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?

Brezhoneg
Буряад
Català
Čeština
Cymraeg
Dansk
★ Deutsch
Eesti
Ελληνικά
Español
Esperanto
Euskara
فارسی
Fiji Hindi
Føroyskt
Français
Frysk
Gaeilge
Gàidhlig
Galego
ગુજરાતી
客家語/Hak-kâ-ngî
한국어
Հայերեն
हिन्दी
Hrvatski
Ido
Ilokano
Bahasa Indonesia
Íslenska
Italiano
עברית
Basa Jawa
ಕನ್ನಡ
ქართული
Қазақша
Kiswahili
Kurdî
Latina
Latviešu
Lëtzebuergesch
Lietuvių

Early life



The house in which Pasteur was born, Dole

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 Besançon and earned his *baccalauréat* (BA) degree in 1840. He was appointed teaching assistant at the Besançon college while continuing a degree science course with special mathematics. He failed his first examination in 1841. He managed to pass the *baccalauréat 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 ès sciences* (Bachelor of Science) degree. In 1846, he was appointed professor of physics at the Collège de Tournon at Ardèche, but Antoine Jérome Balard (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 Dijon Lycée 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 Lille 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])



Pasteur in 1857

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.[2] 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 1867. In Paris, he established the Pasteur Institute in 1887, 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.[16] 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 Lille. In 1856 a local wine manufacturer, M. Bigot, 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

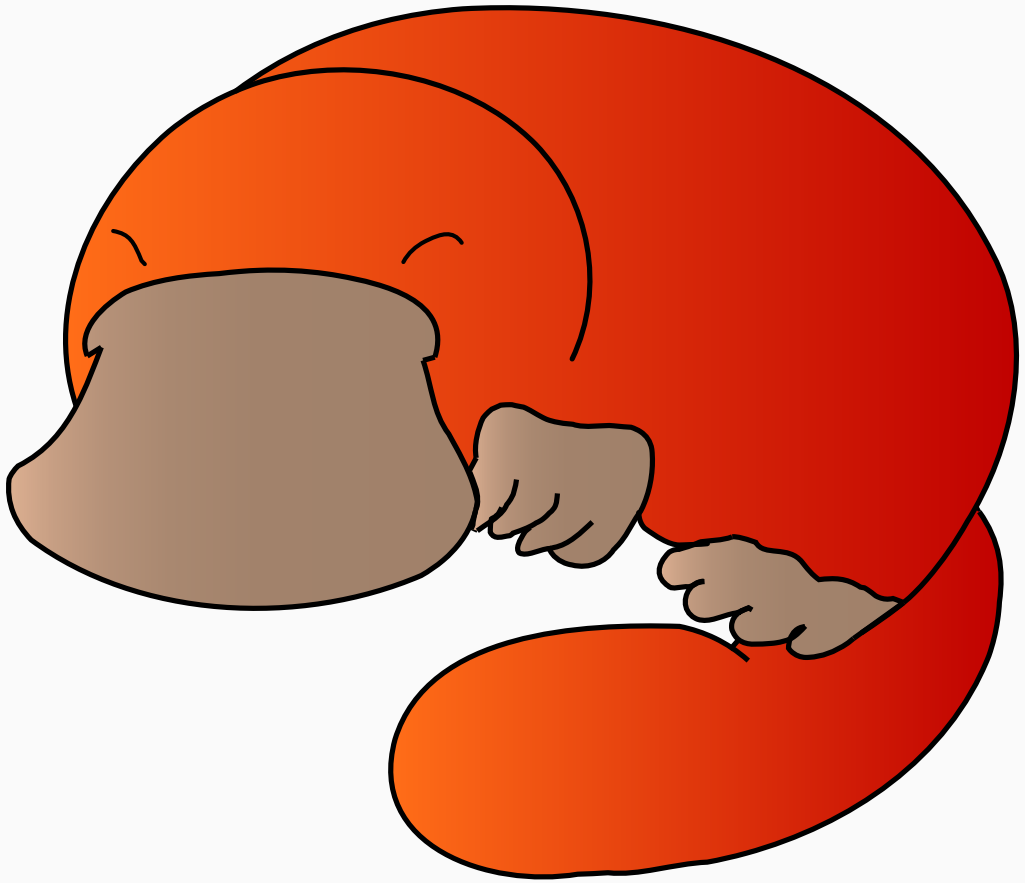
Existing tools

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://askplatyp.us/



Demo time!

PPP

Project website

Ask me anything alpha version

Who is Alan Turing?

Query

Link to this request

Who is Alan Turing?

Result

Show internal results

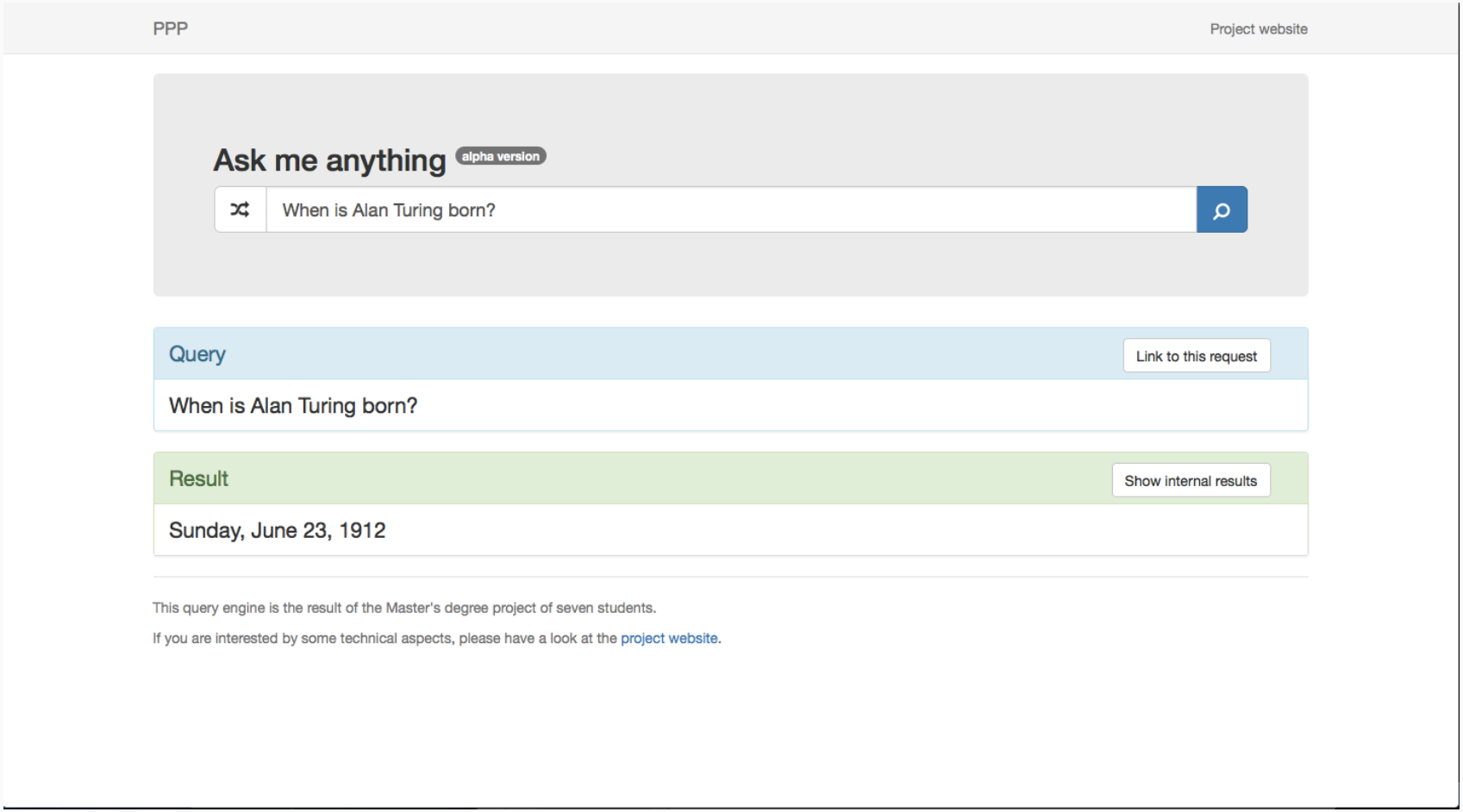
Alan Turing

Alan Mathison Turing, OBE, FRS (/ˈtjʊərɪŋ/ TEWR-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. Wikipedia



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.


PPP
Project website
Ask me anything
alpha version
When is Alan Turing born?
Query
Link to this request
When is Alan Turing born?
Result
Show internal results
Sunday, June 23, 1912
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.

PPP

Project website

Ask me anything

alpha version

Where is ENS Lyon?

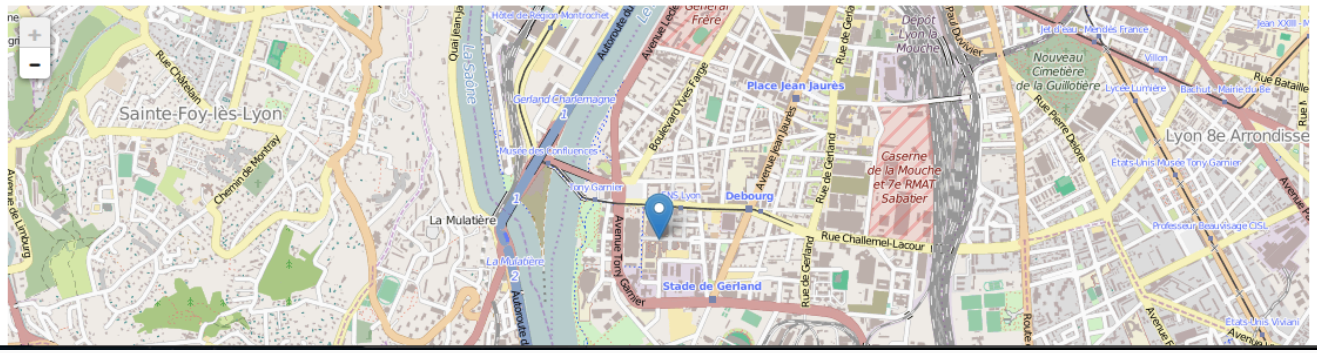
Query

Link to this request

Where is ENS Lyon?

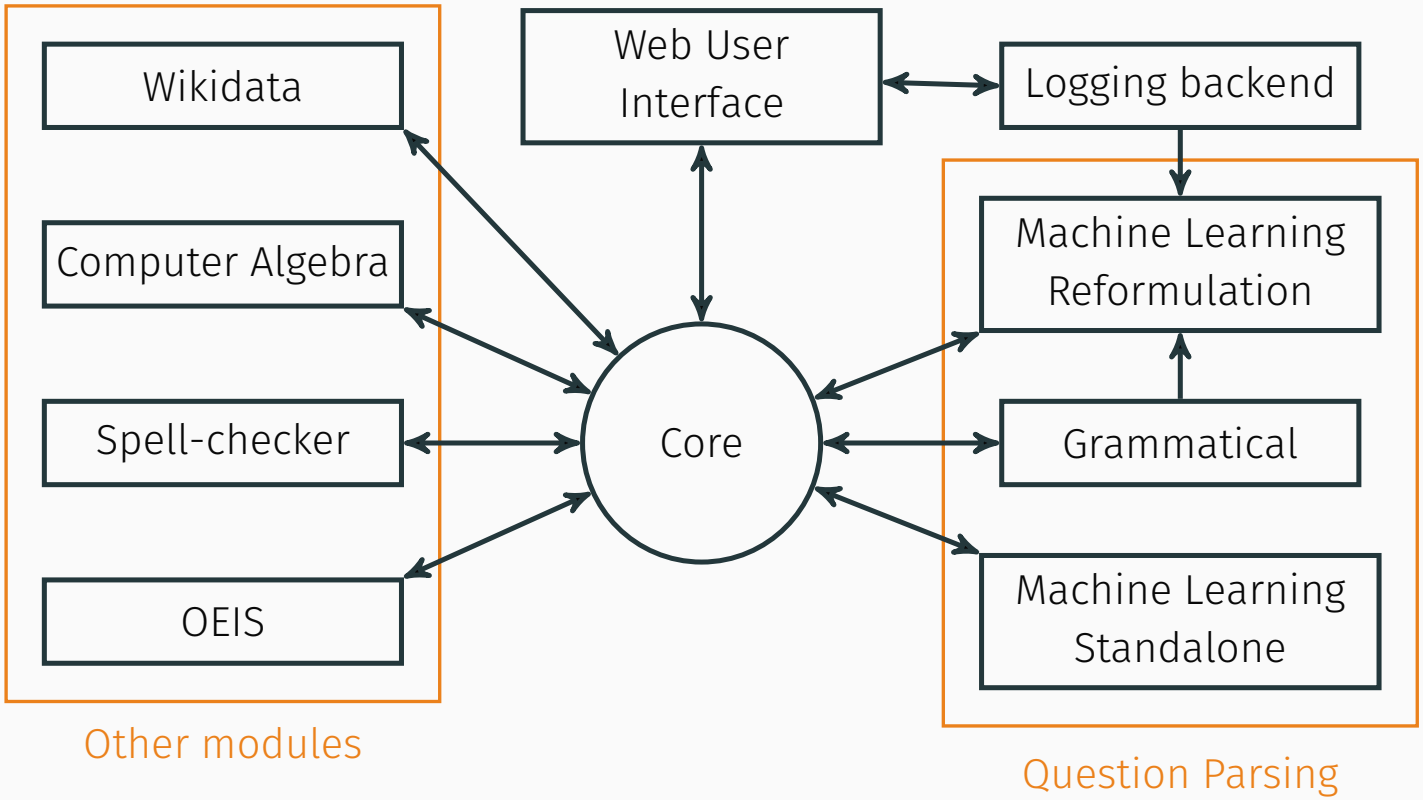
Result

Show internal results



OVERVIEW

Architecture



Datamodel

How to represent formally a question asked in natural language?

