

# Information Extraction

Lecture 1: Motivation &  
Knowledge Representation

Fabian M. Suchanek

# Schedule

Monday 20 April

9:00-10:30 Motivation & Knowledge Representation

10:45-12:15 Named Entity Recognition

lunch break

14:00-15:30 Disambiguation

15:45-17:15 Fact Extraction

Tuesday 21 April

9:00-10:30 Semantic Web

10:45-12:15 Lab (**graded!**)

lunch break

14:00-17:00 Julien Leproust from Exalead

# Organization

- Grading is based on lab work
- Labs are done alone
- All material and schedule  
is available online at

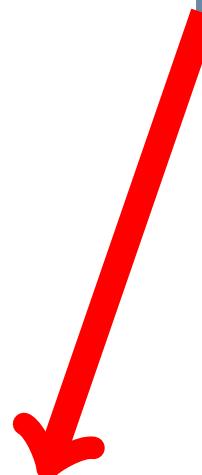
<http://suchanek.name/work/teaching/ces2015>

# Lecturer: Fabian Suchanek

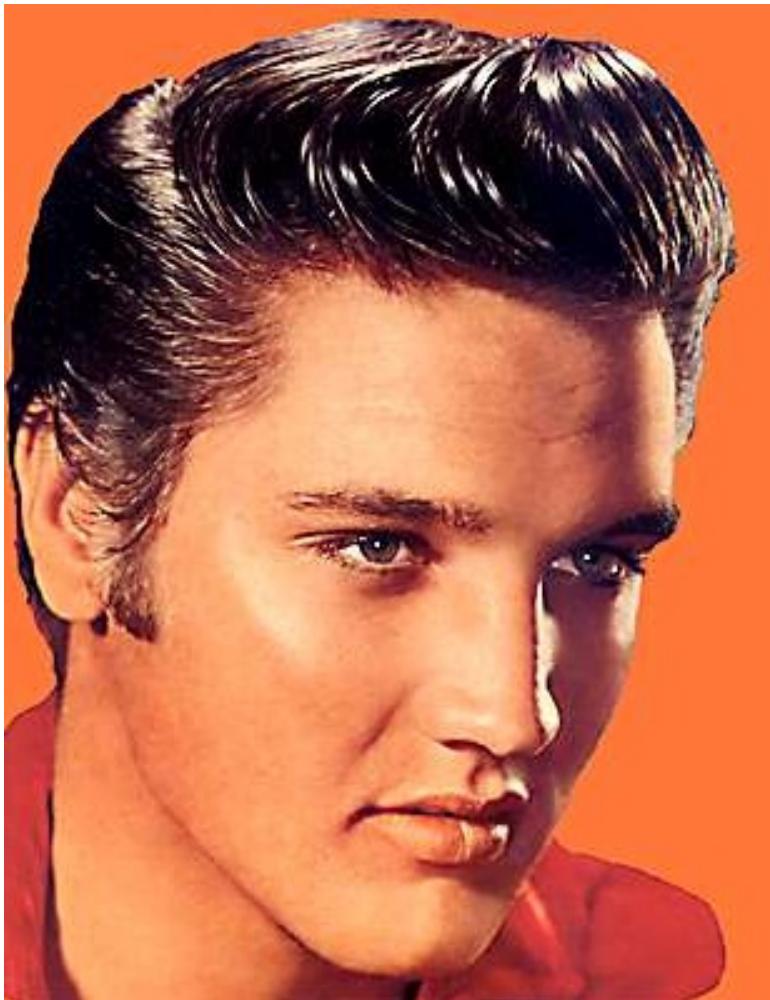


Associate Professor at  
Télécom ParisTech  
DBWeb team, office C201-6

<http://suchanek.name>



# Elvis Presley



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

Is Elvis still alive?

# Elvis Alive: Google

The image shows a Google search results page. At the top left is the Google logo. In the search bar, the query "is elvis alive?" is typed. Below the search bar, the "Web" tab is selected, followed by "Images", "Maps", "Shopping", "Videos", "More ▾", and "Search tools". A red horizontal line is under the "Web" tab. Below the tabs, it says "About 21,700,000 results (0.22 seconds)". The first result is a link to "WebSpawner - IS ELVIS DEAD OR ALIVE ???" with the URL "www.webspawner.com/users/elviselvis/". The text next to the link states: "As it turns out, there are many concrete reasons to believe that the King is still alive. The Gravesite. For Starters, **Elvis's** name is spelled wrong on his headstone." At the bottom right of the search results area, the word "Google" is written.

is elvis alive?

Web Images Maps Shopping Videos More ▾ Search tools

About 21,700,000 results (0.22 seconds)

[WebSpawner - IS ELVIS DEAD OR ALIVE ???](http://www.webspawner.com/users/elviselvis/)  
www.webspawner.com/users/elviselvis/  
As it turns out, there are many concrete reasons to believe that the King is still alive. The Gravesite. For Starters, **Elvis's** name is spelled wrong on his headstone.

Google

Google: undecided

# Elvis Alive: Bing



is elvis alive?



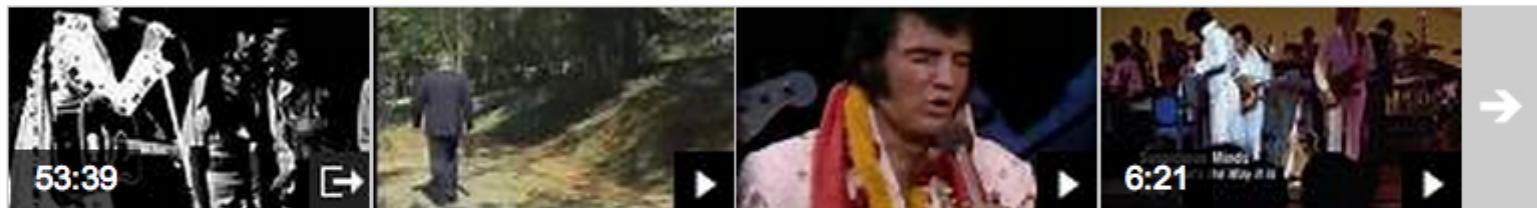
23 200 000 ERGEBNISSE

Einschränken nach Sprache ▾

Einschränken nach Region

## Videos von is elvis alive?

[bing.com/videos](http://bing.com/videos)



[Elvis Live at  
Madison Square  
YouTube](#)

[Elvis Alive  
YouTube](#)

[Elvis Presley-An  
American Trilog...  
WAt.tv](#)

[Elvis Presley  
Suspicious Mind...  
YouTube](#)

## elvis is alive! - YouTube Diese Seite übersetzen

[www.youtube.com/watch?v=1vMcEJww8os](http://www.youtube.com/watch?v=1vMcEJww8os) ▾

10 Min. · 139 457 Ansichten · Hinzugefügt am 09/03/2007

We are not interested in fame. Our goal is to show our work and let the world know it:  
**Elvis is alive!!!!**

Bing

Bing: yes

# Elvis Alive: Ask.com

Réponses

Images



is elvis alive?

[Proof Elvis is Alive by J. Parra - ...](#)

[www.youtube.com/watch?v=TRBm6q-3SP0](http://www.youtube.com/watch?v=TRBm6q-3SP0)

7 Jan 2012 ... In this video, it will show that Elvis is still alive and a doctor who treated him claims that he is still alive.

[Ask.com](#)

Ask.com: yes

# Elvis Alive: Evi



Is elvis alive?



You asked: Is elvis alive?

Elvis Presley died in Memphis on August 16th 1977. He was killed by a myocardial infarction. He was 42 years, 7 months and 8 days old when he died.



Elvis Presley



[website](#)



[wikipedia](#)

Evi

Evi: no

# Elvis Alive: Wolfram Alpha



is elvis alive?



Ex

Input Interpretation:

Elvis Presley alive?

Result:

No

Wolfram Alpha: no

Wolfram Alpha

# How do they know?



Sources



- The New York Times Company. "Obituaries." *The New York Times*. »

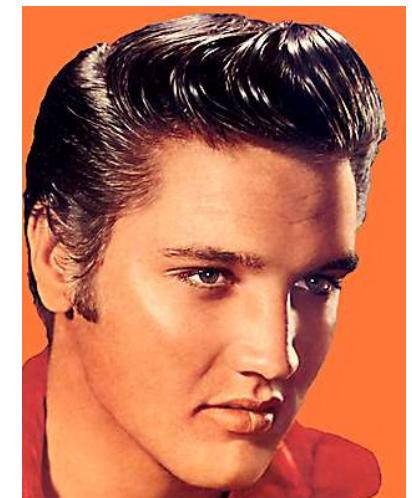


Elvis Presley, the first  
and greatest American  
rock-and-roll star,  
died yesterday at  
the age of 42.

# Text understanding

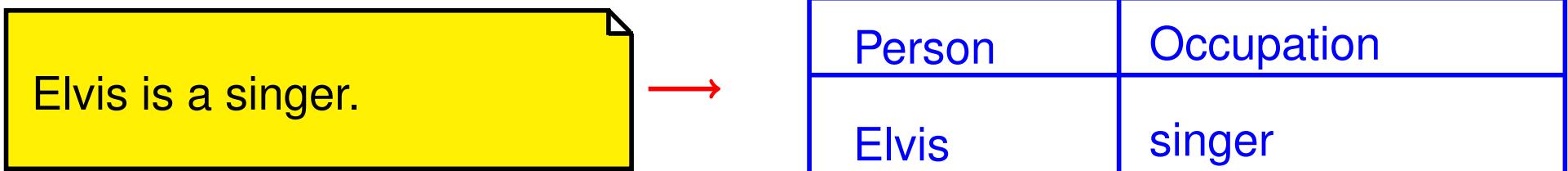
To answer the question, Wolfram Alpha has to understand this text.

Elvis Presley, the first  
and greatest American  
rock-and-roll star,  
died yesterday at  
the age of 42.

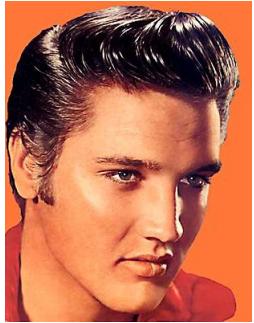


# Information Extraction

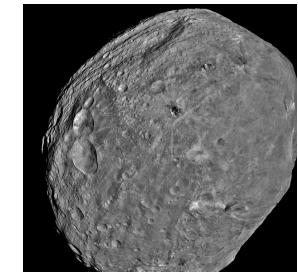
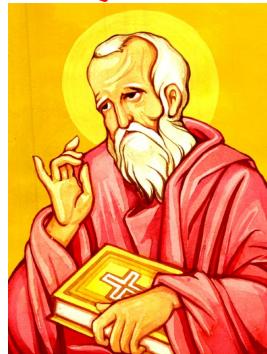
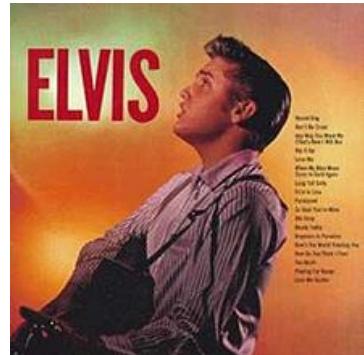
Information Extraction (IE) is the process of deriving structured factual information from digital text documents.



