NLP For Economists

NLP: A quick introduction

Sowmya Vajjala

Munich Graduate School of Economics - LMU Munich

Guest Course, October 2020

Goals for this session

- Introduce some basic ideas about natural language processing
- Describe why NLP can be challenging
- Talk about some commonly seen NLP tasks (not all of these may be relevant to you now or later)

note: images without source attribution are taken from our book: practicalnlp.ai

What is NLP?

1. NLP is a sub-field of Artificial intelligence that is concerned with analyzing, modeling and understanding human language using computational methods.
2. It explores how humans can interact with computers in human languages
3. The eventual goal is to make computers understand (and generate) human languages, and make them communicate with humans like humans
4. Because of its role in the process of human-computer interaction, NLP has a wide range of technological applications
5. It is also becoming popular as a research method in a broad range of disciplines in social sciences.

Inter-disciplinary by nature

NLP is very inter-disciplinary. Draws from research in Computer Science, Linguistics, Mathematics, Statistics, Psychology etc.,

Computational Linguistics vs NLP

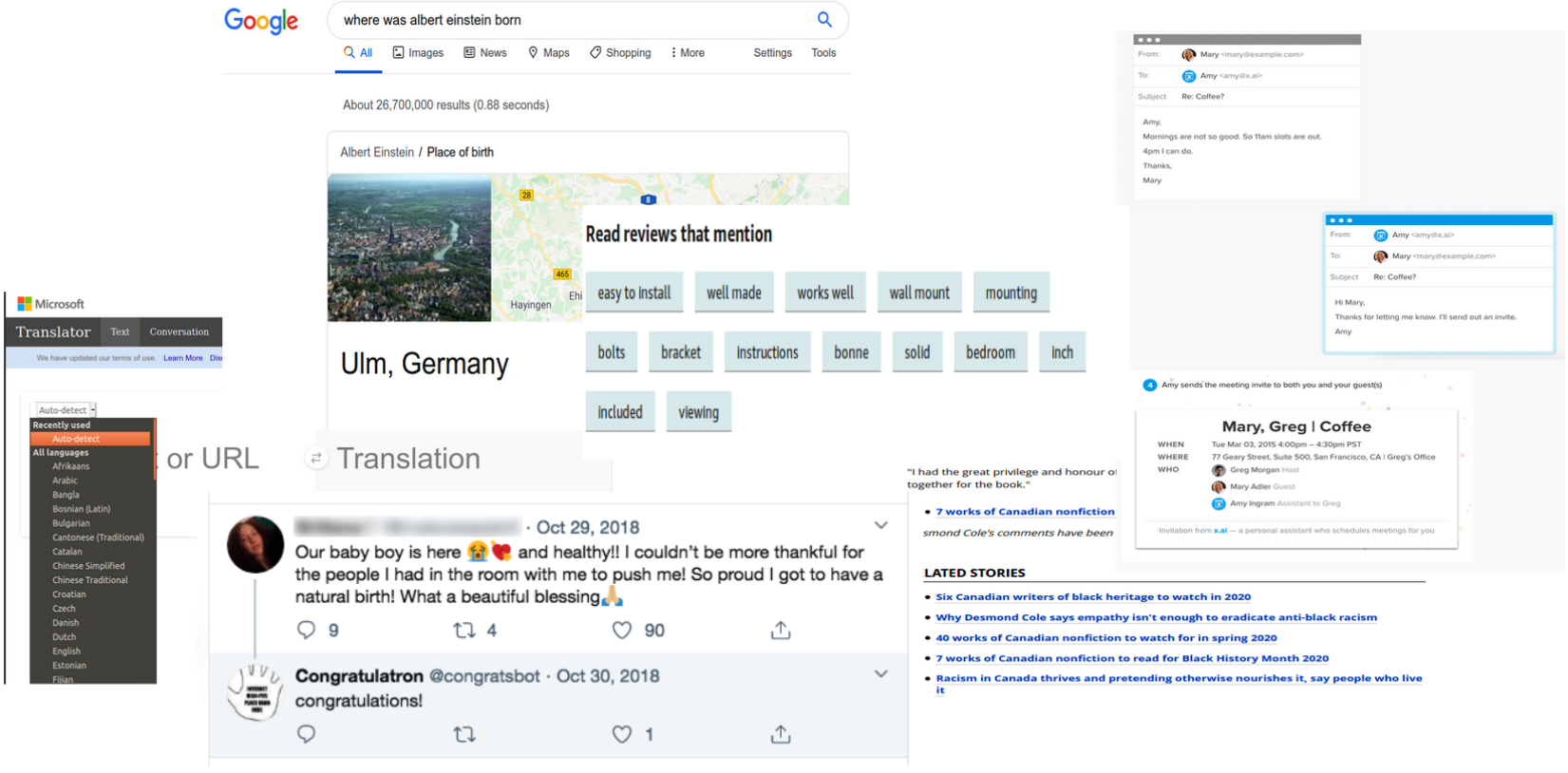
The terms are used synonymously. However, generally, NLP is typically used by people involved in engineering and technology development, and CL is typically used by traditionally linguistics groups who adapted computational methods.