Tree structure.

Datamodel – Leaves: resources

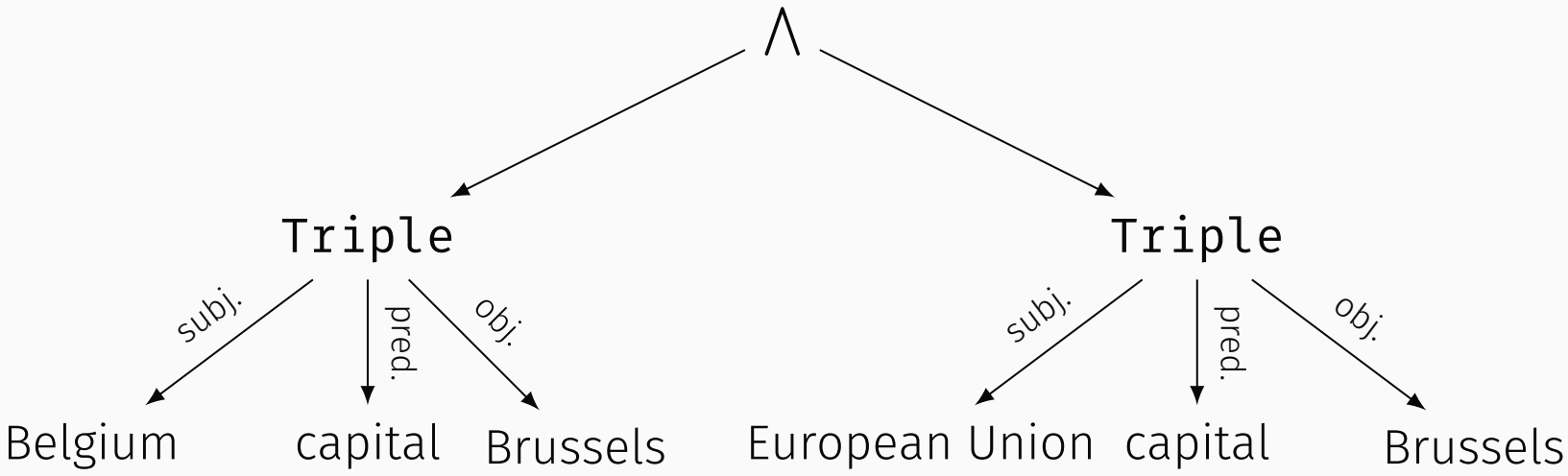
- Strings: ”Isaac Newton”
- Dates: 25 December 1642
- Geographic coordinates: 47° 30’ 18” N ; 9° 44’ 57” E
- ...

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

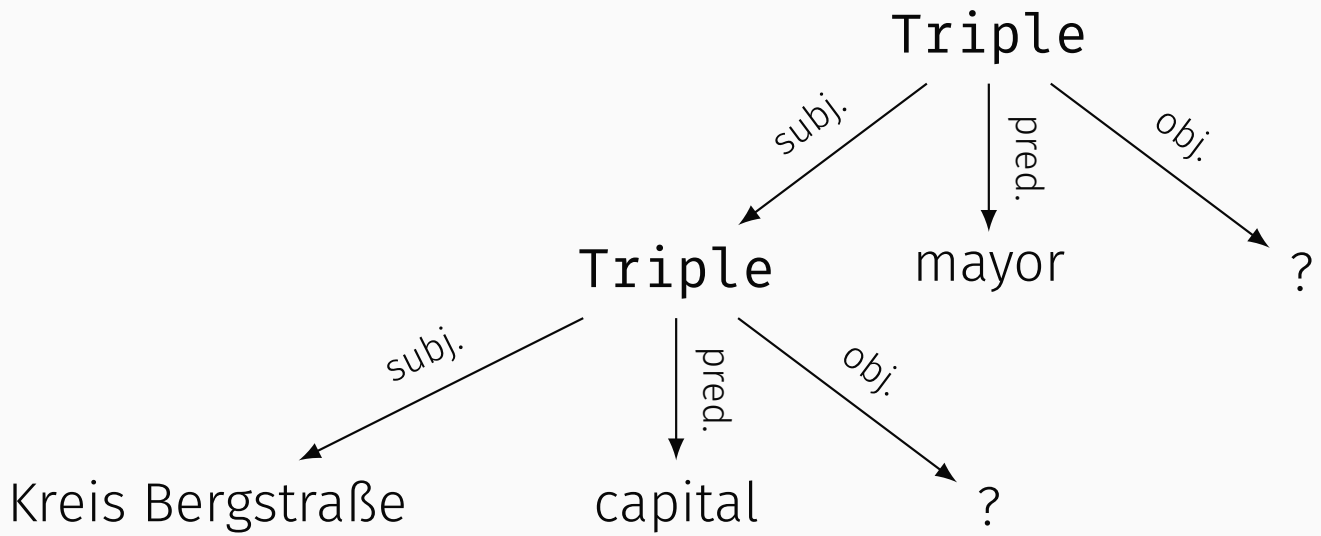
Datamodel – Examples

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



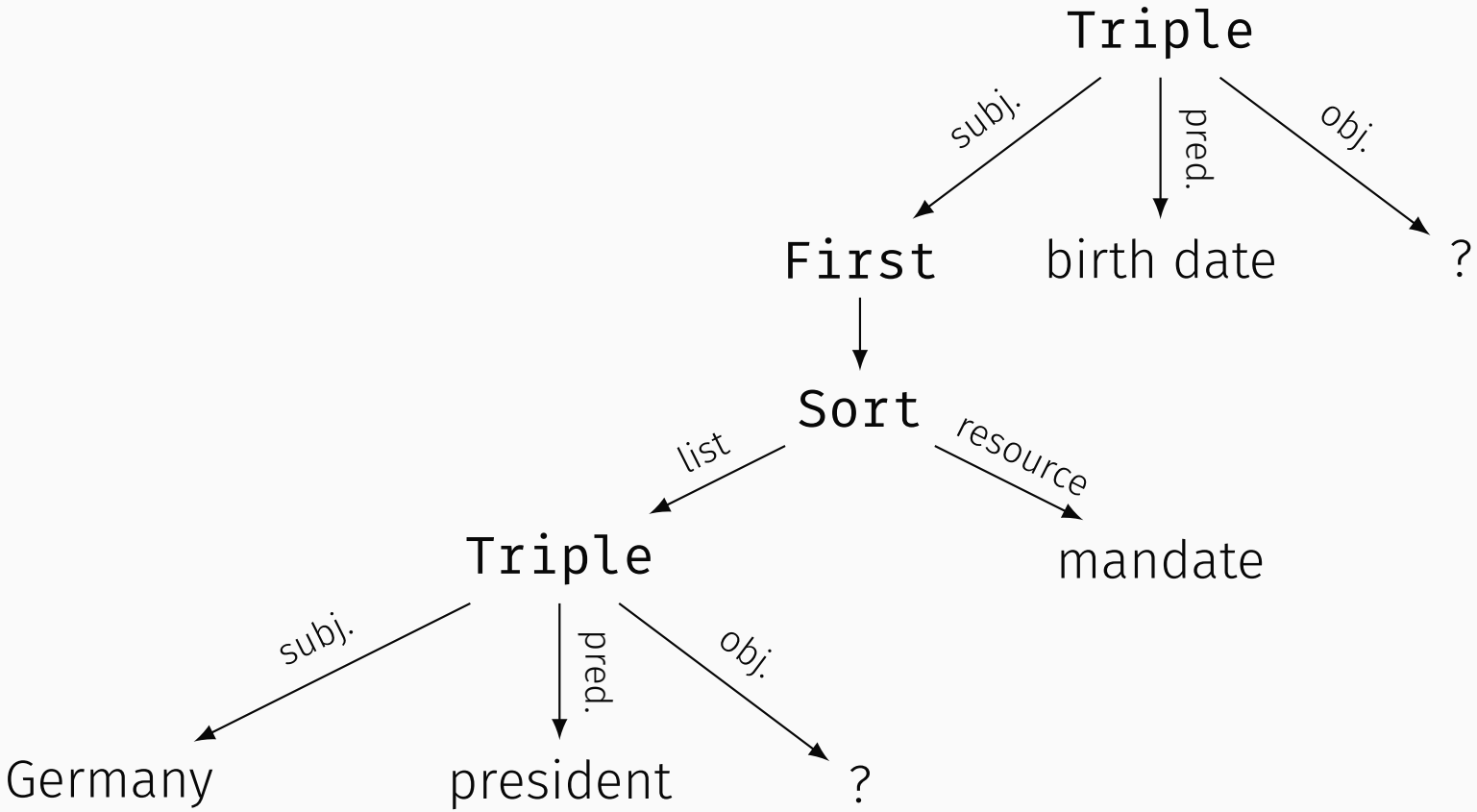
Datamodel – Examples

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



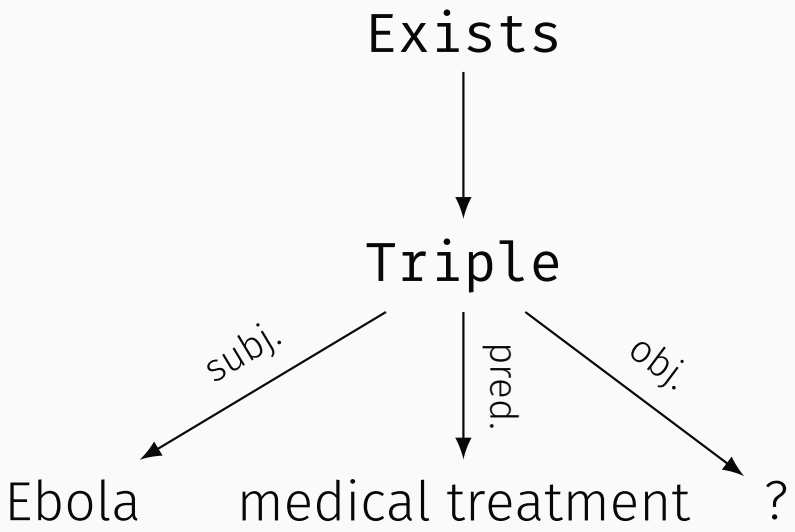
Datamodel – Examples

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



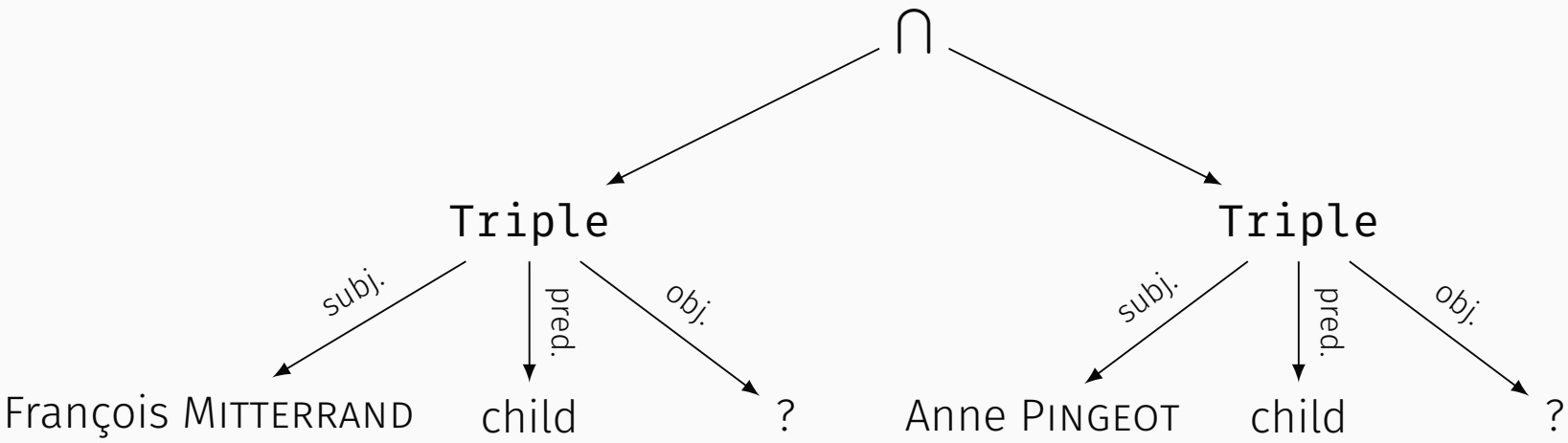
Datamodel – Examples

"Is there a medical treatment for Ebola?"



Datamodel – Examples

"Who are the children of François Mitterrand and Anne Pingeot?"

