



# Intelligence artificielle et les Bases de Connaissances

Fabian M. Suchanek



I am an Elvis Fan!

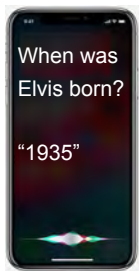


# We visited Elvis in New Zealand



Elvis' lodge in Murchison

# New Technology: Knowledge Bases



Apple Siri

# New Technology: Knowledge Bases



Apple Siri

What is the capital  
of New Zealand?  
"Wellington"



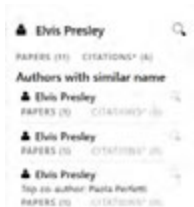
Amazon Echo

Discovered 6 kinase  
proteins that relate  
to cancer

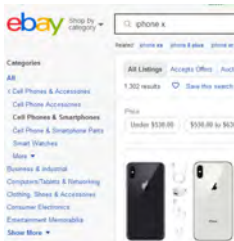


IBM Watson

# New Technology: Knowledge Bases



Microsoft  
Academic

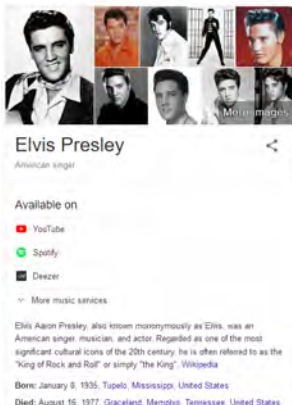


Ebay Knowledge Graph



Facebook

# New Technology: Knowledge Bases



The image shows a Google search result for "Elvis Presley". At the top, there is a grid of nine images of Elvis Presley in various poses and outfits. Below the images, the name "Elvis Presley" is displayed in a large, bold font, followed by the subtitle "American singer". To the right of the name is a share icon. Below this, there is a section titled "Available on" with icons for YouTube, Spotify, and Deezer, and a link to "More music services". Further down, a paragraph of text describes Elvis Aaron Presley, also known mononymously as Elvis, as an American singer, musician, and actor, regarded as one of the most significant cultural icons of the 20th century. It mentions he is often referred to as the "King of Rock and Roll" or simply "the King", and includes a link to Wikipedia. At the bottom, his birth and death information are listed: Born: January 8, 1935, Tupelo, Mississippi, United States; Died: August 16, 1977, Graceland, Memphis, Tennessee, United States.

Elvis Presley

American singer

Available on

YouTube

Spotify

Deezer

More music services

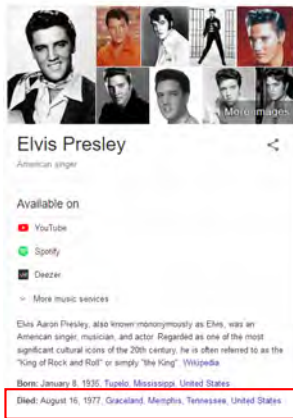
Elvis Aaron Presley, also known mononymously as Elvis, was an American singer, musician, and actor. Regarded as one of the most significant cultural icons of the 20th century, he is often referred to as the "King of Rock and Roll" or simply "the King". [Wikipedia](#)

Born: January 8, 1935, Tupelo, Mississippi, United States

Died: August 16, 1977, Graceland, Memphis, Tennessee, United States

Google Knowledge Graph

# New Technology: Knowledge Bases



A screenshot of a Google search for "Elvis Presley". The search bar on the left contains the text "Elvis Presley". To the right, a knowledge panel displays a collage of eight images of Elvis Presley at the top. Below the images, the name "Elvis Presley" is followed by the subtitle "American singer". A section titled "Available on" lists "YouTube", "Spotify", and "Deezer" with icons. Below this is a paragraph of text about Elvis Aaron Presley, identifying him as an American singer, musician, and actor, and referring to him as the "King of Rock and Roll". At the bottom of the panel, his birth and death information are listed: "Born: January 8, 1935, Tupelo, Mississippi, United States" and "Died: August 16, 1977, Graceland, Memphis, Tennessee, United States". The death information is enclosed in a red rectangular box.

Elvis Presley

American singer

Available on

- YouTube
- Spotify
- Deezer

> More music services

Elvis Aaron Presley, also known mononymously as Elvis, was an American singer, musician, and actor. Regarded as one of the most significant cultural icons of the 20th century, he is often referred to as the "King of Rock and Roll" or simply "the King". [Wikipedia](#)

Born: January 8, 1935, Tupelo, Mississippi, United States

Died: August 16, 1977, Graceland, Memphis, Tennessee, United States

???

