

NATURAL LANGUAGE PROCESSING

A PRACTICAL OVERVIEW OF THE POSSIBILITIES
FOR INFORMATION RETRIEVAL

Christophe SERVAN, PhD – Chief Scientist



Qwant

CHRISTOPHE SERVAN, PHD



WEB SEARCH ENGINE

Web Search engines:

- Google
- Baidu
- Yahoo
- Yandex
- Bing
- Naver
- Seznam
- Qwant

Meta-Web Search Engine:

- Ecosia
- DuckDuckGo
- ...

QWANT: THE EUROPEAN SEARCH ENGINE THAT RESPECTS YOUR PRIVACY



No third-party cookie



Hashed/salted IP address



No search history



Own server farm in EU, algorithms & index



HTTPS



Host of additional services



Qwant



Qwant Sport



Qwant Shopping



Qwant Music



Qwant Games



Qwant Junior



Qwant Maps



Qwant IOT



Qwant Med



Qwant Pay

QWANT IN NUMBERS

- Launched in 2013 (V4 in summer 2018)
- 150 employees with rapid growth
- Located in France (Paris, Rouen, Nice, Ajaccio, Epinal), Germany and Italy
- Investors: Axel Springer, CDC, EIB
- Sole European search engine with its own index and algorithms
- More than 80 million connections per month
- In the Top 50 sites in Europe (Source: similarweb.com)



Qwant

ACTUAL STATE OF NLP IN IR

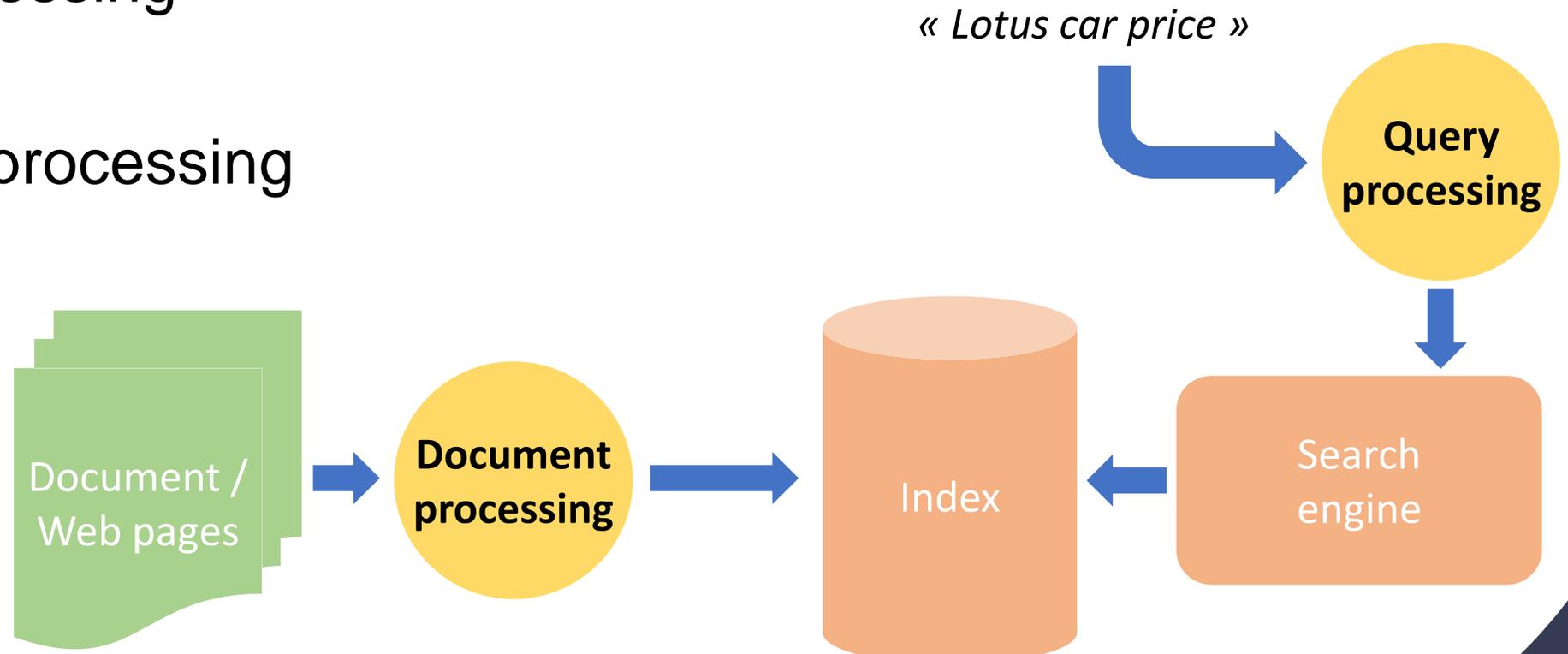
POPULAR ENGINES FOR IR

- LEMUR (INDRI)
 - Lucene
 - Nutch
 - SolR
 - Compass
 - Elasticsearch
- ... and many other...

NATURAL LANGUAGE PROCESSING (NLP) IN IR

Classical view

- Query processing
- Document processing



NLP FOR IR: PRACTICAL STATE OF THE ART

- Query processing
- Features
- Document processing

NLP FOR IR: PRACTICAL STATE OF THE ART

- Query processing
 - Stemming
 - Stop words
 - Suggestions
 - Entailment
- Document processing
- Features

NLP FOR IR: PRACTICAL STATE OF THE ART

- Query processing
 - Stemming
 - Stop words
 - Suggestions
 - Entailment
- Document processing
 - Stemming
 - Stop words
- Features

NLP FOR IR: PRACTICAL STATE OF THE ART

- Query processing

- Stemming
- Stop words
- Suggestions
- Entailment

- Document processing

- Stemming
- Stop words

- Features

- TF-IDF
- BM25

NLP FOR IR: PRACTICAL STATE OF THE ART

- Query processing

- Stemming
- Stop words
- Suggestions
- Entailment

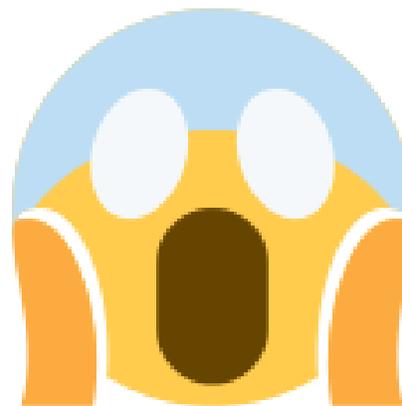
- Document processing

- Stemming
- Stop words

- Features

- TF-IDF
- BM25

⇒ That's All!

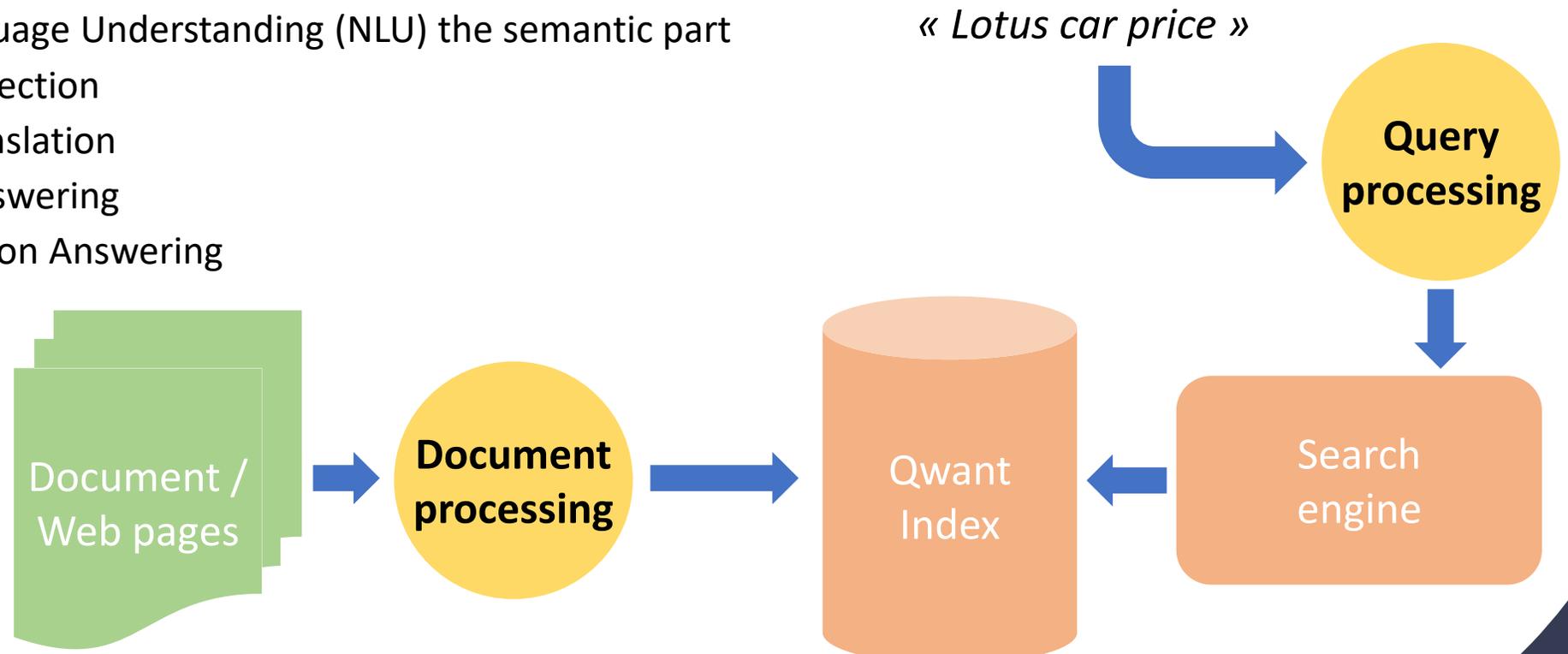


NLP & IR @ QWANT



NATURAL LANGUAGE PROCESSING (NLP)

- **NLP** is crucial for the Qwant search engine
- Integration inside the search engine through **document and query** processing :
 - Name Entity Recognition (NER) like dates, locations, names, etc.
 - Natural Language Understanding (NLU) the semantic part
 - Intention detection
 - Machine Translation
 - Question/Answering
 - Visual Question Answering



INFORMATION EXTRACTION / NER

Query: what is the weather like in Paris?