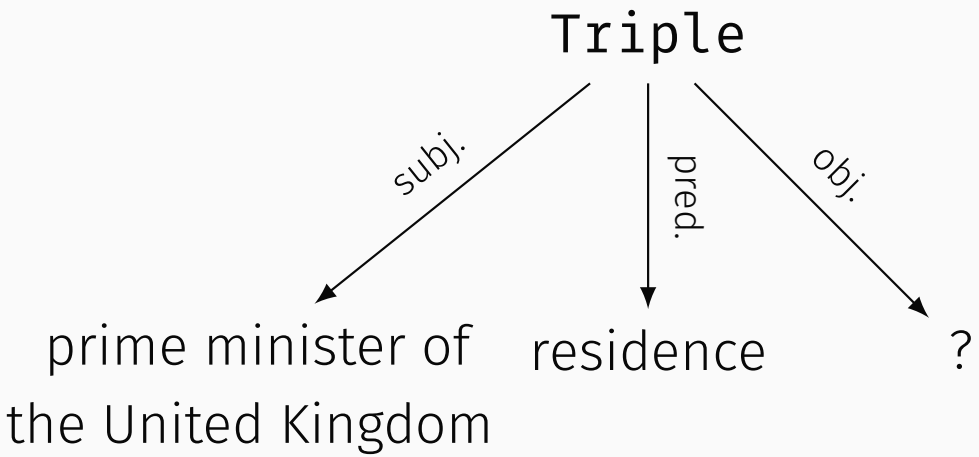


QUESTION PARSING

Map the input question to the most relevant normal form.

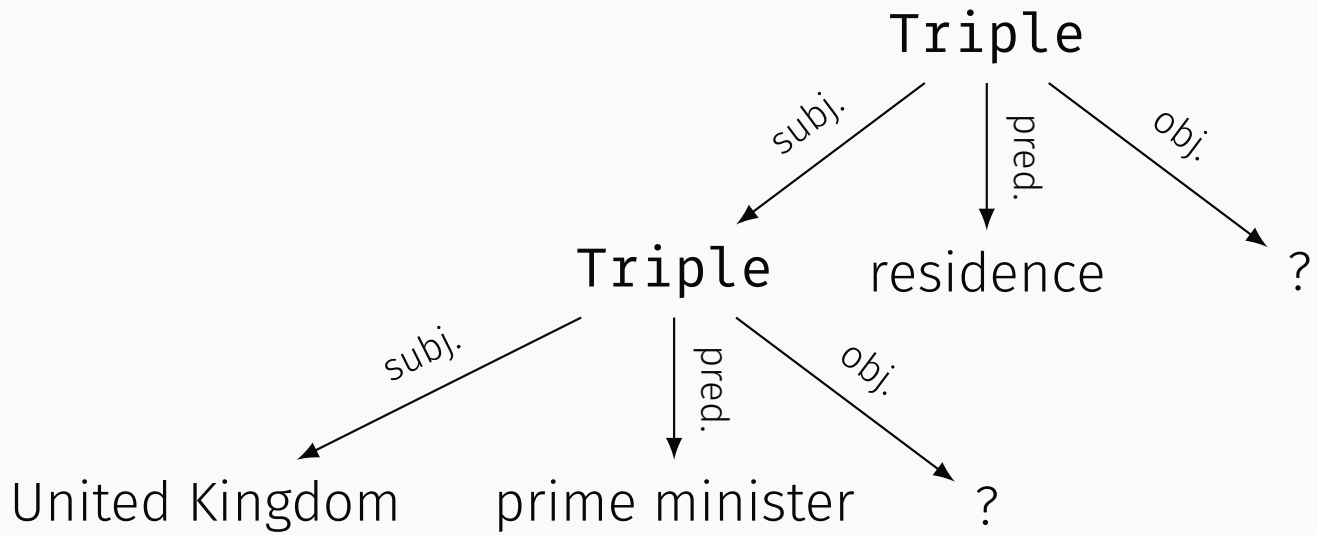
Map the input question to the most relevant normal form.

Where does the prime minister of the United Kingdom live?



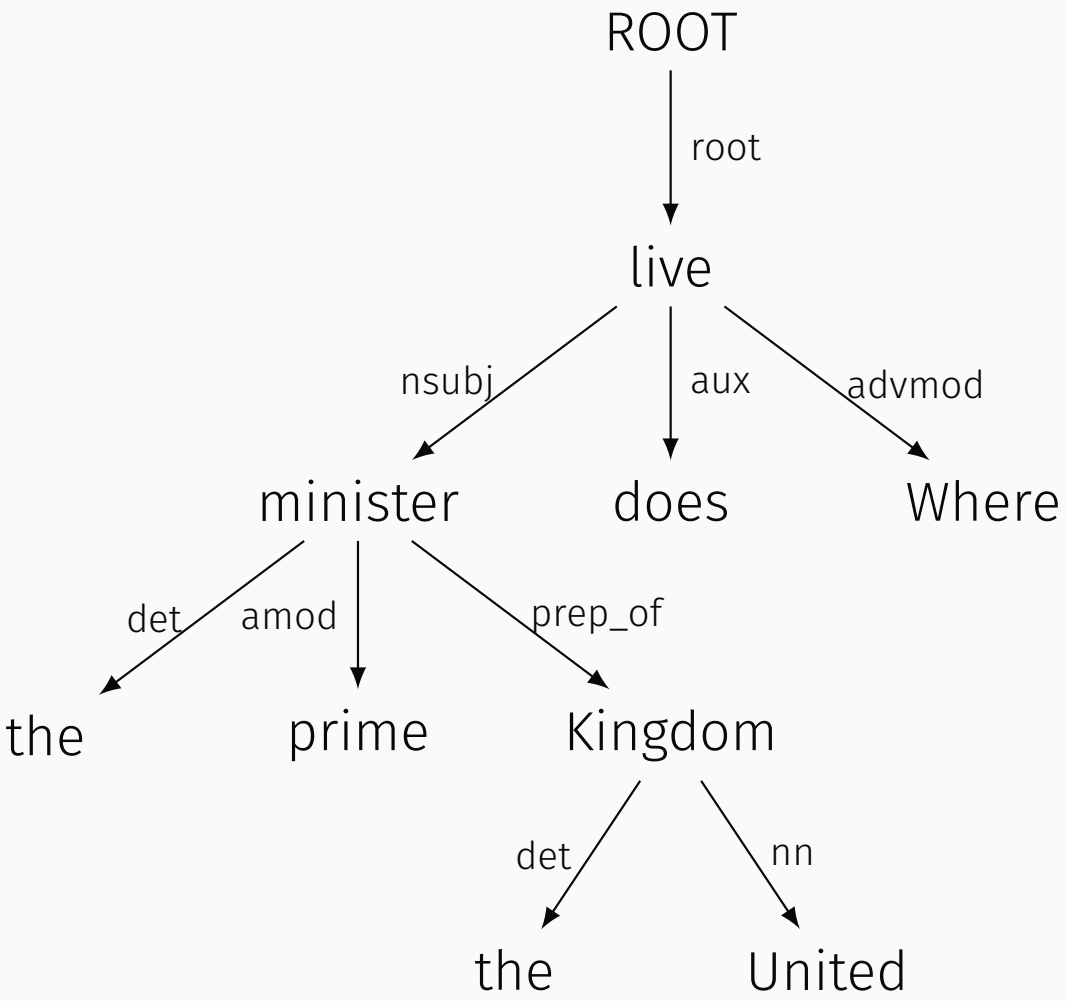
Map the input question to the most relevant normal form.

Where does the prime minister of the United Kingdom live?

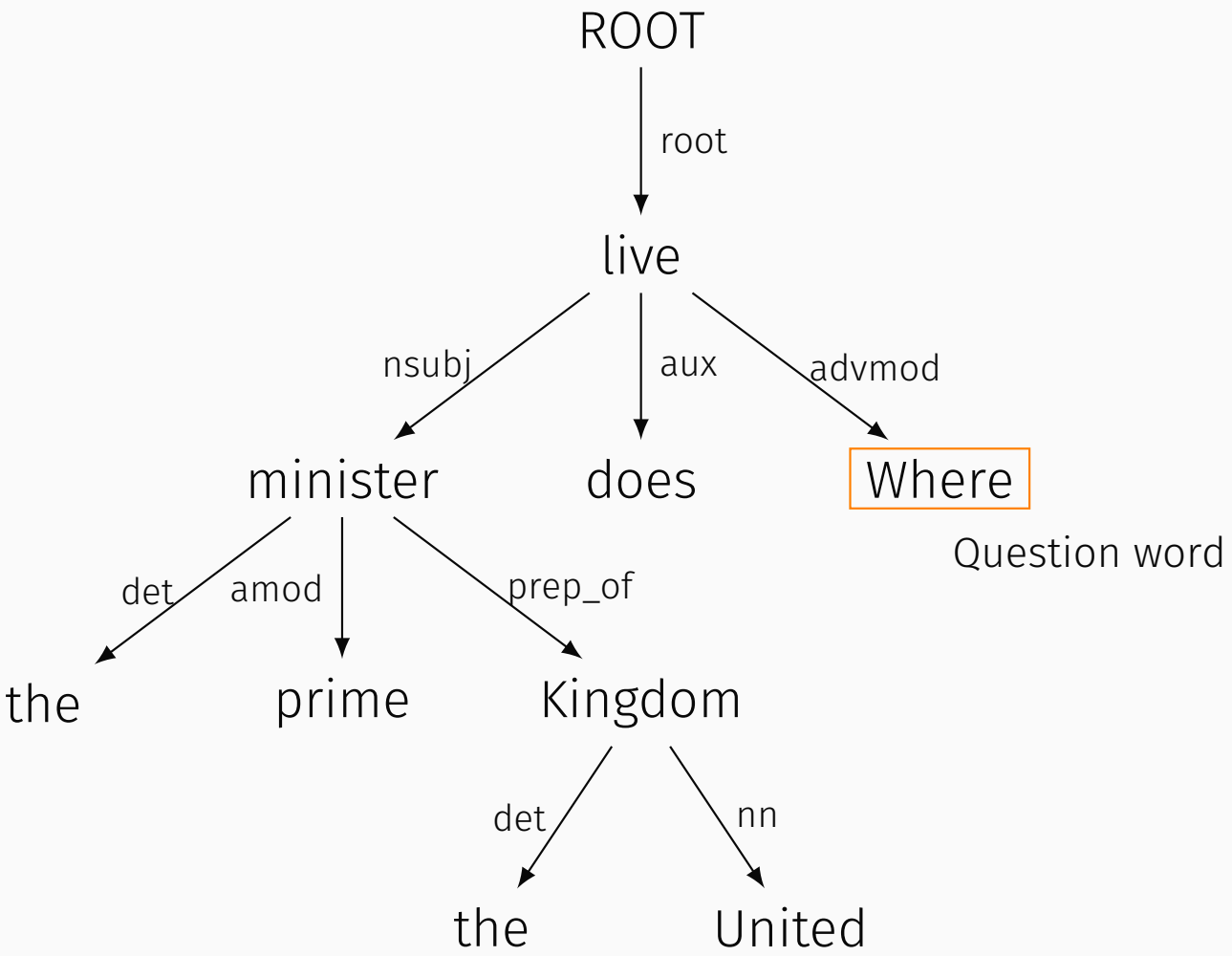


Grammatical approach

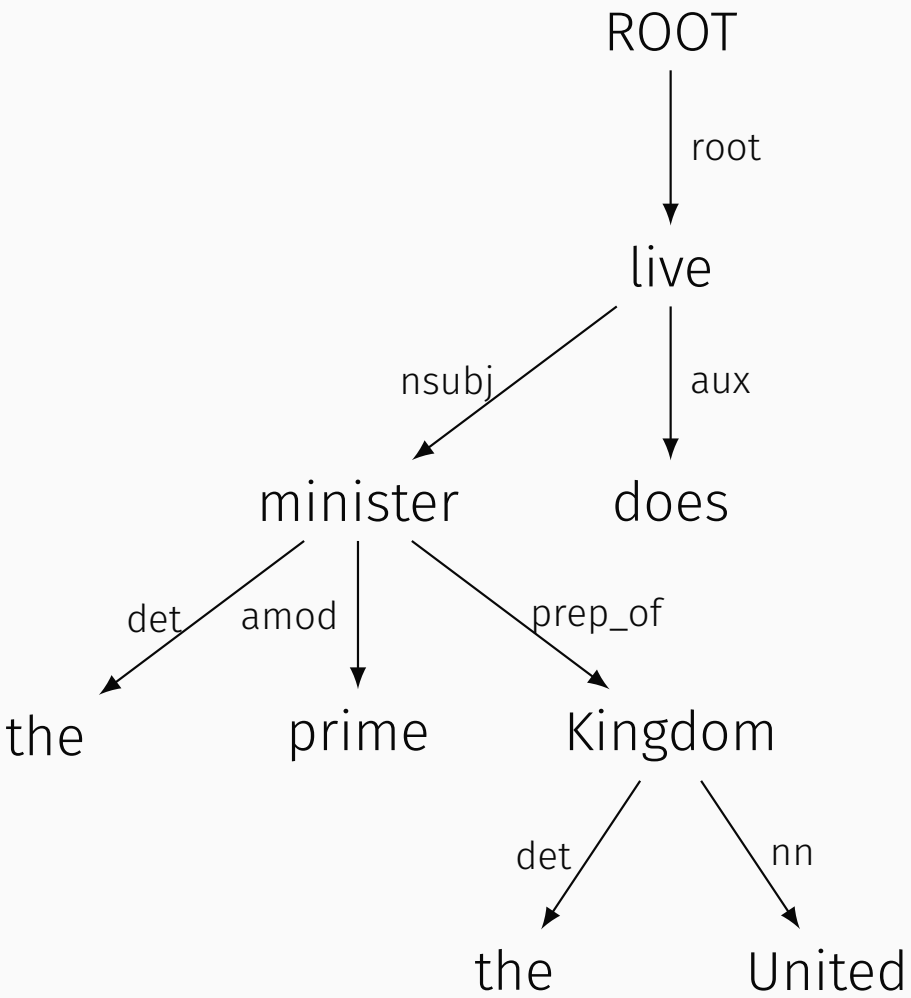
Dependency tree output by the Stanford Parser