# Difficulty:Ambiguity



Elvis is a singer.



# Difficulty: Detect phrases

Elvis is a rock star.



Elvis is a rock star.



Elvis is a real rock star.



# Difficulty:Verbal phrases

Elvis works as a rock star.

Elvis is famous as a rock star.

Elvis claims to be a rock star.

Elvis is not a rock star.

Now is he a rock star or what??

difficulties > 31

# Difficulty: Parsing

Bell, a telecommunication company, which is based in Los Angeles, makes and distributes electronic, computer, and building products.

# Difficulty: Parsing

Bell, a telecommunication company, which  
is based in Los Angeles, makes and distributes  
electronic, computer, and building products.

Subject	Predicate	Object
Bell	is	telecomm. company
Bell	based in	Los Angeles
Bell	makes	electronic products
Bell	distributes	electronic products
Bell	makes	computer products
...		

# Difficulty:Table Interpretation

No.	Country	Capital
1	Afghanistan	Kabul
2	Albania	Tirane
3	Algeria	Algiers
4	Andorra	Andorra la Vella

# Difficulty:Table Interpretation

Afghanistan hasCapital Kabul

No.	Country	Capital
1	Afghanistan	Kabul
2	Albania	Tirane
3	Algeria	Algiers
4	Andorra	Andorra la Vella

# Difficulty:Table Interpretation

Physical	
Dimensions (WxHxD)	3.82 x 2.30 x 0.70"
Weight	2.3kg Camera Only

# Difficulty:Table Interpretation

Physical	
Dimensions (WxHxD)	3.82 x 2.30 x 0.70"
Weight	2.3kg Camera Only

Weight hasPhysical 3.82

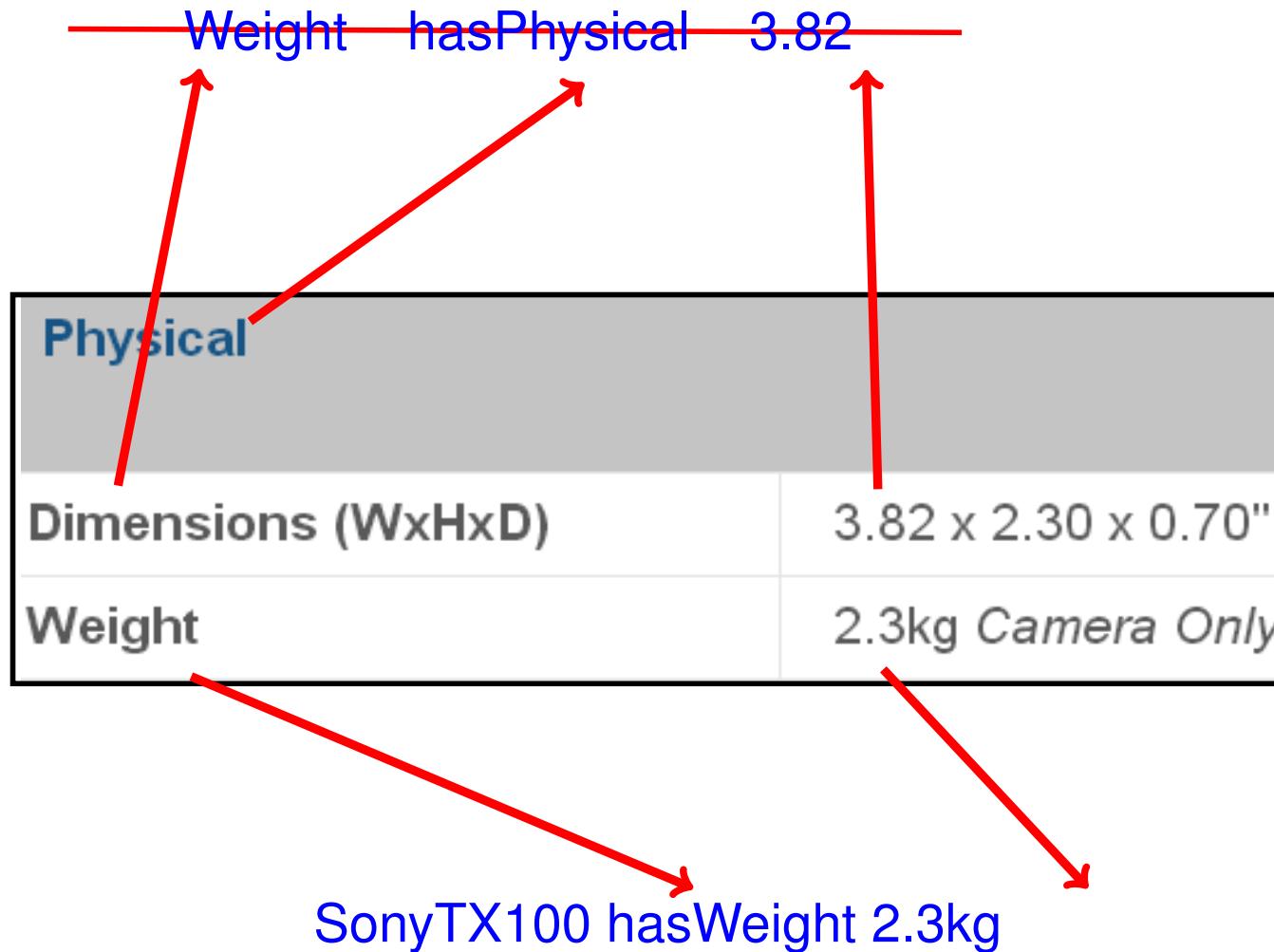
# Difficulty:Table Interpretation

Physical	
Dimensions (WxHxD)	3.82 x 2.30 x 0.70"
Weight	2.3kg Camera Only

Weight hasPhysical 3.82

The diagram illustrates the relationship between the table data and the header information. Three red arrows point from the 'Weight' row in the table to the '3.82' value in the 'hasPhysical' column of the header. One arrow points from the 'Weight' label to the first '3.82'. Another arrow points from the 'Weight' label to the second '3.82'. A third arrow points from the 'Weight' label to the 'hasPhysical' label.

# Difficulty:Table Interpretation



difficulties > 31

# Difficulty: Parsing Literals

## Background information

**Birth name** Elvis Aaron Presley

**Born** January 8, 1935

Tupelo, Mississippi, U.S.

# Difficulty: Parsing Literals

Literals take many different forms:

8th of January 1935

1935-01-08

8/1/1935

8/Jan/1935

08-01-1935

b. 1935

1/8/1935

Jan 8, 1935

# Difficulty:Segmentation

## ADRESSES ET ACCÈS - CONTACTS

**Adresse postale :** 46 rue Barrault F-75634 Paris Cedex 13

**Téléphone :** +33 (0)1 45 81 77 77 - **Fax :** +33 (0)1 45 89 79

# Difficulty:Segmentation

## ADRESSES ET ACCÈS - CONTACTS

Adresse postale : 46 rue Barrault F-75634 Paris Cedex 13

Téléphone : +33 (0)1 45 81 77 77 - Fax : +33 (0)1 45 89 79

# Difficulty:Segmentation

## ADRESSES ET ACCÈS - CONTACTS

Adresse postale : 46 rue Barrault F-75634 Paris Cedex 13

Téléphone : +33 (0)1 45 81 77 77 - Fax : +33 (0)1 45 89 79

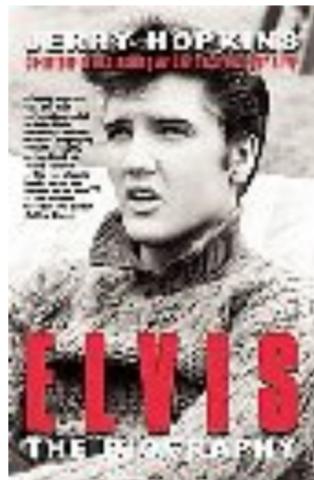


No	Street	Country	ZIP	City
46	Barrault	France	75634	Paris

# Difficulty:Bad HTML Code

```
<table>
```

```
<tr>
```



[View Larger Image](#)

## Elvis : The Biography

**Hopkins, Jerry**

**9780859653916**

ISBN 10: **0859653919 / 0-85965-391-9** <br>

ISBN 13: **9780859653916** <br>

Publication Date: **2007**

```
</table>
```

- Tables are used for layout
- Lists do not have list tags
- HTML code is invalid

# Why would anybody want to do Information Extraction?

The reason is that...

this is abundant...



Elvis is a singer.



...and this is useful.



Person	Occupation
Elvis	singer

# Application: Emails

Hi I'm russian girl with 25 years old  
looking for a faithful partner. Chat with  
me on Monday 2pm at [http://wtf.ru!!!](http://wtf.ru)

Add to calendar?

Added [Chat with Russian girl](#) on Mon 2013-03-11 at 14:00. [Undo](#)

[More examples](#)  
[Even more examples](#)

# Application: Customer care

Dear Apple support,

my iPhone 6 charging cable is not interoperable  
with my fridge. Can you please help me?



Product	Part	Problem
iPhone 6	Charger	Compat.



# Application: Predictions

## How two scientists are using the *New York Times* archives to predict the future

by Laura Hazard Owen FEB. 1, 2013 - 7:24 AM PST

 7 Comments    +1 

A▼ A▲

---

SUMMARY: Researchers are creating software that analyzes 22 years of *New York Times* archives, Wikipedia and about 90 other web resources to predict future disease outbreaks, riots and deaths — and hopefully prevent them.