INFORMATION EXTRACTION / NER

Query: what is the weather like in Paris?

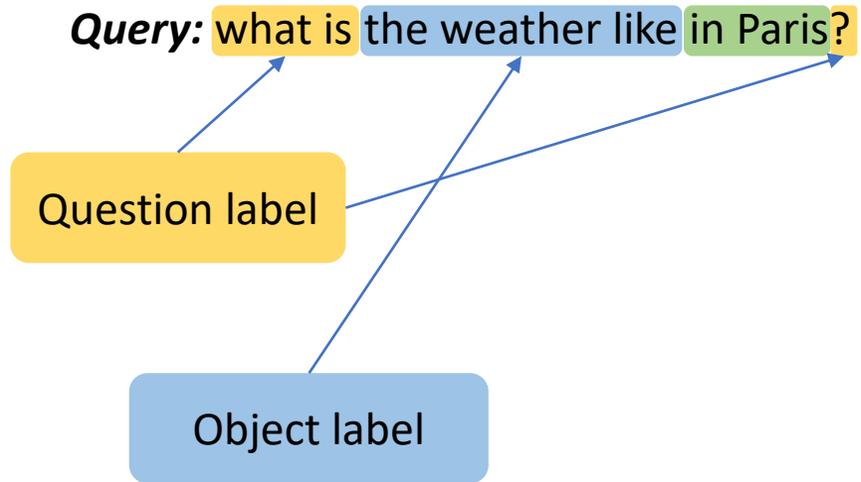
INFORMATION EXTRACTION / NER

Query: what is the weather like in Paris?

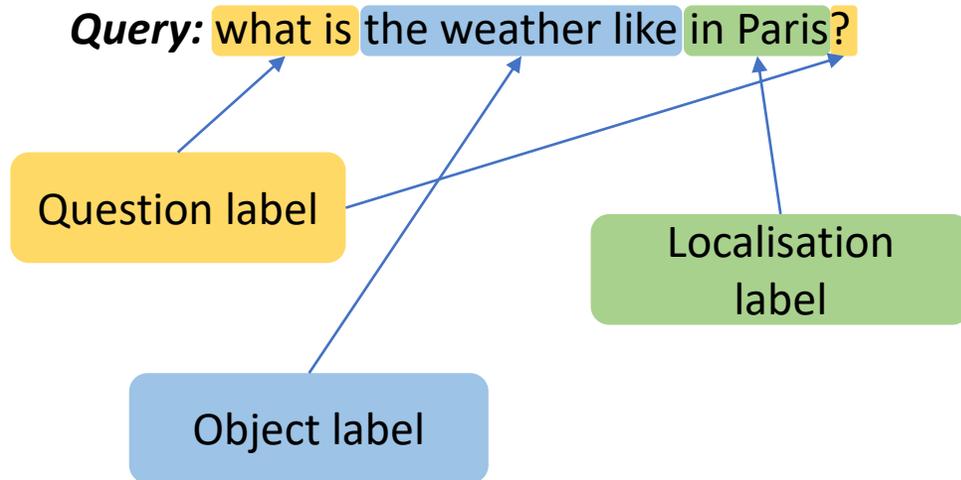
Question label



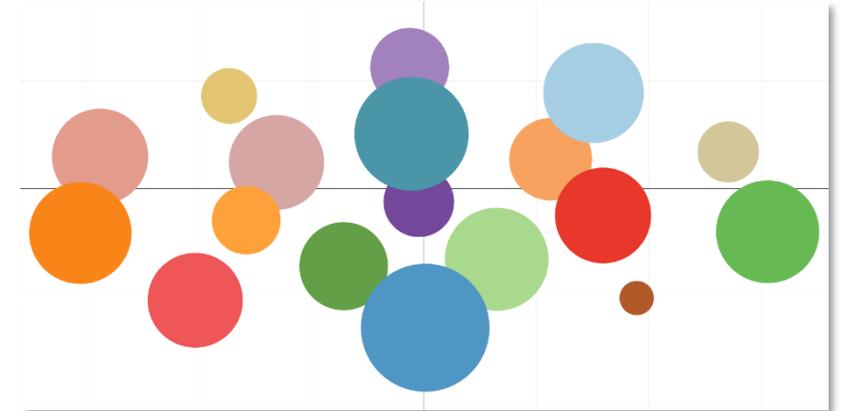
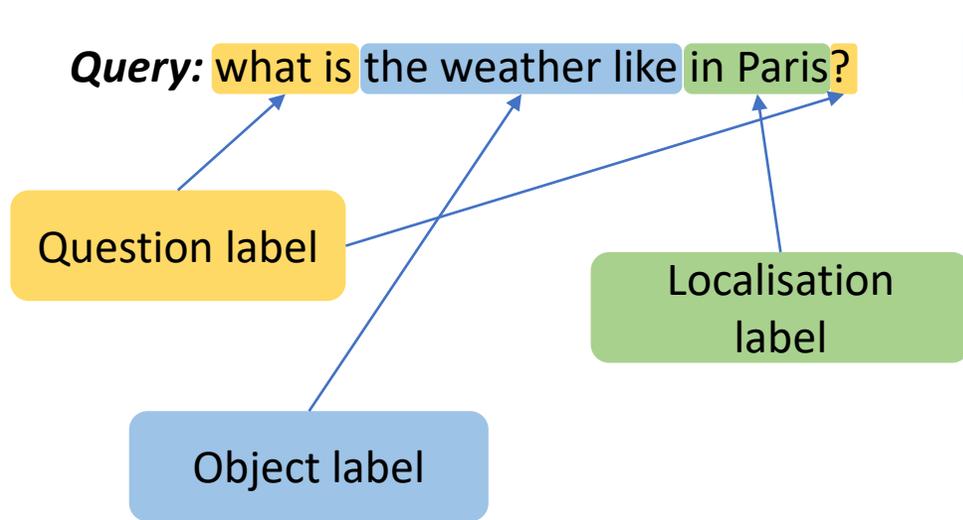
INFORMATION EXTRACTION / NER



