

# Information extraction & automated knowledge graph construction

Lecture 1: Overview

---

Lauriane Aufrant (Inria Paris)

Université Paris Cité – UFRL – M2 LI – 09/11/2023

## Some logistics

- ▶ Thursday, 09:00-11:00
- ▶ 6 lectures: 09/11  $\rightsquigarrow$  14/12
- ▶ Course material will be uploaded each week at <https://who.paris.inria.fr/Lauriane.Aufrant/>
- ▶ Questions, issues: `first.last@inria.fr`

## What you will learn

- ▶ 09/11 – Overview of information extraction
- ▶ 16/11 – Entities, relations...
- ▶ 23/11 – Coreference, linking...
- ▶ 30/11 – From IE to automated knowledge graph construction
- ▶ 07/12 – IE annotations
- ▶ 14/12 – IE in a specialty domain

**Evaluation:** group project (December-January)

# Today: Overview of information extraction

- ▶ Information extraction: goal & applications
- ▶ Knowledge graphs
- ▶ Main IE tasks

# Information extraction

## The goal of information extraction

*Convert unstructured text into a structured format, subject to the user's needs*

## Text in

**Brazil** ranks number 5 in the list of countries by population.

The term "**Ibu Negara**" (Lady/Mother of the State) is used for **wife of the President of Indonesia**.

**Game of Thrones** is an adaptation of A Song of Ice and Fire, George R. R. Martin's series of fantasy novels. It ranks **fourth** among the **IMDB Top Rated TV Shows**

## Data out

### THE COUNTRIES WITH THE LARGEST POPULATION

China	1	1,388,232,693
India	2	1,342,512,706
Unites States	3	326,474,013
Indonesia	4	263,510,146
<b>Brasil</b>	5	174,315,386

### THE COUNTRY'S' FIRST LADIES

Brigitte Macron

- Spouse: Emmanuel Macron, President of France (2017 - )  
Melania Trump

- Spouse: Donald J. Trump, U.S. President (2017- )

**Iriana Widodo**

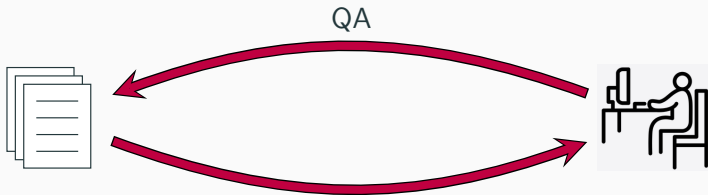
- Spouse: Joko Widodo, **President of Indonesia** (2014 - )

- Also known as: **"Ibu Negara"** (Lady/Mother of the State)

### IMDB TOP RATED TV SHOWS

- 1 Planet Earth II (2016) 9.6.
- 2 Band of Brothers (2001) 9.5.
- 3 Planet Earth (2006) 9.5.
- 4 **Game of Thrones** (2011) 9.4.
- 5 Breaking Bad (2008) 9.4.

# Information extraction $\neq$ question answering





# Use case: Search enrichment

Google


who is Charles' mother

Images News Videos In Peaky Blinders Sobhraj DiLaurentis Babbage Books Maps All filters Tools








About 517,000,000 results (0.34 seconds)

Charles III / Mother

## Elizabeth II



People also search for

 Prince Philip	 Charles III	 Queen Elizabeth The...	 George VI	 Princess Margaret	 Elizabeth II	 Anne, Princess Royal
--	--	---	--	--	---	---

Feedback

People also ask

Who is the biological mother of King Charles III?

Who are the parents of Charles the First?

### Elizabeth II

Former Queen of the United Kingdom

Elizabeth II was Queen of the United Kingdom and other Commonwealth realms from 6 February 1952 until her death in 2022. She was queen regnant of 32 sovereign states over the course of her lifetime and remained the monarch of 15 realms by the time of her death. [Wikipedia](#)

**Born:** April 21, 1926, Bruton Street, London, United Kingdom

**Died:** September 8, 2022, Balmoral Castle, United Kingdom

**Spouse:** Prince Philip (m. 1947–2021)

**Grandchildren:** Prince Harry, Duke of Sussex, [MORE](#)

**Siblings:** Princess Margaret

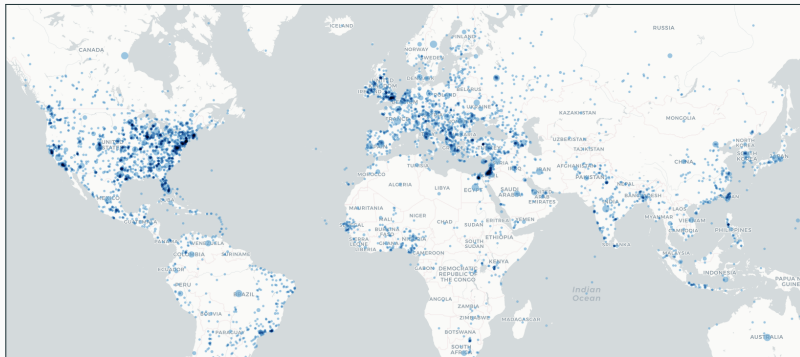
**Grandparents:** George V, Mary of Teck, [MORE](#)