 tweet this

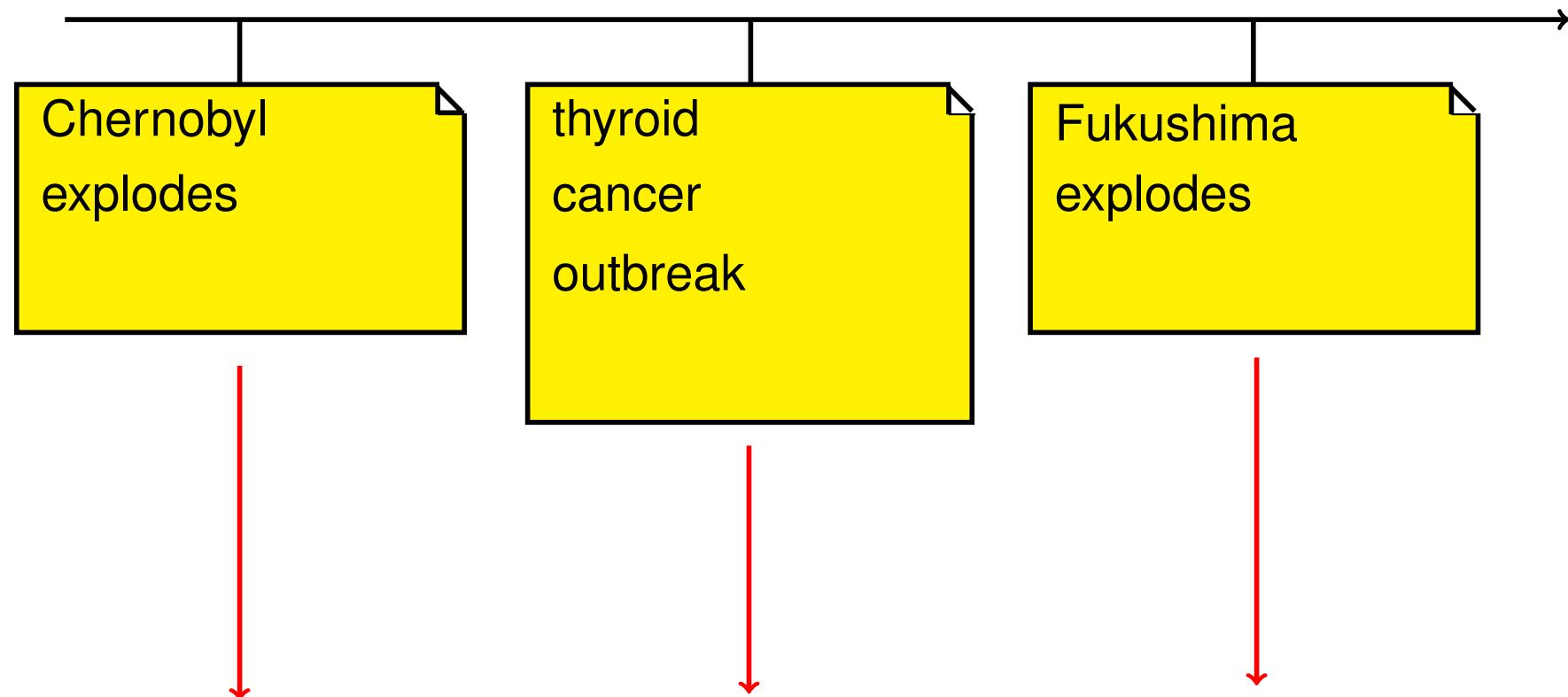


photo: Shutterstock / Ruggiero Scardigno

Gigaom

# Application: Predictions

news articles in time



Nuclear  
catastrophe



Cancer  
outbreak

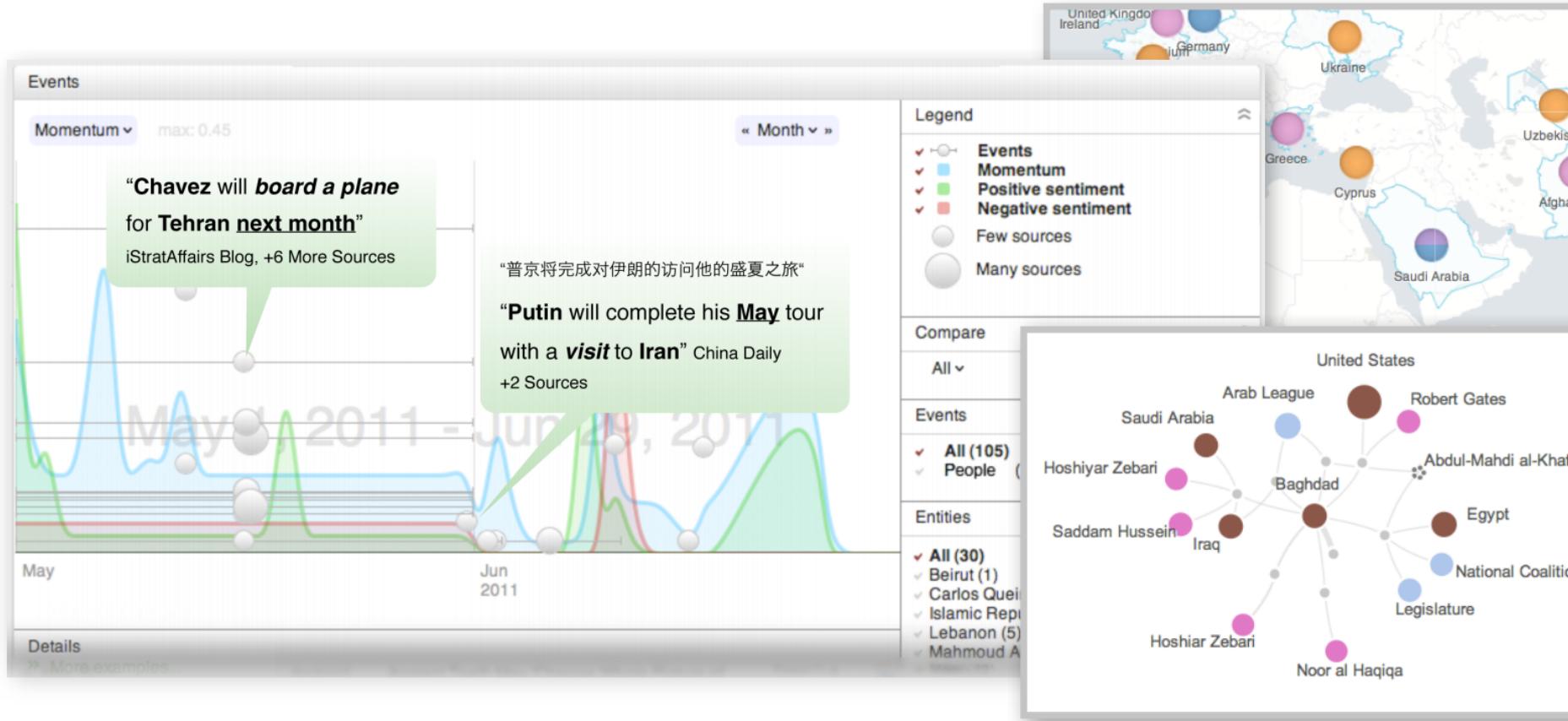
Nuclear  
catastrophe



?

ads,pred>22

# Application: Analysis



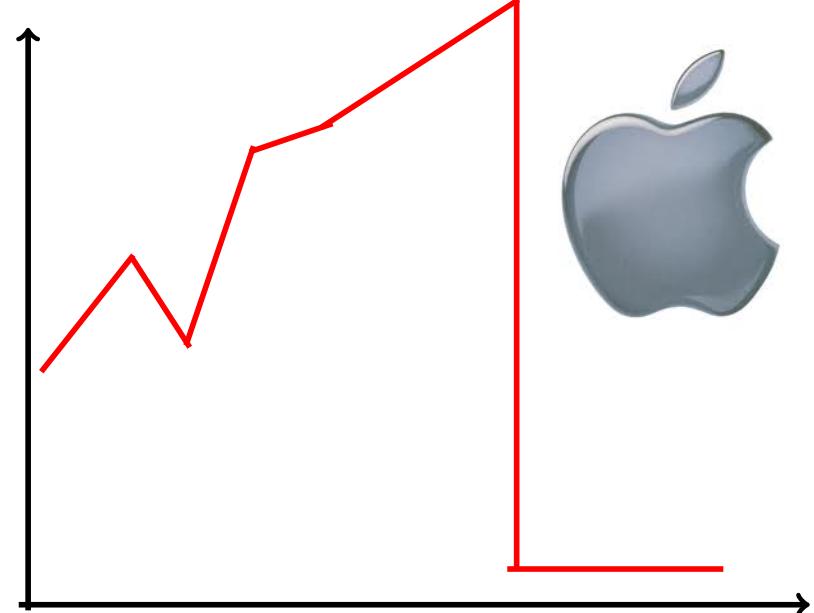
Information extracted from the news can help model and analyze political events.

# Appl.: Opinion Mining

iPhone maps  
doesn't work



iPhone  
+ negative  
opinion



# Appl: Classified Ads

## Internet Explorer 9



### Pros

Internet Explorer is the only brows option we looked for, including dire telephone.

### Cons

Its vast popularity makes it a target for hackers.

### The Verdict: 9.08/10

With competitive speed and functionality, it proves to be a viable competitor.

## Google Ads

Use Chrome!

No, use IE!

Internet Explorer (browser)



# Application: Assistants

## Booking Confirmation

### Guest Details

Name : Mr. Fabian Suchanek (ID#P5402)

Address : Campus E1.4  
Saarbrücken, Saarland, 66111  
Germany

4968193255014 4968193255014  
[f.m.suchanek@gmail.com](mailto:f.m.suchanek@gmail.com)

### Room Type

### Package/Rate Detail

#### Non-Refundable Promotion

#### Sleep Two(S)

Inclusions None  
Amenities Wi-Fi High speed wireless Internet service  
Requires  
100% Deposit

forward email

to assistant

Wipolo



portal.bio > 26

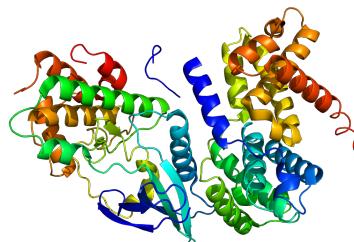
# Application: Biomedical



## PubMed

PubMed comprises more than 22 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

PubMed



Substance	Inhibits
No0ne	UndEr
s1ands	Th <i>i</i> s

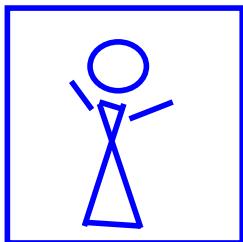
Figure out from biomedical literature  
which substances inhibit which proteins.

# Application: Portals



Extraction Engine

e.g. DBLife



Everything about  
Angelina Jolie

Name, Birthdate, Lovers

# Application: Portals



## DBLife

Fabian Suchanek

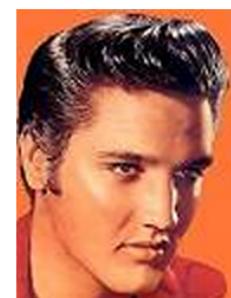
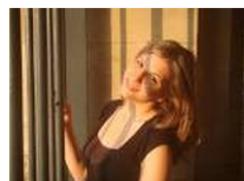
Search

### Fabian M. Suchanek

[Bing](#) [Citeseer](#) [DBLP](#) [Google](#) [Google Scholar](#) [Kosmix](#) [Wikipedia](#) [Yahoo!](#)



from Google Images



[more](#)

### Related People

- [Gerhard Weikum](#)
- [Mauro Sozio](#)
- [Gjergji Kasneci](#)
- [Jian Pei](#)

[more](#)

### Related Topics

- [location](#)
- [ranking](#)
- [privacy](#)
- [query processing](#)

[more](#)

### Related Organizations

- [Microsoft Research](#)
- [Yahoo! Research](#)
- [University of Hong Kong](#)
- [Microsoft](#)

Sorted by Year/Conf, [Year/Citation](#), [Citation](#)

[Community Statistics](#)

2011

31

PIKM 2011: the 4th ACM workshop for Ph.D. students in information and knowledge management. Anisoara Nica, Fabian M. Suchanek. CIKM 2011, 2633-2634. [Web Search](#) [BibTeX](#) [Download](#)

# Application: Shopping



**Original Elvis Presley Kostüm  
Elviskostüm Gr. 48/50 (M), 52/54 (L)**

Verfügbarkeit: sofort ■■■  
versandfertig

Artikel-Nr.: SM29142

Stückzahl:   Gefällt mir 0  

**44,90 EUR**

inklusive 19% Mehrwertsteuer zzgl. Versand

Bitte wählen Sie eine Größe aus:  
Größe bitte wählen... 