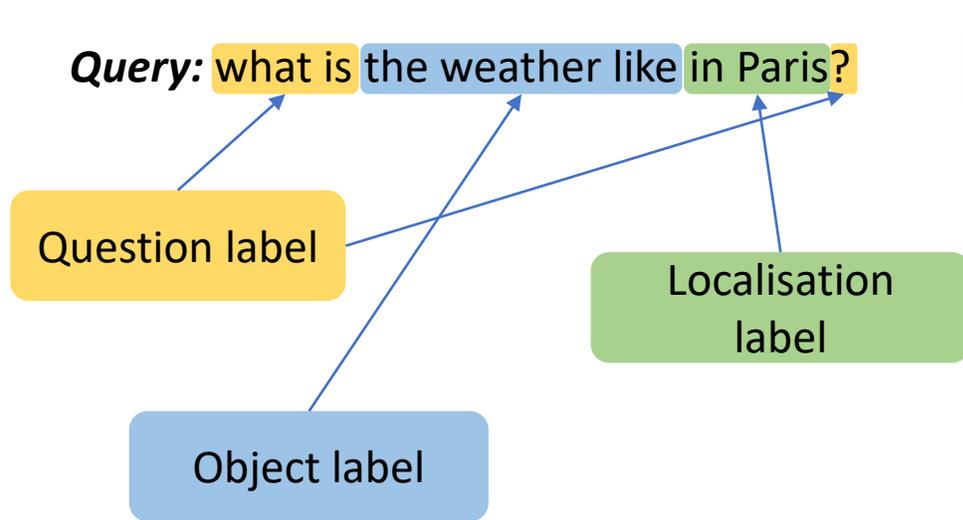
INFORMATION EXTRACTION / NER



INFORMATION EXTRACTION / NER

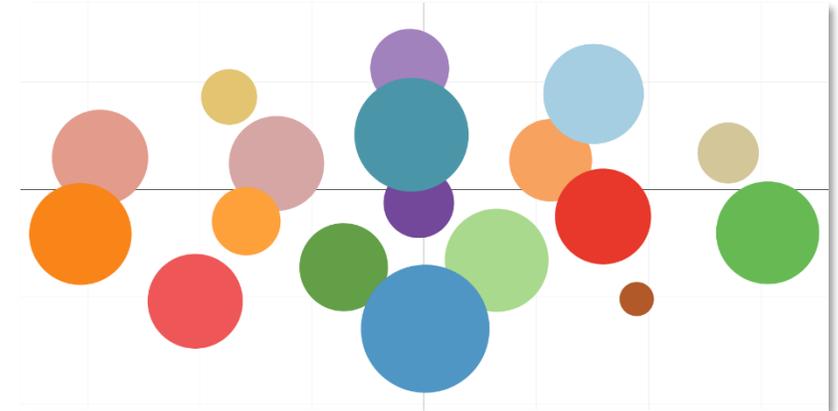
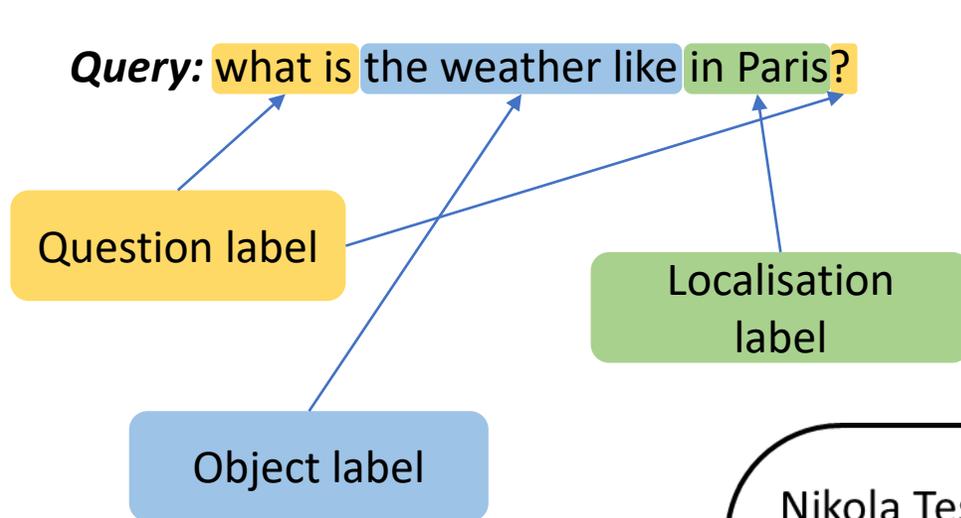


INFORMATION EXTRACTION / NER



Document:

INFORMATION EXTRACTION / NER



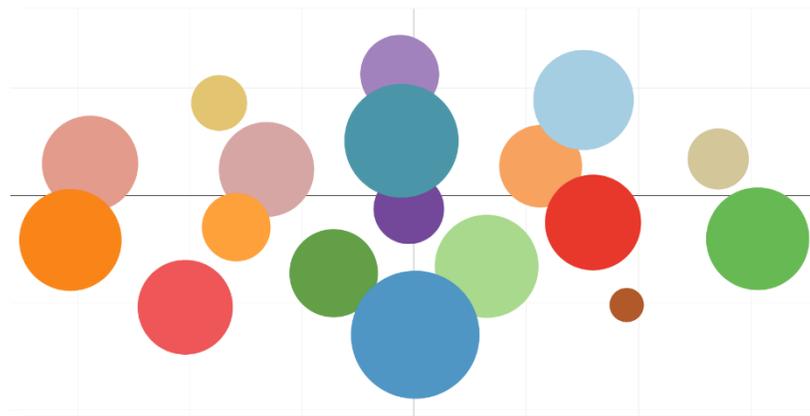
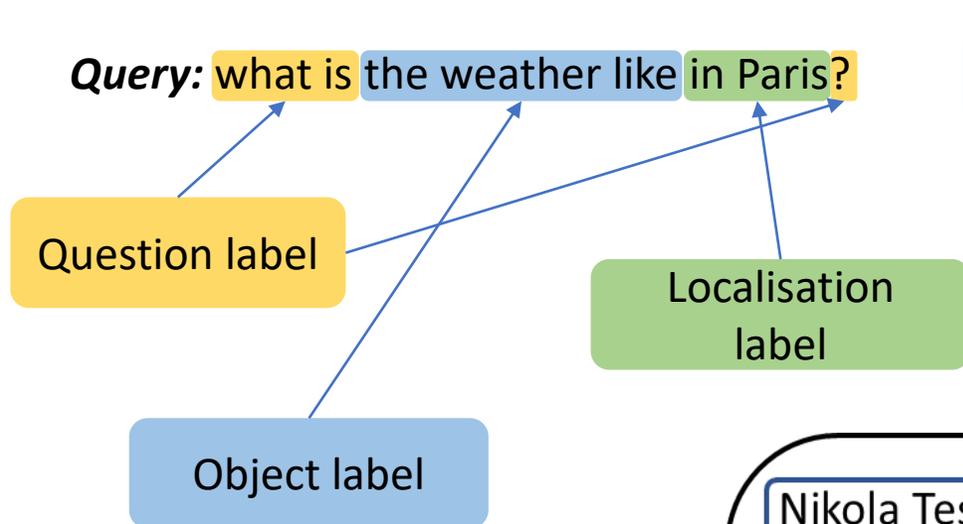
Document:

Nikola Tesla was born an ethnic Serb in the village Smiljan, Lika county, in the Austrian Empire (present day Croatia), on 10 July 1856. His father, Milutin Tesla (1819–1879), was an Eastern Orthodox priest.

Tesla's mother, Đuka Tesla (née Mandić; 1822–1892), whose father was also an Orthodox priest, had a talent for making home craft tools and mechanical appliances and the ability to memorize Serbian epic poems. Đuka had never received a formal education.

Tesla credited his eidetic memory and creative abilities to his mother's genetics and influence. Tesla's progenitors were from western Serbia, near Montenegro

INFORMATION EXTRACTION / NER



Document:

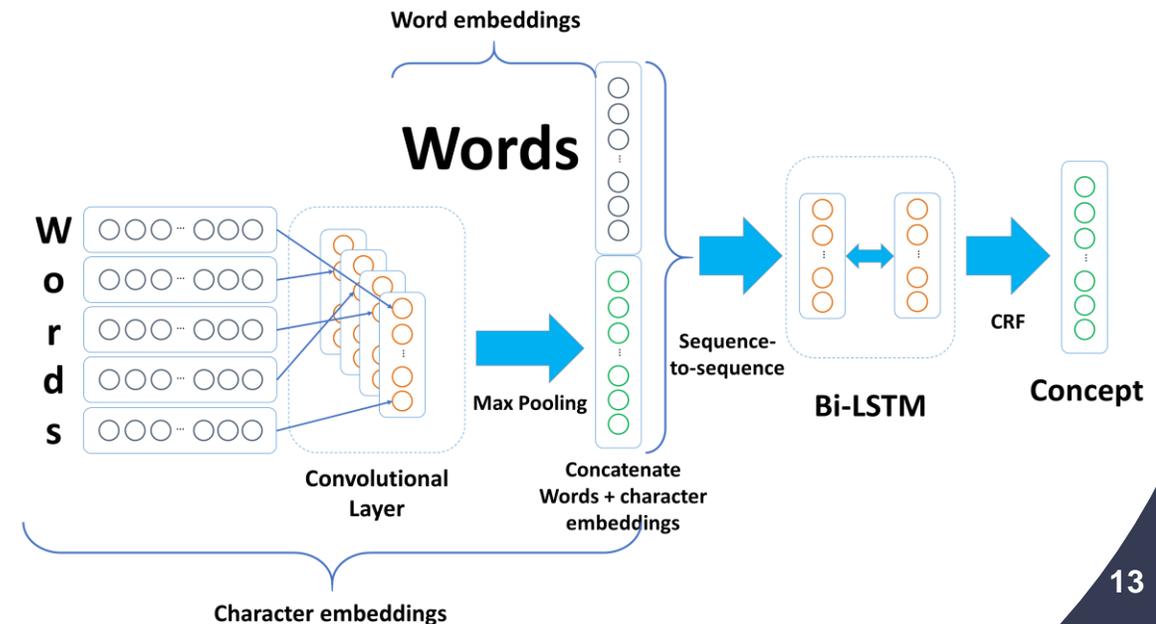
Nikola Tesla was born an ethnic Serb in the village Smiljan, Lika county in the Austrian Empire (present day Croatia), on 10 July 1856. His father, Milutin Tesla (1819–1879), was an Eastern Orthodox priest. Tesla's mother, Đuka Tesla (née Mandić, 1822–1892), whose father was also an Orthodox priest, had a talent for making home craft tools and mechanical appliances and the ability to memorize Serbian epic poems. Đuka had never received a formal education. Tesla credited his eidetic memory and creative abilities to his mother's genetics and influence. Tesla's progenitors were from western Serbia, near Montenegro

INFORMATION EXTRACTION / NER

- Use of OpenNMT toolkit
 - ⇒ MT approach
- Model inspired from [Ma and Hovy, 2016]
- Best performances on SLU SoA tasks (ATIS, MEDIA)

