# PROJET PENSÉES PROFONDES

December 18, 2014

École Normale Supérieure de Lyon

Raphaël CHARRONDIÈRE

Marc CHEVALIER

Quentin CORMIER

Tom Cornebize

Yassine Hamoudi

Thomas Pellissier Tanon

Valentin LORENTZ

Adviser: Eddy Caron

#### A NEED FOR ANSWERS

#### Who was the wife of Louis Pasteur?

Буряад Čeština Deutsch

Français Gaeilge Gaidhlig

Galego 客家III/Hak-kā-ngi

Hrvatski

Italiano daframeno Kurdi

Latina

Latvieti

Lietuviu

Létzebuergesch

Pasteur was born, Dole

#### Early life

Louis Pasteur was born on December 27, 1822, in Dole, Jura, France, to a Catholic family of a poor tanner. He was the third child of Jean-Joseph Pasteur and Jeanne-Etiennette Roqui. In 1827, the family moved to Arbois, where he entered primary school in 1831. He was an average student in his early years, and not particularly academic, as his interests were fishing and sketching. His pastels and portraits of his parents and friends, made when he was 15,

were later kept in the museum of the Pasteur Institute in Paris. In 1838, he left for Paris to join the Institution Barbet, but became homesick and returned in November. In 1839, he entered the Collège Royal de Besancon and earned his baccalauréat (BA) degree in 1840. He was appointed teaching assistant at the Besancon college while continuing a degree science course with special mathematics. He failed his first examination in 1841. He managed to pass the baccalauréal scientifique (general science) degree in 1842 from Dijon but with a poor grade in chemistry. After one failed attempt for the entrance test for the École Normale Supérieure in Paris in 1842, he succeeded in 1844. In 1845 he received the licencié és sciences (Bachelor of Science) degree. In 1846, he was appointed professor of physics at the Collège de Tournon at Ardèche, but Antoine Jérome Baland (one of the discoverers of the element bromine) wanted him back at the École Normale Supérieure as a graduate assistant (préparateur) for chemistry courses. He joined Balard and simultaneously started his research in crystallography and in 1847, he submitted his two theses, one in chemistry and the other in physics. After serving briefly as professor of physics at the Diion Lyone in 1848, he became professor of chemistry at the University of Strasbourg, where he met and courted Marie Laurent, daughter of the university's rector in 1849. They were married on May 29, 1849, and together had five children, only two of whom survived to adulthood; the other three died of typhoid. These personal tragedies were his motivations for curing infectious diseases [2]9]

#### Professional career

Pasteur was appointed to the Chair of Chemistry in the faculty of sciences of the University of Strasbourg, In 1854, he was named dean of the new faculty of sciences at Lilie

University. It was on this occasion that Pasteur uttered his oft-quoted remark: "dans les champs de l'observation, le hasard ne favorise que les esprits préparés." (In the field of observation, chance favors only the prepared mind (10) In 1857, he moved to Paris as the director of scientific studies at the École Normale Supérieure where he took control from 1858 to 1867 and introduced a series of reforms. The

examinations became more rigid, which led to better results, greater competition, and increased prestige. He raised the standard of scientific work, leading to two serious student revolts, [7] In 1862, he was appointed professor of geology, physics, and chemistry at the École nationale supérieure des Beaux-Arts, the position which held until his resignation in 1967. In Paris, he established the Pasteur Institute in 1987, in which he was its director for the rest of his life (3)(4)9

#### Research contributions

#### Molecular asymmetry

In Pasteur's early work as a chemist, he resolved a problem concerning the nature of tartaric acid (1848).[11][12][13][14][15] A solution of this compound derived from living things (specifically, wine lees) rotated the plane of polarization of light passing through it. The mystery was that tartaric acid derived by chemical synthesis had no such effect, even though its chemical reactions were identical and its elemental composition was the same, [18] This was the first time anyone had demonstrated chiral molecules,

#### Fermentation and germ theory of diseases

Pasteur demonstrated that fermentation is caused by the growth of micro-organisms, and the emergent growth of bacteria in nutrient broths is due not to spontaneous generation, but rather to biogenesis (Omne vivum ex vivo "all life from life"). He was motivated to investigate the matter while working at Life. In 1856 a local wine manufacturer, M. Bioot, the father of his student, sought for his advice on the problems of making beetroot alcohol and souring after long storage. [17] He could not immediately pursue the matter mainly



#### **EXISTING TOOLS**

WolframAlpha Marie Pasteur

Google Wikipedia page of Marie Pasteur Bing Wikipedia page of Louis Pasteur

Yahoo answers.com page containing this question

Closed-source, bad on some questions.

