Machine Learning Introduction

Tourism Facing Digital Transition

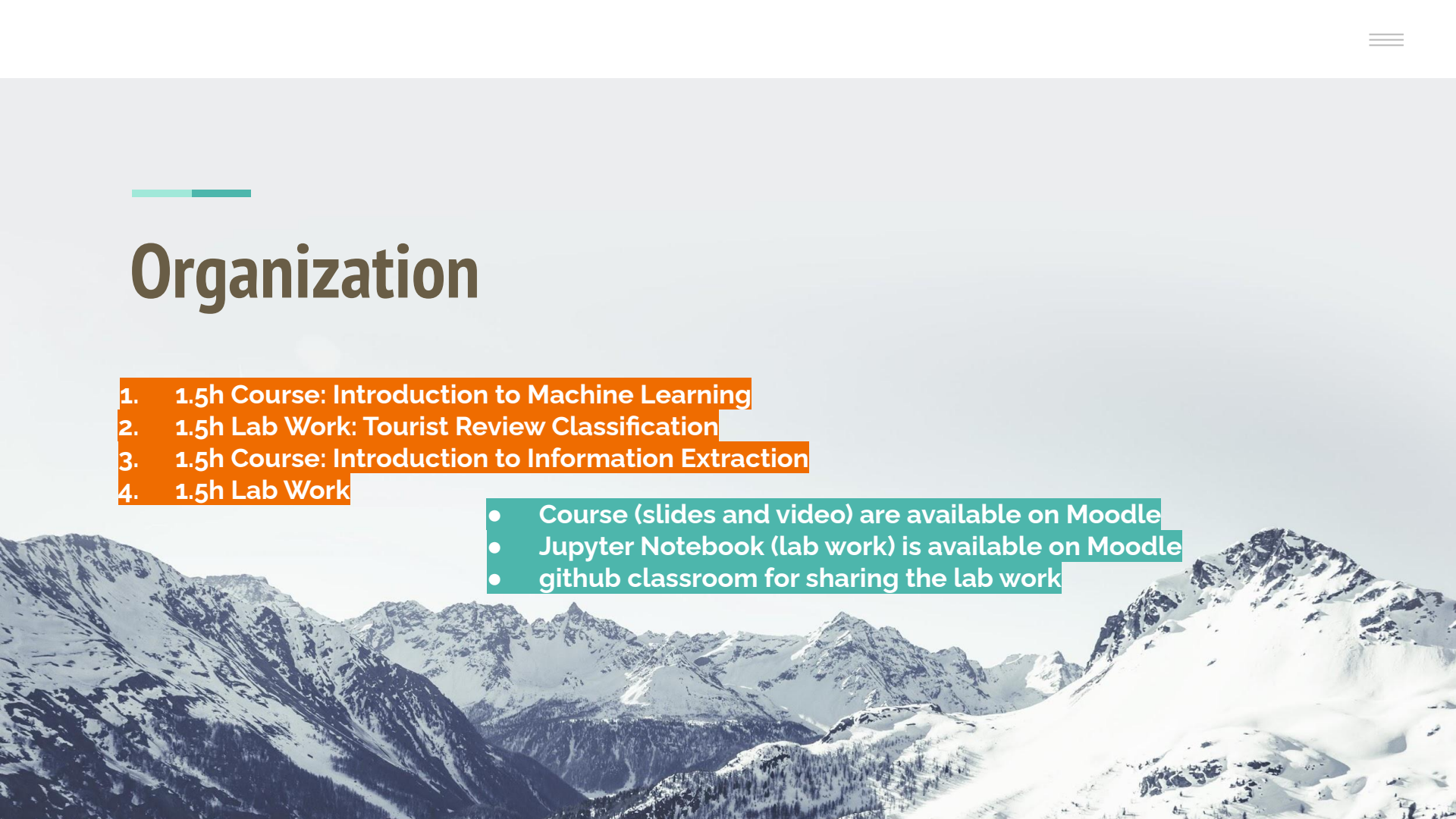
Emanuela Boros
University of La Rochelle, France

22-26 February 2021

Organization

1. 1.5h Course: Introduction to Machine Learning
2. 1.5h Lab Work: Tourist Review Classification
3. 1.5h Course: Introduction to Information Extraction
4. 1.5h Lab Work

- Course (slides and video) are available on Moodle
- Jupyter Notebook (lab work) is available on Moodle
- github classroom for sharing the lab work



Tourism Facing Digital Transition - Text Mining techniques

Text analysis/mining is the process of automatically **organizing** and **extracting** relevant information from **unstructured** text (documents, customer feedback, social media, email, etc.).