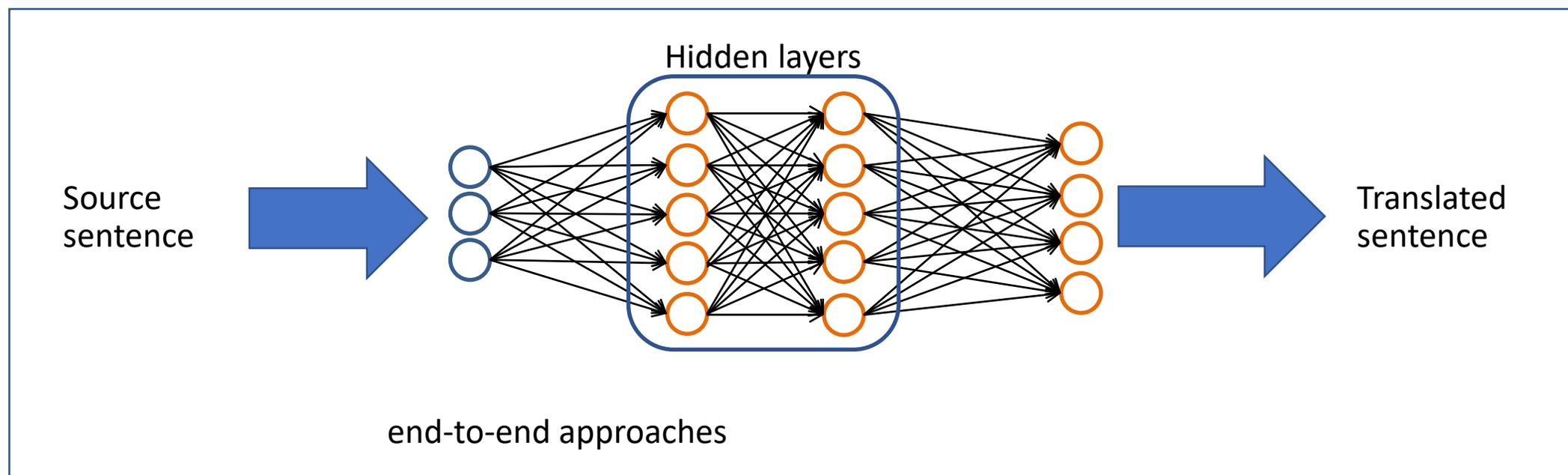
INFORMATION EXTRACTION / NER

- Use of OpenNMT toolkit
⇒ MT approach
- Model inspired from [Ma and Hovy, 2016]
- Best performances on SLU SoA tasks (ATIS, MEDIA)



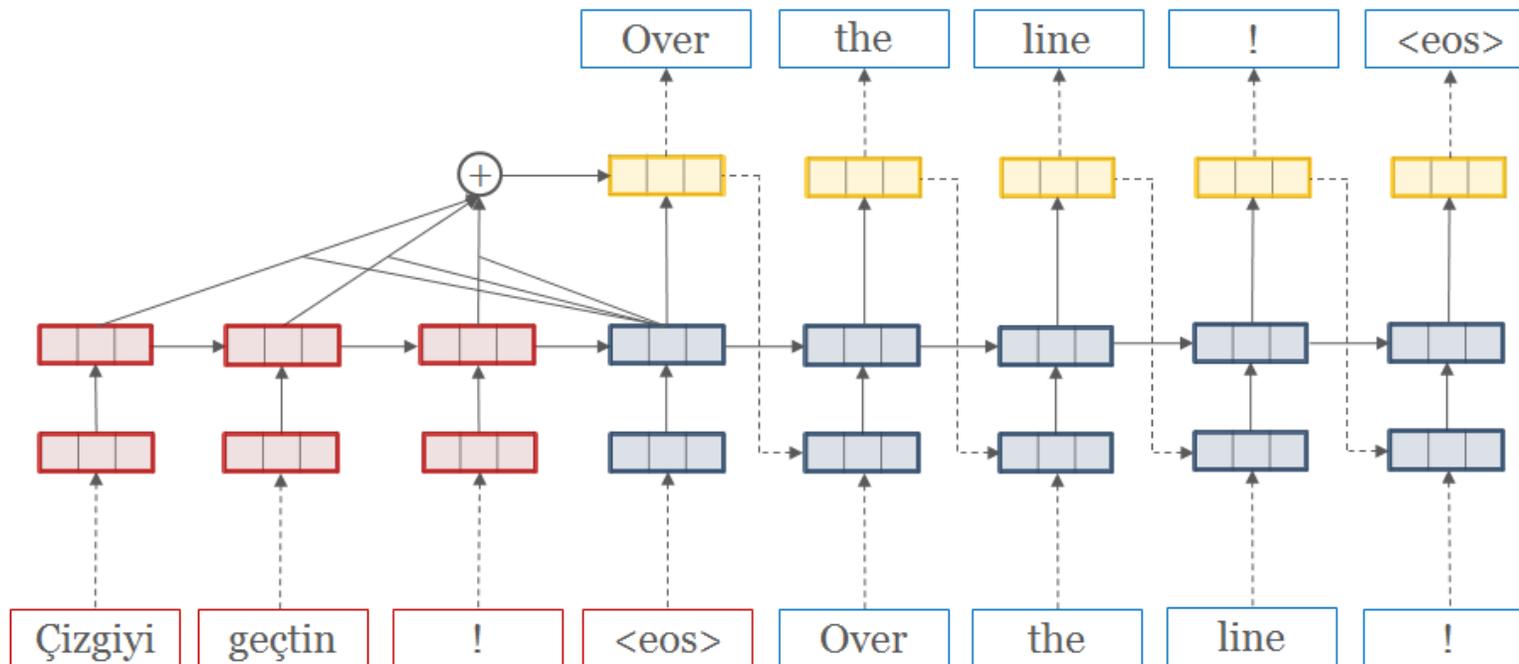
NEURAL MACHINE TRANSLATION

- Replace the TM and LM in the translation process [Kalchbrenner and Blunsom, 2013; Sutskever et al., 2014; Bahdanau et al., 2015; Luong et al., 2015; Jean et al., 2015]



NEURAL MACHINE TRANSLATION

- Sequence-to-Sequence Approach
 - With attention models
- Transformer models (a.k.a full attention model)

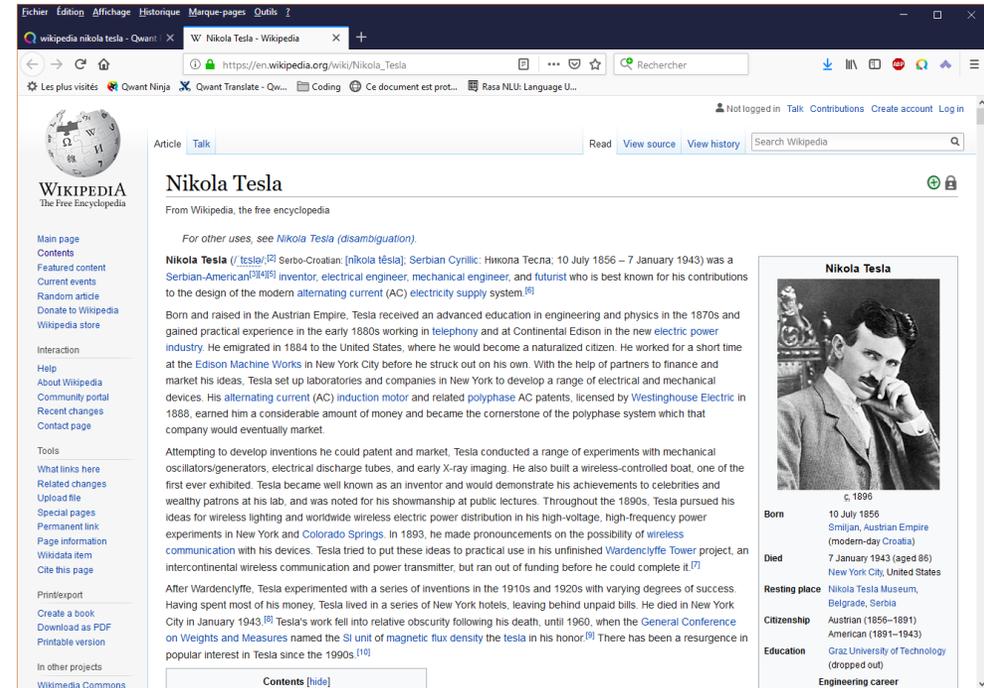


QUERYING IN SEVERAL LANGUAGES

Query: Nikola Tesla wikipedia

QUERYING IN SEVERAL LANGUAGES

Query: Nikola Tesla wikipedia



The screenshot shows a web browser window displaying the English Wikipedia article for Nikola Tesla. The browser's address bar shows the URL https://en.wikipedia.org/wiki/Nikola_Tesla. The page title is "Nikola Tesla" and it includes a search bar and navigation links like "View source" and "View history".

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

In other projects
Wikimedia Commons

Article Talk

Read View source View history Search Wikipedia

Nikola Tesla

From Wikipedia, the free encyclopedia

For other uses, see Nikola Tesla (disambiguation).

Nikola Tesla (/ˈtɛslə/^[a]; Serbo-Croatian: [n̩ikolaˈtɛ̌sa]; Serbian Cyrillic: Никола Тесла; 10 July 1856 – 7 January 1943) was a Serbian-American^[a]^[b] inventor, electrical engineer, mechanical engineer, and futurist who is best known for his contributions to the design of the modern alternating current (AC) electricity supply system.^[b]

Born and raised in the Austrian Empire, Tesla received an advanced education in engineering and physics in the 1870s and gained practical experience in the early 1880s working in telephony and at Continental Edison in the new electric power industry. He emigrated in 1884 to the United States, where he would become a naturalized citizen. He worked for a short time at the Edison Machine Works in New York City before he struck out on his own. With the help of partners to finance and market his ideas, Tesla set up laboratories and companies in New York to develop a range of electrical and mechanical devices. His alternating current (AC) induction motor and related polyphase AC patents, licensed by Westinghouse Electric in 1888, earned him a considerable amount of money and became the cornerstone of the polyphase system which that company would eventually market.

