

Information Extraction: Motivation

Nada Mimouni

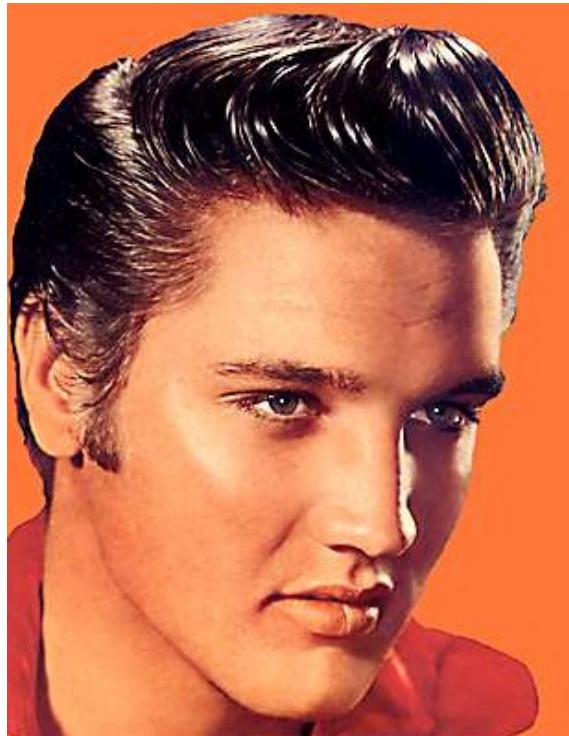
Lecturer: Nada Mimouni



Slides by: Fabian Suchanek
Professor at Télécom ParisTech
DBWeb team, office C201-6
<https://suchanek.name>

nada.mimouni@telecom-paristech.fr

Elvis Presley



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

Is Elvis still alive?

Elvis Alive: Google

is elvis alive 

All Images Videos News Maps More Settings Tools

About 17,700,000 results (0.46 seconds)

Elvis Presley
American singer



Date of death: August 16, 1977
Place of death: Graceland, Memphis, Tennessee, United States
Cause of death: Heart attack

[Google](#)

Google: no

Elvis Alive: Bing

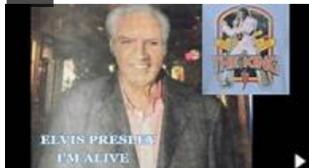
 is elvis alive 

All Images Videos Maps News | My saves

20 700 000 Results Date ▾ Language ▾ Region ▾

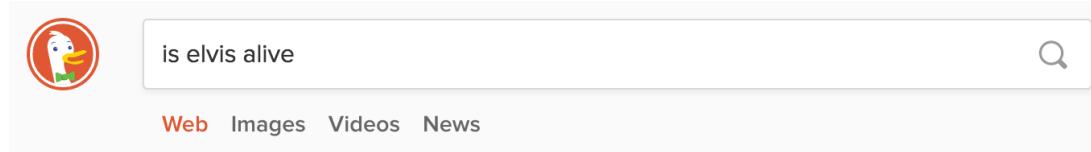
ELVIS PRESLEY - I'M ALIVE - YouTube

5:34 //www.youtube.com/watch?v=YBQlWhYA0lw ▾

 05/01/2014 · The TRUTH Now Declassified. **ELVIS PRESLEY** is **Alive**, and Has made His Comeback. This Information Is Being Kept From The Public Eye , The News Media , And ...
Author: Knights Of The King...
Views: 3,6M

Bing: yes

Elvis Alive: DuckDuckGo



The screenshot shows the DuckDuckGo search interface. At the top left is the DuckDuckGo logo (a red duck head inside a white circle). To its right is a search bar containing the query "is elvis alive". To the right of the search bar is a magnifying glass icon. Below the search bar are four navigation links: "Web" (which is underlined in red), "Images", "Videos", and "News". A horizontal line separates this from the search results. Below the line are three filter options: a toggle switch set to "France", a dropdown menu set to "Safe Search: Strict", and another dropdown menu set to "Any Time".

Evidence Elvis Presley is Alive - Home | Facebook

Evidence Elvis Presley is Alive. 26K likes. WELCOME ELVIS FANS! This page provides evidence that Elvis Presley never left the building!!

 <https://www.facebook.com/evidenceelvispresleyisalive>

DuckDuckGo: 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.* »

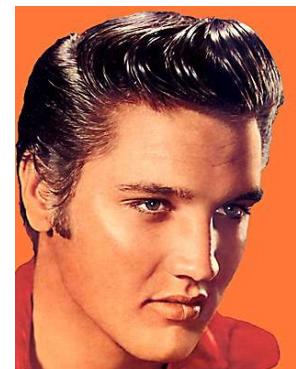


Elvis Presley, the first
and greatest American
rock-and-roll star,
died yesterday at
the age of 42.

Text understanding

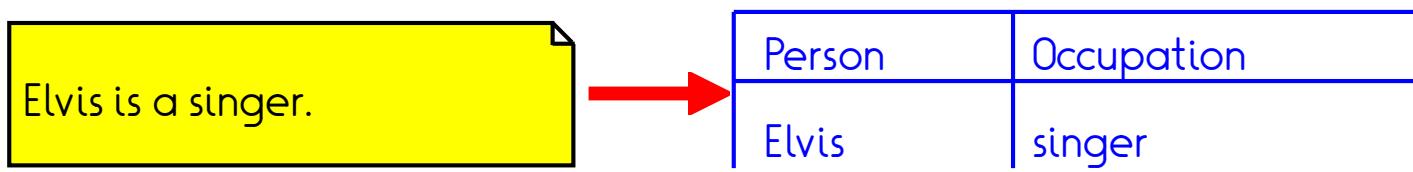
To answer the question, Wolfram Alpha has to understand this text.