# PPP INTRODUCES PLATYPUS

http://ppp.pony.ovh/



# Demo time!

PPP Project website

#### Ask me anything alpha version

Who is Alan Turing?

٥

#### Query

Link to this request

#### Who is Alan Turing?

Result

Show internal results

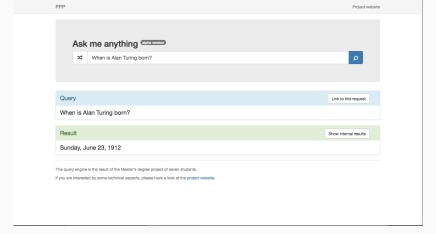
#### Alan Turing III ®

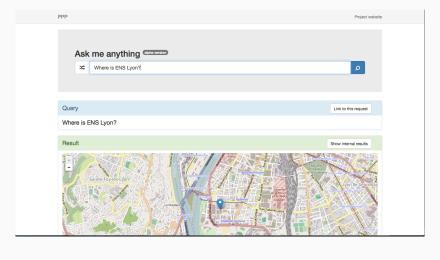
Alan Mathison Turing, OBE, FRS (Ylpsny) TENPT-ing; 23 June 1912 – 7 June 1954) was a British mathematician, logician, cryptanalyst, philosopher, pioneering computer scientist, mathematical biologist, and marathon and ultra distance runner. He was highly influential in the development of computer science, providing a formalisation of the concepts of "algorithm" and "computation" with the Turing machine, which can be considered a model of a general purpose computer. Turing is widely considered to be the father of theoretical computer science and artificial Intelligence. Why considered to the concepts of the control of the cont



This query engine is the result of the Master's degree project of seven students.

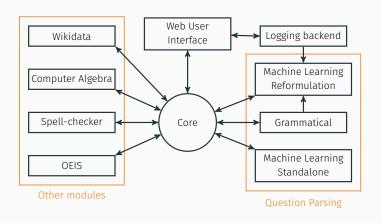
If you are interested by some technical aspects, please have a look at the project website.







#### **ARCHITECTURE**



#### **DATAMODEL**

How to represent formally a question asked in natural language?

Tree structure.

#### DATAMODEL - LEAVES: RESOURCES

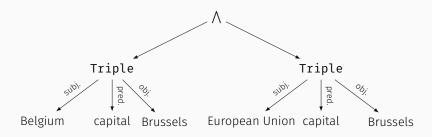
- · Strings: "Isaac Newton"
- · Dates: 25 December 1642
- · Geographic coordinates: 47° 30′ 18″ N; 9° 44′ 57″ E

٠ ..

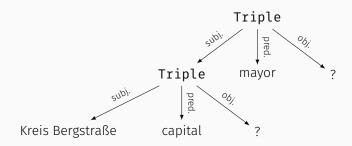
#### DATAMODEL - NODES: OPERATIONS

- Full triples: (Isaac Newton, birth date, 25 December 1642 (Julian))
   → true
- Missing: (?, birth date, 25 December 1642 (Julian)) → [Isaac Newton]
- · Union, Intersection, Difference, ...
- · And, Or, Not
- · Exists
- · Sort, First, Last

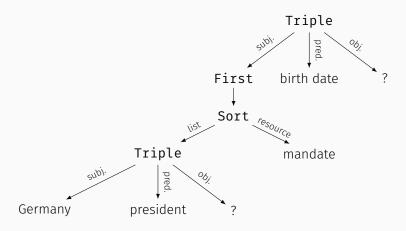
"Is Brussels the capital of Belgium and the European Union?"



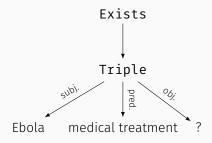
"Who is the mayor of the capital of Kreis Bergstraße?"



