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

Who was the wife of Louis Pasteur?

Brezhoneg Bypnag Catala Čestina Cymraeg Dansk • Deutsch

Eesti EAAnyeeû Español Esperanto Euskara فارسه Fiji Hindi Feroyskt

Français Frysk Gaeige Gaidhlig Galego spreid Will Midd-Ad-noil

한국어 <u)thtb

Hrvatski Ido Ildo

Íslenska Italiano 17139 Basa Jawa IIII Johnggoo

Kurdi Latina Latviešu Létzebuergesc Lietzviu

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-Eliennette Roqui. In 1827, the family moved to Arbois, where he entered primary school in 1831. He was an

average studies in his early years, and ris periodally scadeline, as the interests were folling and describing risk guests and promise of the general sortion in Periodal positions in Periodal positi

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 doen of the new faculty of sciences at Life University, it was on this occasion that Pasteur uttered his of equaled remark: "dans les champs de Tobservation, le hasard ne favorise que les esprits préparés." (in the field of observation, have been sonly the presented midd ⁽¹⁹⁾).

In 1837, he moved to Paris as the director of cointrific studies at the Ecole Homan's Suprience where he took coronal from 1858 to 1817 and introduced a series of reform. The commissions became more rigid, which in lot do better result, pressure corresponds, and increased pressions, the cointribution of the commissions studies records. Find 1852, he was appointed professor of peology, physics, and chemistry at the Ecole nationals suprience des Beasu-Arts, the position which held until his resignation in 1807. In the late, the established the Pasture institute in 1827, it which has as all decrease for the cort of his (En/ReVII).

Research contributions

Molecular asymmetry

In Pasters's early work as a chemist, he resolved a problem concerning the nature of laterirs and (1848). INDENTIFIED A solution of this compound derived from fiving things (specifically, wine less) rotated the plane of polarization of light passing through it. The mystery was that trainar and derived by chemical synthesis had no such effect, even though its chemical reactions were identical and its elemental composition was the same. If I'll his was the first time anyone had demonstrated christ molecules.

Fermentation and germ theory of diseases

Pacteur demonstrated that fementation is caused by the growth of micro-organisms, and the emergent growth of binations in nutrient broths is due not to spontaneous generation, but rather to biogenesis (Orme vivum ex vio "all file from file"), the was motivated to investigate the matter while working at Lifls in 1856 a local vine manufacturer. M Bigot, the father of his student, sought for his advice on the problems of making betwoot acknowled around a souring after long storage. If I'll be could not immediately pursue the matter mainly



EXISTING TOOLS

WolframAlpha Marie Pasteur

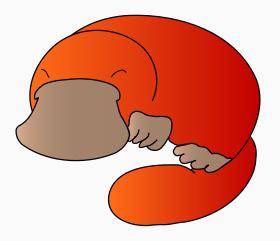
Google Wikipedia page of Marie Pasteur Bing Wikipedia page of Louis Pasteur

Yahoo answers.com page containing this question

Closed-source, bad on some questions.

PPP INTRODUCES PLATYPUS

http://askplatyp.us/



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

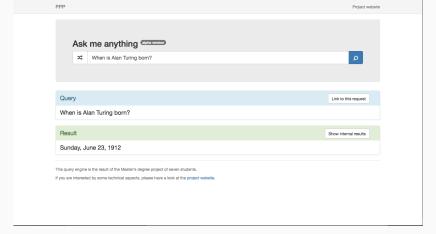
Alan Turing III ®

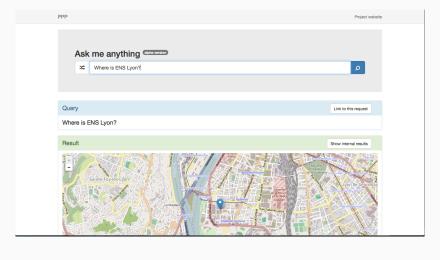
Alan Mathison Turing, OBE, FRS (Ylpsny) TENPT-ing; 23 June 1912 – 7 June 1954) was a British mathematician, logician, cryptanalyst, philosopher, pioneering computer scientist, mathematical biologist, and marathon and ultra distance runner. He was highly influential in the development of computer science, 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.