**Children:** Charles III, Anne, Princess Royal, Prince Edward, Duke of Edinburgh, Prince Andrew, Duke of York

[More about Elizabeth II](#)

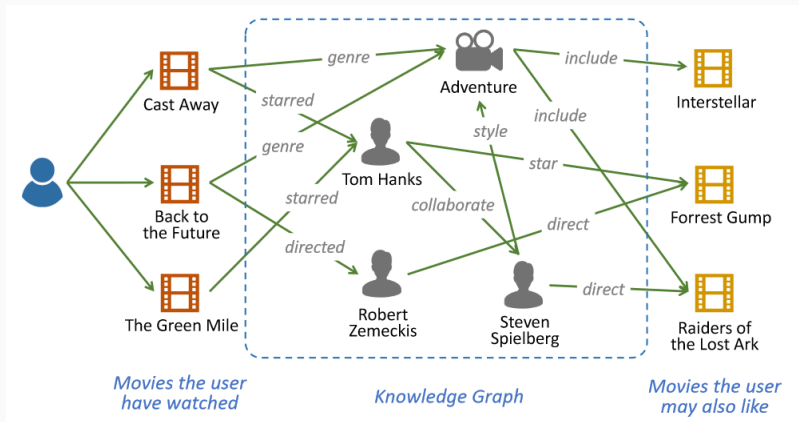
Feedback

# Use case: Geocoding worldwide news



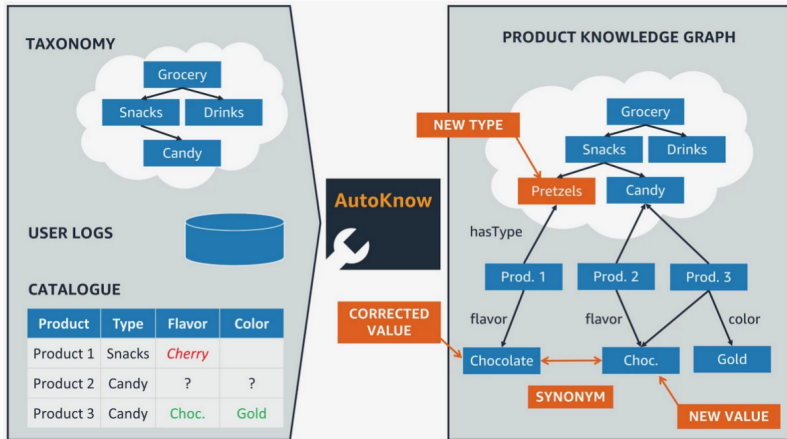


# Use case: Movie recommendation



# Knowledge graphs

# Knowledge graph = knowledge base $\neq$ database



- ▶ People, places, dates...
- ▶ Who did what when to whom?

“Paul is married with Mary”

# Storing knowledge as triples

- ▶ RDF triples (Resource Description Framework)

= subject + predicate + object

↔ flexibility & expressivity

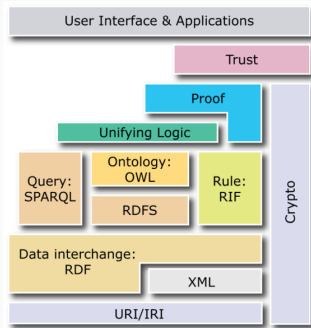
— *has\_spouse*(Paul, Mary)

— Paul *has\_spouse* Mary .

Formats: RDF/XML, Notation3, Turtle...