History of NLP

1. Foundational ideas: 40s and 50s. WWII and Beyond.
2. Main NLP problem of that time (and even now): Machine Translation
3. First few decades: Work focused on the development of speech recognition systems, logic based language understanding systems, creating elaborate grammars to teach human language to computers, and automatic language generation.
4. Late 90s on: Advent of statistical methods and machine learning
5. 2010s: Deep learning
6. Last few years: more into interpretable models, discussion about ethical issues, introspection about the field etc.

Where is NLP used in real-world?

Everyday NLP Applications



Where is NLP used in real-world?

1. Apple Siri, Google assistant and other such software
2. Google Translate and the likes
3. Search Engines
4. Question Answering (e.g., IBM Watson)
5. Sentiment analysis of product reviews on Amazon, for example
6. Spam classification in Gmail, Yahoomail etc
7. Information extraction from text (e.g., identifying calendar entries automatically in gmail)
8. Spelling/Grammar check tools, language learning apps such as DuoLingo etc.

Where is NLP used elsewhere?

1. NLP is used as a method to answer research questions in many disciplines.
2. NLP sometimes plays a major role in discipline specific challenges, going beyond being just a research method.
3. In Google Scholar, I saw mentions of NLP methods in journals as diverse as Asian studies & History to Clinical Oncology.
4. You can observe the use of NLP in many prominent Economics journals such as QJE in the past few years.

What makes NLP challenging?
(what sort of issues pose problems for a computer?)

Language is ambiguous

Some ambiguous sentences

- Newspaper headlines
  - "Children make delicious snacks"
  - "Dead expected to rise"
  - "Republicans grill IRS chief over lost emails"
- Normal, grammatical sentences can be ambiguous too:
  - "I saw a man on a hill with a telescope."
  - "Look at the man with one eye"

We are not even talking about ambiguities involving speech or alternative interpretations due to stress/emphasis on some word.

Some types of ambiguity

1. Lexical ambiguity: due to multiple meanings or senses of word usage
   e.g., He stood near the **bank**
2. Structural ambiguity: due to syntactic structure
   e.g., I saw the man on the hill with telescope.
3. Semantic ambiguity: more interpretations possible
   e.g., John and Mary are married (to each other? or to different people?)
4. Referential ambiguity
   e.g., She dropped the *plate* on the *table* and broke **it**
5. Ambiguity due to the use of non-literal language
   e.g., Time flies like an arrow

Good source to read more:
`http://cs.nyu.edu/faculty/davise/ai/ambiguity.html`

"common" knowledge for humans

Look at these two sentences:
Dog bit man.
Man bit dog.
- For a computer, both of them are linguistically the same. We know only the first one is "normal" English sentence because we have "world knowledge".

Few more challenges

- Language is diverse: many different forms of documents such as news, tweets, legal texts etc.
- Language is creative: its use changes over time, and vocabulary gets richer.
- There are many different languages in the world
- Many spelling variations, slangs, sarcasm etc.