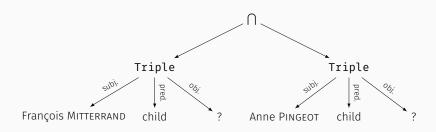
"What is the birth date of the first president of Germany?"



"Is there a medical treatment for Ebola?"



"Who are the children of François MITTERRAND and Anne PINGEOT?"

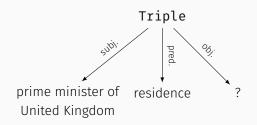


# QUESTION PARSING

Map the input question to the most relevant normal form.

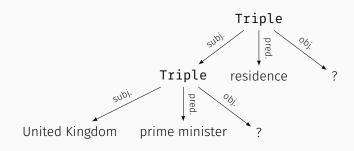
Map the input question to the most relevant normal form.

Where does the prime minister of United Kingdom live?



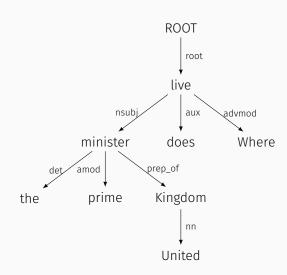
Map the input question to the most relevant normal form.

Where does the prime minister of United Kingdom live?

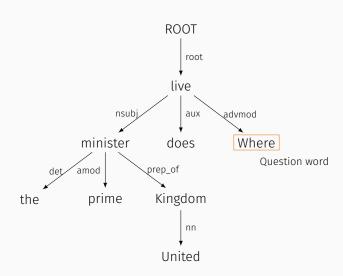


Grammatical approach

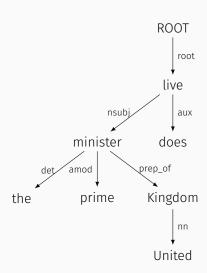
#### DEPENDENCY TREE OUTPUT BY THE STANFORD PARSER

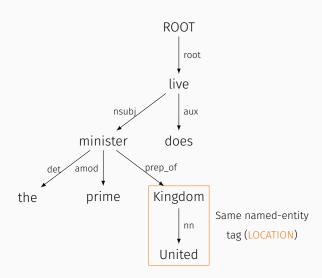


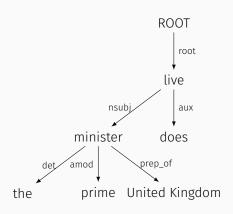
### **IDENTIFY QUESTION WORD**

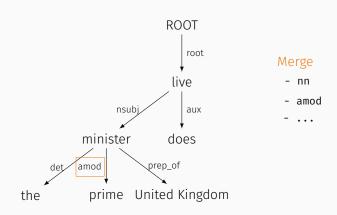


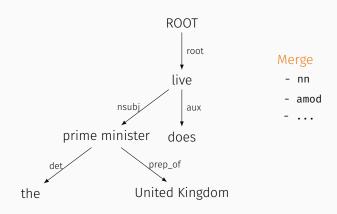
## **IDENTIFY QUESTION WORD**



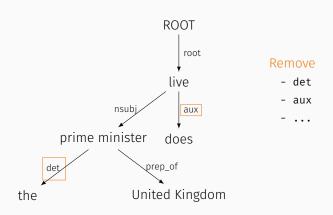




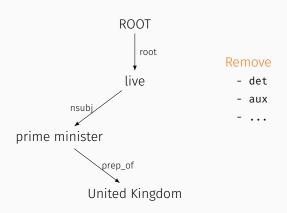




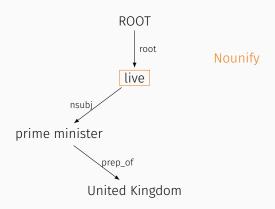
## **REMOVAL**



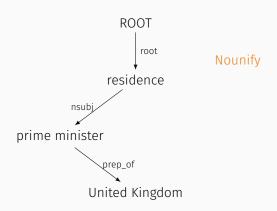
## **REMOVAL**



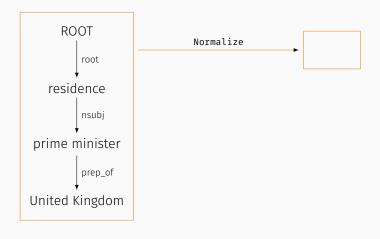
# **NOUNIFICATION**

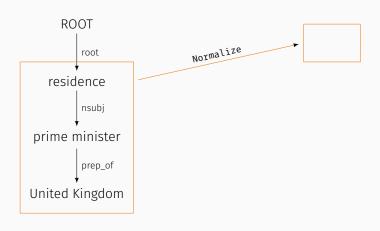


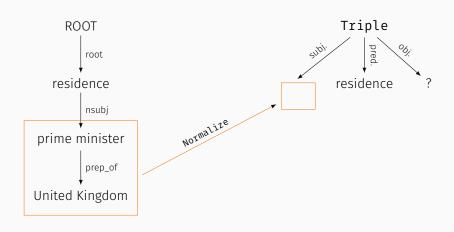