Having the texts classified (**text classification**) and **grouped by type**, the system tries to **extract useful information** from them.

Tourism Facing Digital Transition - Text Mining techniques

- **Text Classification:** the task of choosing correct class label for a given input
  - Deciding whether an is a spam or not (spam detection)
  - Deciding whether the topic of a news article is from a fixed list of topic areas such as “sports”, “technology”, and “politics” (document classification)
- **Sentiment Analysis (opinion mining):** identify and extract subjective information in source material
  - Online reviews, ratings and recommendations in social media sites for businesses looking to market their products, identifying new opportunities and manage their reputations

Information Retrieval

- **Topic Modeling:** algorithms for discovering the main themes that pervade a large and otherwise unstructured collection of documents
- **Named Entity Recognition:** methods for identifying entities in data (names, locations, organizations, etc.)

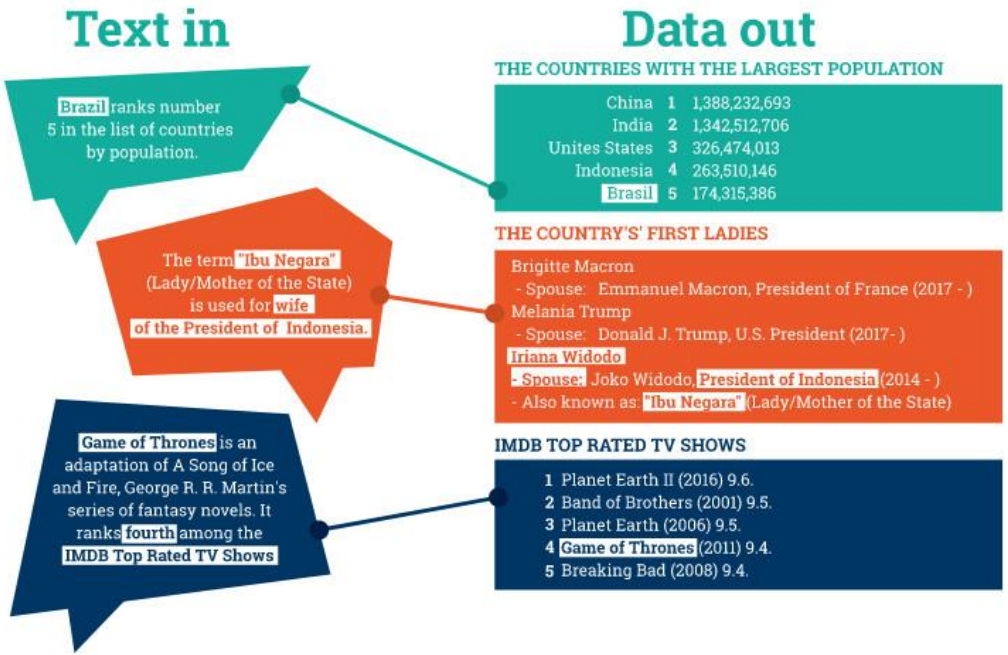
Information Extraction

- **Relation Extraction:** discerns the relationships that exist among the entities detected in a text
- **Event Extraction:** discerns the entities + their relationships that exist among the entities and their roles in events

Information Extraction - examples

Information Extraction systems extract clear, factual information → *Who did what to whom when?*

- identify relevant information from documents (financial news, or tourist information, medical documents), pulling information from a variety of sources and aggregates it into a structured form



Information Extraction - examples

Named Entity Recognition (NER)

A very important task: find and classify entities in text, for example:

"We had a very nice stay, the hotel is nicely situated in walking distance to the Eiffel tower and the exhibition halls of Porte de Versailles. Clean, nice hotel with a great neighborhood. We were about an hour walk to most of the big attractions (Eiffel Tower, Notre Dame, The Louvre). Some days we walked and some days we ubered or took the metro. Both Uber and Metro are easily navigated here. There is a metro very close by that runs conveniently through the city. The breakfast buffet has a large choice of fruits, bread etc. Can only recommend!"