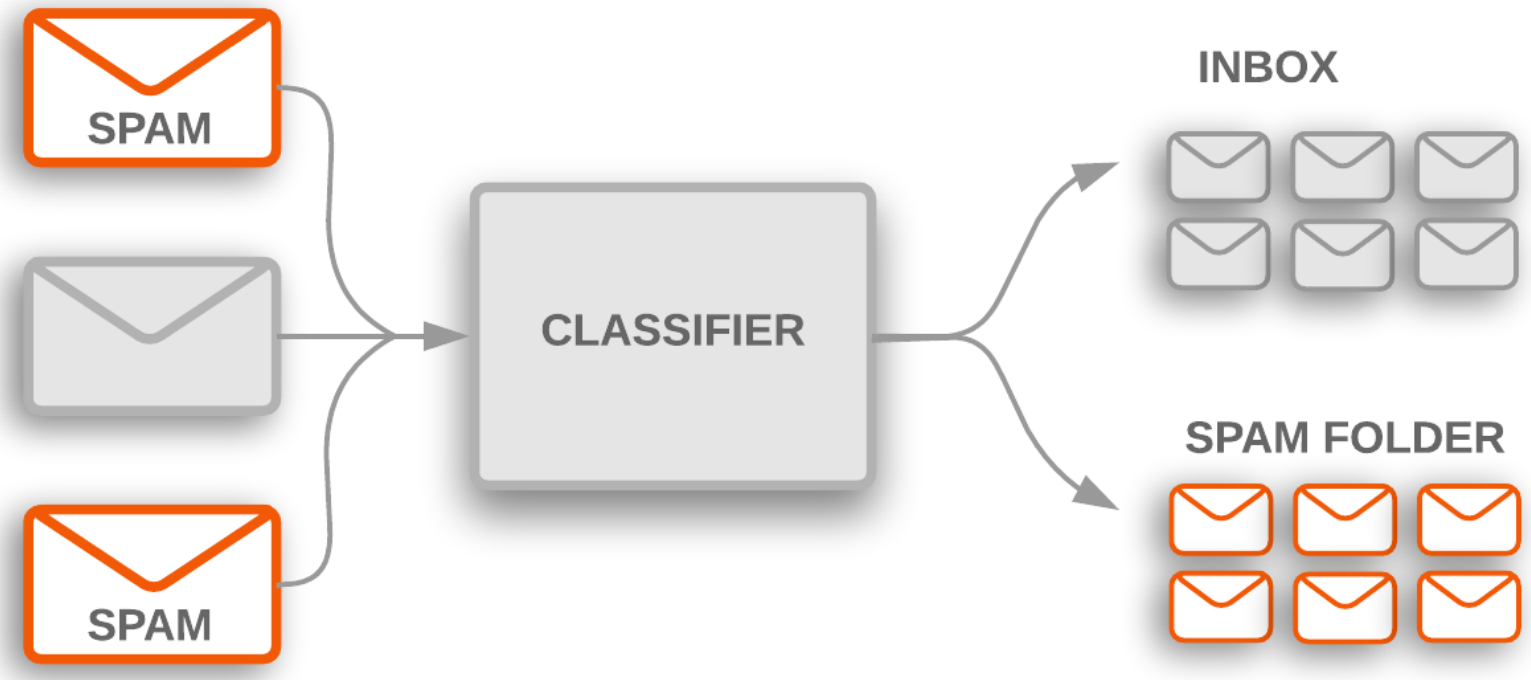
- NLP solutions should account for all these things!

So, the summary is:

perfect NLP is hard to achieve because of all these issues that come up when we start using computers to analyze language!

Some NLP tasks

Text Classification



¿¿

source: https://developers.google.com/machine-learning/guides/text-classification

Information Extraction

SAN FRANCISCO — Shortly after Apple used a new tax law last year to bring back most of the $252 billion it had held abroad, the company said it would buy back $100 billion of its stock.

On Tuesday, Apple announced its plans for another major chunk of the money: It will buy back a further $75 billion in stock.

"Our first priority is always looking after the business and making sure we continue to grow and invest," Luca Maestri, Apple's finance chief, said in an interview. "If there is excess cash, then obviously we want to return it to investors."

Apple's record buybacks should be welcome news to shareholders, as the stock price is likely to climb. But the buybacks could also expose the company to more criticism that the tax cuts it received have mostly benefited investors and executives.

¿¿

- Who is Luca Mestri?

Information Extraction

SAN FRANCISCO — Shortly after Apple used a new tax law last year to bring back most of the $252 billion it had held abroad, the company said it would buy back $100 billion of its stock.

On Tuesday, Apple announced its plans for another major chunk of the money: It will buy back a further $75 billion in stock.

"Our first priority is always looking after the business and making sure we continue to grow and invest," Luca Maestri, Apple's finance chief, said in an interview. "If there is excess cash, then obviously we want to return it to investors."

Apple's record buybacks should be welcome news to shareholders, as the stock price is likely to climb. But the buybacks could also expose the company to more criticism that the tax cuts it received have mostly benefited investors and executives.

¿¿

- Who is Luca Mestri? needs: Named Entity Recognition and Linking, Relation extraction
- What is the article about?