# **NOUNIFICATION**

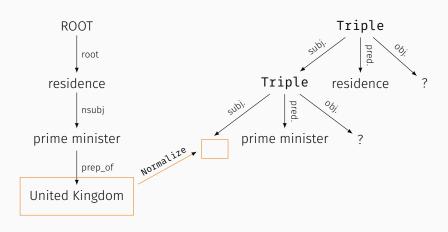


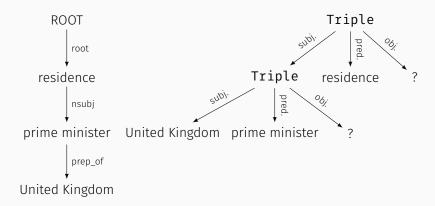
## **NORMALIZATION**







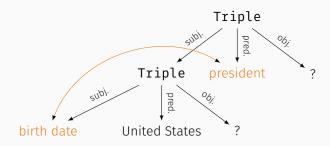






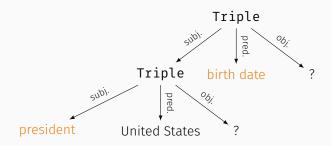
#### REFORMULATION: A TRANSLATION MODULE

When was the president of the United States born?



#### REFORMULATION: A TRANSLATION MODULE

When was the president of the United States born?

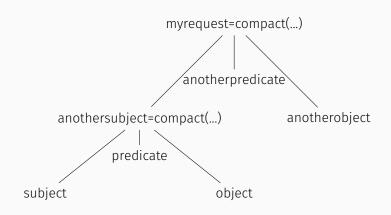


#### **OPERATION**

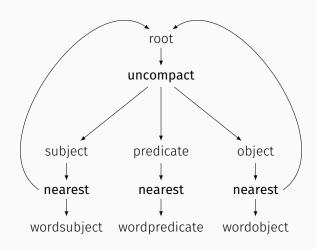
Projection: trees  $\mapsto$  vector space.

The dictionary associates a vector triple to a vector.

# **PROJECTION**



#### FROM VECTOR TO TREE





# **KEYWORDS QUESTIONS**

United States president

president United States

### GOAL OF THE MODULE

Classify each word into one of the four categories:

{to ignore, subject, predicate, object }

#### **EXAMPLE OF CLASSIFICATION**

```
(United, subject) (States, subject) (president, predicate) (president, predicate) (United, subject) (States, subject)
```

#### LOOK-UP TABLE

Word 
$$w \mapsto \text{Vector } V_w \in \mathbb{R}^{25}$$

If two words  $w_1$  and  $w_2$  are synonymous then,

$$\|V_{w_2} - V_{w_1}\|_2 \approx 0$$

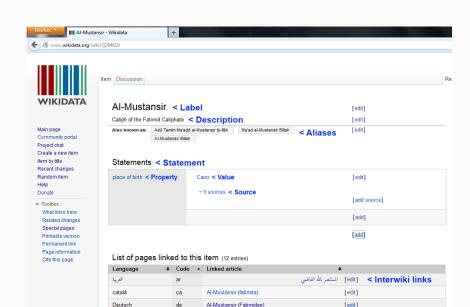
#### **BACK-END**



# WIKIDATA (http://wikidata.org)?

- · A free knowledge base
- · Structured version of Wikipedia
- · 12 millions of entries
- · Multilingual