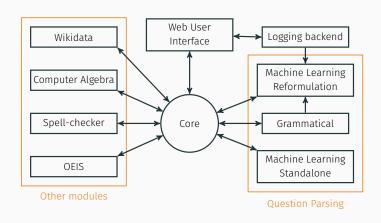
If you are interested by some technical aspects, please have a look at the project website.





OVERVIEW

ARCHITECTURE



DATAMODEL

How to represent formally a question asked in natural language? Tree structure.

DATAMODEL - LEAVES: RESOURCES

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

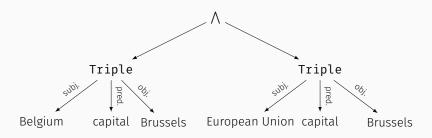
٠ ..

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

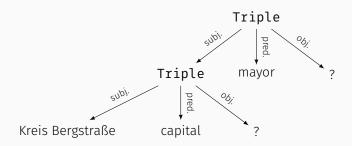
DATAMODEL - EXAMPLES

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

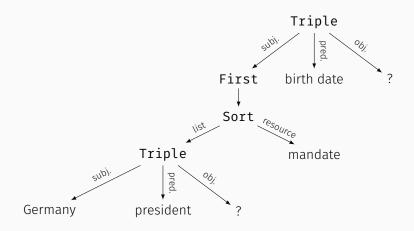


DATAMODEL - EXAMPLES

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

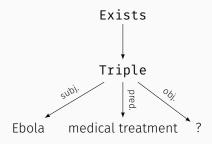


"What is the birth date of the first president of Germany?"



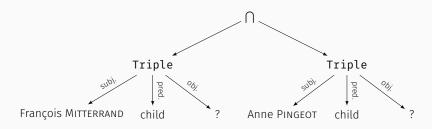
DATAMODEL - EXAMPLES

"Is there a medical treatment for Ebola?"



DATAMODEL - EXAMPLES

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

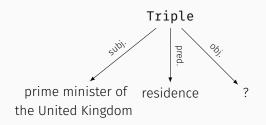


QUESTION PARSING

Map the input question to the most relevant normal form.

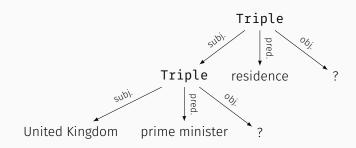
Map the input question to the most relevant normal form.

Where does the prime minister of the United Kingdom live?

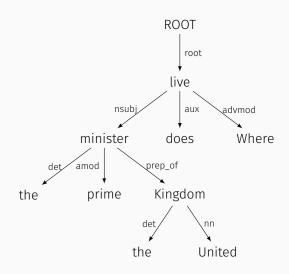


Map the input question to the most relevant normal form.

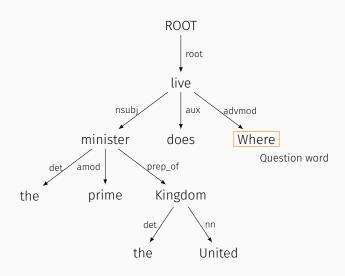
Where does the prime minister of the United Kingdom live?