Information Extraction

SAN FRANCISCO — Shortly after Apple used a new tax law last year to bring back most of the $252 billion it had held abroad, the company said it would buy back $100 billion of its stock.

On Tuesday, Apple announced its plans for another major chunk of the money: It will buy back a further $75 billion in stock.

"Our first priority is always looking after the business and making sure we continue to grow and invest," Luca Maestri, Apple's finance chief, said in an interview. "If there is excess cash, then obviously we want to return it to investors."

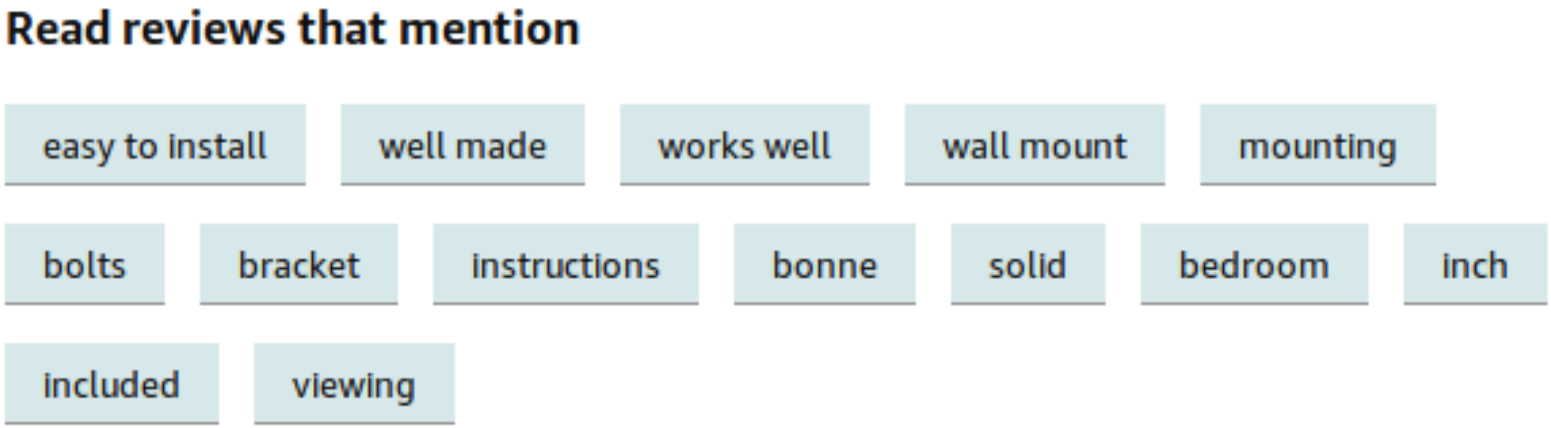
Apple's record buybacks should be welcome news to shareholders, as the stock price is likely to climb. But the buybacks could also expose the company to more criticism that the tax cuts it received have mostly benefited investors and executives.