**Kaufen** MERKZETTEL WUNSCHLISTE

# Application: Shopping

The screenshot shows a product listing for an "Original Elvis Presley Kostüm". The product image features a person dressed as Elvis Presley in a white jumpsuit with rhinestones, holding a microphone. The product title is "Original Elvis Presley Kostüm" and the subtitle is "Elviskostüm Gr. 48/50 (M), 52/54 (L)". Availability information includes "Verfügbarkeit: sofort versandfertig" and "Artikel-Nr: SM29142". A "Stückzahl" input field is set to 1. Social sharing icons for Facebook, Twitter, and Google+ are present. The price is listed as "44,90 EUR" with a note about VAT and shipping. A dropdown menu for size selection is labeled "bitte wählen...". A large red arrow points from the bottom left towards the "Kaufend" button. Another red arrow points vertically down from the top right towards the "WUNSCHLISTE" button.

Title	Type	Price
Elvis Kostüm	Costume	44.90

# Application: Shopping

pricegrabber  
Smart Shopping Anytime Anywhere

Shop by Department ▾ elvis costumes

Women's Boots | Men's Wallets | Women's Handbags | Women's Dresses | north face jackets

Home > Clothing

## Featured Brand

See Products from:



Filter By :

Price Range: \$1 - \$50

[Clear All](#)

## Elvis Costumes Clothing - 325 Matches

[Quick Compare](#)

[Top 3 Products](#)



Rubie's Costume Co. Adult  
Elvis Costume



Elvis Presley Eagle  
Jumpsuit Rock Gold...

## Category Matches

[Adult Costumes \(116\)](#)

[Kids' Costumes \(127\)](#)

[Adult Costume Accessories \(74\)](#)

[Men's Tees \(3\)](#)

[Women's Headwear \(2\)](#)

[Sunglasses \(1\)](#)

## Price Range

\$1 - \$50

\$  to \$

[Update](#)

As low as  
**\$47.99**

[SEE IT](#)

As low as  
**\$24.00**

[SEE IT](#)

Price Grabber

# Application: Search

Google Search Volume by Language

This is a  
Chrome  
Experiment

The WebGL Globe is an open platform for visualizing geographic data. · Get the code · By the Google Data Arts Team

Google Globe

# Application: Search



Google built its “knowledge graph”- a collection of factual knowledge

# Application: Search

Google when was elvis born? 1 + Share

Web Images Maps Shopping Videos More ▾ Search tools SafeSearch ▾

About 33,500,000 results (0.23 seconds)

**January 8, 1935**

Elvis Presley, Date of birth

[Elvis Presley - Wikipedia, the free encyclopedia](#)  
[en.wikipedia.org/wiki/Elvis\\_Presley](http://en.wikipedia.org/wiki/Elvis_Presley)

Elvis Presley was born on **January 8, 1935**, in Tupelo, Mississippi, to 18-year-old Vernon Elvis and 22-year-old Gladys Love Presley, in the two-room shotgun ...

[Elvis Presley discography - Personal relationships of Elvis - Elvis Presley \(album\)](#)

**Elvis Presley**

Elvis Aaron Presley was an American singer and actor. A cultural icon, he is commonly known by the single name Elvis. One of the most popular musicians of the 20th century, he is often referred to as the "King of Rock and Roll" or "the King". [Wikipedia](#)

**Born:** January 8, 1935, [Tupelo](#)

**Died:** August 16, 1977, [Memphis](#)

**Height:** 1.83 m

**Albums:** [Elvis Presley](#), [Elvis' Christmas Album](#), [Christmas Duets](#), [More](#)

**Songs**

<a href="#">Jailhouse Rock</a>	1957	Billboard Top Ro...
<a href="#">Suspicious Minds</a>	1969	From Elvis in Me...
<a href="#">Always on My Mind</a>	1973	Separate Ways

Factual information  
helps Google  
answer user queries.

# Knowledge Graph

Person	Profession
Elvis	Singer
Bob	Painter



The knowledge graph is based on

- IE from Wikipedia
- IE from the Web (“Knowledge Vault”)

[See Wikipedia](#)

# Appl.: Question Answering

- When is Madonna's next concert in Europe?
- Which protein inhibits atherosclerosis?
- Who was king of England when Napoleon I was Emperor of France?
- Has any scientist ever won the Nobel Prize in Literature?
- Which countries have a HDI comparable to Sweden's?
- Which scientific papers have led to patents?

# Question Answering

Questions of this kind can be answered with structured information.

Person	Occupation	Awards
Russell	Mathematician	NP Literature
Bohr	Physicist	NP Physics
Elvis	Singer, Hero	Grammy

# Question Answering

Structured information can be extracted from digital documents.

WIKIPEDIA  
The Free Encyclopedia

Create account Log in

Article Talk Read Edit View history Search

## Bertrand Russell

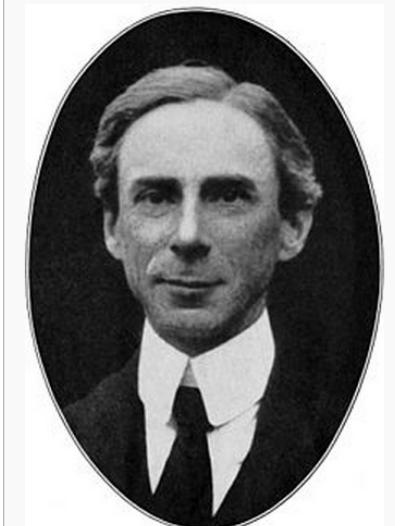
From Wikipedia, the free encyclopedia

**Bertrand Arthur William Russell, 3rd Earl Russell, OM, FRS**<sup>[1]</sup> (18 May 1872 – 2 February 1970) was a British philosopher, logician, mathematician, historian, and social critic.<sup>[2]</sup> At various points in his life he considered himself a liberal, a socialist, and a pacifist, but he also admitted that he had never been any of these in any profound sense.<sup>[3]</sup> He was born in Monmouthshire, into one of the most prominent aristocratic families in Britain.<sup>[4]</sup>

Russell led the British "revolt against idealism" in the early 20th century.<sup>[5]</sup> He is considered one of the founders of analytic philosophy along with his predecessor Gottlob Frege and his protégé Ludwig Wittgenstein. He is widely held to be one of the 20th century's premier logicians.<sup>[2]</sup> He co-authored, with A. N. Whitehead, *Principia Mathematica*, an attempt to ground mathematics on logic. His philosophical essay "On Denoting" has been considered a "paradigm of philosophy."<sup>[6]</sup> His work has had a considerable influence on logic, mathematics, set theory, linguistics, computer science (see type theory and type system), and philosophy, especially philosophy of language, epistemology, and metaphysics.

Russell was a prominent anti-war activist; he championed anti-imperialism<sup>[7][8]</sup> and went to prison for his pacifism during World War I.<sup>[9]</sup> Later, he campaigned against Adolf Hitler, then criticised Stalinist totalitarianism, attacked the United States of America's involvement in the Vietnam War, and was an outspoken proponent of nuclear disarmament.<sup>[10]</sup> In 1950 Russell was awarded the Nobel Prize in Literature, "in recognition of his varied and significant writings in which he champions humanitarian ideals and freedom of thought."<sup>[11]</sup>

**Bertrand Russell**



**Born** 18 May 1872  
Trellech, Monmouthshire, UK

**Died** 2 February 1970 (aged 97)  
Penrhyndeudraeth, Wales, UK

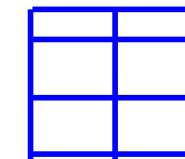
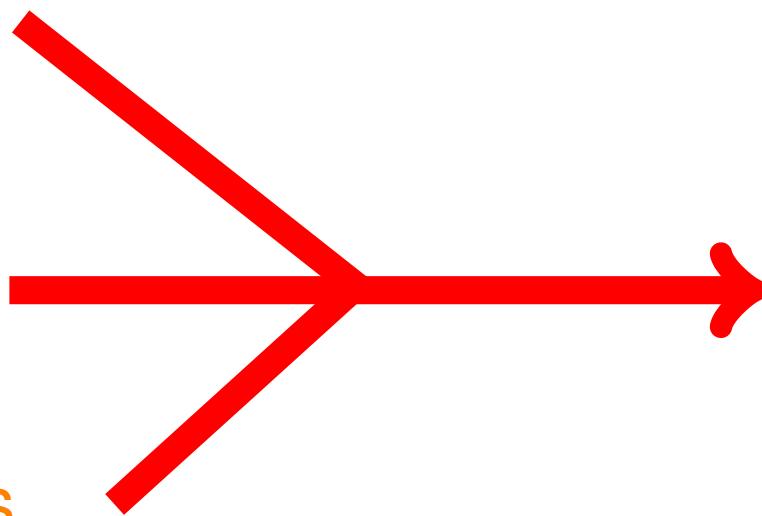
# YAGO

The YAGO project extracts information from Wikipedia and other sources.



WordNet

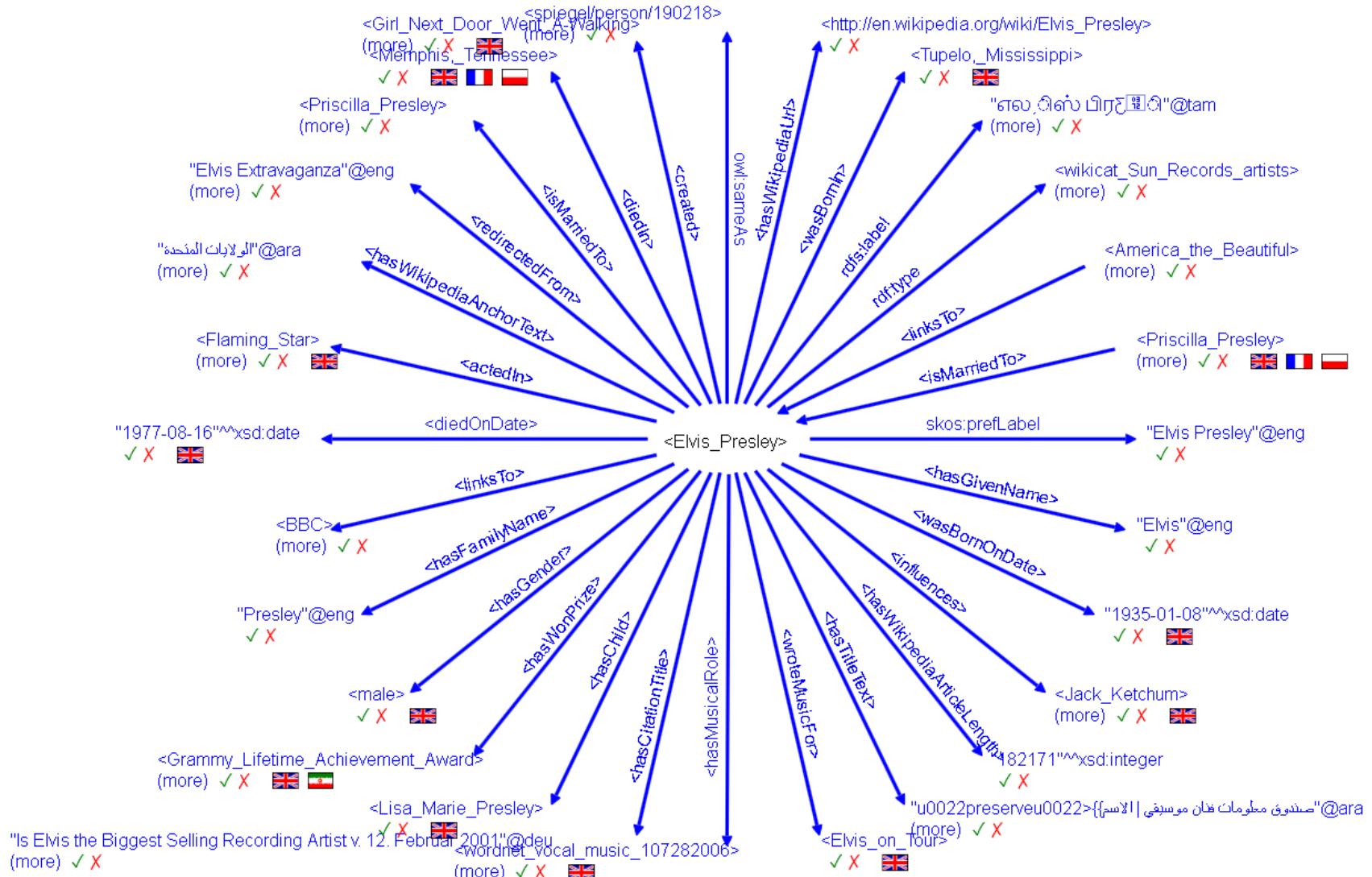
GeoNames



YAGO

# YAGO

YAGO is a large knowledge graph built from Wikipedia.