- ▶ OWL2 (Web Ontology Language):

full language + tools + interoperability



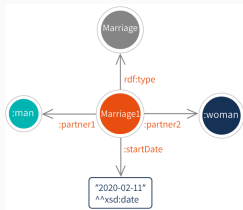


## Storing all knowledge as triples

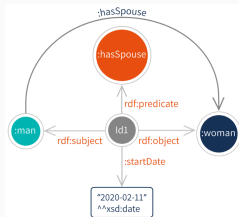
“Paul married Mary on 11 February 2020”

# Storing all knowledge as triples

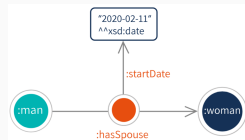
“Paul married Mary on 11 February 2020”



n-ary relation



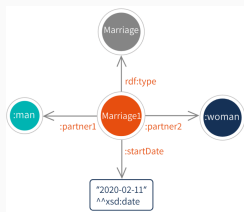
reification



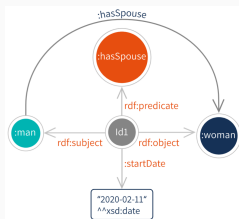
RDF\*

# Storing all knowledge as triples

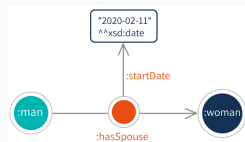
“Paul married Mary on 11 February 2020”



n-ary relation



reification

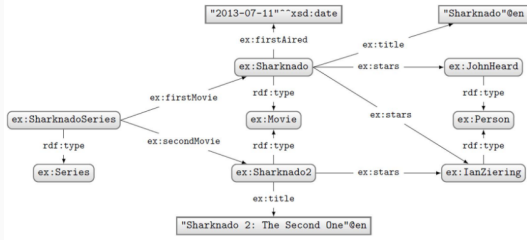


RDF\*

“I believe Paul married Mary”

“Paul married Mary after he moved to Paris”

# SPARQL Query Language for RDF



*Who stars in 'Sharknado'?*

```
PREFIX ex: <http://ex.org/voc#>
SELECT *
WHERE {
  ex:Sharknado ex:stars ?star .
}
```

?star

ex:JohnHeard  
ex:IanZiering

*What movies did the stars of 'Sharknado' also star in?*

```
PREFIX ex: <http://ex.org/voc#>
SELECT *
WHERE {
  ex:Sharknado ex:stars ?star .
  ?movie ex:stars ?star .
}
```

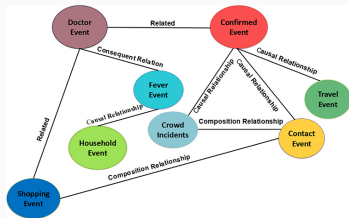
?star

?movie

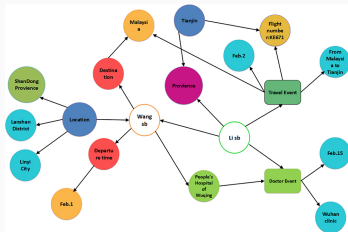
ex:IanZiering      ex:Sharknado2

# Knowledge graphs $\neq$ ontologies

Ontology: concepts, vocabulary...



Knowledge graph: facts  
= **populated** ontology





# Example: Google Knowledge Graph

Eddie Aikau

Search filters: All, News, Videos, Images, Shopping, More, Settings, Tools

About 350,000 results (0.52 seconds)

en.wikipedia.org · wiki · Eddie\_Aikau ·

## Eddie Aikau - Wikipedia

Edward Ryon Makuahana'i Aikau (Kahului, Hawaii, May 4, 1946 – March 17, 1978) was a Hawaiian lifeguard and surfer. As the first lifeguard at Waimea Bay on ...

Favorite waves: Waimea Bay (North Shore, Oa... Height: 5 ft 11 in (1.80 m)  
Years active: 1959–1978 Born: May 4, 1946; Kahului, Territory of Hawaii, ...

Life · Memorial surfing invitational · Popular culture · References

### People also ask

- How did Eddie Aikau die?
- What year did Eddie Aikau die?
- Is Eddie Aikau still alive?
- What is the story of Eddie Aikau?


Feedback

### Videos

- How Eddie Aikau Became One of Surfing's Most Legendary ...**  
Comedy Central  
YouTube - Aug 22, 2019
- Who Was Eddie Aikau And What Happened To Him? Google ...**  
Dadakoglu  
YouTube - May 4, 2019
- Hawaiian : The Legend of Eddie Aikau Official Trailer - 30 for 30**  
Quiksilver  
YouTube - Sep 11, 2013

## Eddie Aikau

Lifeguard



Edward Ryon Makuahana'i Aikau was a Hawaiian lifeguard and surfer. As the first lifeguard at Waimea Bay on the island of Oahu, he saved over 500 people and became famous for surfing the big Hawaiian surf, winning several awards including the 1977 Duke Kahanamoku Invitational Surfing Championship. [Wikipedia](#)

**Born:** May 4, 1946, Kahului, HI  
**Died:** March 17, 1978, Hawaii  
**Spouse:** Linda Crosswhite (m. 1972–1978)  
**Residence:** Kahului, Hawaii, United States  
**Siblings:** Clyde Aikau, Myra Aikau  
**Parents:** Sol Aikau, Henrietta Aikau