¿¿

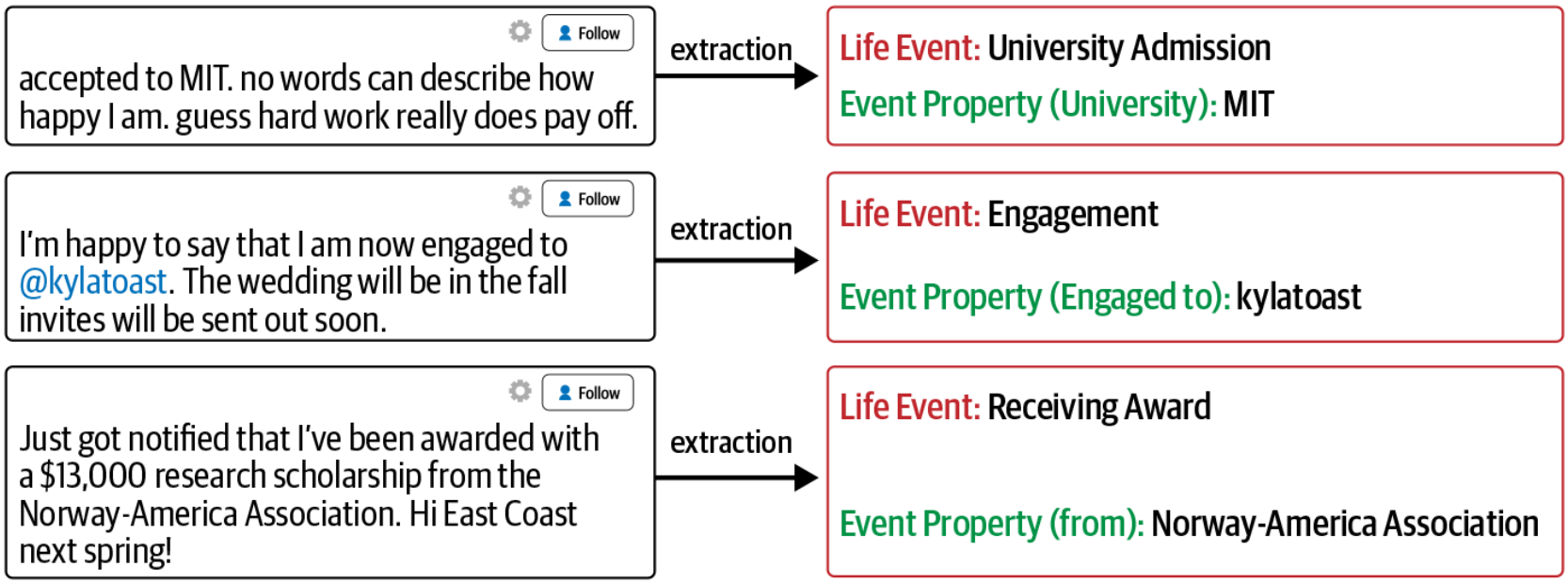
- Who is Luca Mestri?
  needs: Named Entity Recognition and Linking, Relation extraction
- What is the article about?
  needs: Key phrase extraction, event extraction