Attempting to develop inventions he could patent and market, Tesla conducted a range of experiments with mechanical oscillators/generators, electrical discharge tubes, and early X-ray imaging. He also built a wireless-controlled boat, one of the first ever exhibited. Tesla became well known as an inventor and would demonstrate his achievements to celebrities and wealthy patrons at his lab, and was noted for his showmanship at public lectures. Throughout the 1890s, Tesla pursued his ideas for wireless lighting and worldwide wireless electric power distribution in his high-voltage, high-frequency power experiments in New York and Colorado Springs. In 1893, he made pronouncements on the possibility of wireless communication with his devices. Tesla tried to put these ideas to practical use in his unfinished Wardenclyffe Tower project, an intercontinental wireless communication and power transmitter, but ran out of funding before he could complete it.^[c]

After Wardenclyffe, Tesla experimented with a series of inventions in the 1910s and 1920s with varying degrees of success. Having spent most of his money, Tesla lived in a series of New York hotels, leaving behind unpaid bills. He died in New York City in January 1943.^[d] Tesla's work fell into relative obscurity following his death, until 1960, when the General Conference on Weights and Measures named the SI unit of magnetic flux density the tesla in his honor.^[d] There has been a resurgence in popular interest in Tesla since the 1990s.^[10]

Contents [hide]

Nikola Tesla



c. 1896

Born	10 July 1856 Smiljan, Austrian Empire (modern-day Croatia)
Died	7 January 1943 (aged 86) New York City, United States
Resting place	Nikola Tesla Museum, Belgrade, Serbia
Citizenship	Austrian (1856–1891) American (1891–1943)
Education	Graz University of Technology (dropped out)

Engineering career

QUERYING IN SEVERAL LANGUAGES

Query: Nikola Tesla wikipedia

The image displays two overlapping browser windows showing the Wikipedia article for Nikola Tesla. The top window shows the English version of the article, and the bottom window shows the French version. Both windows show the article title, a search bar, and a sidebar with navigation options. The French version includes a detailed biography section with a portrait of Tesla and a table of his life events.

English Wikipedia Article:

Nikola Tesla

Article Talk Read View source View history Search Wikipedia

Not logged in Talk Contributions Create account Log in

January 1943) was a own for his contributions

ysics in the 1870s and ew electric power worked for a short time rs to finance and and mechanical Westinghouse Electric in tem which that

with mechanical trolled boat, one of the to celebrities and ds. Tesla pursued his equency power if wireless nctlyffe Tower project, an complete it [7]

degrees of success. He died in New York General Conference s been a resurgence in

Nikola Tesla

c. 1899

Born	10 July 1856 Smiljan, Austrian Empire (modern-day Croatia)
Died	7 January 1943 (aged 86) New York City, United States
Resting place	Nikola Tesla Museum, Belgrade, Serbia
Citizenship	Austrian (1856–1891) American (1891–1943)
Education	Graz University of Technology (dropped out)
Engineering career	

French Wikipedia Article:

Nikola Tesla

Article Discussion Lire Voir le texte source Afficher l'historique Rechercher dans Wikipédia

Non connecté Discussion Contributions Créer un compte Se connecter

⚡ Pour les articles homonymes, voir *Nikola Tesla* (homonymie).

Ци Џи Cette page contient des caractères cyrilliques. En cas de problème, consultez Aide:Unicode ou testez votre navigateur.

Nikola Tesla (en serbe cyrillique : Никола Тесла), né le 10 juillet 1856 à Smiljan dans l'Empire d'Autriche (actuelle Croatie) et mort le 7 janvier 1943 à New York, est un inventeur et ingénieur naturalisé américain (en 1891) d'origine serbe. Il a principalement œuvré dans le domaine de l'électricité, mais était également ingénieur mécanique.

Tesla a d'abord travaillé dans la téléphonie et l'ingénierie électrique avant d'émigrer aux États-Unis en 1884 pour travailler avec Thomas Edison, avant de collaborer avec George Westinghouse qui enregistra un grand nombre de ses brevets. Considéré comme l'un des plus grands scientifiques dans l'histoire de la technologie, pour avoir déposé quelque 300 brevets couvrant au total 125 inventions² (qui seront pour beaucoup attribuées à tort à Edison)³ et avoir décrit de nouvelles méthodes pour réaliser la « conversion de l'énergie », Tesla est reconnu comme l'un des ingénieurs les plus créatifs de la fin du xx^e et du début du xx^e siècle. Quant à lui, il préférerait plutôt se définir comme un découvreur.

Ses travaux les plus connus et les plus largement diffusés portent sur l'énergie électrique. Il a mis au point les premiers alternateurs permettant la naissance des réseaux électriques de distribution en courant alternatif, dont il est l'un des pionniers. Tesla s'est beaucoup intéressé aux technologies modernes se focalisant sur l'électricité qui était le noyau de ses inventions. Il est connu pour avoir su mettre en pratique la découverte du caractère ondulatoire de l'électromagnétisme (théorisé par James Clerk Maxwell en 1864), en utilisant les fréquences propres des composants des circuits électriques afin de maximiser leur rendement.

De son vivant, Tesla était renommé pour ses inventions ainsi que pour son sens de la mise en scène, faisant de lui un archétype du « savant fou ». Grand humaniste qui aurait notamment aimé que l'électricité soit gratuite et accessible à tous⁴, il resta malgré tout dans

Nikola Tesla

Nikola Tesla

Tesla en 1890 (à 34 ans), photographié par Napoléon Sarony.

Biographie	
Naissance	10 juillet 1856 Smiljan (Empire d'Autriche)

QUERYING IN SEVERAL LANGUAGES

Query: Nikola Tesla wikipedia

The image displays three overlapping browser windows showing the Wikipedia article for Nikola Tesla in different languages:

- Top window (English):** Shows the English Wikipedia article for Nikola Tesla. The URL is https://en.wikipedia.org/wiki/Nikola_Tesla. The article title is "Nikola Tesla".
- Middle window (French):** Shows the French Wikipedia article for Nikola Tesla. The URL is https://fr.wikipedia.org/wiki/Nikola_Tesla. The article title is "Nikola Tesla".
- Bottom window (German):** Shows the German Wikipedia article for Nikola Tesla. The URL is https://de.wikipedia.org/wiki/Nikola_Tesla. The article title is "Nikola Tesla".

The German article includes a table of contents (Inhaltsverzeichnis) and a biographical section (Leben) with a portrait of Nikola Tesla from 1890.

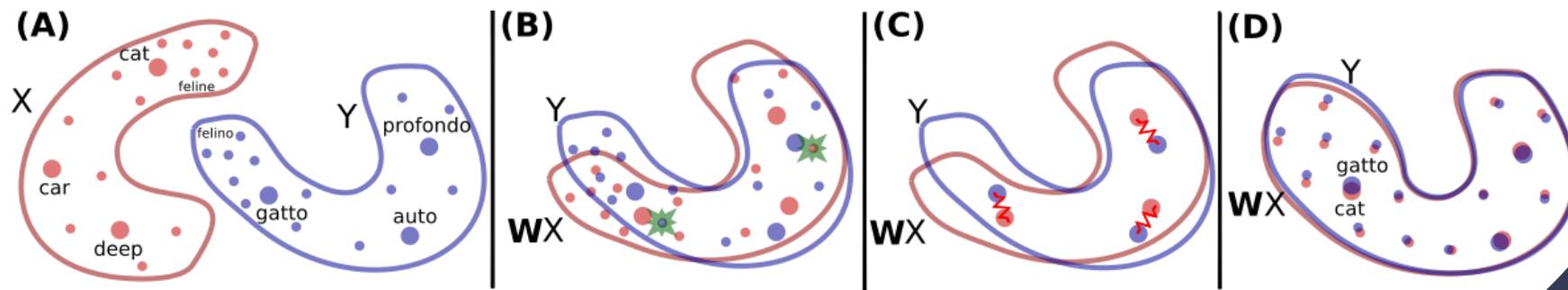
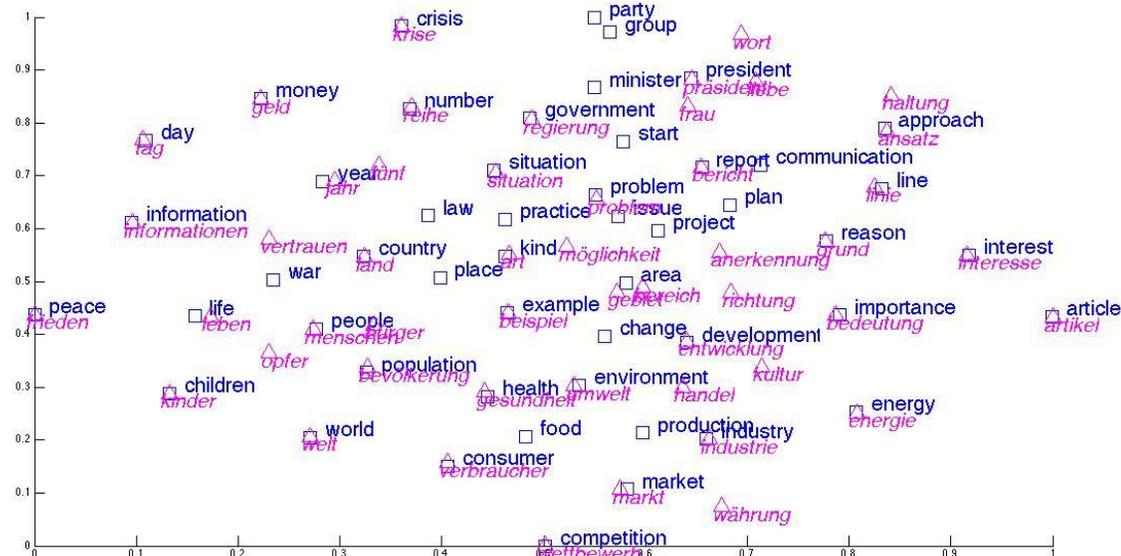
Inhaltsverzeichnis [Verbergen]	
1	Leben
1.1	Herkunft und Ausbildung
1.2	Erste berufliche Anstellungen
1.3	Umszug nach New York
1.4	1889 bis 1900
1.5	1900 bis 1912
1.6	1913 bis 1943
1.7	Auszeichnungen und Ehrendoktorwürden
2	Ehrungen
2.1	Denkmäler
2.2	Tesla als Namensgeber
2.3	Tesla-Museen
2.4	Banknoten und Gedenkmünzen
2.5	Erwähnungen in der Populärkultur
3	Schriften
4	Literatur
5	Weblinks
	Einzelnachweise