### People also search for

View 10+ more

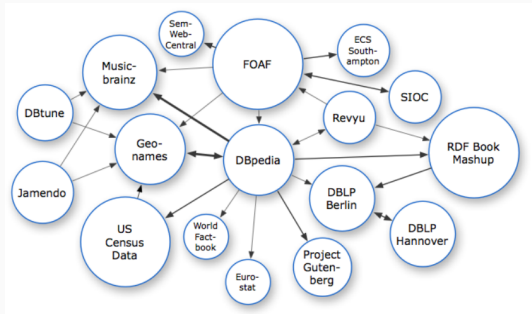
- Clyde Aikau Brother
- Duke Kahanam...
- Mark Foo
- Nainoa Thompson
- Greg Noll

Feedback

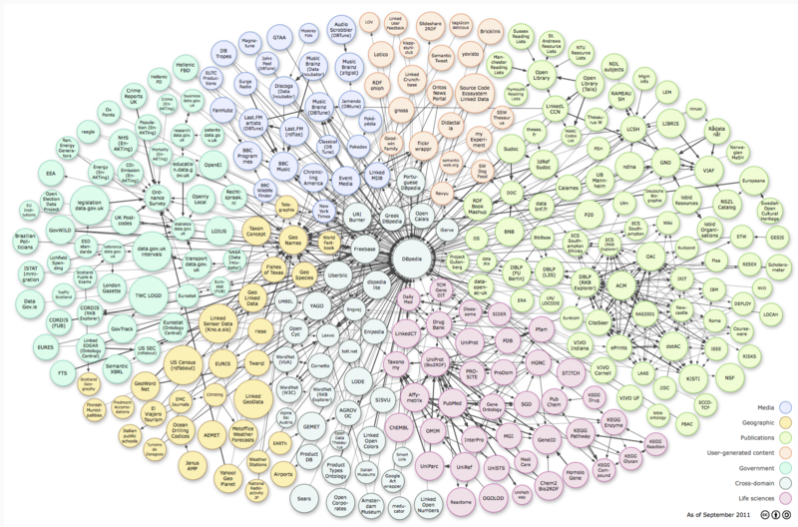




# Linking several knowledge graphs



# Linking MANY knowledge graphs



## IE tasks in practice

## In 1991: MUC-3

TST1-MUC3-0080

BOGOTA, 3 APR 90 (INRAVISION TELEVISION CADENA 1) -- [REPORT] [JORGE ALONSO SIERRA VALENCIA] [TEXT] LIBERAL SENATOR FEDERICO ESTRADA VELEZ WAS KIDNAPPED ON 3 APRIL AT THE CORNER OF 60TH AND 48TH STREETS IN WESTERN MEDELLIN, ONLY 100 METERS FROM A METROPOLITAN POLICE CAI [IMMEDIATE ATTENTION CENTER]. THE ANTIOQUIA DEPARTMENT LIBERAL PARTY LEADER HAD LEFT HIS HOUSE WITHOUT ANY BODYGUARDS ONLY MINUTES EARLIER. AS HE WAITED FOR THE TRAFFIC LIGHT TO CHANGE, THREE HEAVILY ARMED MEN FORCED HIM TO GET OUT OF HIS CAR AND GET INTO A BLUE RENAULT.

HOURS LATER, THROUGH ANONYMOUS TELEPHONE CALLS TO THE METROPOLITAN POLICE AND TO THE MEDIA, THE EXTRADITABLES CLAIMED RESPONSIBILITY FOR THE KIDNAPPING. IN THE CALLS, THEY ANNOUNCED THAT THEY WILL RELEASE THE SENATOR WITH A NEW MESSAGE FOR THE NATIONAL GOVERNMENT.

LAST WEEK, FEDERICO ESTRADA VELEZ HAD REJECTED TALKS BETWEEN THE GOVERNMENT AND THE DRUG TRAFFICKERS.