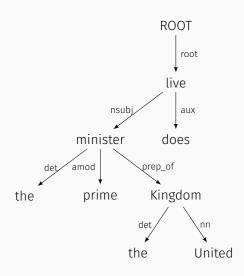
DEPENDENCY TREE OUTPUT BY THE STANFORD PARSER

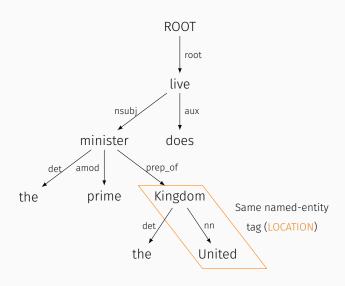


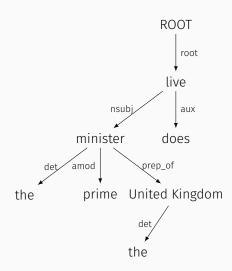
IDENTIFY QUESTION WORD

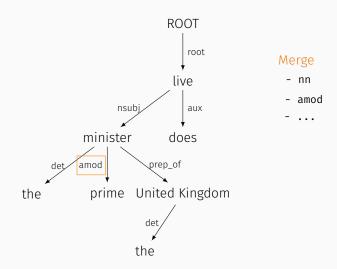


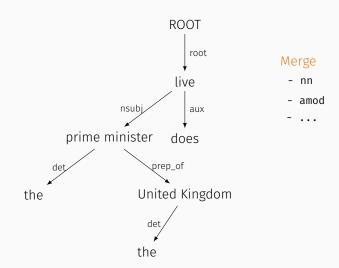
IDENTIFY QUESTION WORD

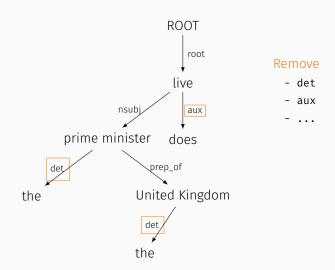


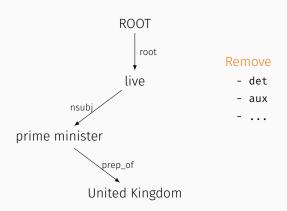




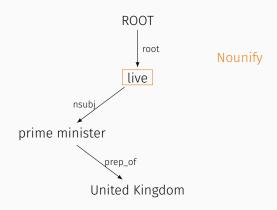




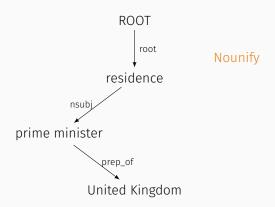




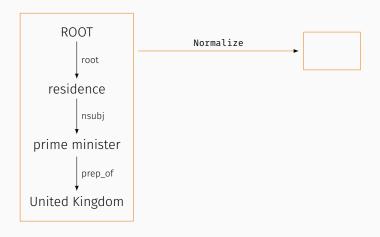
NOUNIFICATION



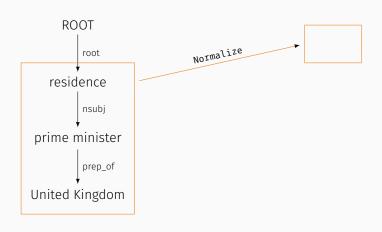
NOUNIFICATION



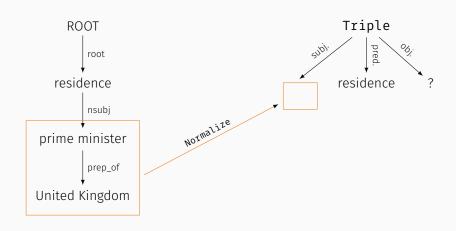
NORMALIZATION



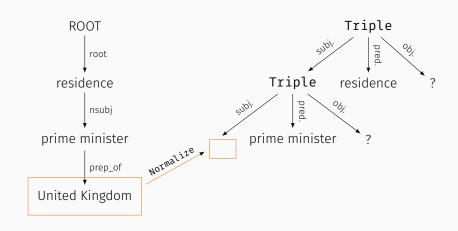
NORMALIZATION



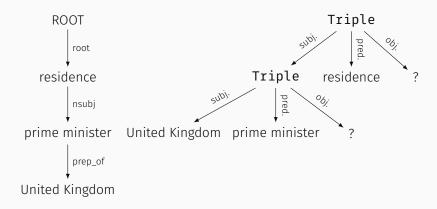
NORMALIZATION



NORMALIZATION

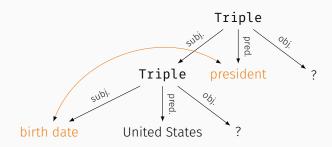


NORMALIZATION



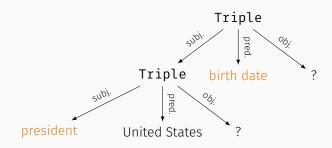
REFORMULATION: A TRANSLATION MODULE

When was the president of the United States born?