Google Knowledge Graph



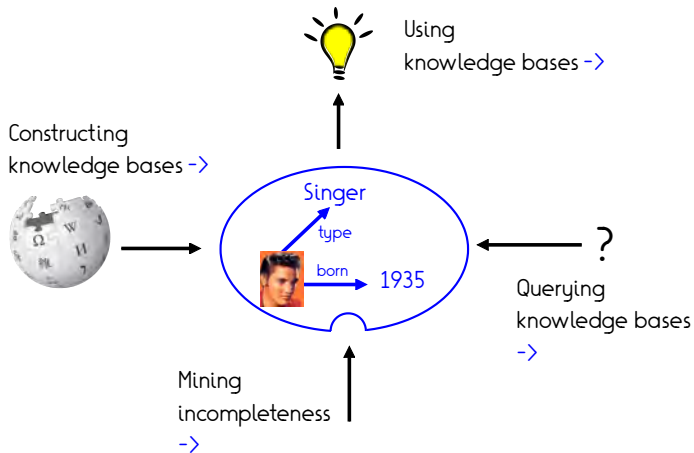
# Knowledge Bases



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 A-Box.)

# Knowledge Base Life Cycle



# Extracting from Wikipedia

Elvis Presley



WIKIPEDIA  
The Free Encyclopedia

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



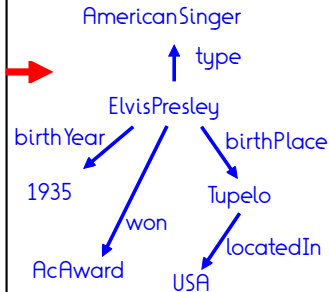
Born: 1935

In: Tupelo

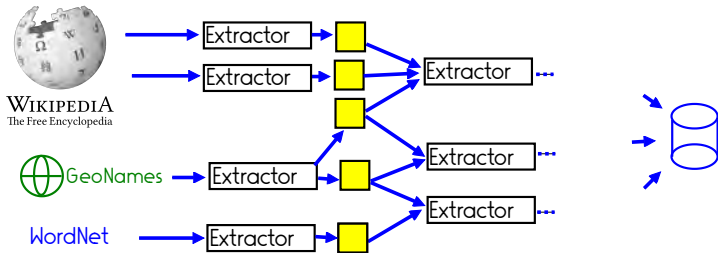
...

Categories:

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



# Creating a large knowledge base

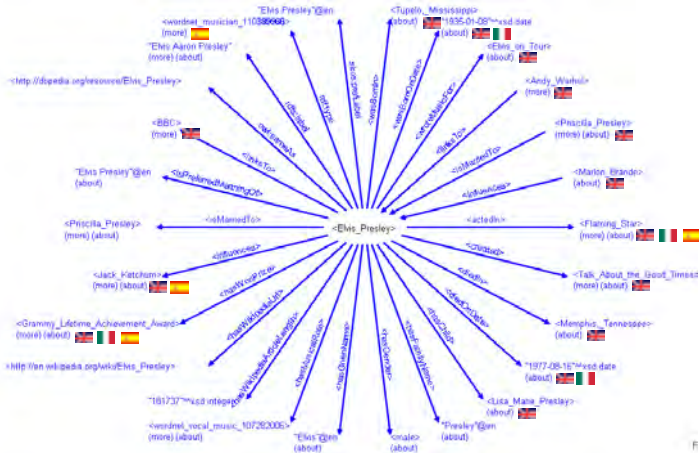


Intermediate extractors

- clean facts
- deduplicate facts and entities
- check consistency

ensuring high quality (95%)

# Example: YAGO about Elvis



# YAGO: a large knowledge base



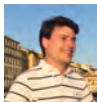
<http://yago-knowledge.org>  
open code and open data

Wikipedia + WordNet  
time and space  
10 languages  
100 relations  
100m facts  
10m entities  
95% accuracy  
used by DBpedia  
and IBM Watson

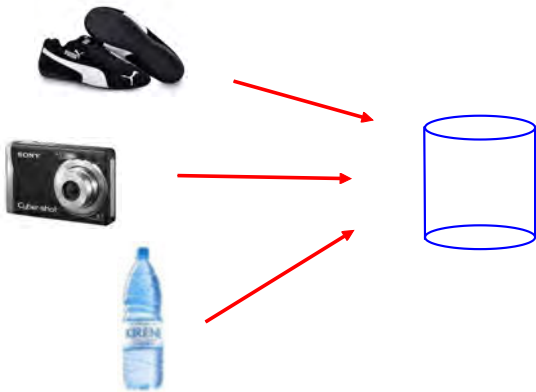


[WWW'07, JWS'08, WWW'11 demo, AIJ'13, WWW'13 demo, CIDR'15, ISWC'16]