Information Extraction - examples

Named Entity Recognition (NER)

A very important task: find and classify entities in text, for example:

"We had a very nice stay, the hotel is nicely situated in walking distance to the exhibition halls of **Porte de Versailles**. Clean, nice hotel with a great neighborhood. We were about an hour walk to most of the big attractions (**Eiffel Tower**, **Notre Dame**, **Louvre**). Some days we walked and some days we ubered or took the metro. Both **Uber** and **Metro** are easily navigated here. There is a metro very close by that runs conveniently through the city. The breakfast buffet has a large choice of fruits, bread etc. Can only recommend!"

Information Extraction - examples

Named Entity Recognition (NER)

A very important task: find and classify entities in text, for example:

"We had a very nice stay, the hotel is nicely situated in walking distance to the exhibition halls of **Porte de Versailles**. Clean, nice hotel with a great neighborhood. We were about an hour walk to most of the big attractions (**Eiffel Tower**, **Notre Dame**, **Louvre**). Some days we walked and some days we ubered or took the metro. Both **Uber** and **Metro** are easily navigated here. There is a metro very close by that runs conveniently through the city. The breakfast buffet has a large choice of fruits, bread etc. Can only recommend!"

Location
Organization

Information Extraction - examples

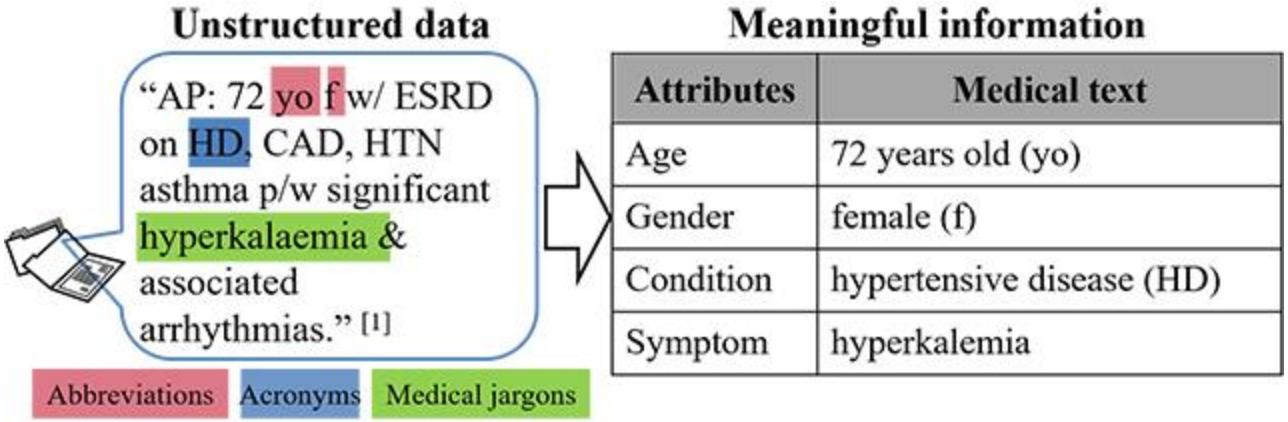
Named Entity Recognition (NER)

Type	Sample Categories	Example
People	Individuals, fictional Characters	***Turing*** is often considered to be the father of modern computer science.
Organization	Companies, parties	***Amazon*** plans to use drone copters for deliveries.
Location	Mountains, lakes, seas	The highest point in the ***Catalinas*** is ***Mount Lemmon*** at an elevation of 9,157 feet above sea level.
Geo-Political	Countries, states, provinces	The Catalinas, are located north, and northeast of ***Tucson***, ***Arizona***, ***United States.***
Facility	Bridges, airports	In the late 1940s, ***Chicago Midway*** was the busiest airport in the United States by total aircraft operations.
Vehicles	Planes, trains, cars	The updated ***Mini Cooper*** retains its charm and agility.