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

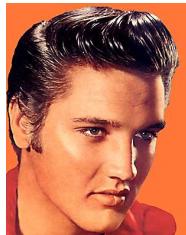


Information Extraction

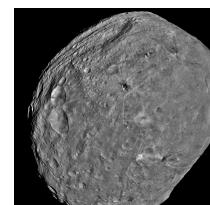
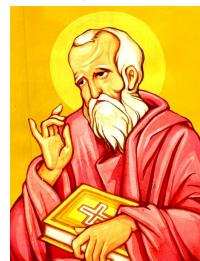
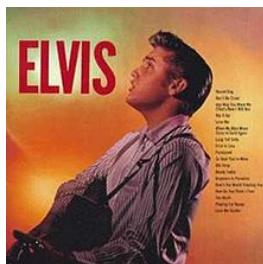
Information Extraction (IE) is the process of deriving structured factual information from digital text documents.



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??

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: 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 06

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 06

>difficulties

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 06



No	Street	Country	ZIP	City
46	Barrault	France	75634	Paris

Why would anybody want to do Information Extraction?

The reason is that...

this is abundant...

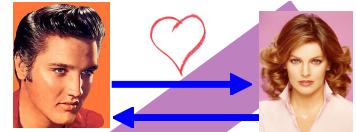
...and this is useful.

A red arrow points from the yellow box towards a blue-bordered table. The table has two columns: "Person" and "Occupation".

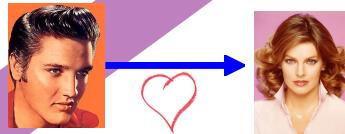
Person	Occupation
Elvis	singer

Semantic IE

IE by Reasoning



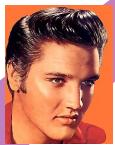
Fact Extraction



Instance Extraction

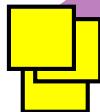


Entity Disambiguation



Entity Recognition

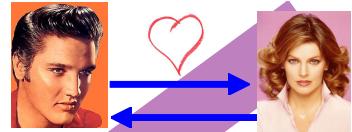
singer **Elvis**



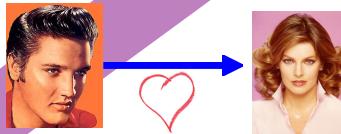
Preparation

Semantic IE

IE by Reasoning



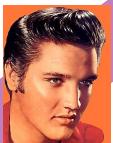
Fact Extraction



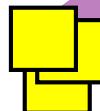
Instance Extraction



Entity Disambiguation



singer **Elvis**

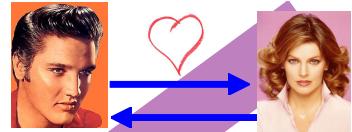


Preparation

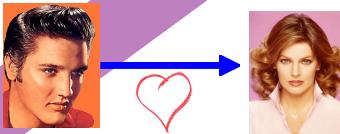
How do we represent world knowledge?
Where do we find the documents?

Semantic IE

IE by Reasoning



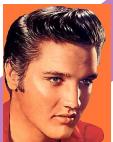
Fact Extraction



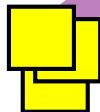
Instance Extraction



Entity Disambiguation



singer **Elvis**



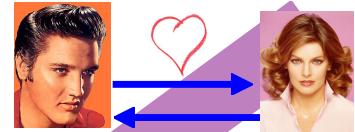
Preparation

Named Entity Recognition (Tries)

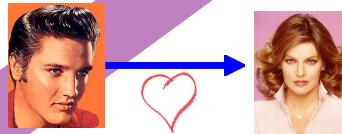
Named Entity Annotation (Rule-based
or by machine learning)

Semantic IE

IE by Reasoning



Fact Extraction



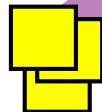
Instance Extraction



Entity Disambiguation



singer **Elvis**

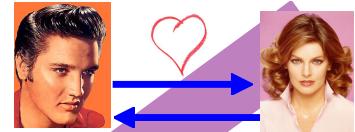


Preparation

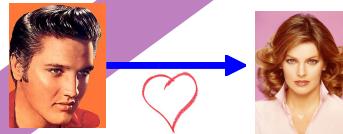
Finding which entity a text refers to,
for example with AIDA

Semantic IE

IE by Reasoning



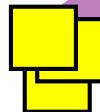
Fact Extraction



Instance Extraction



singer [Elvis]

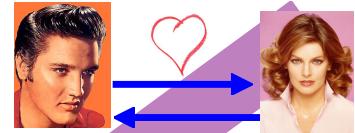


Preparation

Extracting which class an entity belongs to (from the Web and from text documents), for example as in NELL

Semantic IE

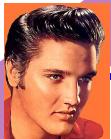
IE by Reasoning



Fact Extraction

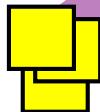


Instance Extraction



singer

singer Elvis

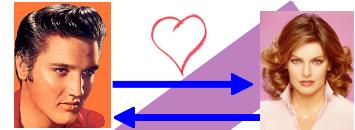


Preparation

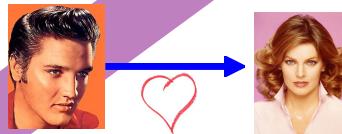
DIPRE Algorithm, POS tagging,
Dependency Parsing.
Example: [YAGO](#)

Semantic IE

IE by Reasoning



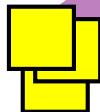
Fact Extraction



Instance Extraction



singer Elvis



Preparation

Taking into account
background knowledge.
Rule mining in KBs.

References

Sunita Sarawagi: Information Extraction

Serge Abiteboul et al: Web data management

Staab and Studer: Ontology Handbook

W3C: The Semantic Web

->applications-of-nlp

->knowledge-representation

->introduction-nlp