>IBEX



# Goal: Harvest entities from the Web



# IBEX: Collect unique ids

Puma PowerTech Blaze Running Shoe - Mens - Black/Silver/Select With-M - UPC # 887128476661



AUD99.02

Please Choose:

Size

- ☐ (EUR)40/(US)8.5/(UK)6
- ☐ (EUR)41/(US)9.5/(UK)7
- ☐ (EUR)42/(US)10.5/(UK)7.5
- ☐ (EUR)43/(US)11.5/(UK)8.5
- ☐ (EUR)44/(US)12/(UK)9
- ☐ (EUR)45/(US)13/(UK)10
- ☒ (EUR)46/(US)14/(UK)11

Add to Cart



Item # 207112

UPC # 887128476661

Bring or go home in the men's Puma PowerTech Blaze running shoe! This performance trainer has upgraded technical features with its OrthoLite® footbed, PowerTech™ cushioning and PowerLite™ abrasion-resistant outsole.

This men's running shoe is best for the neutral runner. What is a pronator?

Smoothable mesh upper with synthetic overlay for structured support.

Lace up closely.

887128476661

Unique identifiers can be found by a regular expression, and verified by a checksum.



# IBEX: Collect unique ids

Puma PowerTech Blaze Running Shoe - Mens - Black/Silver/Select With-M - UPC # 887128476661



AUD99.02

Please Choose:

Size

- ☐ (EUR)40/(US)8.5/(UK)6
- ☐ (EUR)41/(US)9.5/(UK)7
- ☐ (EUR)42/(US)10.5/(UK)7.5
- ☐ (EUR)43/(US)11.5/(UK)8.5



Item # 207112

UPC # 887128476661

Go to gym or go home in the men's Puma PowerTech Blaze running shoe! This performance trainer has upgraded technical features with its OrthoLite® footbed, PowerTech™ cushioning and PowerFlex™ abrasion-resistant outsole.

This men's running shoe is best for the neutral runner. What is pronator?

Smoothable mesh upper with synthetic overlay for structured support

Lace-up closure

id	name	URL
1234	Puma PowerTech Blaze	u1
1234	Please choose	u1
1234	Puma PowerTech Blaze	u2
1234	Puma Shoe	u2
5678	Please choose	u3
5678	Sony Cybershot TS100	u3
...	...	...



# IBEX: analyses

## Found

- 13m email addresses with their name
- 235K chemical products
- 1.4m books
- 1.1m products
- ... with an accuracy of 73%-96%



## Analyzed

- Global trade flow
- frequent email providers
- frequent people names and more

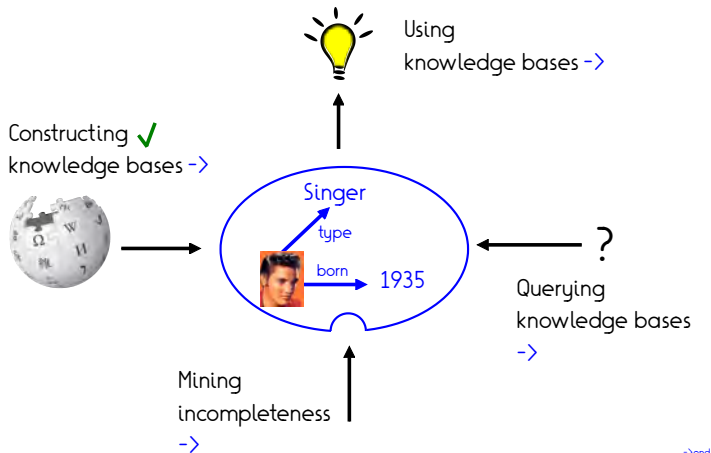


[WebDB 2015]

All data available online at

<http://resources.mpi-inf.mpg.de/d5/ibex/>

# Knowledge Base Life Cycle



# Are all people married?



# Are all people married?



✓

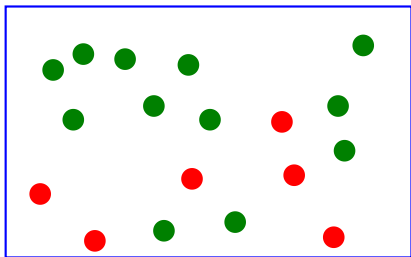


✓



?