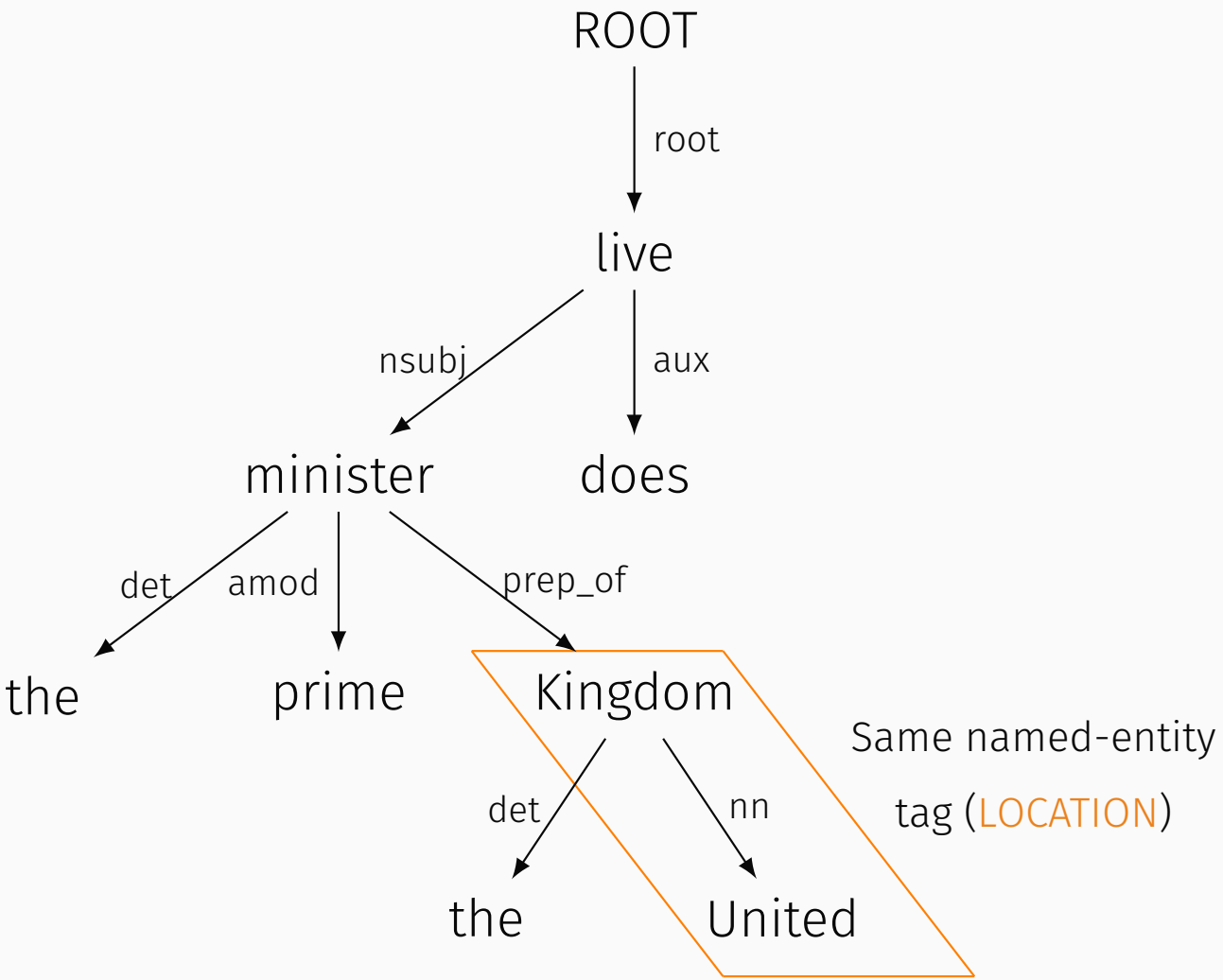
Identify question word



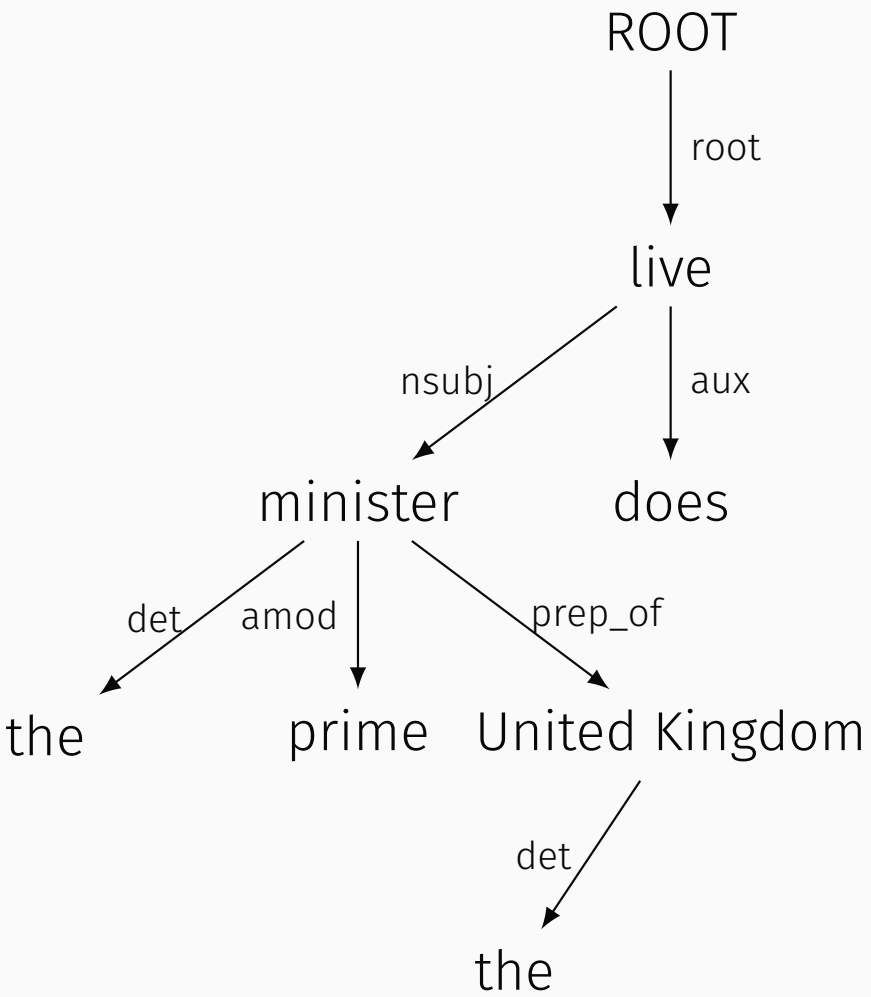
Identify question word



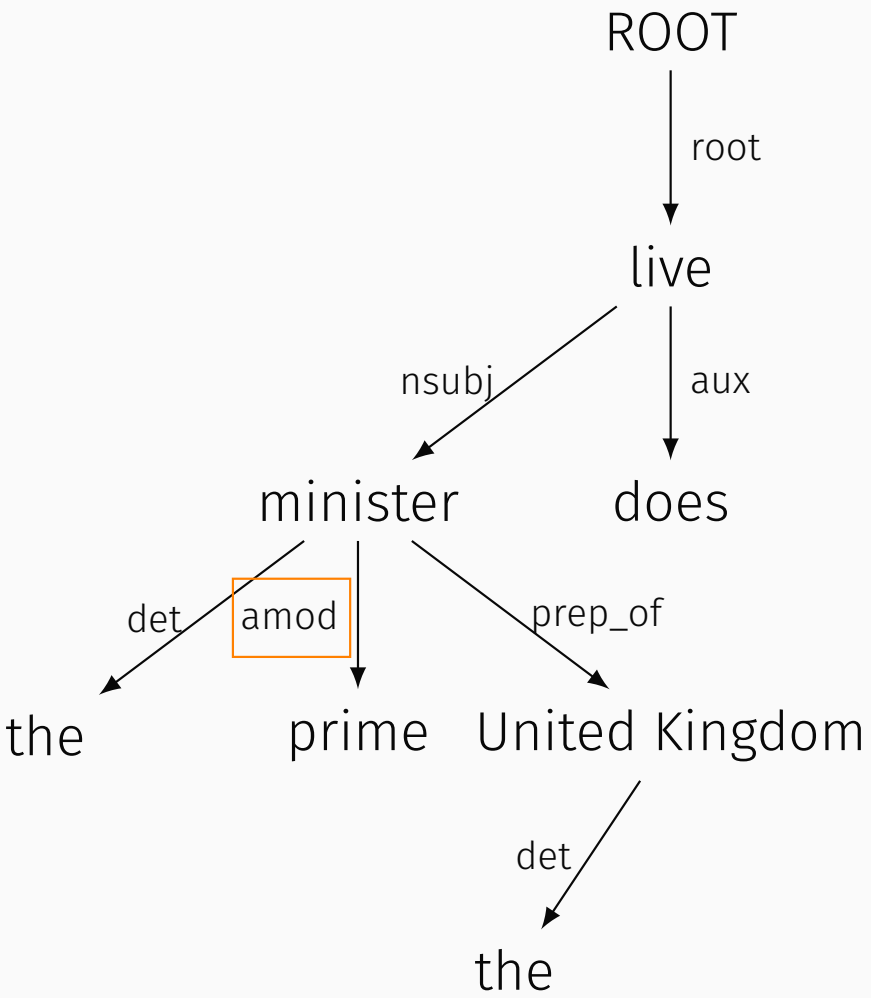
Merging


ROOT
root
live
nsubj
aux
minister
does
det
amod
prep_of
the
prime
Kingdom
det
nn
the
United
Same named-entity
tag (LOCATION)

Merging


ROOT
root
live
nsubj
aux
minister
does
det
amod
prep_of
the
prime
United Kingdom
det
the

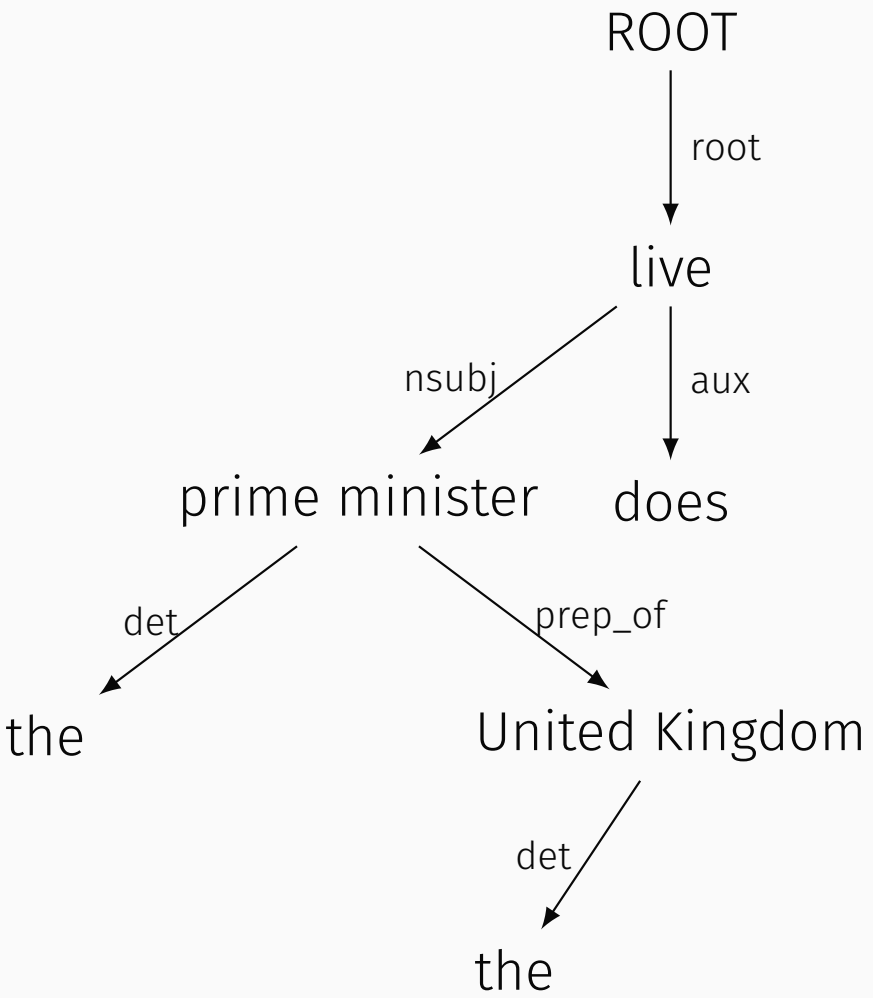
Merging



Merge

- nn
- amod
- ...