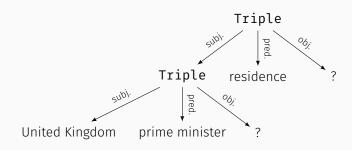
#### AN ITEM

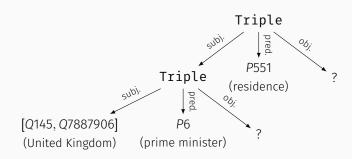


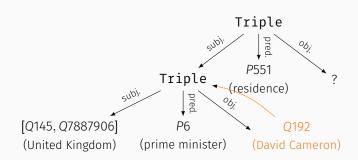
# **ANSWER EXTRACTION: QUESTION**

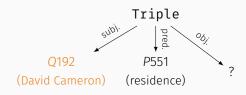
Where does the prime minister of United Kingdom live?

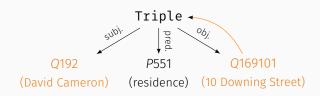
# ANSWER EXTRACTION: NORMAL FORM









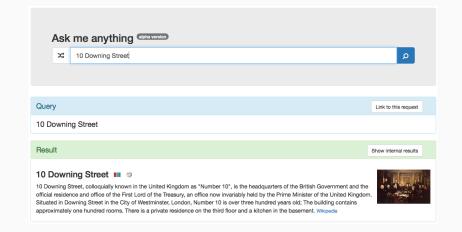


Q169101 (10 Downing Street)

#### ANSWER EXTRACTION: FINAL OUTPUT



#### SIMPLE RESEARCH



CAS

Based on Sympy (math) and PLY (parsers).

Based on Sympy (math) and PLY (parsers).

Two syntaxes:

Based on Sympy (math) and PLY (parsers).

#### Two syntaxes:

· Mathematica

Based on Sympy (math) and PLY (parsers).

#### Two syntaxes:

- · Mathematica
- · Intuitive syntax with permissive notations

LET'S SUM UP









VS

## **NESTED QUESTION**

Who is the wife of the president of the United States?

WolframAlpha Barack Obama Platypus Michelle Obama

### **NESTED QUESTION**

Who is the wife of the president of the United States?

WolframAlpha Barack Obama Platypus Michelle Obama

What are the birth dates of the daughters of the wife of the president of the United States?

WolframAlpha Barack Obama

Platypus Saturday, July 4, 1998 & Sunday, June 10, 2001

# CONJUNCTION

Who is the actor of Inception and Titanic?

WolframAlpha all the actors of the two movies

Platypus Leonardo DiCaprio

Future work

## BETTER DATABASE

#### Not answered by Wikidata:

- · "How fast is the TGV?"
- · "How wide is a tennis court?"
- $\rightarrow$  Improve Wikidata?
- $\rightarrow$  Use another database?

## BETTER QUESTION PARSING

#### Not parsed correctly:

- · "What is the date of birth of Isaac Newton?"
- · "In which band does Bono sing?"
- → Train the Stanford CoreNLP library?
- → Improve the algorithm of the Grammatical module?
- → Better datasets for the ML modules?

## **NEW MODULES**

cooking recipes train timetable HAL programming language interpreter literature meteo cinema music translation sport statistics and predictions

### OTHER IDEAS...

- · Other languages support (French...)
- · Improve user experience
- · Advertise Platypus

#### Conclusion

### SOME FACTS

#### 23 repositories & 2313 commits

6	PHP	Wikidata libraries and module
12	Python	Other modules, core, and libraries
1	C++	ML-Reformulation
1	Shell	Deployment scripts
1	et <sub>e</sub> x	This presentation and the report
1	Markdown	The specification
1	HTML/CSS/Javascript	The Web User Interface

26k lines of code (13k in PHP, 10k in Python)

4.3k lines of Latex & 2.9k lines of Markdown

4 accepted pull requests to libraries we use (corenlp-python, aspell-python, python-datautil)

### THE PPP?

- $\cdot$  A powerful and open source question answering framework
- · Innovative question parsing algorithms
- · A demo, Platypus, with general knowledge and math

# QUESTIONS?

# http://projetpp.github.io/



0 https://github.com/ProjetPP

ppp@pony.ovh