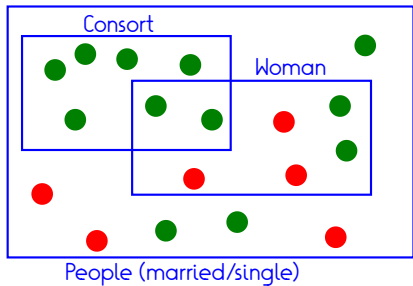
# Are all people married?



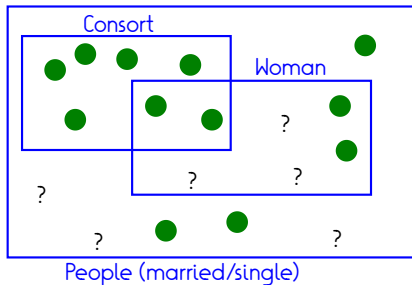
Real World

People (married/single)

# Are all people married?



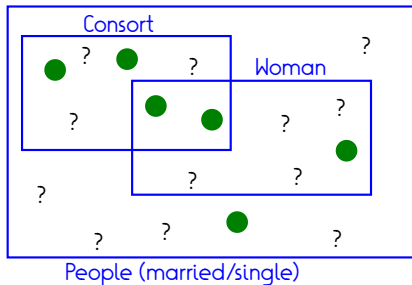
# Are all people married?



Knowledge base  
without negative  
information

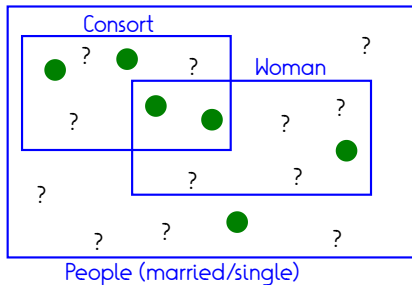


# Are all people married?



Knowledge base  
without negative  
information and  
with incompleteness

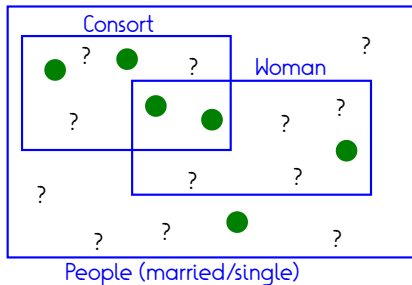
# Are all people married?



Knowledge base  
without negative  
information and  
with incompleteness

Baseline 1: Obligatory if all instances have it

# Are all people married?

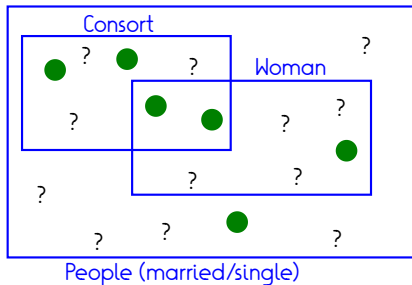


Knowledge base  
without negative  
information and  
with incompleteness

Baseline 1: Obligatory if all instances have it ✗

Baseline 2: Obligatory if at least n% of instances have it

# Are all people married?



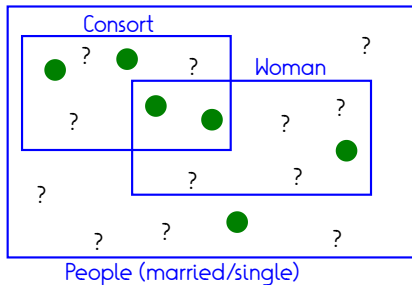
Knowledge base  
without negative  
information and  
with incompleteness

Baseline 1: Obligatory if all instances have it ✗

Baseline 2: Obligatory if at least n% of instances have it => Woman ✗

Baseline 3: Obligatory if all instances that have it fall in the class

# Are all people married?



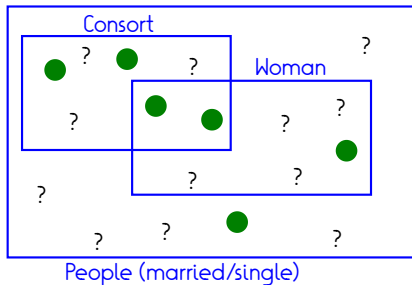
Knowledge base  
without negative  
information and  
with incompleteness

Baseline 1: Obligatory if all instances have it ✗

Baseline 2: Obligatory if at least n% of instances have it => Woman ✗

Baseline 3: Obligatory if all instances that have it fall in the class  
=> Person ✗

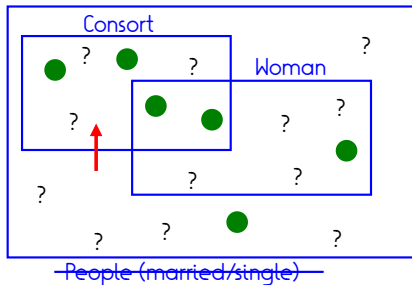
# Are all people married?