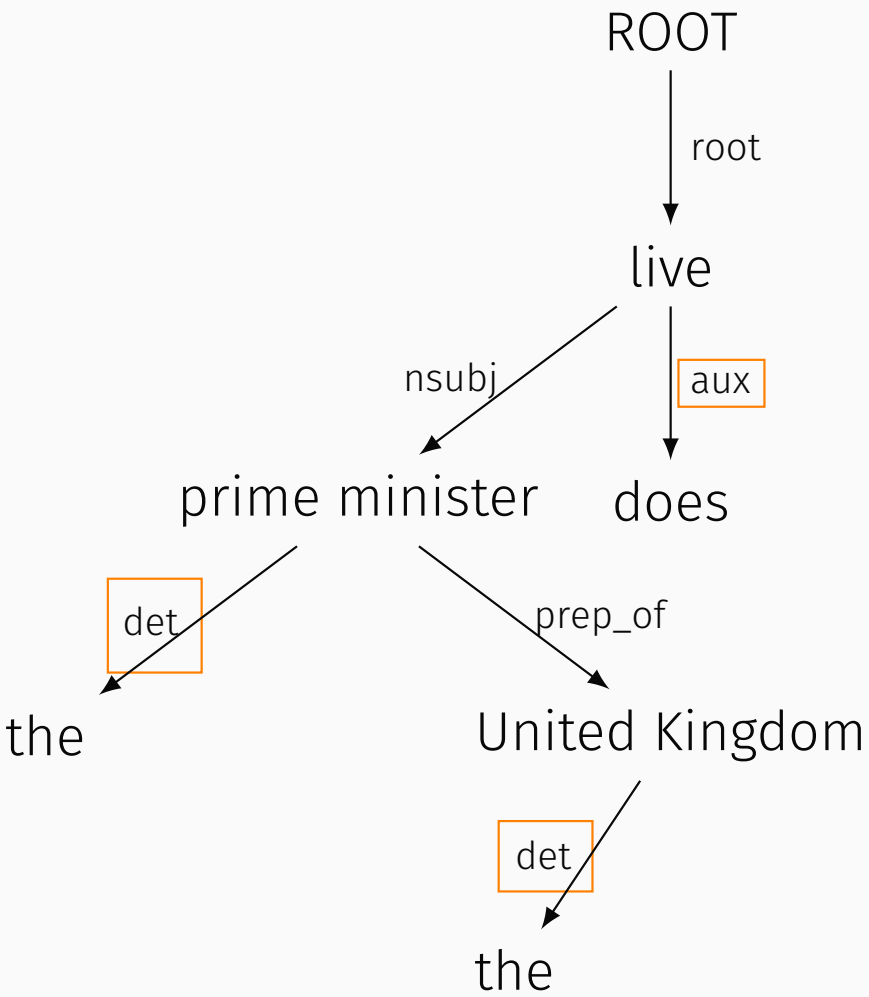
Merging



Merge

- nn
- amod
- ...