## In 1991: MUC-3

0. MESSAGE ID	TST1-MUC3-0080
1. TEMPLATE ID	1
2. DATE OF INCIDENT	03 APR 90
3. TYPE OF INCIDENT	KIDNAPPING
4. CATEGORY OF INCIDENT	TERRORIST ACT
5. PERPETRATOR: ID OF INDIV(S)	"THREE HEAVILY ARMED MEN"
6. PERPETRATOR: ID OF ORG(S)	"THE EXTRADITABLES" / "EXTRADITABLES"
7. PERPETRATOR: CONFIDENCE	CLAIMED OR ADMITTED: "THE EXTRADITABLES" / "EXTRADITABLES"
8. PHYSICAL TARGET: ID(S)	*
9. PHYSICAL TARGET: TOTAL NUM	*
10. PHYSICAL TARGET: TYPE(S)	*
11. HUMAN TARGET: ID(S)	"FEDERICO ESTRADA VELEZ" ("LIBERAL SENATOR", "ANTIOQUIA DEPARTMENT LIBERAL PARTY LEADER" / "SENATOR" / "LIBERAL PARTY LEADER" / "PARTY LEADER")
12. HUMAN TARGET: TOTAL NUM	1
13. HUMAN TARGET: TYPE(S)	GOVERNMENT OFFICIAL / POLITICAL FIGURE: "FEDERICO ESTRADA VELEZ"
14. TARGET: FOREIGN NATION(S)	-
15. INSTRUMENT: TYPE(S)	*
16. LOCATION OF INCIDENT	COLOMBIA: MEDELLIN (CITY)
17. EFFECT ON PHYSICAL TARGET(S)	*

- ▶ Named entity recognition
- ▶ Coreference resolution
- ▶ Entity linking
- ▶ Relation extraction
- ▶ Open information extraction
- ▶ Event extraction
- ▶ ...

“In December 1903 the Royal Swedish Academy of Sciences awarded Marie and Pierre Curie, along with Henri Becquerel, the Nobel Prize in Physics”

## Named entity recognition

“In December 1903 the Royal Swedish Academy of Sciences awarded Marie and Pierre Curie, along with Henri Becquerel, the Nobel Prize in Physics”

In December 1903 DATE the Royal Swedish Academy of Sciences ORG awarded Marie PERSON and Pierre Curie PERSON, along with Henri Becquerel ORG, the Nobel Prize in Physics WORK\_OF\_ART



**Ana** is a Graduate Student at **UT Dallas**.

**She** loves working in **Natural Language Processing** at **the institute**.

**Her** hobbies include blogging, dancing and singing.

## Coreference resolution

**Ana** is a Graduate Student at **UT Dallas**.



**She** loves working in **Natural Language Processing** at **the institute**.



**Her** hobbies include blogging, dancing and singing.

“Floyd revolutionized rock with the Wall”

## Entity linking

“Floyd revolutionized rock with the Wall”

↓  
.../wiki/**Pink\_Floyd**

.../wiki/Floyd\_(name)

.../wiki/Floyd,\_Iowa

↓  
.../wiki/Rock\_(geology)

.../wiki/The\_Rock

.../wiki/**Rock\_Music**

↓  
.../wiki/Defensive\_Wall

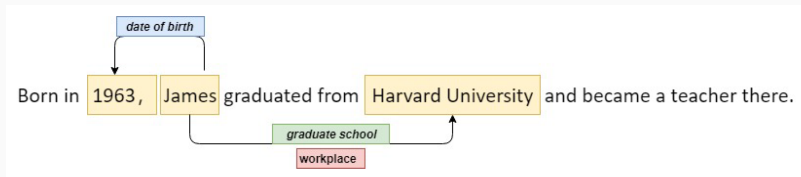
.../wiki/Berlin\_Wall

.../wiki/**The\_Wall\_(album)**

“Born in 1963, James graduated from Harvard University and became a teacher there.”

## Relation extraction

“Born in 1963, James graduated from Harvard University and became a teacher there.”



# Open information extraction

# Open information extraction

**Bill Gates, Microsoft co-founder**, stepped down as **CEO** in January 2000. **Gates** was included in the **Forbes wealthiest list** since 1987 and **was the wealthiest** from 1995 to 2007...

It was announced that **IBM** would buy **Ciao** for an undisclosed amount. The **CEO, MacLorraine** has occupied the **corner office of the Hopkinton**, company

The company's storage business is also threatened by new, born-on-the Web could providers like Dropbox and Box, and ...

relation extraction



Co-founder(Bill Gates, Microsoft)  
Director-of (MacLorraine, Ciao)  
Employee-of (MacLorraine, Ciao)  
...

open IE



(Bill Gate, be, Microsoft co-founder)  
(Bill Gates, stepped down as, CEO)  
(Bill Gates, was included in, the Forbes wealthiest list)  
(Bill Gates, was, the wealthiest)  
(IBM, would buy, Ciao)  
(MacLorraine, has occupied, the corner office of the Hopkinton)  
...



“Yesterday, some demonstrators threw stones at soldiers in Israeli.”

## Event extraction

“Yesterday, some demonstrators threw stones at soldiers in Israeli.”



*See you next week!*

`first.last@inria.fr`