Knowledge base  
without negative  
information and  
with incompleteness

Theorem: If the KB is sampled randomly uniformly from the real world, and if the density of an attribute changes when we go into an intersecting class, then the attribute cannot be obligatory.

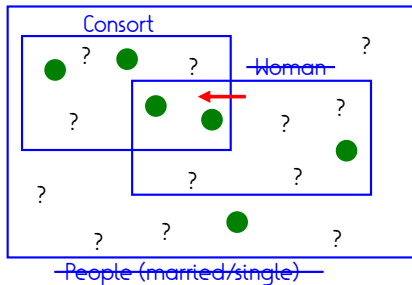
# Are all people married?



Knowledge base  
without negative  
information and  
with incompleteness

Theorem: If the KB is sampled randomly uniformly from the real world, and if the density of an attribute changes when we go into an intersecting class, then the attribute cannot be obligatory.

# Are all people married?



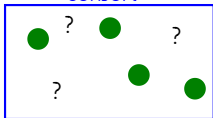
Knowledge base  
without negative  
information and  
with incompleteness

Theorem: If the KB is sampled randomly uniformly from the real world, and if the density of an attribute changes when we go into an intersecting class, then the attribute cannot be obligatory.



# Determining obligatory attributes

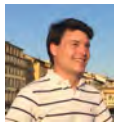
Consort



We can predict obligatory attributes of classes with up to 80% precision (at 40% recall).

Caveat:

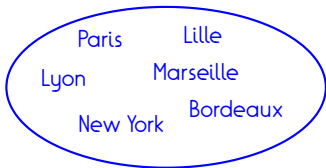
We do not actually predict, but exclude.



[WWW 2018]

# Incompleteness: Missing entities

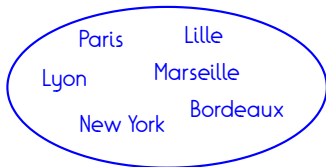
We have the following cities in our knowledge base:



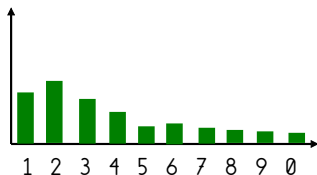
Are there any cities missing?

# Incompleteness: Missing entities

We have the following cities in our knowledge base:



Are there any cities missing?



- 1) Take the number of inhabitants of each city
- 2) Take the first digit
- 3) Plot the number of cities per first digit

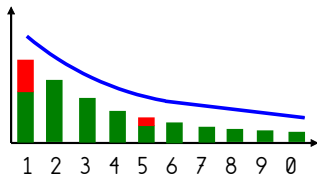
# Incompleteness: Missing entities

Benford's law says that the first digit  $d$  appears with probability

$$\log_{10}\left(1 + \frac{1}{d}\right)$$

=> We can give a minimum numbers of cities that are missing to make the distribution representative of the real world.

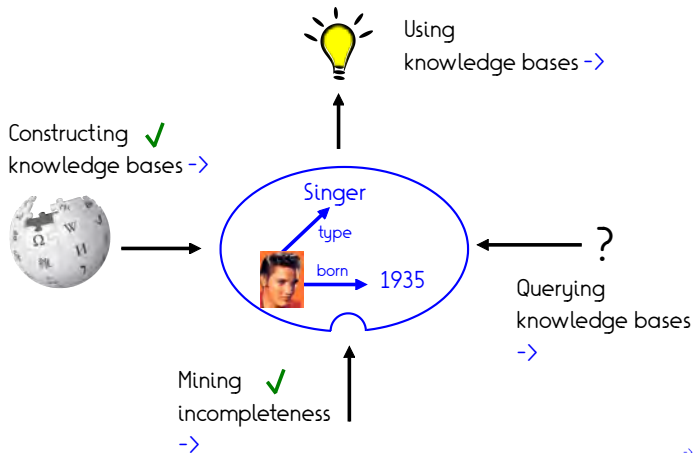
(For other classes, we can learn a parameter for a variant of the law.)