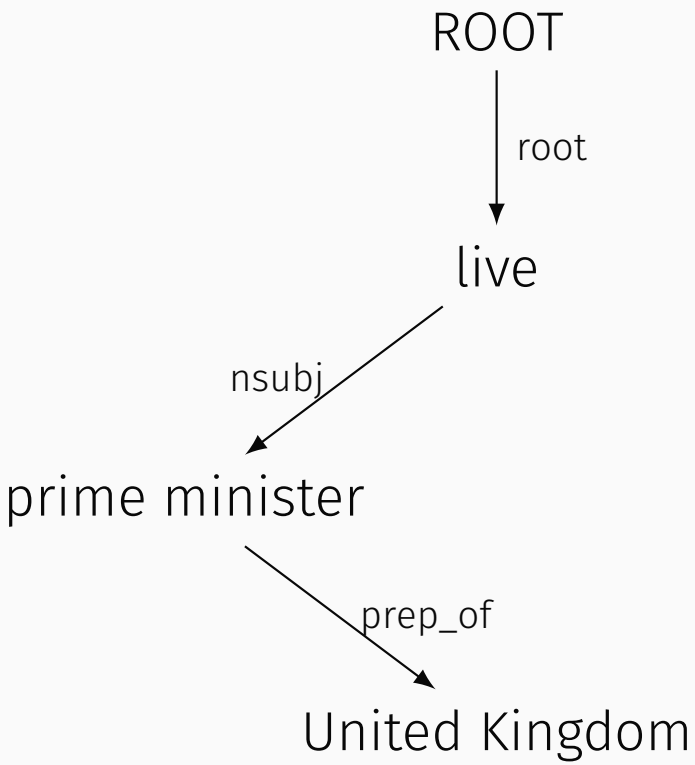
Removal



Remove

- `det`
- `aux`
- `...`

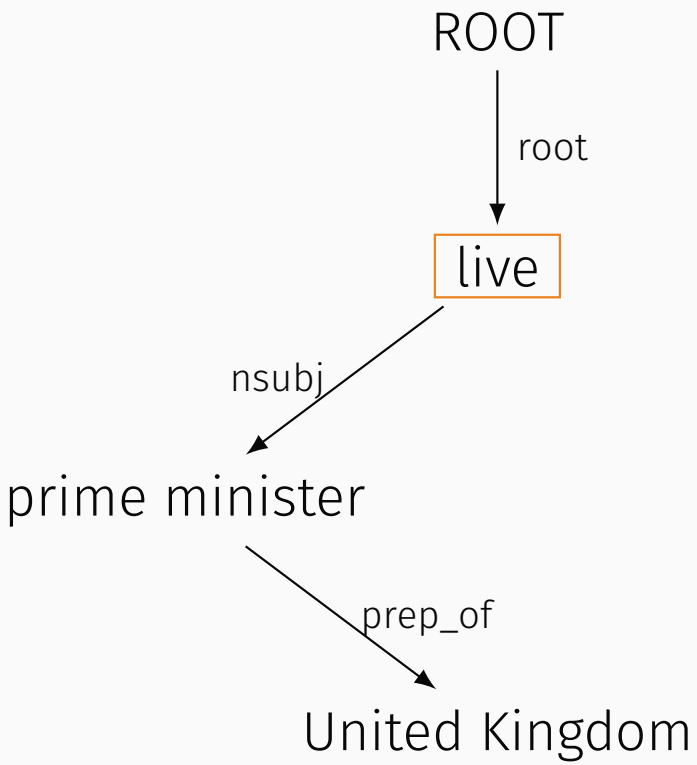
Removal



Remove

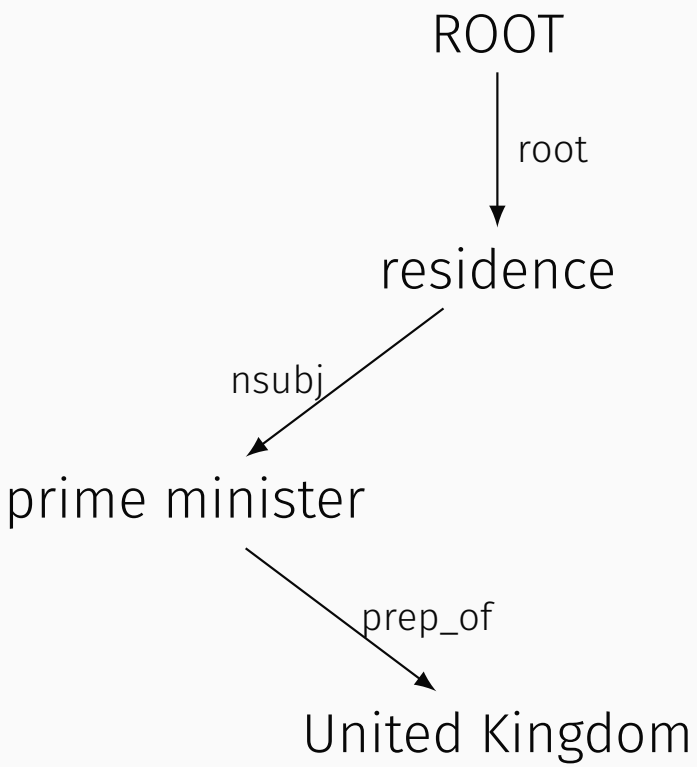
- `det`
- `aux`
- `...`

Nounification



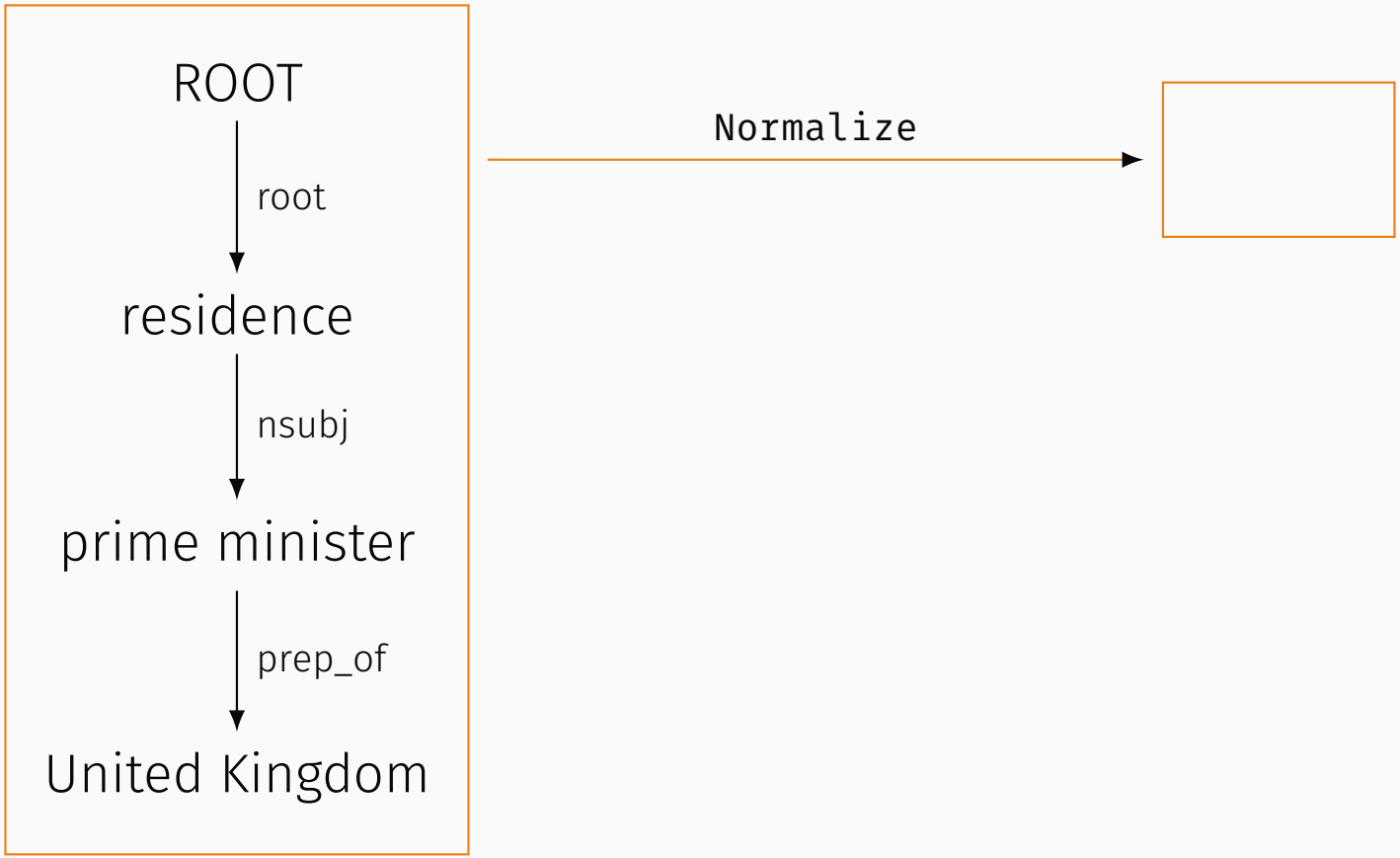
Nounify

Nounification

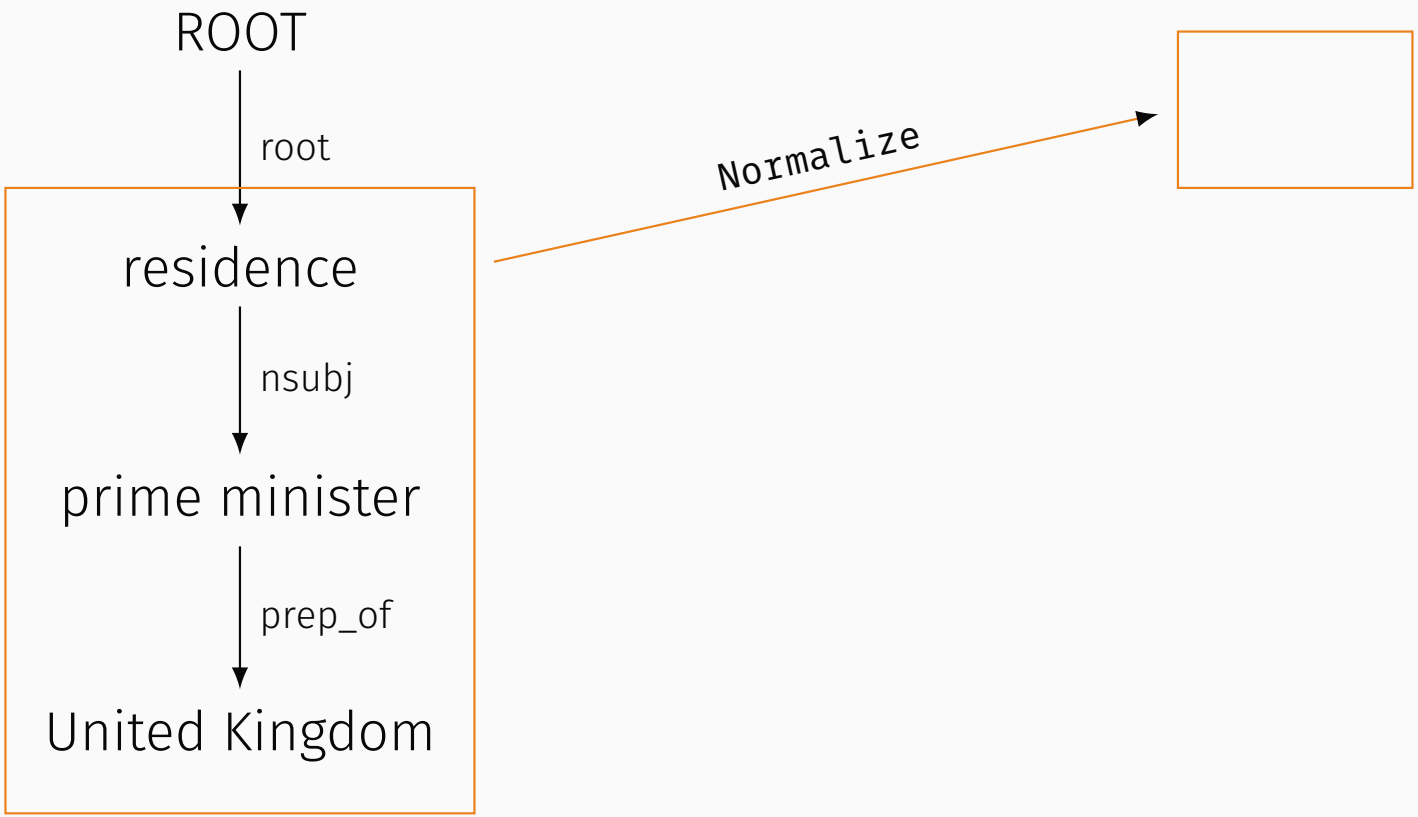


Nounify

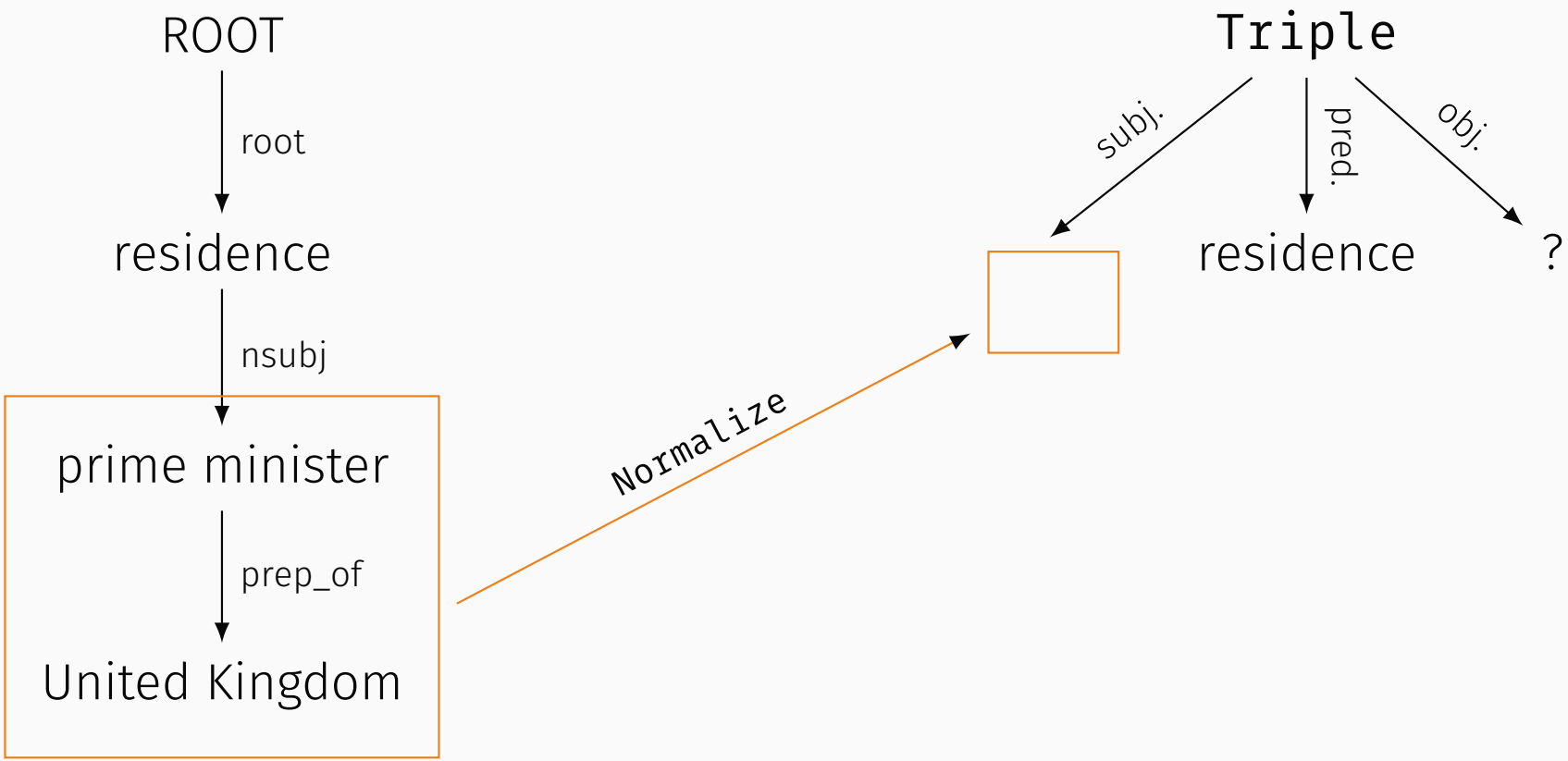
Normalization



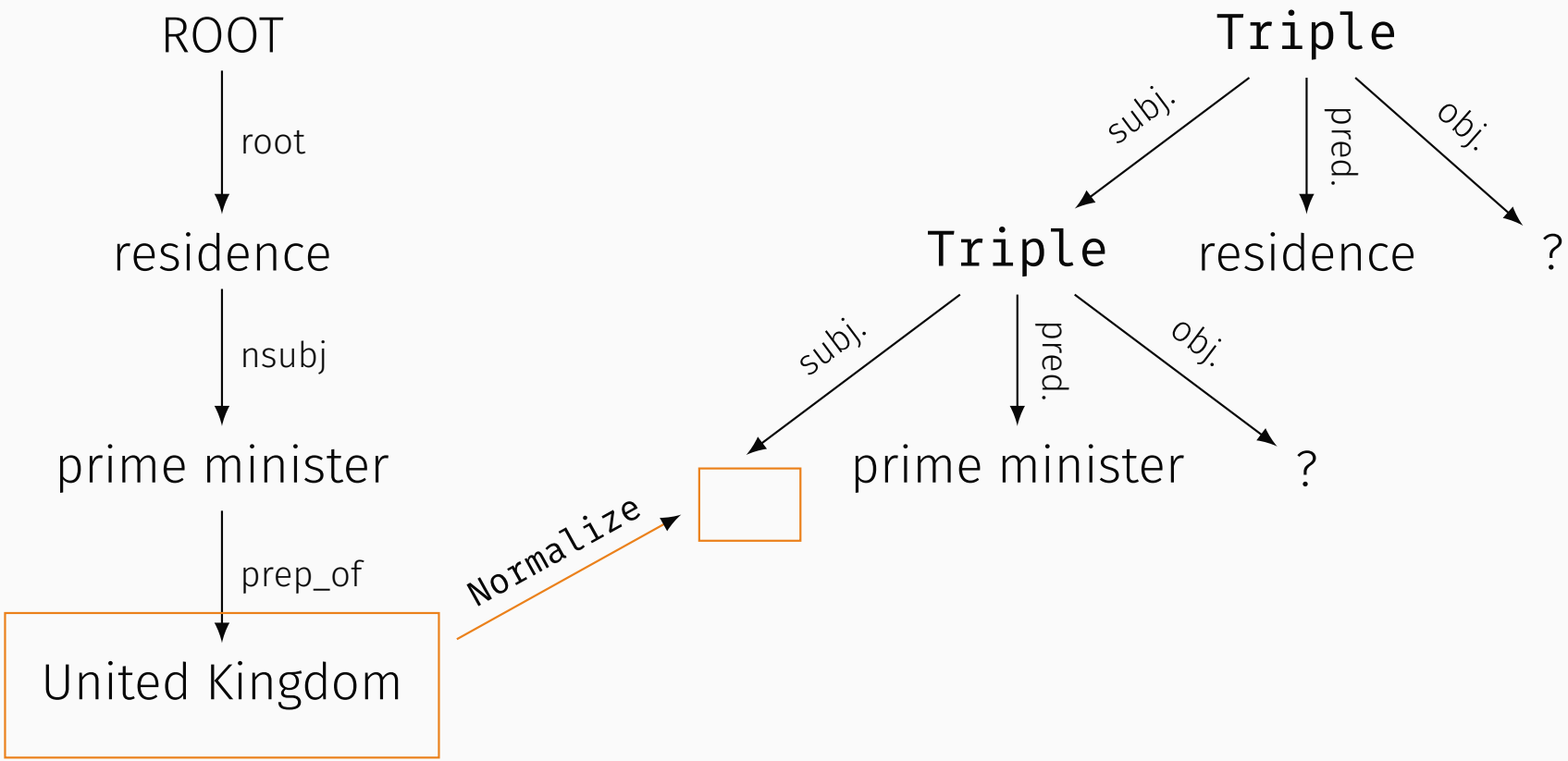
Normalization



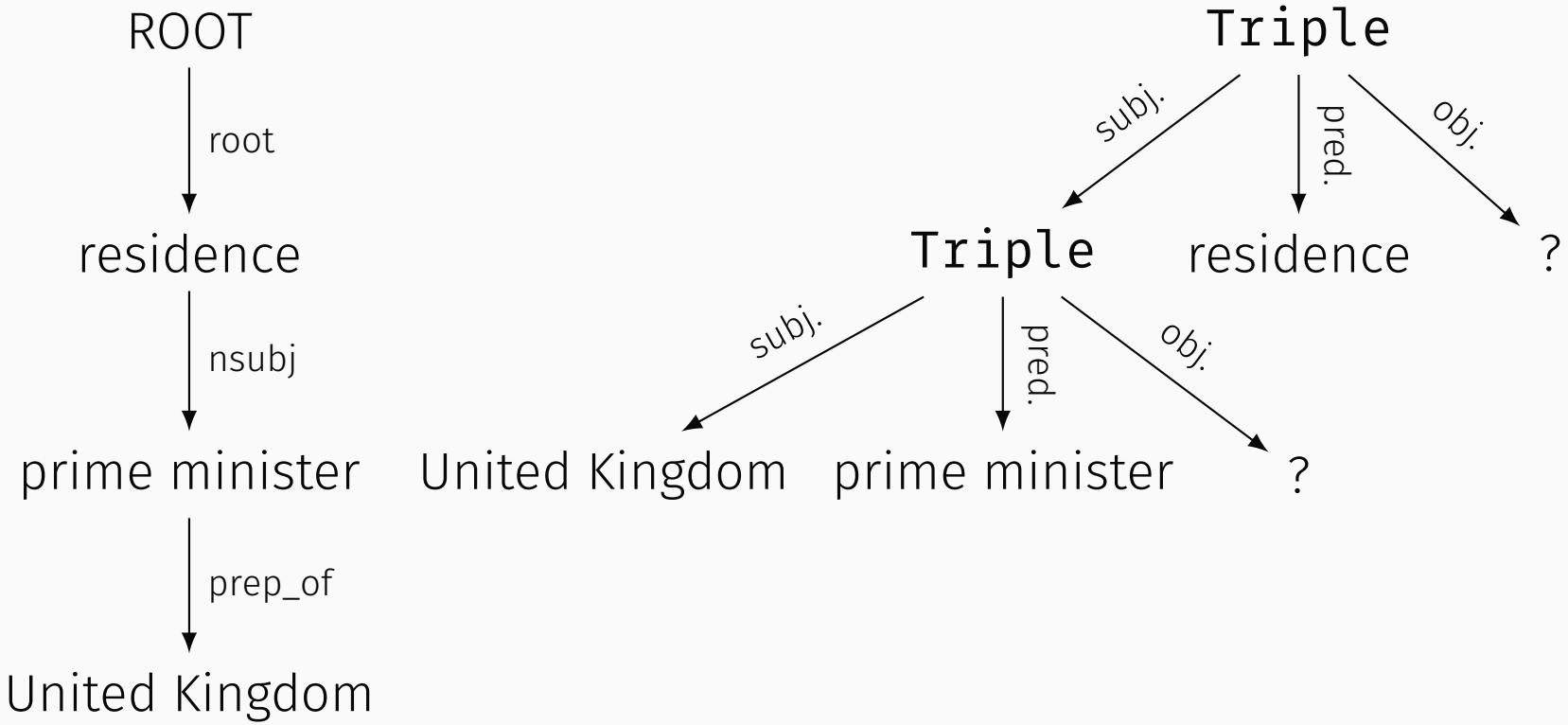
Normalization