Biographie	
Naissance	10 juillet 1856 Smiljan (Empire d'Autriche)

QUERYING IN SEVERAL LANGUAGES

Embeddings multilingues:

- BIVEC
 - Joint learning of word embeddings for two languages
- MUSE
 - Independent learning of several language
 - Cross-lingual alignment



QUERYING IMAGES IN SEVERAL LANGUAGES

Query: eine Frau Geige spielt

QUERYING IMAGES IN SEVERAL LANGUAGES

Query: eine Frau Geige spielt



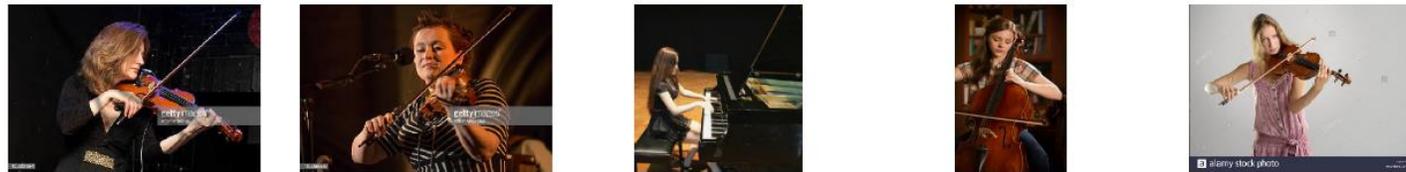
(a) "eine Frau Geige spielt"

QUERYING IMAGES IN SEVERAL LANGUAGES

Query: eine Frau Geige spielt



(a) "eine Frau Geige spielt"



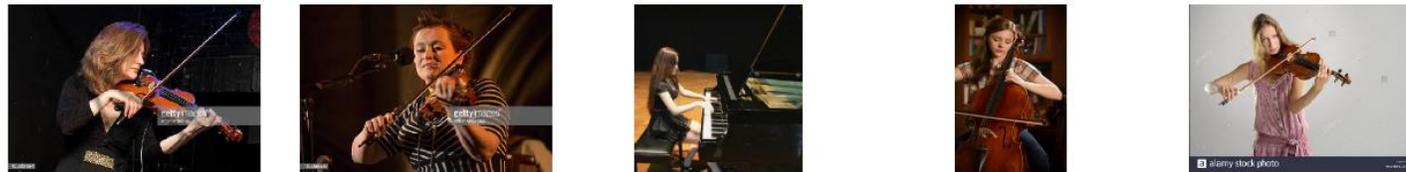
(b) "Une femme jouant du violon"

QUERYING IMAGES IN SEVERAL LANGUAGES

Query: eine Frau Geige spielt



(a) "eine Frau Geige spielt"

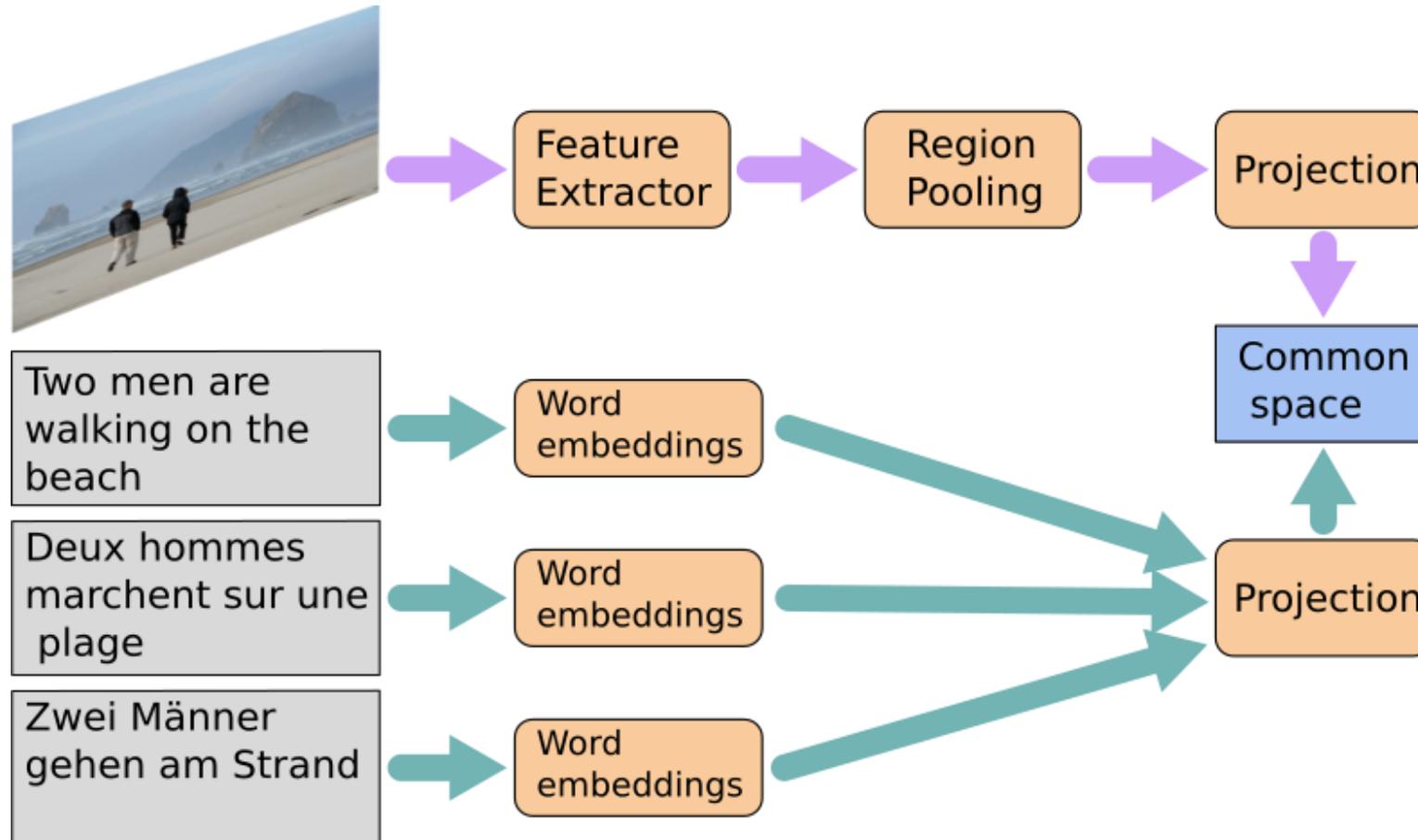


(b) "Une femme jouant du violon"



(c) "A woman playing violon"

QUERYING IN SEVERAL LANGUAGES



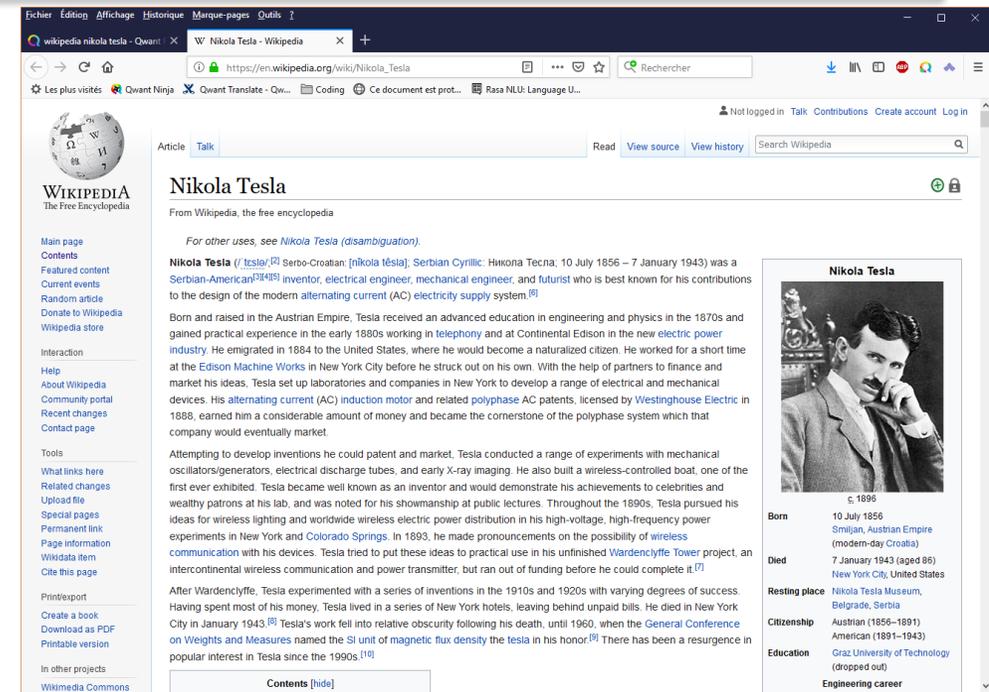
QUESTION ANSWERING IN NATURAL LANGUAGE

- **Query:** When was born Nikolas Tesla?