[ISWC 2018]



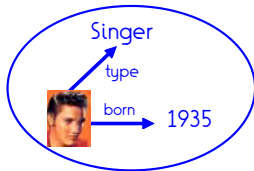
# Knowledge Base Life Cycle



->end

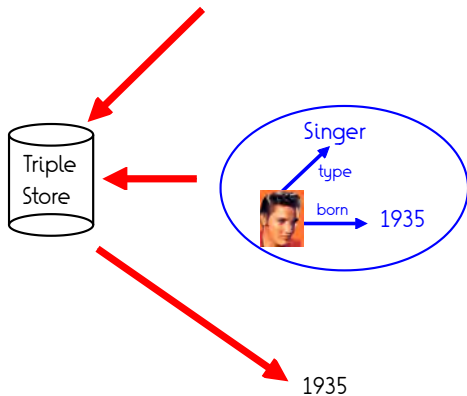
# Querying Knowledge Bases

When was Elvis born?



# Querying Knowledge Bases

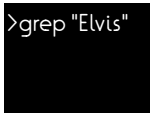
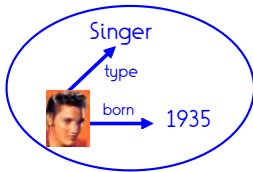
When was Elvis born?



# Querying Knowledge Bases

When was Elvis born?

If I have only  
a single query



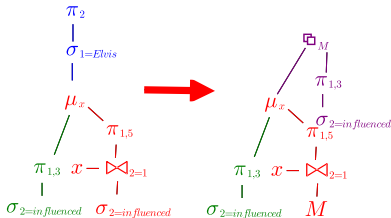
1935



# Querying Knowledge Bases

Who were the people that  
Elvis influenced (transitively)?

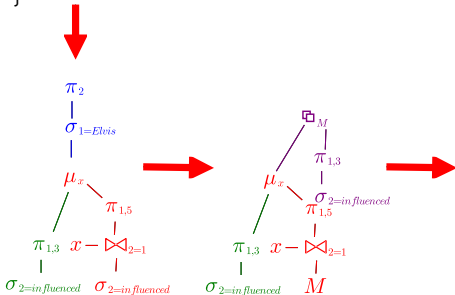
SELECT ?a {  
 <Elvis> <influenced>+ ?a  
}



# Querying Knowledge Bases

Who were the people that  
Elvis influenced (transitively)?

```
SELECT ?a {  
  <Elvis> <influenced>+ ?a  
}
```



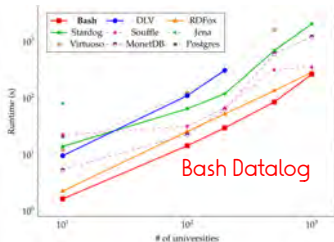
```
while sort -t '$\t' -k 1 -k 1  
<(cat tmp/lock_mat1 1  
join -t '$\t' -1 2 -2 1 -o 1  
<(sort -t '$\t' -k 1 tmp/d  
l comm -23 - tmp/full2 >  
mv tmp/new2 tmp/delta2  
sort -u -merge -o tmp/  
[ -s tmp/delta2 ];  
do continue; done
```

# Querying Knowledge Bases

VLDB reviewer: Unix Shell commands to run queries is not very innovative.  
The DB community exists to overcome the shortcomings of this approach.

# Querying Knowledge Bases

VLDB reviewer: Unix Shell commands to run queries is not very innovative.  
The DB community exists to overcome the shortcomings of this approach.



Bash Datalog

including  
the loading time

	Bash	RDFox	BigDatalog	Stardog
LiveJournal (69M edges)	2min	1min	9min	15min
Orkut (117M edges)	2.5min	2min	30min	18min
Friendster (2586M edges)	4:30h	OOM	OOS	>10h
WikiData (2100M edges)	1h	OOM		

# Convert SPARQL to Bash online

**Bash Datalog**  
Answering Datalog Queries with Unix Shell Commands

AboutDatalog modeSPARQL/OWL modeAPI

**SPARQL query**

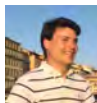
```
BASE <http://yago-knowledge.org/resource/>
SELECT ?a WHERE {
  <Elvis_Presley> <influenced>+ ?a.
}
```