# DEANNA

Which mathematician won the Nobel Prize in Literature?

[Show Sample Questions](#) • [Show Advanced Options](#) • Time: 19006 ms

## Structured Query

```
?x type wordnet_mathematician_110301261 ;  
?x hasWonPrize Nobel_Prize_in_Literature
```

Try it out 



```
?x type wordnet_mathematician_110301261 ; ?x hasWonPrize Nobel_Prize_in_Literature
```

Submit

# IBM WATSON

IBM Watson is a question answering system based on many sources (i.a. YAGO).



[Watson](#)

[New York Times article](#)

It outperformed the 74-fold human winner of the Jeopardy! quizz show.

# Overview

## Motivation

## Knowledge Representation

## Instance Extraction

## Fact Extraction

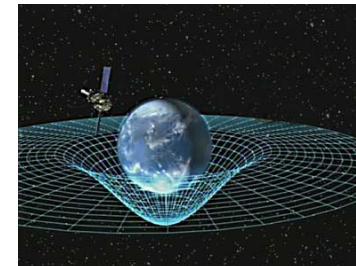
## The Semantic Web

## Linked Data

## RDFa

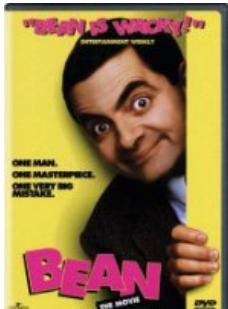
# Entity

An entity (also: instance, resource)  
is any particular object of the world  
or imagination, be it abstract or concrete.



(The exact definition of 'instance' is a philosophical conundrum. Don't worry about it.)

# Digression: Entities

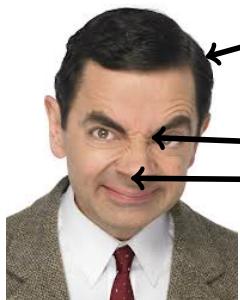


Is the movie “Mr. Bean” an entity?  
Or is every DVD of the movie an entity?

# Digression: Entities



Is the movie “Mr. Bean” an entity?  
Or is every DVD of the movie an entity?



Is this an entity?

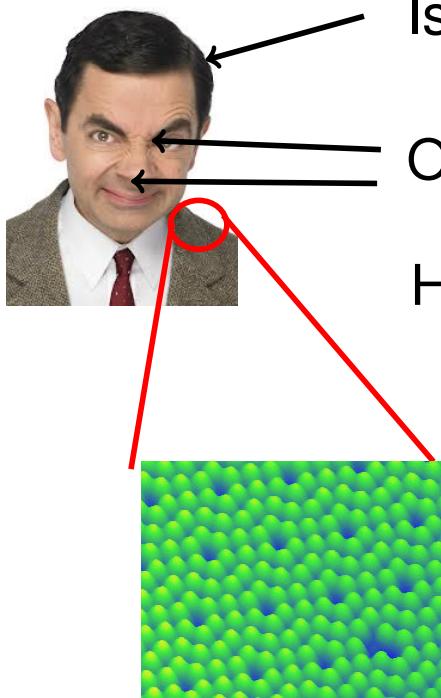
Or this?

How many entities are there?

# Digression: Entities



Is the movie “Mr. Bean” an entity?  
Or is every DVD of the movie an entity?



Is this an entity?

Or this?

How many entities are there?

Isn’t everything just atoms?

# Digression: Identity



Over time, all parts of a ship are replaced at some point of time.  
Then, is it still the same ship?

[see: Theseus's ship on Wikipedia](#)

# Digression: Identity



Over time, all parts of a ship are replaced at some point of time.  
Then, is it still the same ship?

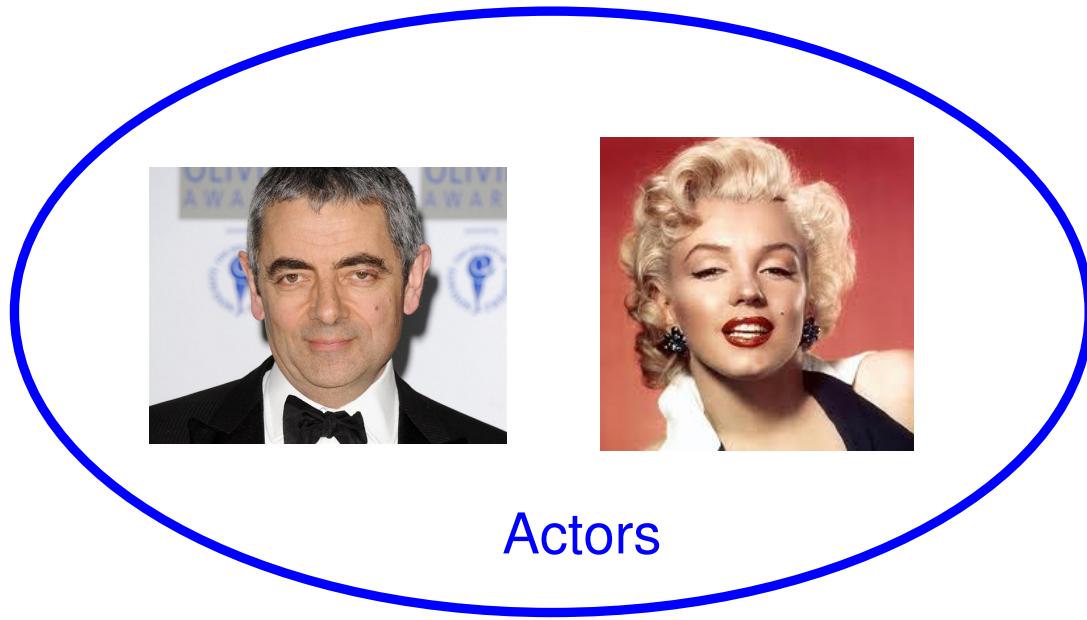
[see: Theseus's ship on Wikipedia](#)

Humans replace their cells every 7 years.

[New York Times](#)

# Class

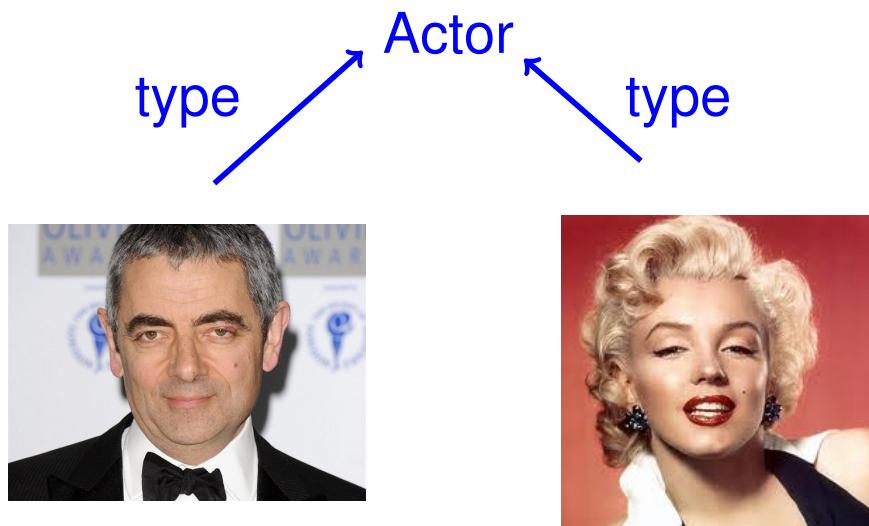
A class (also: concept) is a set of similar entities.



Cars  
Cities  
Rivers  
Universities  
Theories  
...