Normalization



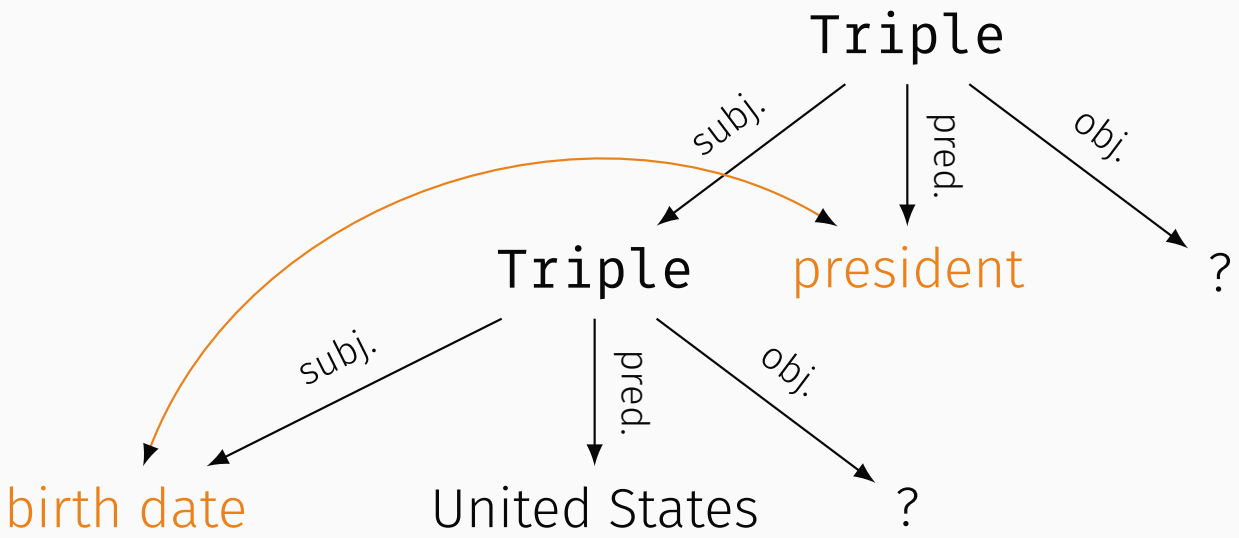
Normalization



Machine Learning – Reformulation

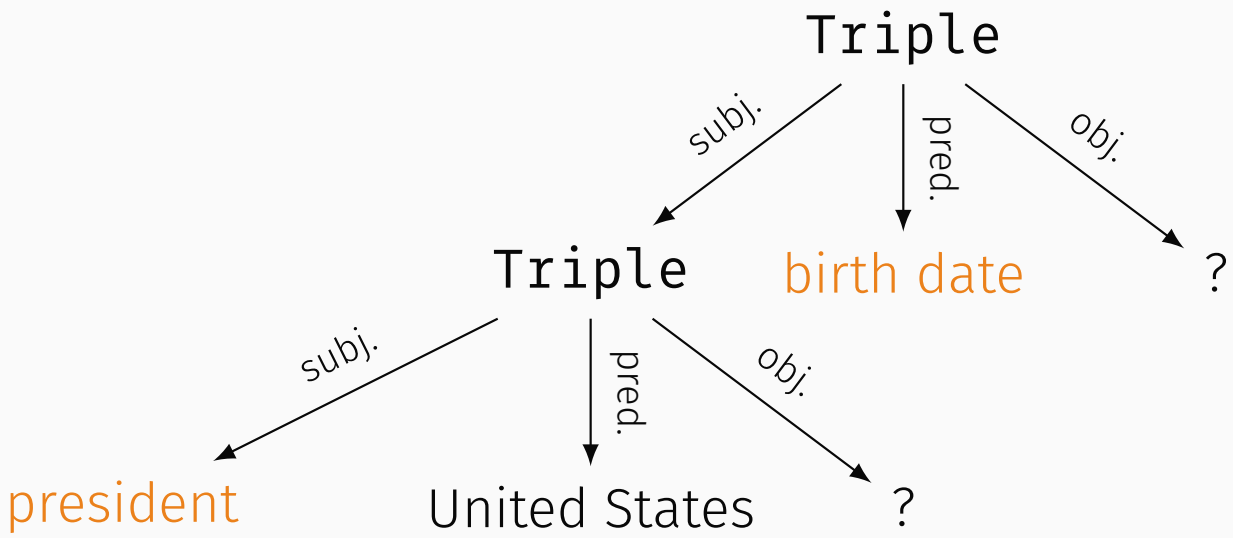
Reformulation: a translation module

When was the president of the United States born?



Reformulation: a translation module

When was the president of the United States born?

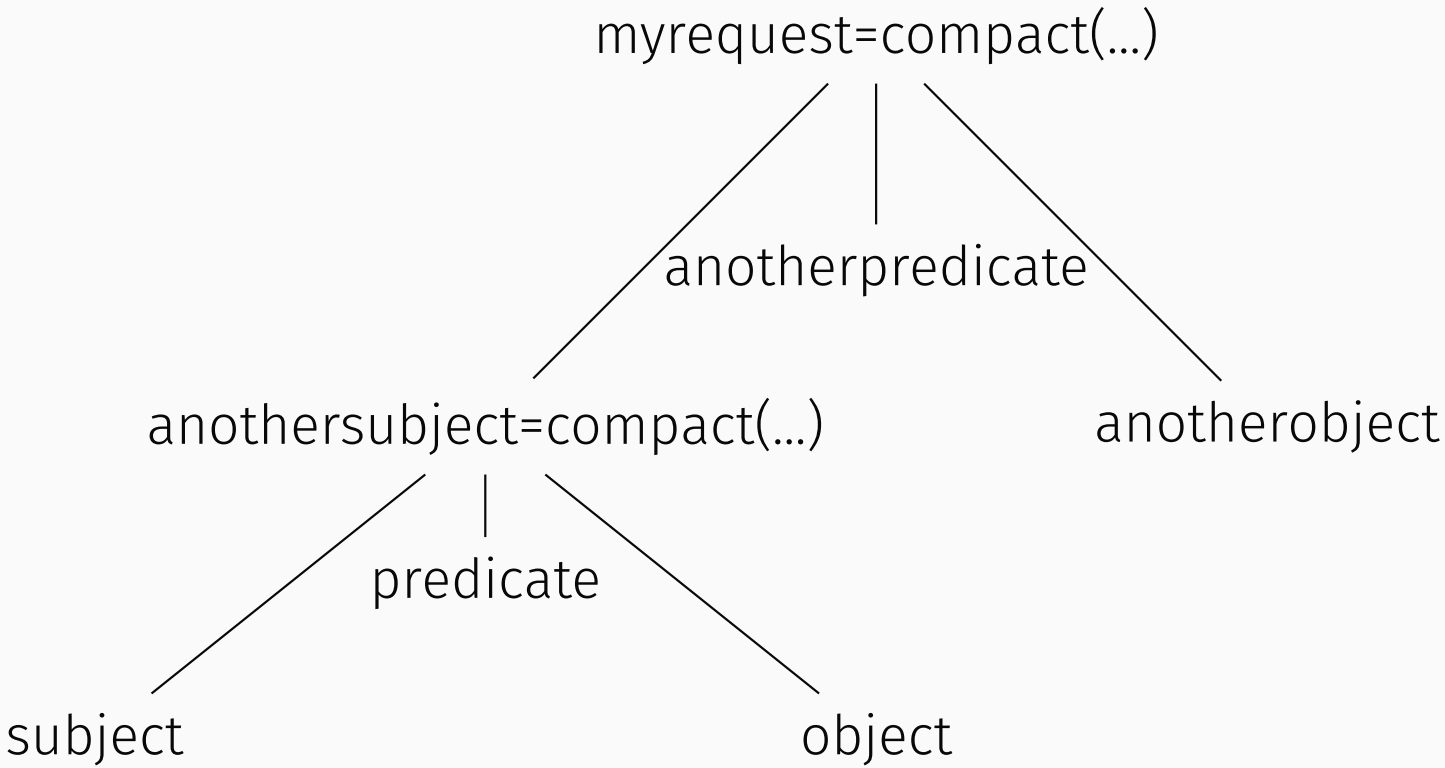


Operation

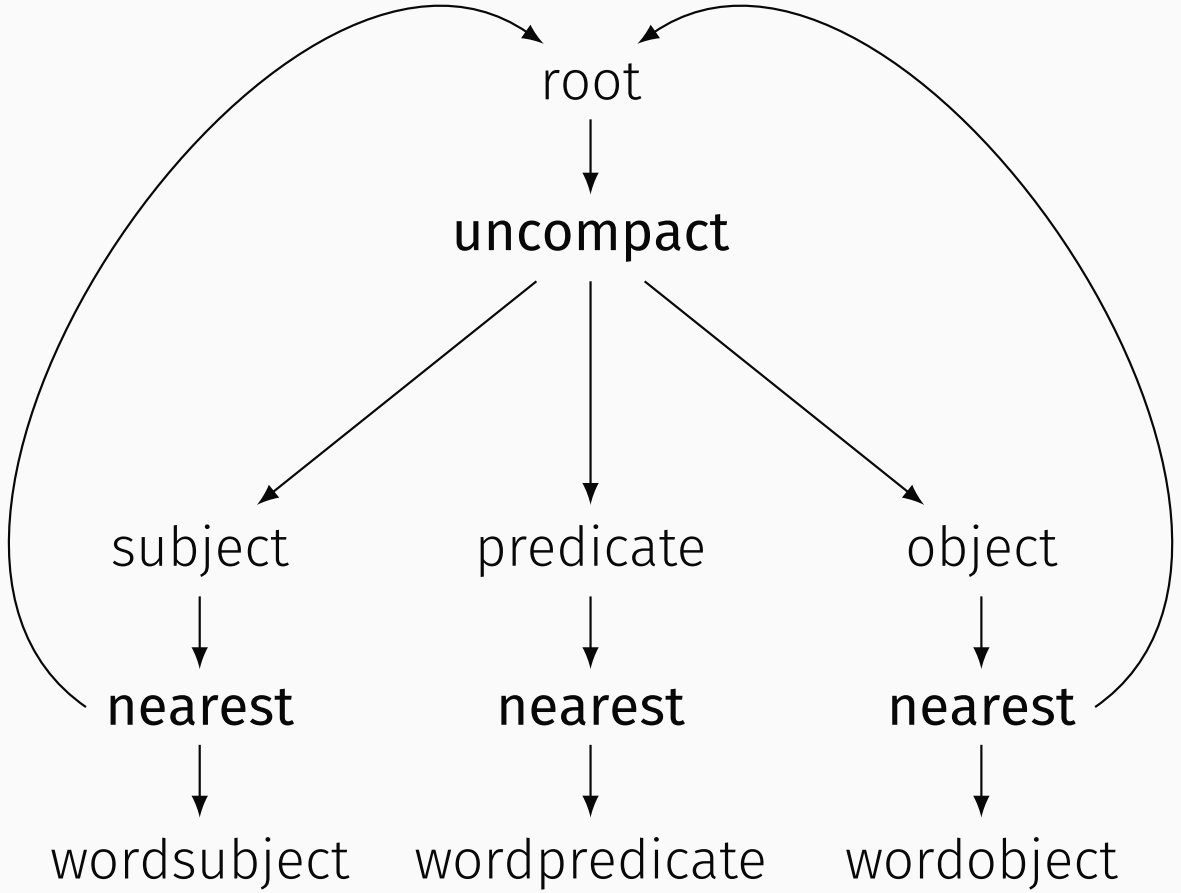
Projection: trees $\mapsto$ vector space.

The dictionary associates a vector triple to a vector.