Convert to bash scriptDownload script

**Bash script**

```
read_ntriples() { $awk -F" '" '{ sub(" ", "\t"); sub(" ",
"\t"); sub(/ \.\/, ""); print $0 }' "$@"; }
conv_ntriples() { $awk -F'\t' '{ print $1 " " $2 " " $3 "
." }'; }

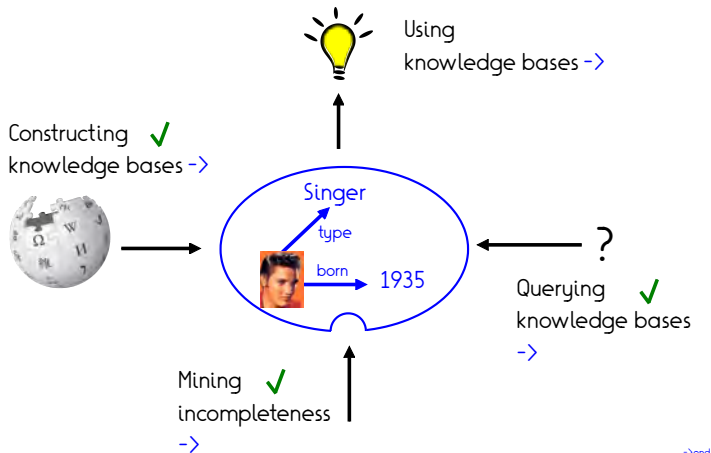
touch tmp/mat0 tmp/mat1
$awk -v FS='\t' '{ $1 == "<http://yago-knowledge.org
/resource/Elvis_Presley>" && $2 == "<http://yago-
knowledge.org/resource/influenced>" { print $3 >>
"tmp/mat0" } }
```



[ISWC 2018]

<https://thomasrebele.org/projects/bashlog/>

# Knowledge Base Life Cycle



->end

# Combinatorial Creativity



# Description Logics do not work

$Mop \equiv Tool \sqcap \exists has.Stick \sqcap \exists has.Strings$

$BabyMop \equiv$

$Romper \sqcap \exists has.(Mop \sqcap \neg \exists has.Stick) \sqcap \exists has.Baby$

$\equiv \dots \sqcap \exists has.\perp \sqcap \dots$





# Language for Combinatorial Creativity

$$Mop \equiv Tool \sqcap \exists has.Stick \sqcap \exists has.Strings$$

Subtraction:  $Mop - \exists has.\top \equiv Tool \sqcap \exists has.Strings$

Addition:  $Mop + \exists has.\top \equiv Mop$

Succession:  $Mop \rightarrow \exists u.\top \equiv Stick$

Selection\*:  $Mop \uparrow \exists has.\top \equiv \exists has.Stick$

$$Romper + \exists has.(Mop - \exists has.Stick) + \exists has.Baby \\ \equiv BabyMop$$



# Language for Combinatorial Creativity

$Mop \equiv Tool \sqcap \exists has.Stick \sqcap \exists has.Strings$

Subtraction:  $Mop - \exists has.\top \equiv Tool \sqcap \exists Strings$

Addition:  $Mop + \exists has.\top \equiv Mop$

Succession:  $Mop \rightarrow \exists r.\top \equiv Stick$

Selection\*:  $Mop \uparrow \exists has.\top \equiv \exists has.Stick$

## 1) Descriptive experiments



## 2) Generative experiments

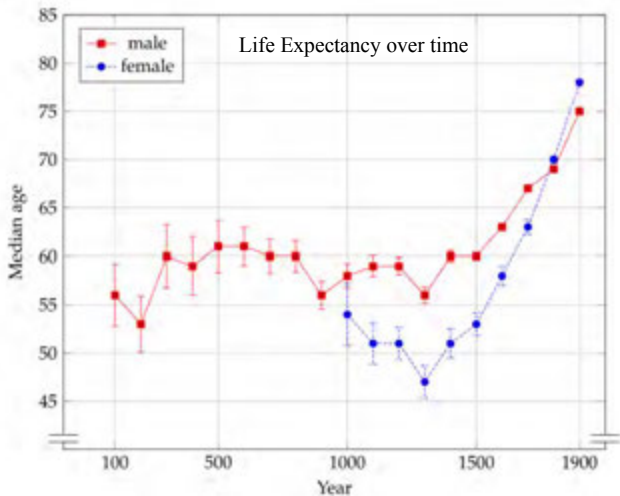
1/3 nonsense, 1/3 exists,  
1/3 "imaginable"



[ISWC 2016  
paper & demo]