In practice, named entity recognition can be extended to types that are not in the table above, such as temporal expressions (time and dates), genes, proteins, medical related concepts (disease, treatment and medical events) and etc.

Information Extraction - examples

Named Entity Recognition (NER)



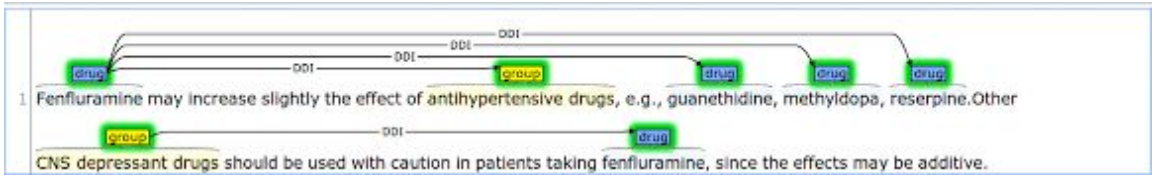
- medical related concepts (disease, treatment and medical events) and etc.
- online patient forums can provide valuable supplementary information on drug effectiveness and side effects

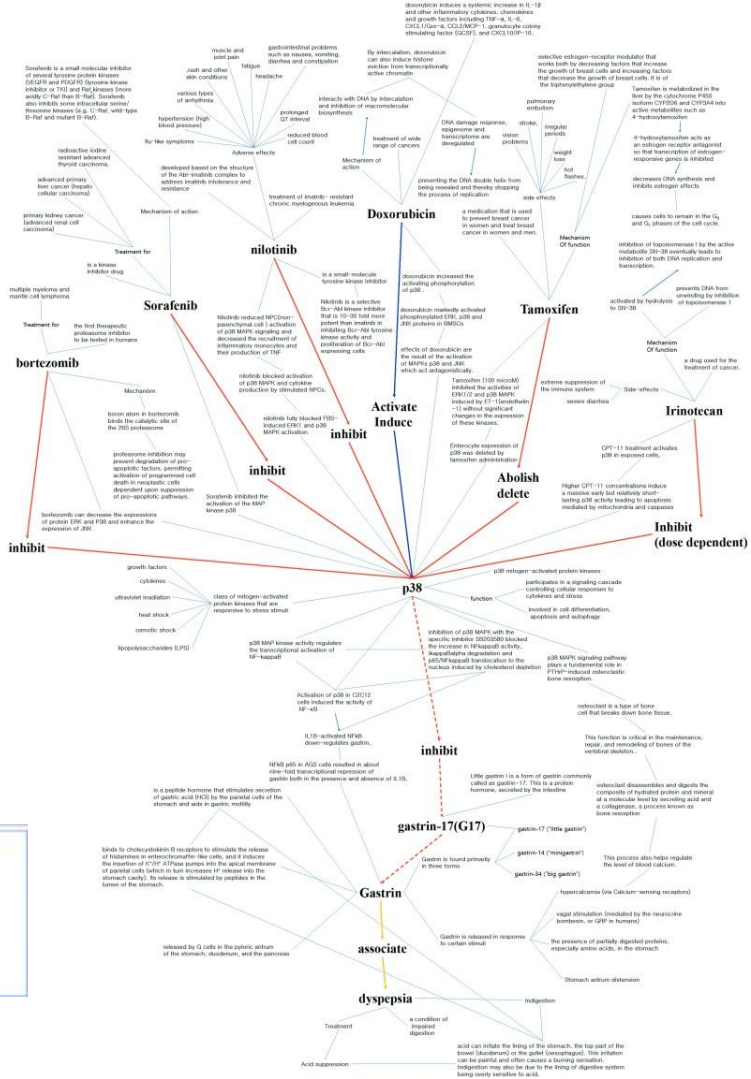
Information Extraction - examples

Relation Extraction (RE)

Relation extraction discerns the relationships that exist among the entities detected in a text.