REFORMULATION: A TRANSLATION MODULE

When was the president of the United States born?

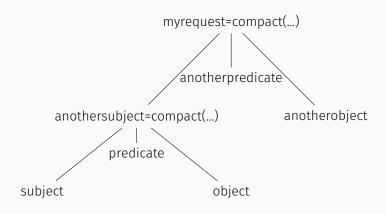


OPERATION

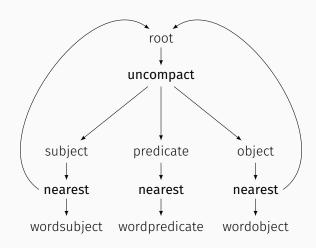
Projection: trees \mapsto vector space.

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,

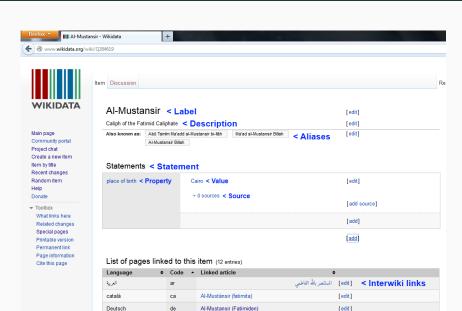
$$\|V_{w_2} - V_{w_1}\|_2 \approx 0$$

BACK-END

WIKIDATA (http://wikidata.org)?

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

AN ITEM



Al-Mustansir Billah

en

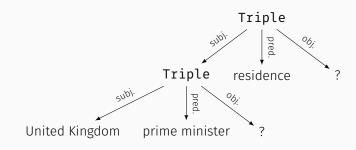
[edit]

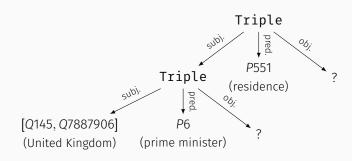
English

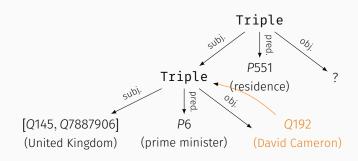
ANSWER EXTRACTION: QUESTION

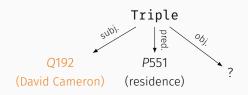
Where does the prime minister of the United Kingdom live?

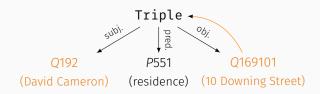
ANSWER EXTRACTION: NORMAL FORM











Q169101 (10 Downing Street)

Answer extraction: Final output



Result

Show internal results

10 Downing Street 💵 🗇

10 Downing Street, colloquially known in the United Kingdom as "Number 10", is the headquarters of the British Government and the official residence and office of the First Lord of the Treasury, an office now invariably held by the Prime Minister of the United Kingdom. Situated in Downing Street in the City of Westminster, London, Number 10 is over three hundred years old and contains approximately one hundred rooms. There is a private residence on the third floor and a kitchen in the basement. Wispedia



SIMPLE RESEARCH



Result Show internal results

10 Downing Street III 🧐

10 Downing Street, colloquially known in the United Kingdom as "Number 10", is the headquarters of the British Government and the official residence and office of the First Lord of the Treasury, an office now invariably held by the Prime Minister of the United Kingdom. Situated in Downing Street in the City of Westminster, London, Number 10 is over three hundred years old; The building contains approximately one hundred rooms. There is a private residence on the third floor and a kitchen in the basement. Wikipedia



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







VS



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

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

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 _E 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

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

THE PPP?

- · 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/