>DigitalHum

# YAGO for the Digital Humanities



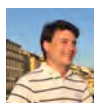
## Presence of foreign companies



red: many foreign companies mentioned

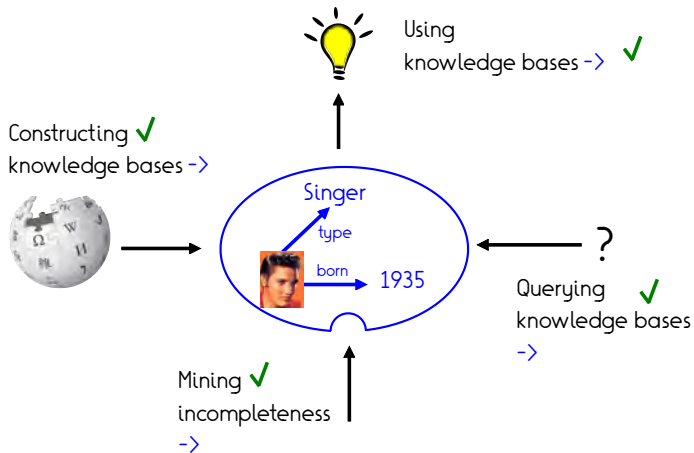
blue: few foreign companies mentioned

Le Monde




[AKBC 2013,  
VLDB 2014 vision,  
LDOW 2018]

# Knowledge Base Life Cycle



# Is Elvis dead?



**Elvis Presley**  
*American singer*

Available on

- YouTube
- Spotify
- Deezer

> More music services

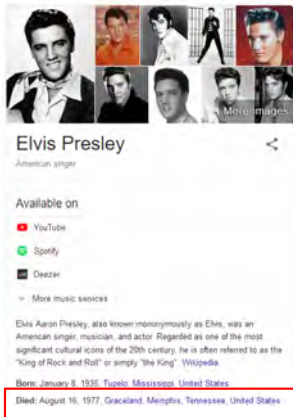
Elvis Aaron Presley, also known mononymously as *Elvis*, was an American singer, musician, and actor. Regarded as one of the most significant cultural icons of the 20th century, he is often referred to as the "King of Rock and Roll" or simply "the King". [Wikipedia](#)

Born: January 8, 1935, Tupelo, Mississippi, United States

Died: August 16, 1977, Graceland, Memphis, Tennessee, United States

???

# Is Elvis dead?



A screenshot of a web profile for Elvis Presley. At the top is a collage of eight images of him in various poses and outfits. Below the images is the name "Elvis Presley" and the subtitle "American singer". Underneath are icons for "Available on" services: YouTube, Spotify, and Deezer, followed by a link to "More music services". A short biography follows, describing him as an American singer, musician, and actor, and referring to him as the "King of Rock and Roll". At the bottom, his birth and death information are listed: "Born: January 8, 1935, Tupelo, Mississippi, United States" and "Died: August 16, 1977, Graceland, Memphis, Tennessee, United States". The death information is enclosed in a red rectangular border.

Elvis Presley  
*American singer*

Available on

YouTube Spotify Deezer

> More music services

Elvis Aaron Presley, also known mononymously as Elvis, was an American singer, musician, and actor. Regarded as one of the most significant cultural icons of the 20th century, he is often referred to as the "King of Rock and Roll" or simply "the King". [Wikipedia](#)

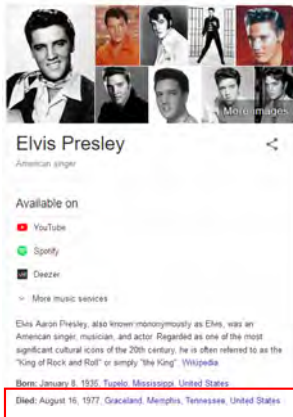
Born: January 8, 1935, Tupelo, Mississippi, United States

Died: August 16, 1977, Graceland, Memphis, Tennessee, United States



???

# Is Elvis dead?



Elvis Presley

American singer

Available on

YouTube

Spotify

Deezer

More music services

Elvis Aaron Presley, also known mononymously as Elvis, was an American singer, musician, and actor. Regarded as one of the most significant cultural icons of the 20th century, he is often referred to as the "King of Rock and Roll" or simply "the King". [Wikipedia](#)

Born: January 8, 1935, Tupelo, Mississippi, United States

Died: August 16, 1977, Graceland, Memphis, Tennessee, United States



100m statements

95% accuracy

-> 5m wrong statements

???



# Knowledge Base Life Cycle