- **Mining medical texts**: discharge summaries, narrative patient records
  - Protein binding relations useful for drug discovery
  - Detection of gene-disease relations from biomedical literature
  - Finding drug-side effect relations in health records



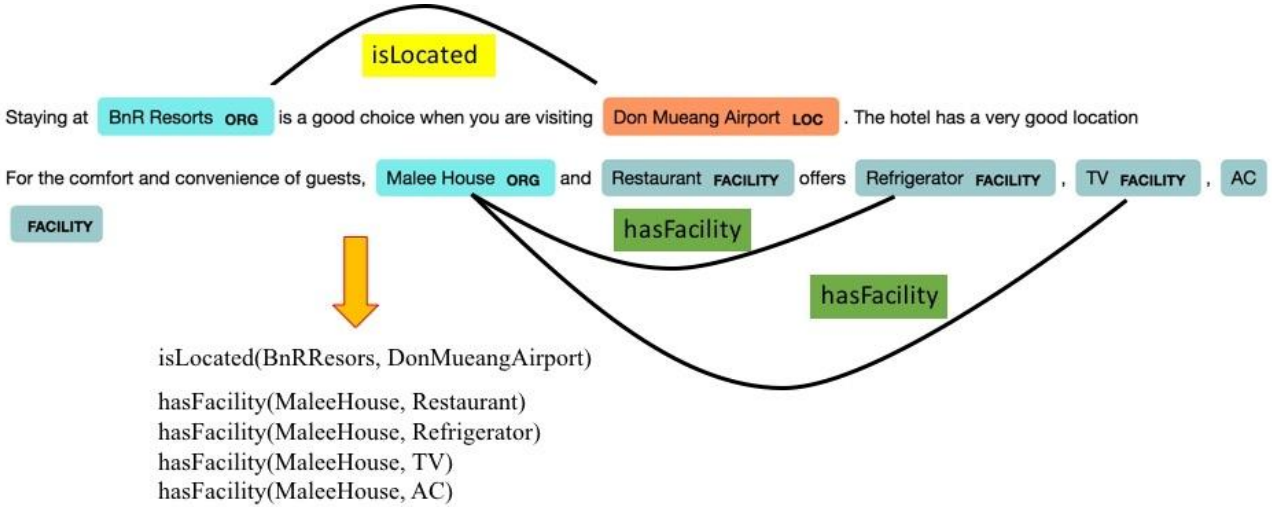


Information Extraction - examples

Relation Extraction (RE)

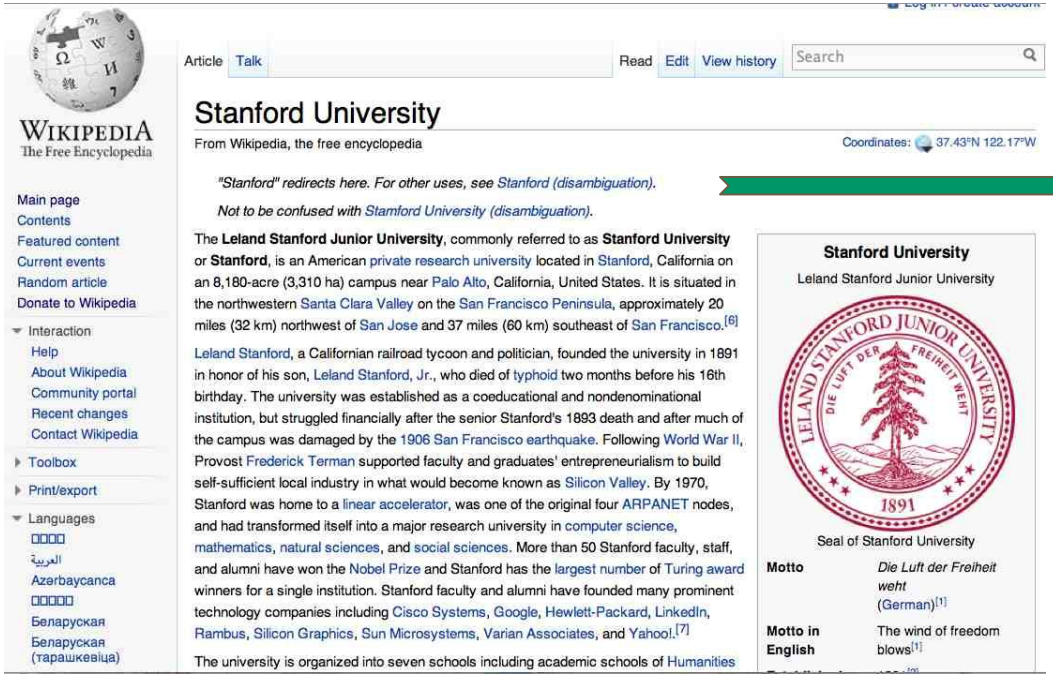
Relation extraction discerns the relationships that exist among the entities detected in a text.