QUESTION ANSWERING IN NATURAL LANGUAGE

- **Query:** When was born Nikolas Tesla?

« **Nikola Tesla** (/ˈtɛslə/^[2]; Serbo-Croatian: [n̩kɔla t̩ɛsla]; **Serbian Cyrillic**: Никола Тесла; 10 July 1856 – 7 January 1943) was a **Serbian-American**^{[3][4][5]} **inventor**, **electrical engineer**, **mechanical engineer**, and **futurist** who is best known for his contributions to the design of the modern **alternating current** (AC) **electricity supply** system. »



The screenshot shows the Wikipedia article for Nikola Tesla. The article text is as follows:

Nikola Tesla (/ˈtɛslə/^[2]; Serbo-Croatian: [n̩kɔla t̩ɛsla]; **Serbian Cyrillic**: Никола Тесла; 10 July 1856 – 7 January 1943) was a **Serbian-American**^{[3][4][5]} **inventor**, **electrical engineer**, **mechanical engineer**, and **futurist** who is best known for his contributions to the design of the modern **alternating current** (AC) **electricity supply** system.^[6]

Born and raised in the Austrian Empire, Tesla received an advanced education in engineering and physics in the 1870s and gained practical experience in the early 1880s working in telephony and at Continental Edison in the new electric power industry. He emigrated in 1884 to the United States, where he would become a naturalized citizen. He worked for a short time at the Edison Machine Works in New York City before he struck out on his own. With the help of partners to finance and market his ideas, Tesla set up laboratories and companies in New York to develop a range of electrical and mechanical devices. His **alternating current** (AC) **induction motor** and related **polyphase AC** patents, licensed by Westinghouse Electric in 1888, earned him a considerable amount of money and became the cornerstone of the polyphase system which that company would eventually market.

Attempting to develop inventions he could patent and market, Tesla conducted a range of experiments with mechanical oscillators/generators, electrical discharge tubes, and early X-ray imaging. He also built a wireless-controlled boat, one of the first ever exhibited. Tesla became well known as an inventor and would demonstrate his achievements to celebrities and wealthy patrons at his lab, and was noted for his showmanship at public lectures. Throughout the 1890s, Tesla pursued his ideas for wireless lighting and worldwide wireless electric power distribution in his high-voltage, high-frequency power experiments in New York and Colorado Springs. In 1893, he made pronouncements on the possibility of **wireless communication** with his devices. Tesla tried to put these ideas to practical use in his unfinished **Wardenclyffe Tower** project, an intercontinental wireless communication and power transmitter, but ran out of funding before he could complete it.^[7]

After Wardenclyffe, Tesla experimented with a series of inventions in the 1910s and 1920s with varying degrees of success. Having spent most of his money, Tesla lived in a series of New York hotels, leaving behind unpaid bills. He died in New York City in January 1943.^[8] Tesla's work fell into relative obscurity following his death, until 1960, when the **General Conference on Weights and Measures** named the SI unit of magnetic flux density the tesla in his honor.^[9] There has been a resurgence in popular interest in Tesla since the 1990s.^[10]

Contents [hide]

Biographical information

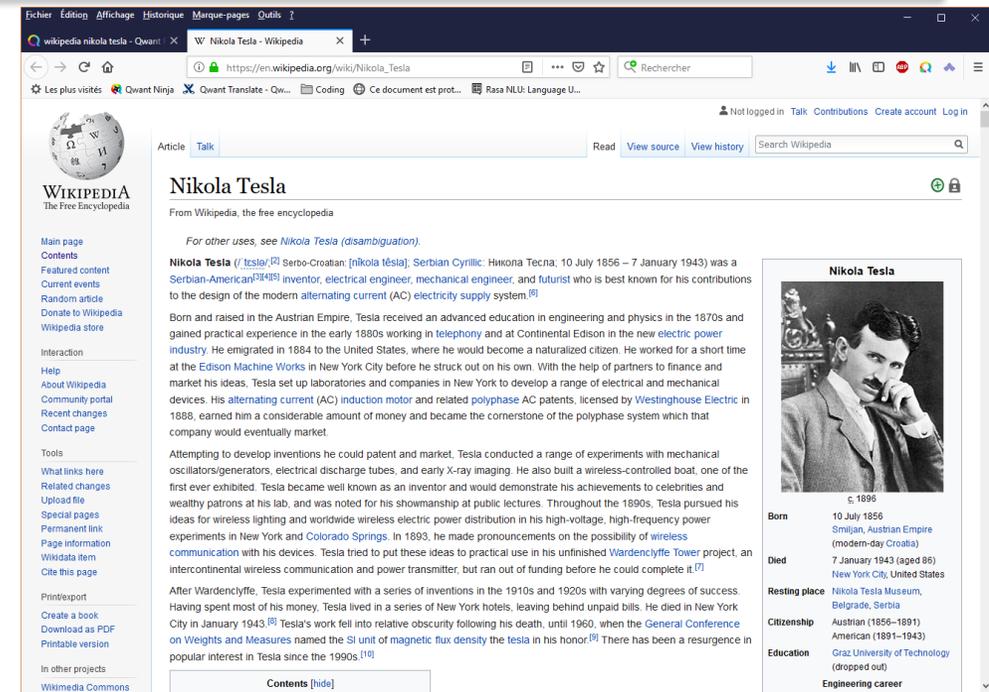
Born	10 July 1856 Smiljan, Austrian Empire (modern-day Croatia)
Died	7 January 1943 (aged 86) New York City, United States
Resting place	Nikola Tesla Museum, Belgrade, Serbia
Citizenship	Austrian (1856–1891) American (1891–1943)
Education	Graz University of Technology (dropped out)
Engineering career	

QUESTION ANSWERING IN NATURAL LANGUAGE

- **Query:** When was born Nikolas Tesla?

« **Nikola Tesla** (/ˈtɛslə/^[2] Serbo-Croatian: [n̩kɔla t̩ɛsla]; **Serbian Cyrillic:** Никола Тесла; 10 July 1856 – 7 January 1943) was a **Serbian-American**^{[3][4][5]} **inventor**, **electrical engineer**, **mechanical engineer**, and **futurist** who is best known for his contributions to the design of the modern **alternating current** (AC) **electricity supply** system. »

10 July 1856



The screenshot shows the Wikipedia page for Nikola Tesla. The title is "Nikola Tesla" and the subtitle is "From Wikipedia, the free encyclopedia". The main text begins with "Nikola Tesla (/ˈtɛslə/^[2] Serbo-Croatian: [n̩kɔla t̩ɛsla]; Serbian Cyrillic: Никола Тесла; 10 July 1856 – 7 January 1943) was a Serbian-American^{[3][4][5]} inventor, electrical engineer, mechanical engineer, and futurist who is best known for his contributions to the design of the modern alternating current (AC) electricity supply system.^[6]"

Below the text is a portrait of Nikola Tesla, captioned "Nikola Tesla" and "c. 1895". To the right of the portrait is a table with the following information:

Born	10 July 1856 Smiljan, Austrian Empire (modern-day Croatia)
Died	7 January 1943 (aged 86) New York City, United States
Resting place	Nikola Tesla Museum, Belgrade, Serbia
Citizenship	Austrian (1856–1891) American (1891–1943)
Education	Graz University of Technology (dropped out)
Engineering career	

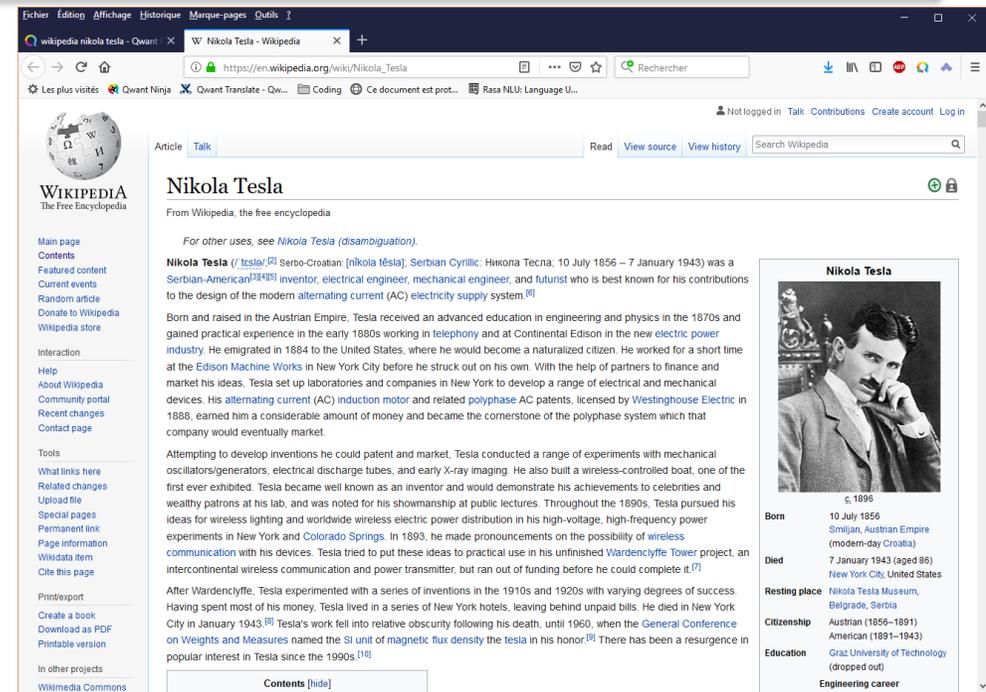
QUESTION ANSWERING IN NATURAL LANGUAGE

- **Query:** When was born Nikolas Tesla?

