extracts by Reasoning

Elvis died in 528

Einstein died in 1955

occurs(“Elvis”, “died in”, 528)
occurs(“Einstein”, “died in”, 1955)
died(Einstein, 1955), born(Elvis, 1935)
occurs(X’, P, Y) & means(X’, X) & R(X, Y) => pattern(P, R)
occurs(X’, P, Y) & means(X’, X) & pattern(P, R) => R(X, Y)
born(X, Y) & died(X, Z) => Z > Y
...
Elvis died in 528

Solving a Weighted MAX SAT problem at scale

occurs("Elvis","died in",528)
occurs("Einstein","died in",1955)
died(Einstein,1955), born(Elvis, 1935)
occurs(X’,P,Y) & means(X’,X) & R(X,Y) => pattern(P,R)
occurs(X’,P,Y) & means(X’,X) & pattern(P,R) => R(X,Y)
born(X,Y) & died(X,Z) => Z>Y
...
SOFIE extracts by Reasoning

Elvis died in 528

occurs(“Elvis”, “died in”, 528)
occurs(“Einstein”, “died in”, 1955)
died(Einstein, 1955), born(Elvis, 1935)
occurs(X’, P, Y) & means(X’, X) & R(X, Y) => pattern(P, R)
occurs(X’, P, Y) & means(X’, X) & pattern(P, R) => R(X, Y)
born(X, Y) & died(X, Z) => Z > Y
...

[WWW 2009]
Product extraction
Work on Ontologies

Constructing ontologies

Mining ontologies

Applying ontologies

Protecting ontologies

Aligning ontologies

A \land B \Rightarrow C
Rule Mining finds patterns

\[\text{married}(x, y) \land \text{hasChild}(x, z) \Rightarrow \text{hasChild}(y, z)\]
Rule Mining finds patterns

married(x, y) ∧ hasChild(x, z) ⇒ hasChild(y, z)

But: Rule mining needs counter examples and RDF ontologies are positive only
Assumption:
If we know \( r(x,y_1), \ldots, r(x,y_n) \), then all other \( r(x,z) \) are false.
Partial Completeness Assumption

Assumption:
If we know $r(x,y_1),..., r(x,y_n)$, then all other $r(x,z)$ are false.
Partial Completeness Assumption

Assumption:
If we know $r(x,y_1),..., r(x,y_n)$, then all other $r(x,z)$ are false.
Partial Completeness Assumption

Assumption:
If we know $r(x,y_1),..., r(x,y_n)$, then all other $r(x,z)$ are false.

+ an efficient implementation
AMIE finds rules in ontologies

\[ AMIE \quad \text{(5min)} \quad \Rightarrow \quad type(x, pope) \Rightarrow diedIn(x, Rome) \]
AMIE finds rules in ontologies

\[ \text{type}(x, \text{pope}) \Rightarrow \text{diedIn}(x, \text{Rome}) \]

[WWW 2013]
Work on Ontologies

- Mining ontologies
- Constructing ontologies
  - Wikipedia
- Applying ontologies
- Protecting ontologies
- Aligning ontologies

Le Monde

\[ A \land B \Rightarrow C \]
Ontologies are complementary

Singer

- type
- plays
- born 1935-01-08

YAGO

RockSinger

- type
- married
- birthDate 1935

ElvisPedia

1935-01-08

ElvisPedia
Wo is the spouse of the guitar player?

YAGO

ElvisPedia

No Links => No Use
Match classes, entities, & relations

- Singer
  - type
  - plays
  - born 1935-01-08

- RockSinger
  - type
  - married
  - birthDate 1935

- subClassOf
- sameAs
- subPropertyOf

YAGO

ElvisPedia
Match classes, entities, & relations

“Elvis”

name

“Elvis”

label
Match classes, entities, & relations

1. Match literals

“Elvis”

name

“Elvis”

label
Match classes, entities, & relations

1. Match literals
Match classes, entities, & relations

2. Assume small equivalence of all relations
Match classes, entities, & relations

2. Assume small equivalence of all relations
3. If entities share a relation that is highly inverse functional, and object is matched, match them.
3. If entities share a relation that is highly inverse functional, and object is matched, match them.
Match classes, entities, & relations

4. If relations share many pairs, increase their match
Match classes, entities, & relations

4. If relations share many pairs, increase their match
5. Iterate

\[ P(e_1 \equiv e_2) = \prod_{1}^{42} \alpha^\beta \ldots P(r_1 \subseteq r_2) \ldots \]

\[ P(r_1 \subseteq r_2) = 42\phi \ldots P(e_1 = e_2) \ldots \]
Match classes, entities, & relations

6. Compute class subsumption

\[ P(c_1 \subseteq c_2) = \arcsin(4.1125) \times P(e_1 \equiv e_2) \times . \]
Match classes, entities, & relations

singer

AmericanSinger

type

6. Compute class subsumption
PARIS: match entities, classes, relations

PARIS matches DBpedia & YAGO

- in 2 hours
- with 90% accuracy

[VLDB 2012]
1. Coalesce the KBs

- bornInCountry: USA
- bornInCity: Tupelo
- locatedIn: Tupelo
2. Mine rules

\[ \text{bornInCountry}(x, z) \Rightarrow \text{bornInCountry}(x, z) \]

“ROSA rule”
ROSA rules match ontologies

\[ \text{bornInCity}(x, y) \land \text{locatedIn}(y, z) \Rightarrow \text{bornInCountry}(x, z) \]

“ROSA rule”
Work on Ontologies

Constructing ontologies

Mining ontologies

Applying ontologies

Protecting ontologies

Aligning ontologies

A $\land B \Rightarrow C$
Plagiarism

People may “steal” from other ontologies without giving due credit. Most ontologies have licenses that require attribution.
Additive Watermarking

By adding a few fake facts to the source ontology, one can prove theft in the target ontology.

[WWW 2012 demo]
Subtractive Watermarking

One can also prove theft by selectively removing facts from the source ontology.

[ISWC 2011]
Work on Ontologies

Le Monde

Constructing ontologies

Mining ontologies

Applying ontologies

Protecting ontologies

Aligning ontologies

A ∧ B ⇒ C
Mining Le Monde

Le Monde

1944 2013
Mining Le Monde

Le Monde

1944  2013

yago
select knowledge
<table>
<thead>
<tr>
<th>time</th>
<th>place</th>
<th>entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1967</td>
<td>USA</td>
<td>Elvis Presley</td>
</tr>
</tbody>
</table>
Mining Le Monde

Average age of people mentioned

Average age of people mentioned
Mining Le Monde

[AKBC 2013]
Work on Ontologies

Mining ontologies

Constructing ontologies

Protecting ontologies

Aligning ontologies

Applying ontologies

\[ A \land B \Rightarrow C \]
YAGO can answer our question

Find x such that
x is called “Elvis”
x is a singer
x was born after 1970

http://yago-knowledge.org
YAGO can answer our question

Find $x$ such that
$x$ is called “Elvis”
$x$ is a singer
$x$ was born after 1970

Elvis Crespo,
singer,
born 1971

http://yago-knowledge.org
Thank you for your attention!

Applying ontologies

Protecting ontologies

Constructing ontologies

Mining ontologies

Aligning ontologies

A \land B \Rightarrow C

WIKIPEDIA

The Free Encyclopedia

Slides done with Powerline, my free SVG slide editor with LaTeX support