- **Mining tourism texts:** documents, customer feedback, social media, email



Information Extraction - examples

Relation Extraction (RE)



Stanford **LOC-IN** California
Stanford **IS-A** research university
Stanford **LOC-NEAR** Palo Alto
Stanford **FOUNDED-IN** 1891
Stanford **FOUNDER** Leland Stanford

Information Extraction - examples

Event Extraction = Named Entity Recognition + Relation Extraction

[S1] ... by special urban troops, four terrorists have been arrested in soacha.

[S2] They are responsible for the car bomb attack on the Newspaper El Espectador, to a series of bogota dynamite attacks, to the freeing of a group of paid assassins.

[S3] The terrorists are also connected to the murder of Teofilo Forero Castro, ...

[S4] General Ramon is the commander of the 13th infantry brigade.

[S5] He said that at least two of those arrested have fully confessed to having taken part in the accident of Luis Carlos Galan Sarmiento in soacha, Cundinamarca.

[S6] .. triumph over organized crime, its accomplices and its protectors.

...

Perpetrator Individual	four terrorists
Perpetrator Organization	-
Target	Newspaper El Espectador
Victim	Teofilo Forero Castro, Luis Carlos Galan Sarmiento
Weapon	car bomb, dynamite

Attack event:
- entities
- relations

Information Extraction - examples

Named Entity Recognition (NER)

- Game of Thrones: **TV Show**
- A Song of Ice and Fire: **Book**
- George R. R. Martin: **Author**

Relation Extraction (RE)

$$\left(\text{A Song of Ice and Fire} \xrightarrow{\text{Book-Author}} \text{George R. R. Martin}\right)$$

Event Extraction (EE)

$$\left(\text{Game of Thrones} \xrightarrow{\text{TV Show-Author}} \text{George R. R. Martin} \xrightarrow{\text{Author-Book}} \text{A Song of Ice and Fire}\right)$$

Links

https://www.kaggle.com/competitions

Machine learning Coursera famous courses, Andrew Ng, https://www.coursera.org/learn/machine-learning

Machine learning Coursera (on youtube), Andrew Ng, https://www.youtube.com/watch?v=PPLop4L2eGk

The most famous book on deep learning: https://www.deeplearningbook.org/ (Ian Goodfellow, Yoshua Bengio and Aaron Courville)

ML and DL People

Andrew Ng, Founder and CEO of Landing AI, Founder of deeplearning.ai.
Fei-Fei Li, Professor of Computer Science at Stanford University.
Andrej Karpathy, Senior Director of Artificial Intelligence at Tesla.
Demis Hassabis, Founder and CEO of DeepMind.
Ian Goodfellow, Director of Machine Learning at Apple.
Yann LeCun, Vice President and Chief AI Scientist at Facebook.
Jeremy P. Howard, Founding Researcher at fast.ai, Distinguished Research Scientist at the University of San Francisco.
Ruslan Salakhutdinov, Associate Professor at Carnegie Mellon University, Director of AI Research at Apple.
Geoffrey Hinton, Professor of Computer Science at the University of Toronto, VP and Engineering Fellow at Google
Rana el Kaliouby, CEO and Co-Founder of Affectiva.
Daphne Koller, Founder and CEO of insitro, Co-Founder of Coursera, Adjunct Professor of Computer Science and Pathology at Stanford.
Alex Smola, Director, Amazon Web Services.