« **Nikola Tesla** (/ˈtɛslə/^[2] Serbo-Croatian: [n̩kɔla t̩ɛsla]; **Serbian Cyrillic:** Никола Тесла; 10 July 1856 – 7 January 1943) was a **Serbian-American**^{[3][4][5]} **inventor**, **electrical engineer**, **mechanical engineer**, and **futurist** who is best known for his contributions to the design of the modern **alternating current** (AC) **electricity supply** system. »

10 July 1856

« Nikolas Tesla was born on the 10 July 1856 »



The screenshot shows the Wikipedia article for Nikola Tesla. The article text includes: "Nikola Tesla (/ˈtɛslə/^[2] Serbo-Croatian: [nikola t̩ɛsla]; Serbian Cyrillic: Никола Тесла; 10 July 1856 – 7 January 1943) was a Serbian-American^{[3][4][5]} inventor, electrical engineer, mechanical engineer, and futurist who is best known for his contributions to the design of the modern alternating current (AC) electricity supply system.^[6]" The article also features a portrait of Tesla and a biographical sidebar with details such as his birth date (10 July 1856), death date (7 January 1943), and various awards and affiliations.



LEARNING PROCESSES IN A PRACTICAL WAY

LEARNING PROCESSES

- Lightly supervised learning
- Bootstrapping
- Continuous learning

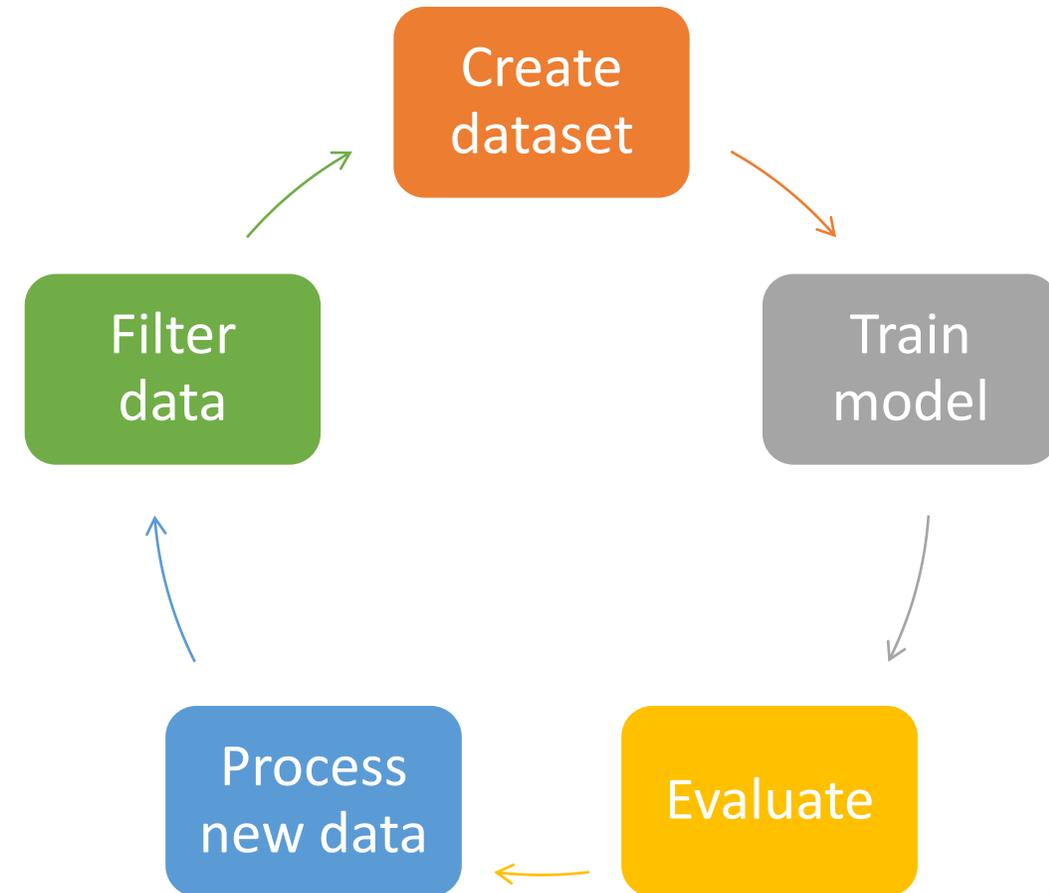
LIGHTLY SUPERVISED LEARNING

- Automatic extraction of information
 - Query & clic logs, databases, etc.
- Automatic generation of example using templates
 - I want a hotel in \${CITY} for 3 days
- Clustering
 - Automatic extraction of interesting clusters
 - Selection of interesting clusters (manually done)

⇒ Human in the loop

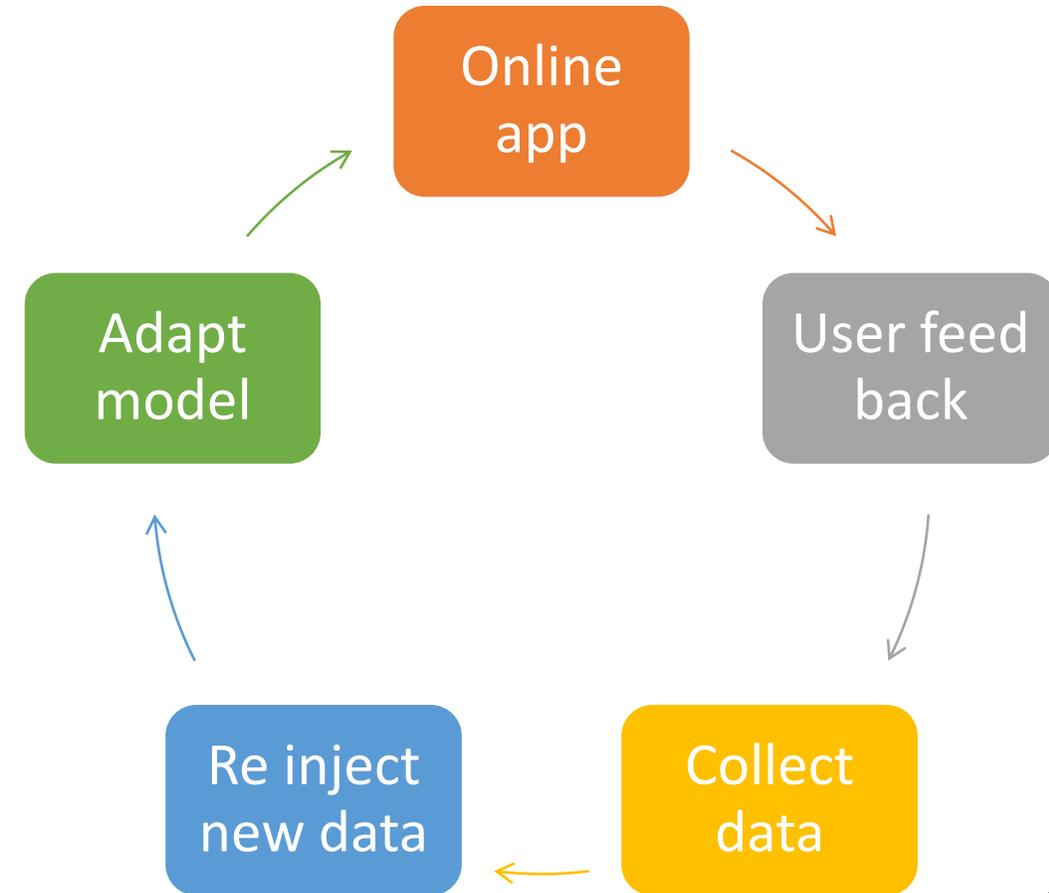
BOOTSTRAPPING

- Start from a small amount of data
- Train and evaluate it
- Apply the new model on new data
- Filter them / correction
- Add them into the new data



CONTINUOUS LEARNING

- From running data & models
- Request user feedback
- Collect data
- Re-inject them into the model
- Adapt model



CONCLUSION



NLP 4 IR

- Language Models
- Machine Translation
- Intention Detection
- NER
- Spoken and Natural Language Understanding
- Word Embeddings
- ...and many others!

PROJECTS

• **Projets internes:**

- Détection d'Intention
- Shopping
- Agent conversationnel (Machine Reading, Thèse)
- Qwant Junior (QR, calcul, définitions, conjugaisons...)
- Traduction Automatique

• **Projets Externes:**

- Cosy Cloud: indexation
- Caliopen: classification de courrier électronique

• **Projets Subventionnés:**

- INRIA CodeScope: Analyse de code & indexation
- INRIA Fake NewsPIA: Fact checking, Comprehension, Classification
- H2020 Answers: Analyse de sentiment
- H2020 AI4EU: Indexation de contenu web
- PIA Social Truth: Analyse de sentiment, reseaux sociaux

NLP 4 IR CHALLENGES

- Improving IR Precision
 - Scaling (searching in a billion document collection)
 - Speed (answering in less than ms)
- ⇒ improve user satisfaction
- ⇒ improve user experience
- ⇒ github.com/QwantResearch/

THE END!
QUESTIONS?

[c.servan\[AT\]qwant\[DOT\]com](mailto:c.servan@qwant.com)