Named Entity Extraction/Linking

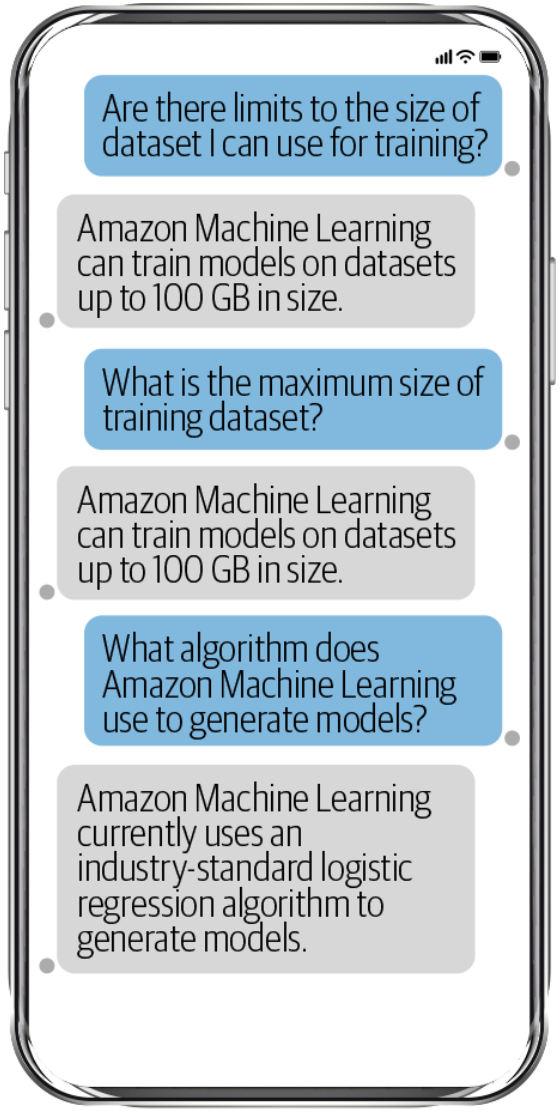
Key Phrase Extraction



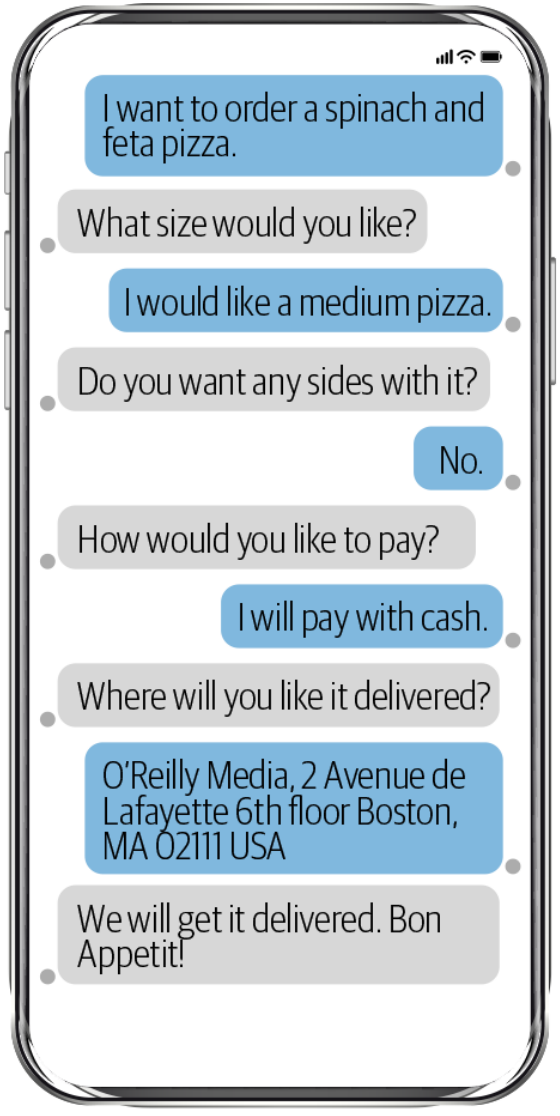
Event/Relation Extraction



Chatbots



FAQ Bot



Flow-Based Bot

Open-Ended Bot

Search

Google

Mary Curie

All Images News Videos Maps More Settings Tools

About 8,520,000 results (1.08 seconds)

5

Did you mean: ***Marie*** Curie 1

3 **Marie Curie - Wikipedia**

https://en.wikipedia.org › wiki › Marie_Curie

Marie Sklodowska Curie was a Polish and naturalized-French physicist and chemist who conducted pioneering research on radioactivity. She was the first ...

Cause of death: Aplastic anemia from exposure ... **Fields**: Physics, chemistry
Children: Irène Joliot-Curie (1897–1956); Ève ... **Doctoral advisor**: Gabriel Lippmann

Irène Joliot-Curie · Ève Curie · Pierre Curie · Curie Institute (Paris)

2 **People also ask**

What is Marie Curie famous for?

How did Marie Curie die?

What is Marie Curie discover?

What impact did Marie Curie have on society?

Feedback

Marie Curie - Biographical - NobelPrize.org

https://www.nobelprize.org › prizes › physics › marie-curie › biographical

Marie Curie, née Maria Sklodowska, was born in Warsaw on November 7, 1867, the daughter of a secondary-school teacher. She received a general education ...

Marie Curie - Facts, Quotes & Nobel Prize - Biography

https://www.biography.com › scientist › marie-curie

Marie Curie became the first woman to win a Nobel Prize and the first person — man or woman — to win the award twice. With her husband Pierre **Curie**, **Marie's** efforts led to the discovery of polonium and radium and, after Pierre's death, the further development of X-rays.

Death Date: July 4, 1934 **Education**: Sorbonne
Birth Date: November 7, 1867

Marie Curie the scientist | Biog, facts & quotes

https://www.mariecurie.org.uk › who › our-history › marie-curie-the-scien...

Marie Curie is remembered for her discovery of radium and polonium, and her huge contribution to

Marie Curie

French-Polish physicist

Mari[illegible]owska Curie was a Polish and naturalized-French ph[illegible]chemist who conducted pioneering research on rad[illegible]. She was the first woman to win a Nobel Prize, is the only woman to win the Nobel prize twice, and is the only person to win the Nobel Prize in two different scientific fields. Wikipedia

4

Born: November 7, 1867, Warsaw, Poland

Died: July 4, 1934, Sancellemoz

Discovered: Radium, Polonium

Education: University of Paris (1903), University of Paris (1894), University of Paris (1891–1893), Flying University, Curie Institute

Awards: Nobel Prize in Physics, Nobel Prize in Chemistry, MORE

Quotes

View 7+ more

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.