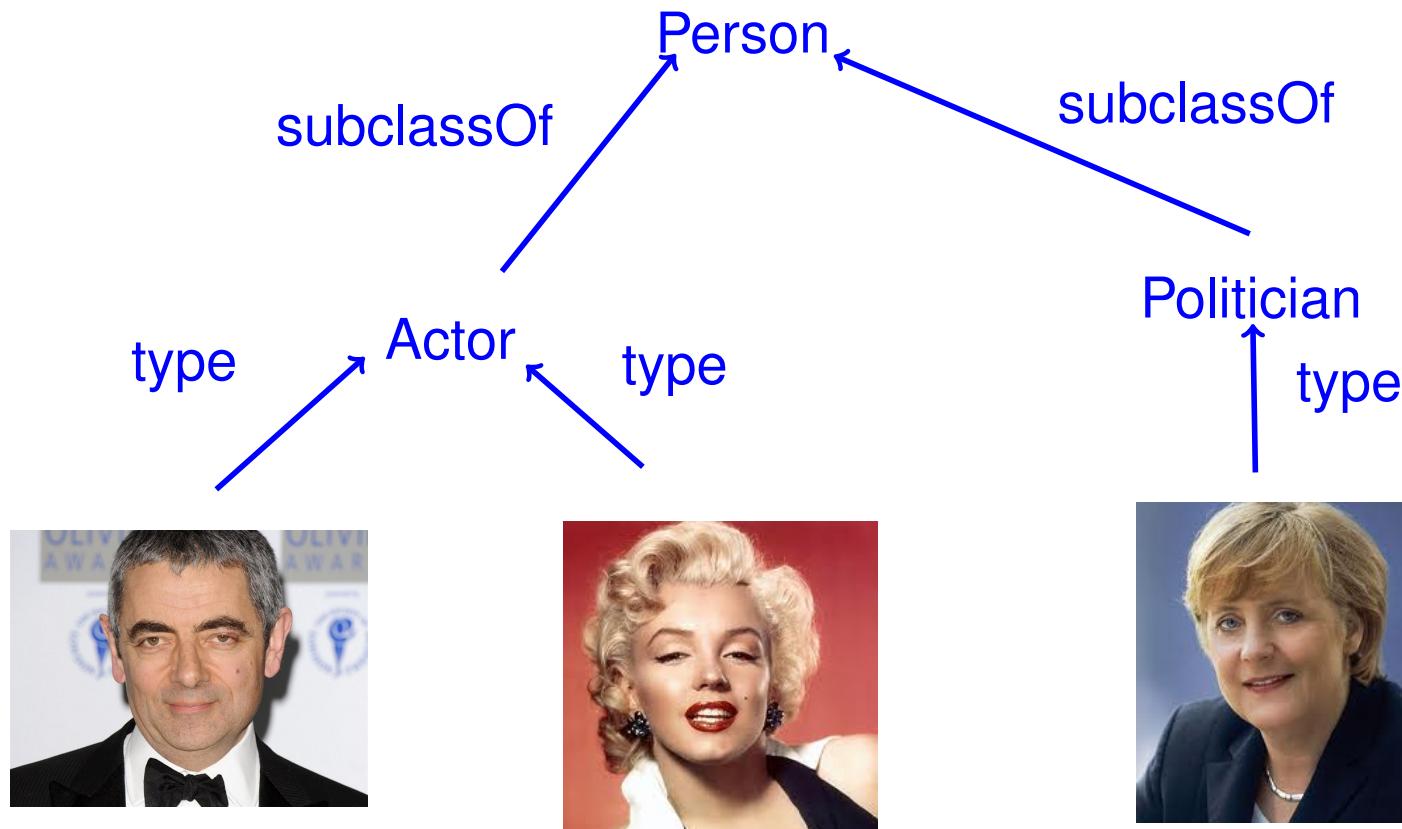
# Instance of a class

An entity is an instance of a class  
(also: belongs to a class, has the type),  
if the entity is an element of that class.



# Subclass

A class is a subclass of another class, if all instances of the first class are also instances of the second class.



# Cheat Sheet

If we can say...

then...

- “a/an X”, “every X”      X is a class
- “Xs” (plural)                X is a class
- “This is X”                 X is an entity
- “X is a Y”                 X is an inst. of Y
- “Every X is a Y”          X is a subclass of Y

# Cheat Sheet

If we can say...                           then...

- “a/an X”, “every X”               X is a class
- “Xs” (plural)                       X is a class
- “This is X”                       X is an entity
- “X is a Y”                       X is an inst. of Y
- “Every X is a Y”               X is a subclass of Y

Try it out: car, Mr. Bean, Coli bacteria, Ford, time

# Task: Subclasses

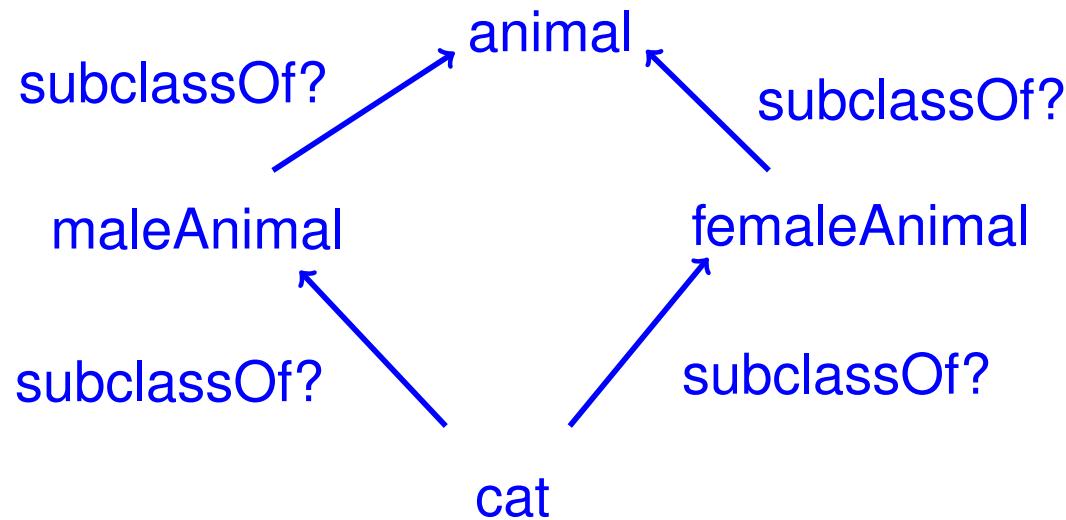
Draw a picture of instances and subclasses in the animal kingdom.

How do you deal with “male” and “female”?

# Task: Subclasses

Draw a picture of instances and subclasses in the animal kingdom.

How do you deal with “male” and “female”?



# Intuition: Relations

A relation is like a table.

Relation “born”:

Person	City	Year
Atkinson	Consett	1955
Monroe	Los Angeles	1926
...		

# Def:Relation

A relation (also: predicate) over classes  
is a subset of their cartesian product.

The classes are called the domains of the relation.

The number of classes is called the arity of the relation.

$$R \subseteq C_1 \times C_2 \times \dots \times C_n$$

$$\text{born} \subseteq \text{person} \times \text{city} \times \text{year}$$

$$\begin{aligned}\text{born} = & \{ < \text{Atkinson}, \text{Consett}, 1955 >, \\ & < \text{Monroe}, \text{LosAngeles}, 1926 >, \dots \}\end{aligned}$$

# Digression: Semantics

A relation is **any** subset of the cartesian product. It does not have to correspond to a real-world relation. Its name is arbitrary.

*born* = {< Atkinson, Consett, 1955 >,  
< Monroe, LosAngeles, 1926 >, ...}

# Digression: Semantics

The semantics/denotation of a symbol is  
the “meaning” of the symbol,  
i.e., the real world entity or tuples.

$$D(\text{Atkinson}) = \text{Rowan Atkinson}$$

---

Syntax

Semantics

$$D(\text{born}) = \{ \langle x, y, z \rangle \mid x \text{ was born in } y \quad \text{in year } z \}$$

# Digression: Semantics

But these are again symbols.

$$D(\text{born}) = \{ \langle x, y, z \rangle \mid x \text{ was born in } y \quad \text{in year } z \}$$



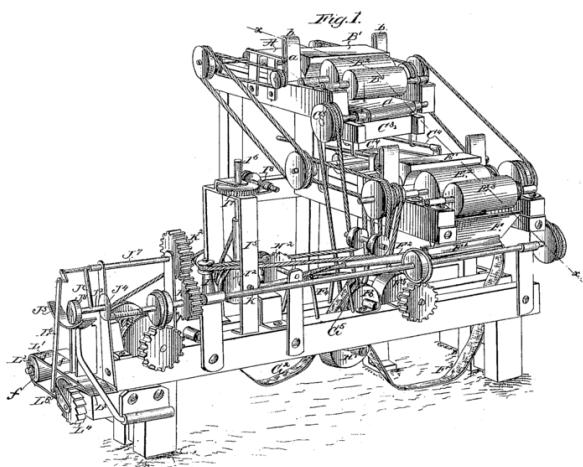
What is their semantics?

# Digression: Semantics

*D(Atkinson was born in Consett)*

# Digression: Semantics

*D(Atkinson was born in Consett)*

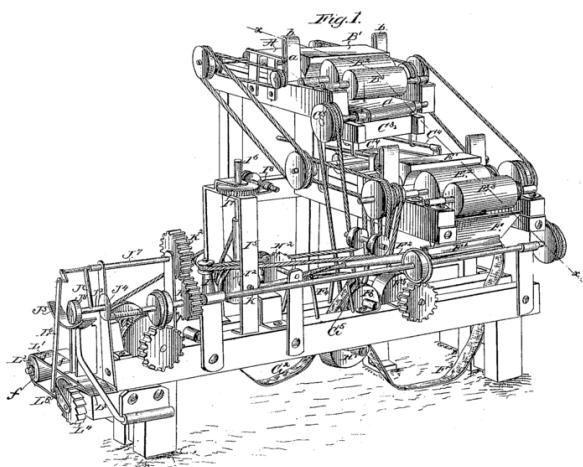


Linguistic parsing  
+ Montague semantics

=  $(\lambda xy \langle x, y \rangle \in \text{born}) \text{Atkinson} \text{Consett}$

# Digression: Semantics

*D(Atkinson was born in Consett)*

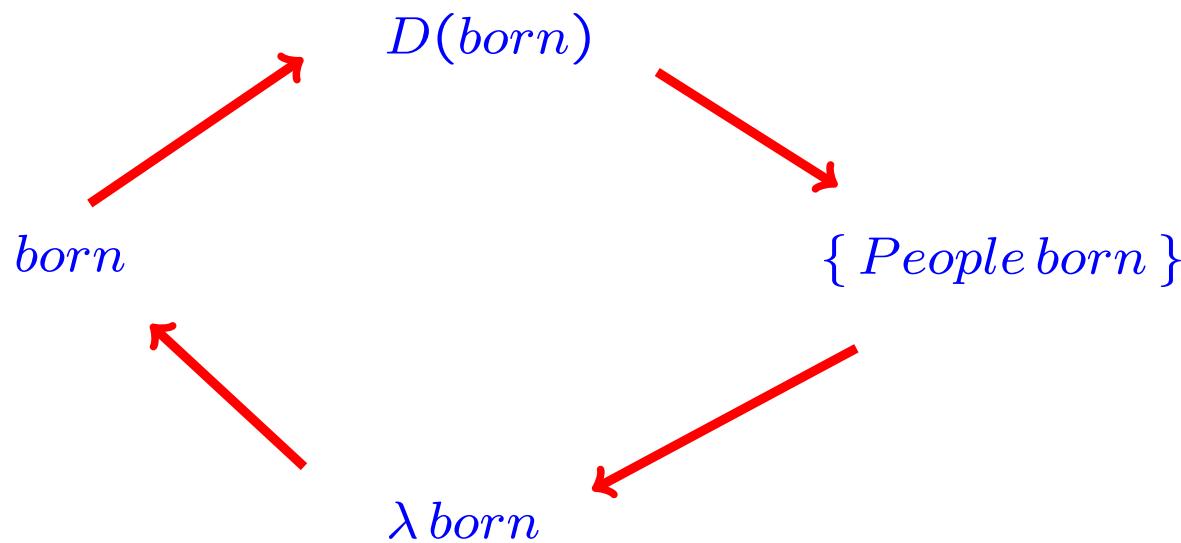


Linguistic parsing  
+ Montague semantics

=  $(\lambda xy \langle x, y \rangle \in \text{born}) \text{ Atkinson Consett}$

=  $\langle \text{Atkinson}, \text{Consett} \rangle \in \text{born}$

# Digression: Semantics



One man's semantics  
is another man's syntax.  
All we do is manipulating symbols.  
There is no connection to the real world.



# Def: Binary Relation

A binary relation is a relation of arity 2.

$$\text{bornInCity} \subseteq \text{person} \times \text{city}$$

For binary relations, the first class is called the **domain** and the second class is called the **range**.