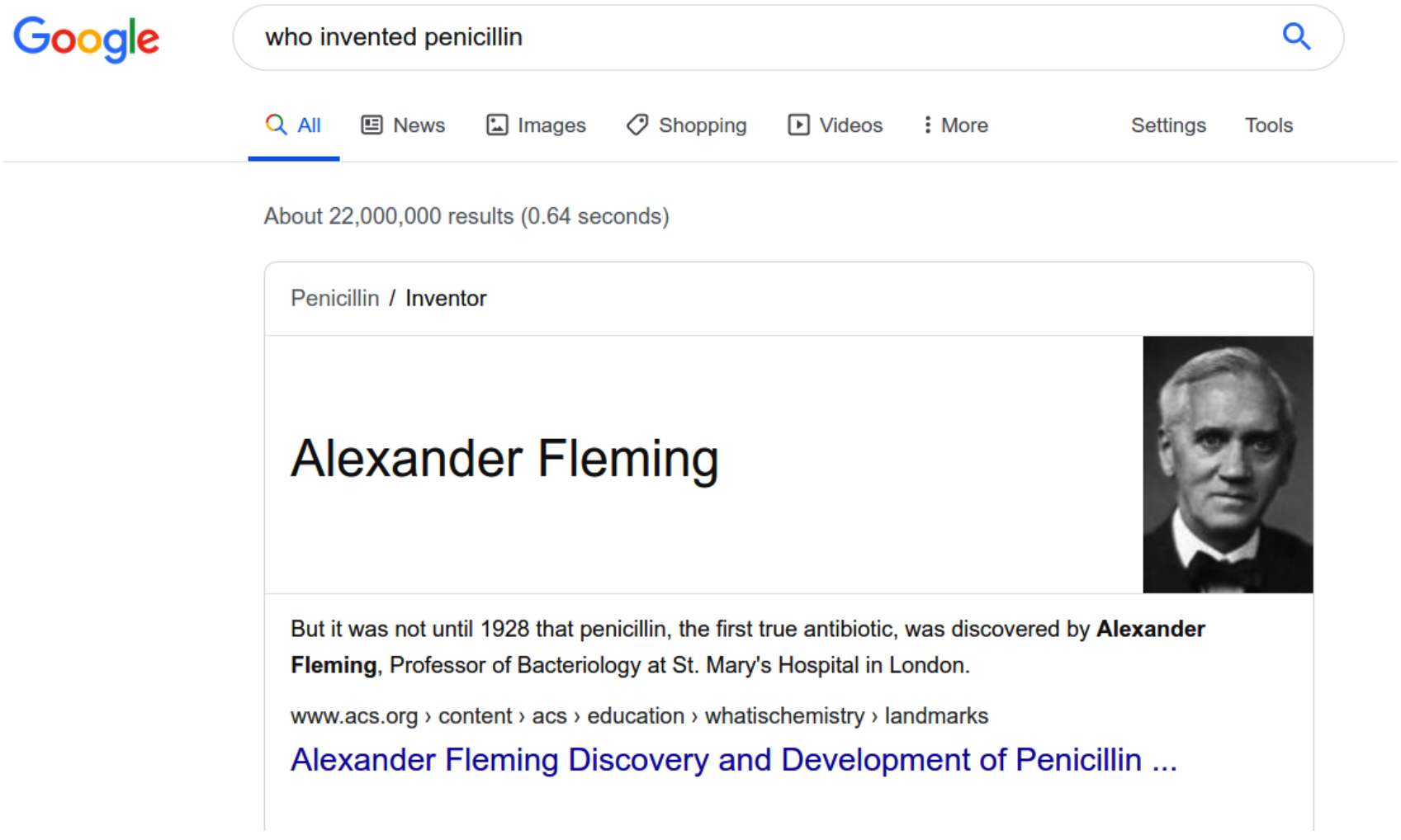
Be less curious about people and more curious about ideas.

One never notices what has been done; one can only see what remains to be done.

People also search for

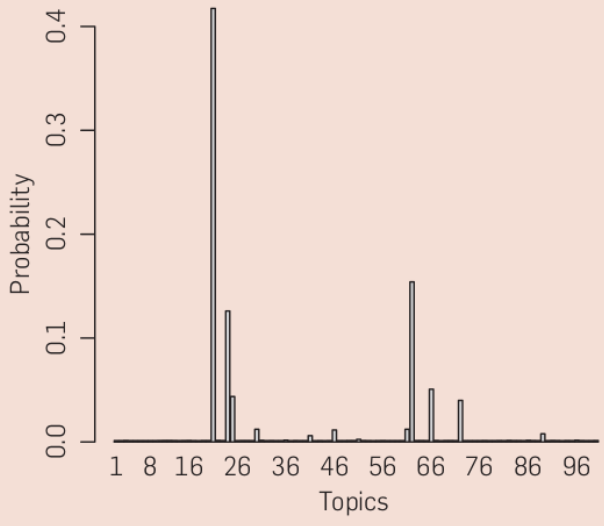
View 15+ more

Question Answering



Topic Modeling

Figure 2. Real inference with LDA. We fit a 100-topic LDA model to 17,000 articles from the journal *Science*. At left are the inferred topic proportions for the example article in Figure 1. At right are the top 15 most frequent words from the most frequent topics found in this article.