We usually visualize a binary relation by an arrow:

$$\text{bornInCity}(\text{Atkinson}, \text{Consett})$$



# Def:Inverse

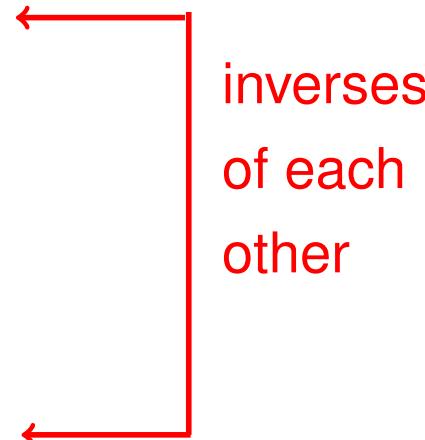
The inverse of a binary relation  $r$  is a relation  $r'$ , such that  $r'(x,y)$  iff  $r(y,x)$ .

$drives \subseteq person \times car$

$drives(MrBean, Mini1000)$

$hasDriver \subseteq car \times person$

$hasDriver(Mini1000, MrBean)$



# Def:Function

A function (also: functional relation) is a binary relation that has at most 1 object for each subject.

$$r \text{ functional} \equiv \forall x : |\{y : r(x, y)\}| \leq 1$$

hasNationality

hasSpouse (in medieval ages)

hasNumberOfTeeth

# Def:Inverse Functional Rel.

An inverse functional relation is a relation whose inverse is functional.

$$r \text{ inv. functional} \equiv \forall y : |\{x : r(x, y)\}| \leq 1$$

hasCitizen

hasSpouse (in medieval ages)

# Digression: Equality

If two entities share the same object of an inverse functional relation, they are equal.

*hasEmail(Bean, me@ bean.com)*

*hasEmail(MrBean, me@ bean.com)*

$\Rightarrow MrBean = Bean$

*born(Bean, 1955)*

*born(MrBean, 1955)*

(Nothing follows)

# Classes as binary relations

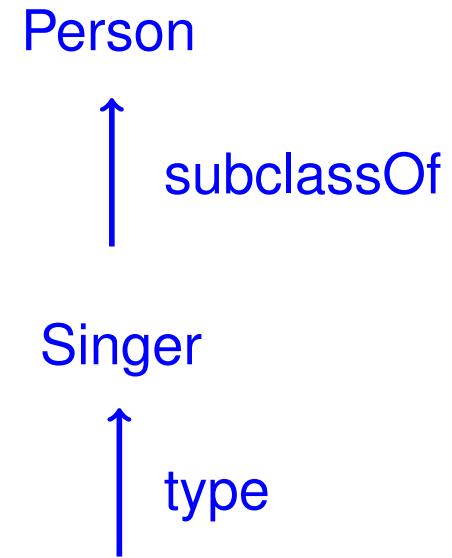
One way to represent a class is by the binary relations  $type$ ,  $subclassOf$

$$type \subseteq entity \times class$$

$$type(Atkinson, singer)$$

$$subclassOf \subseteq class \times class$$

$$subclassOf(singer, person)$$



(Watch out: here, “entity” is the class of entities, “class” is the set of class names (!))

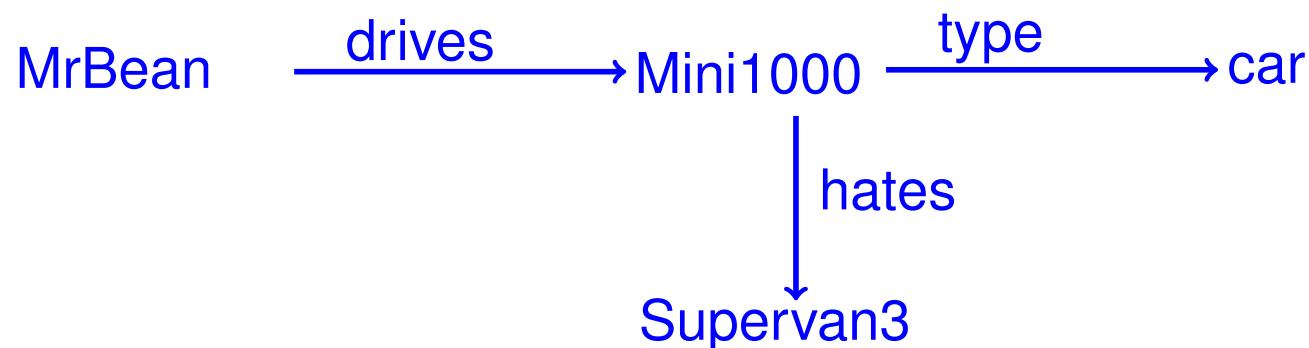
# Def:Knowledge Graph

A knowledge graph (also: Entity-Relationship graph, Knowledge base, KB) is a directed labeled multi-graph that has an edge  $x \rightarrow y$  with label  $r$ , iff  $r(x,y)$ .

$drives(MrBean, Mini1000)$

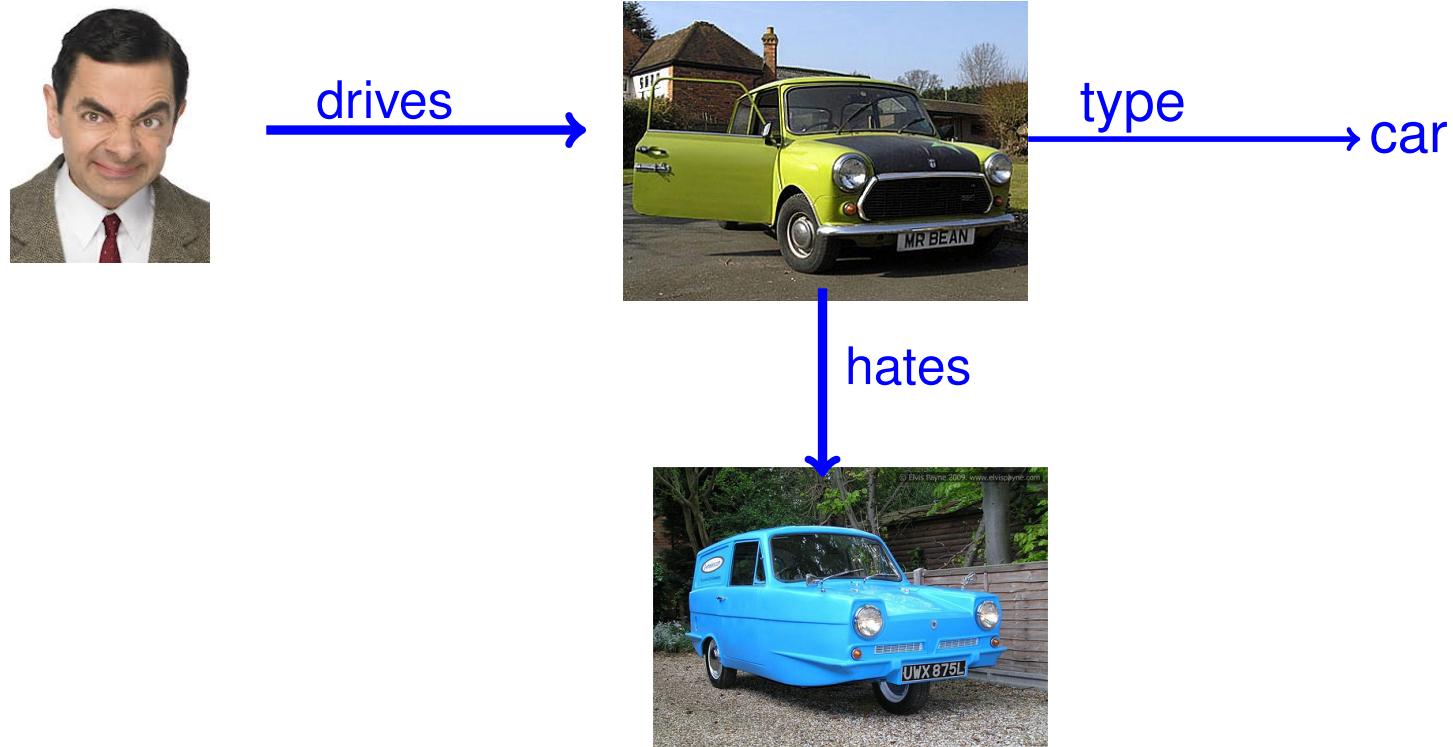
$hates(Mini1000, Supervan3)$

$type(Mini1000, car)$



# Def:Knowledge Graph

A knowledge graph (also: Entity-Relationship graph, Knowledge base, KB) is a directed labeled multi-graph that has an edge  $x \rightarrow y$  with label  $r$ , iff  $r(x,y)$ .

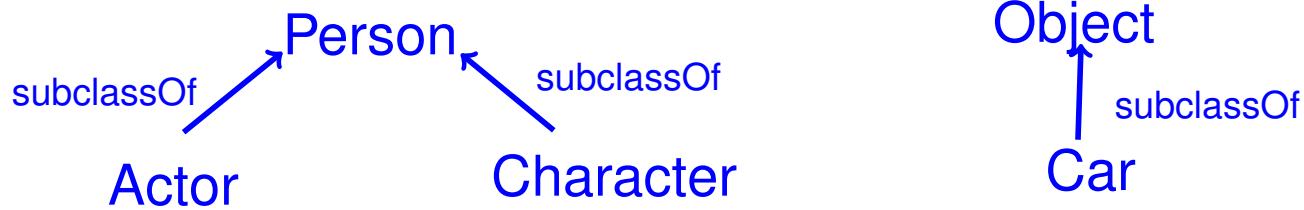


# Schema

A schema is

- a set of classes with their subclassOf-links
  - a set of relations with their domains
- ... that are necessary to model a domain of interest.

Classes:



Relations:

$$drives \subseteq person \times car$$

$$plays \subseteq actor \times character$$

# Task: Schema

Define a schema for the domain of politics.

In that schema, express that Hollande is the president of France, Merkel is the chancellor of Germany, and that Merkel loves Hollande.

# Def: Triple Store

A triple store is a table that contains a KB of binary relations in the form of 3 columns: subject, relation, object.