"Genetics"	"Evolution"	"Disease"	"Computers"
human	evolution	disease	computer
genome	evolutionary	host	models
dna	species	bacteria	information
genetic	organisms	diseases	data
genes	life	resistance	computers
sequence	origin	bacterial	system
gene	biology	new	network
molecular	groups	strains	systems
sequencing	phylogenetic	control	model
map	living	infectious	parallel
information	diversity	malaria	methods
genetics	group	parasite	networks
mapping	new	parasites	software
project	two	united	new
sequences	common	tuberculosis	simulations

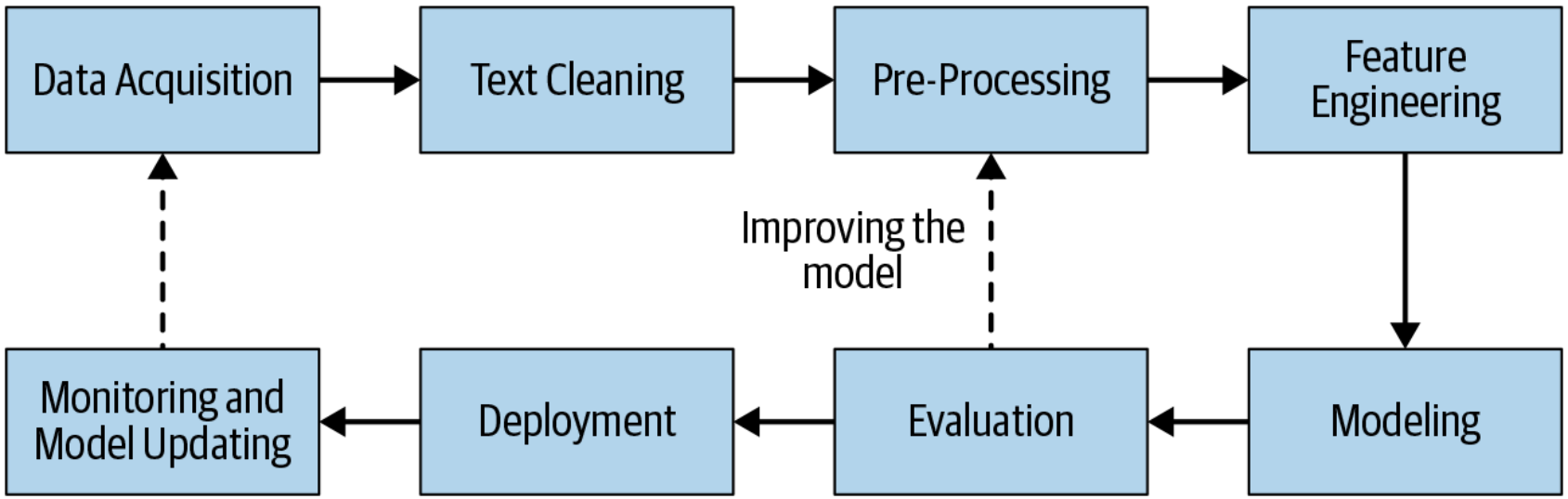
source: `http://www.cs.columbia.edu/~blei/papers/Blei2012.pdf`

Many more

- text summarization
- text recommendation
- machine translation
- text to speech/speech to text conversion
- language generation (e.g., automatically generating weather reports, chatbots, etc.)

.... and so on.

How does one build these systems?



How much should we know about this now?

- We will talk about cleaning, pre-processing, features, models and evaluation.
- Deployment/monitoring are relevant for industry professionals
- I will provide the data for exercises, but you have to think about how to get the data for future projects.

A topic to discuss during our meeting: What are the sources of textual data for economics researchers??

Before I wrap up...

- Don't worry, we won't do everything in this short course!
- The goal of this lecture is just to provide a very broad overview so that you get a bigger picture beyond the boundaries of this course!

ToDo for you

- You can check out the intro chapter of any of the recommended NLP textbooks, if you are curious and have time.
- You can read "Text as Data" paper to get a contextual intro to NLP as economists.