Subject	Relation	Object
MrBean	drives	Mini1000
Mini1000	hates	Supervan3

(The middle column is often called “Predicate”)

# Binary Rel. Representations

Binary relations can be represented as

- relation  $\text{drives} \subseteq \text{person} \times \text{car}$

$\text{drives}(\text{MrBean}, \text{Mini1000})$

- table

<u>drives</u>	Person	Car
	MrBean	Mini1000

- graph

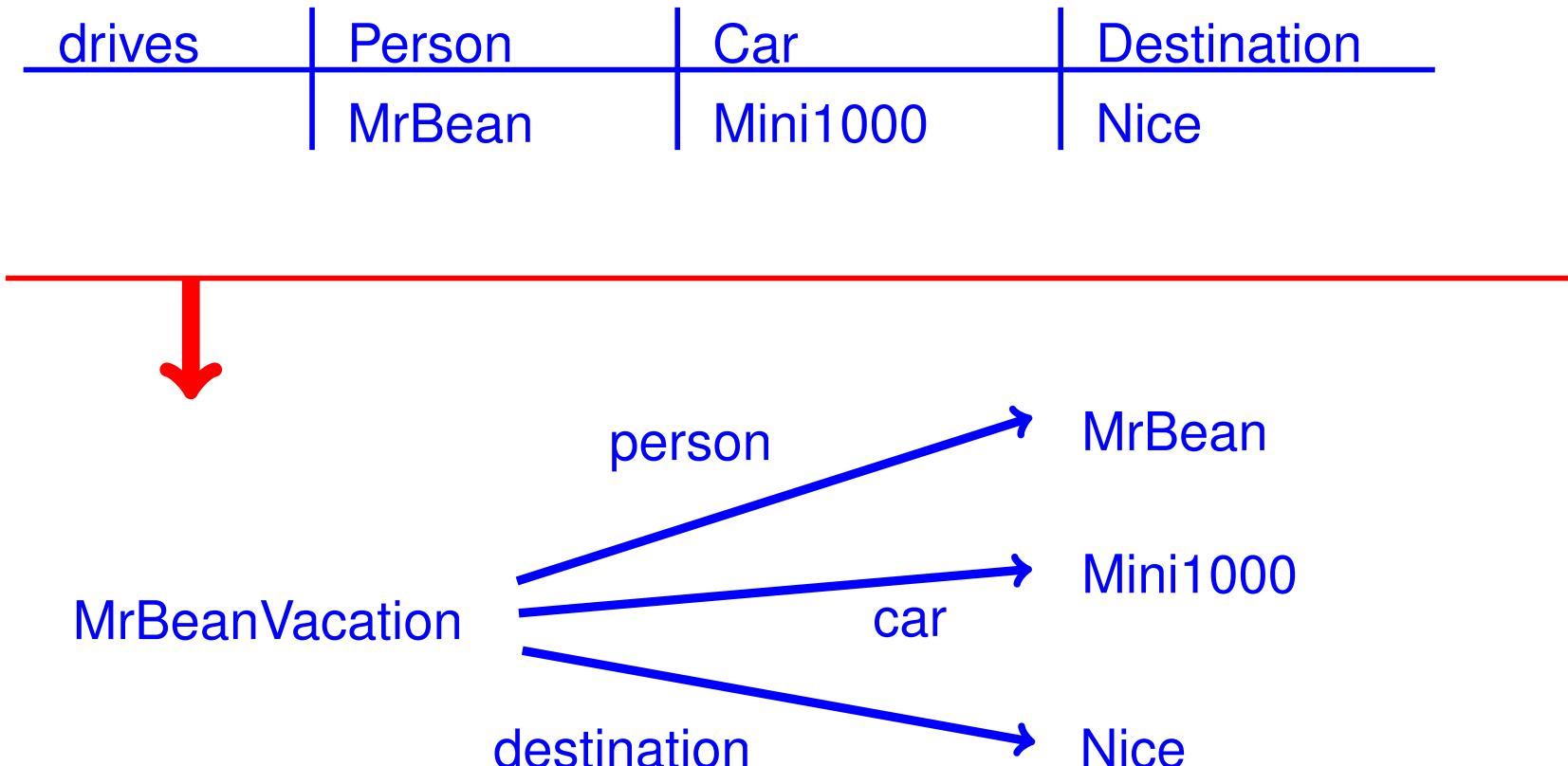


- triple store

<u>Subject</u>	<u>Relation</u>	<u>Object</u>
MrBean	drives	Mini1000

# n-ary facts as binary facts

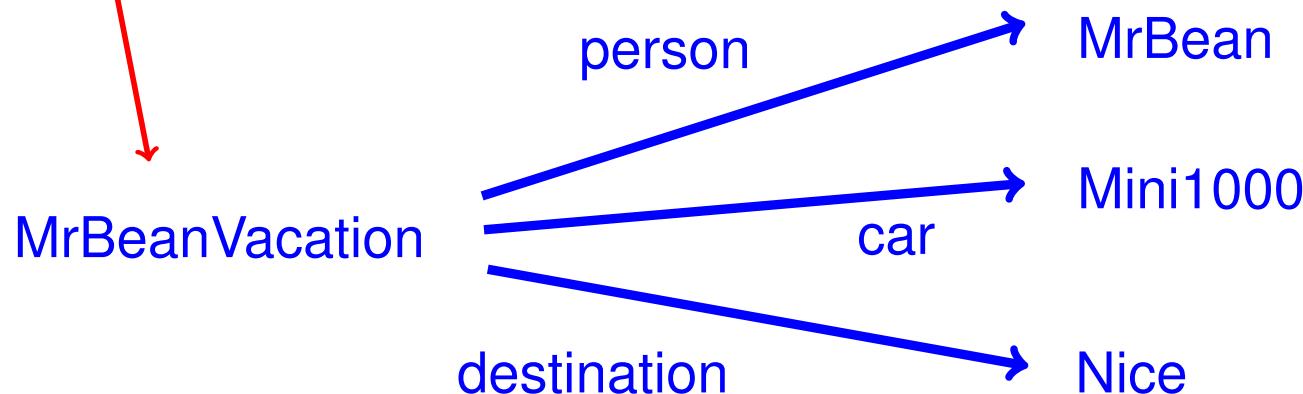
Every n-ary fact can be represented as binary facts.



# Def: Event Entity

An event entity represents an n-ary fact.

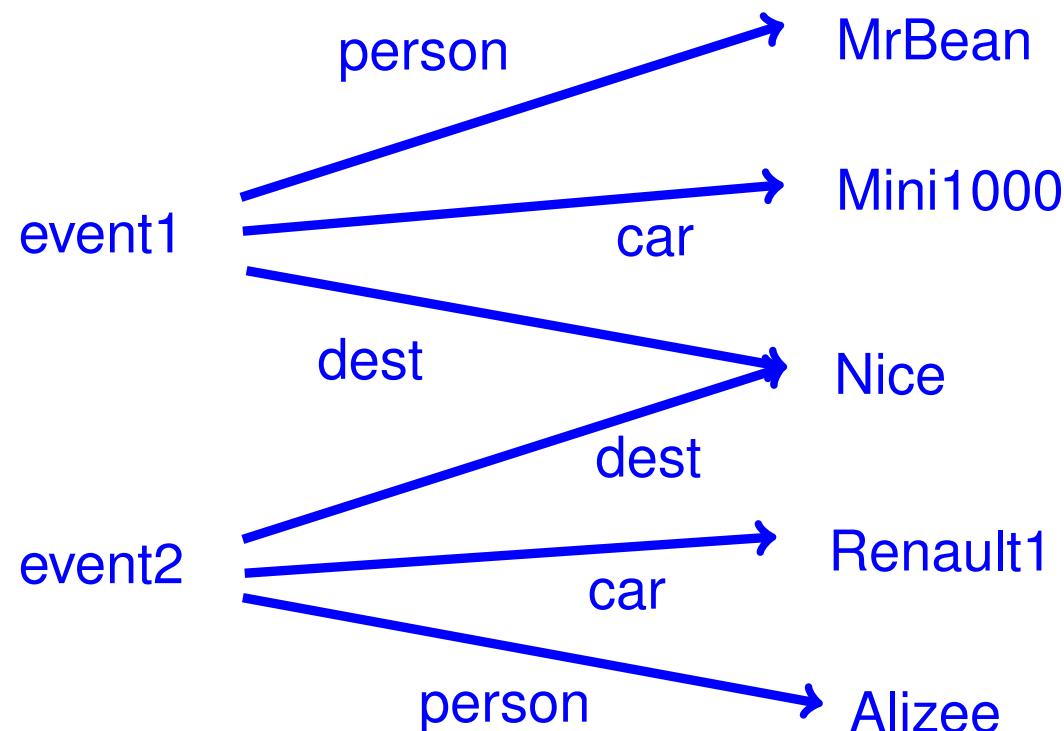
Event entity



# Event Entities

drives	Person	Car	Destination
	MrBean	Mini1000	Nice
	Alizee	Renault1	Nice

T  
Every row  
becomes a  
new event  
entity with  
n outgoing  
edges.



# Task: Event Entities

Draw a knowledge graph for the following facts.

Irma loves Mr. Bean since 1955.

Mr. Bean drives with Irma to the cinema.

Irma and Mr. Bean watch “Titanic”.

The movie is about the trip of the ship  
“Titanic” from Europe to New York.

# Binary relations are flexible

n-ary relations enforce the presence of all arguments:

born	Person	City	Year
	Atkinson	Consett	1955

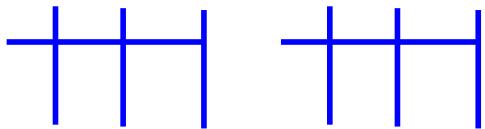
Binary relations don't:



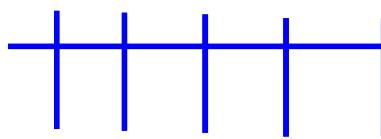
1955

# Binary vs n-ary

Binary and n-ary relations can represent the same facts.



binary



n-ary

- more relations
- less arity
- more flexibility

- less relations
- more arity
- more control

# Def:Name

A name (also: label) of an entity is a human-readable string attached to that entity.



The entity  
is called the  
meaning of  
the name.

# Label

**label** is a binary relation that holds between an entity and its name.

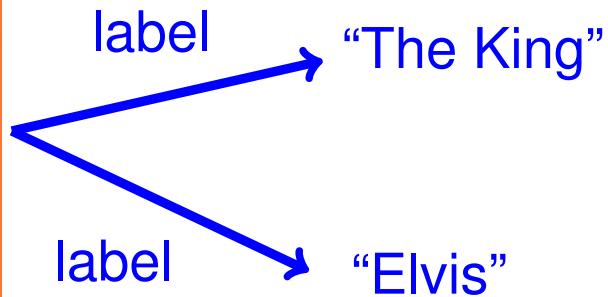
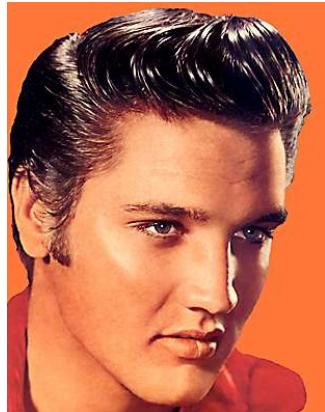


= *label(MrBean, "Mr. Bean")*

# Def:Synonymy

If an entity has multiple names,  
the names are called synonymous.

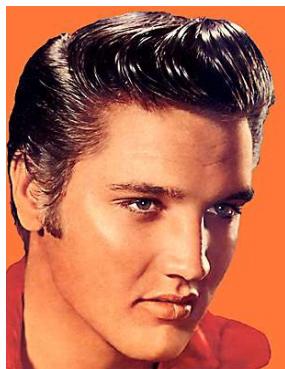
(The adjective for the names is “synonymous”, each name is a “synonym”, the phenomenon is called “synonymy”)



# Def:Ambiguity

If a name is attached to multiple entities,  
the name is called ambiguous.

(The adjective for the names is “ambiguous”, the phenomenon is called “ambiguity”)



label

“The King”

label



# References

The Knowledge Representation model  
we saw today is RDF without URLs:

[RDF Primer](#)

For Information Extraction, see

[Sunita Sarawagi: Information Extraction](#)

[Serge Abiteboul et al: Web data management](#)

[Staab and Studer: Ontology Handbook](#)

[W3C: The Semantic Web](#)