Projection


myrequest=compact(...)
anotherpredicate
anothersubject=compact(...)
anotherobject
predicate
subject
object

From vector to tree



Machine Learning – Standalone

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

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

If two words w_1 and w_2 are synonymous then,

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

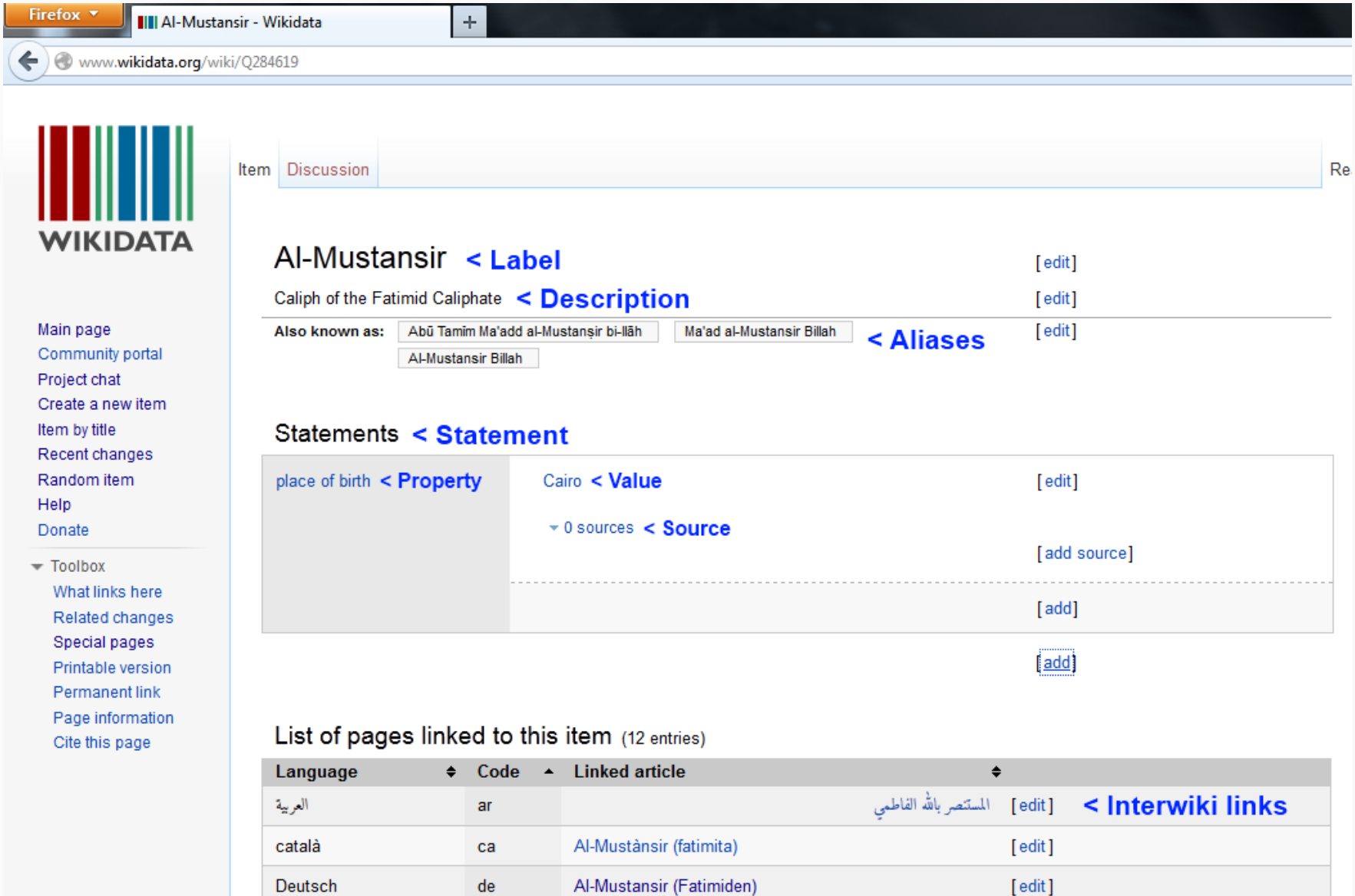
BACK-END

Wikidata

Wikidata (http://wikidata.org)?

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

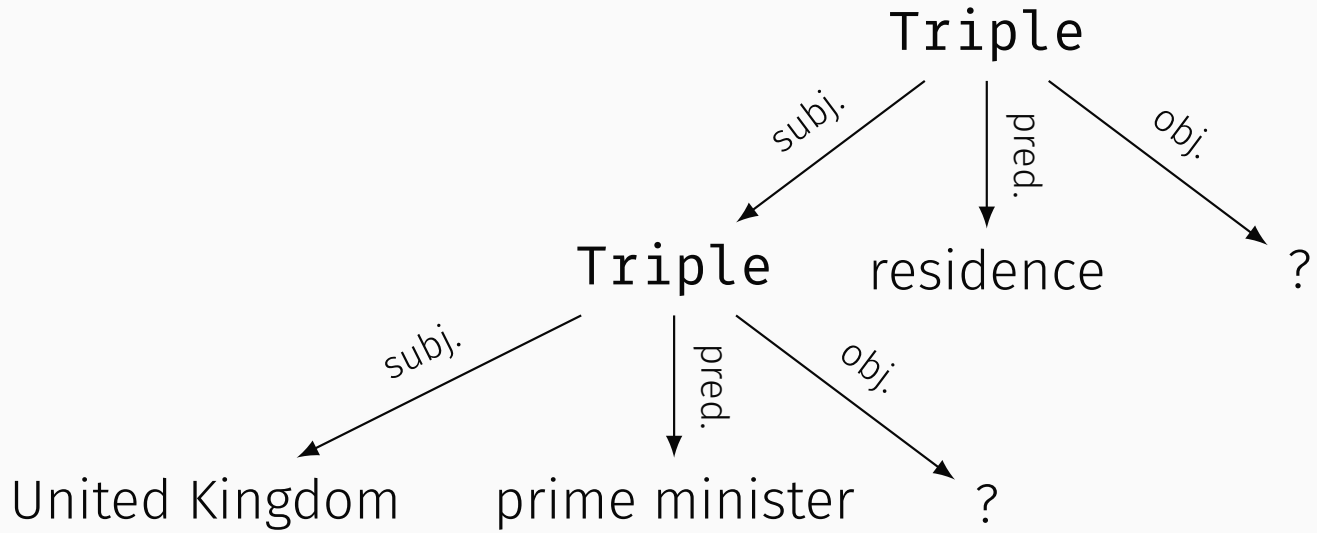
An *item*


Firefox
Al-Mustansir - Wikidata
www.wikidata.org/wiki/Q284619
WIKIDATA
Item
Discussion
Main page
Community portal
Project chat
Create a new item
Item by title
Recent changes
Random item
Help
Donate
Toolbox
What links here
Related changes
Special pages
Printable version
Permanent link
Page information
Cite this page
Al-Mustansir < Label
[edit]
Caliph of the Fatimid Caliphate < Description
[edit]
Also known as:
Abū Tamīm Ma'add al-Mustanṣir bi-llāh
Ma'ad al-Mustansir Billah
Al-Mustansir Billah
< Aliases
[edit]
Statements < Statement
place of birth < Property
Cairo < Value
[edit]
0 sources < Source
[add source]
[add]
[add]
List of pages linked to this item (12 entries)
Language
Code
Linked article
العربية
ar
المستنصر بالله الفاطمي
[edit]
< Interwiki links
català
ca
Al-Mustànsir (fatimita)
[edit]
Deutsch
de
Al-Mustansir (Fatimiden)
[edit]

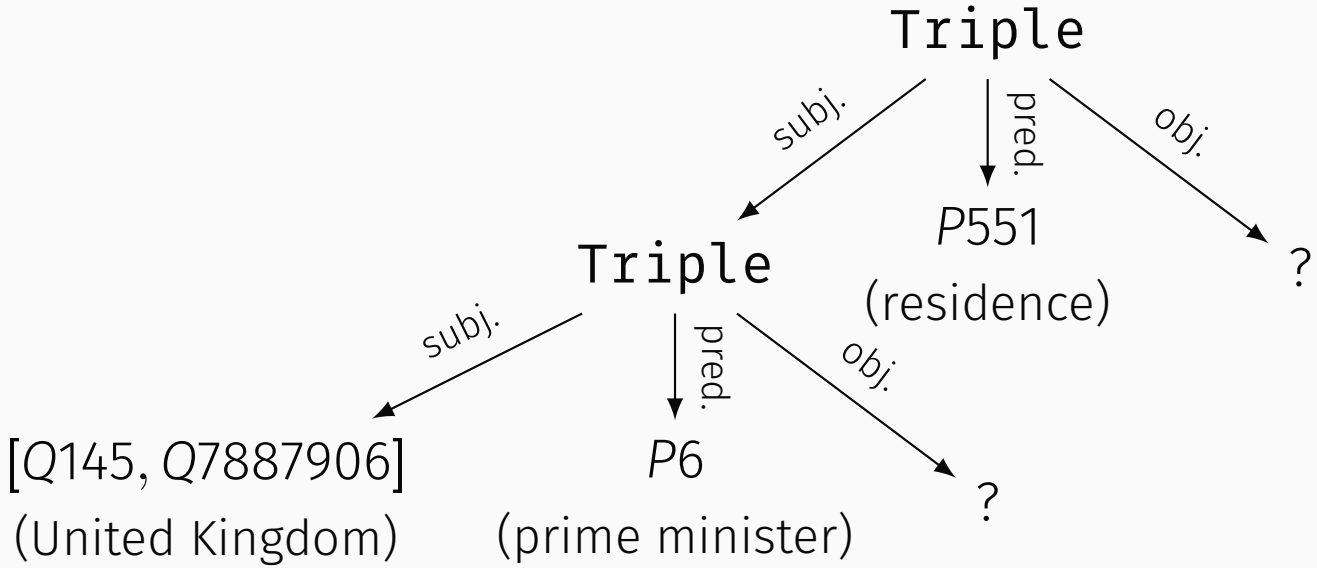
Answer extraction: Question

Where does the prime minister of the United Kingdom live?

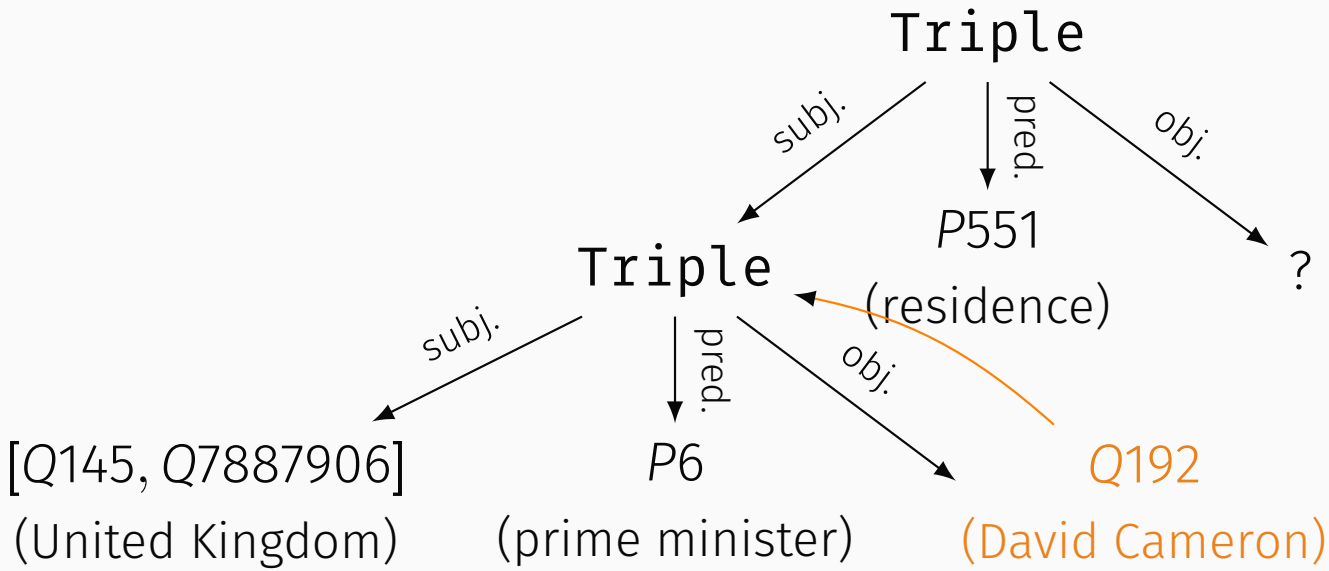
Answer extraction: Normal form



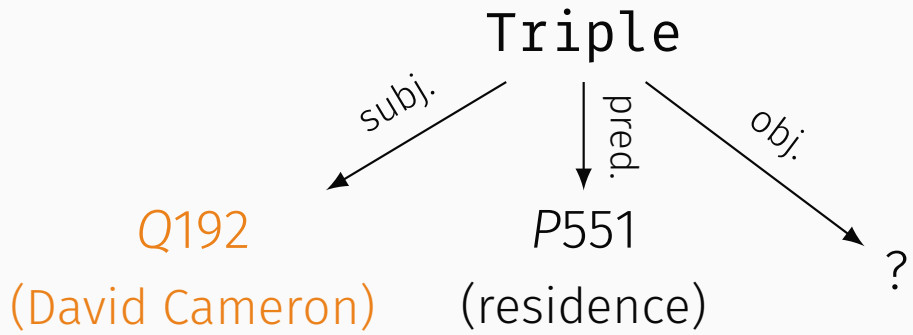
Answer extraction: Module work



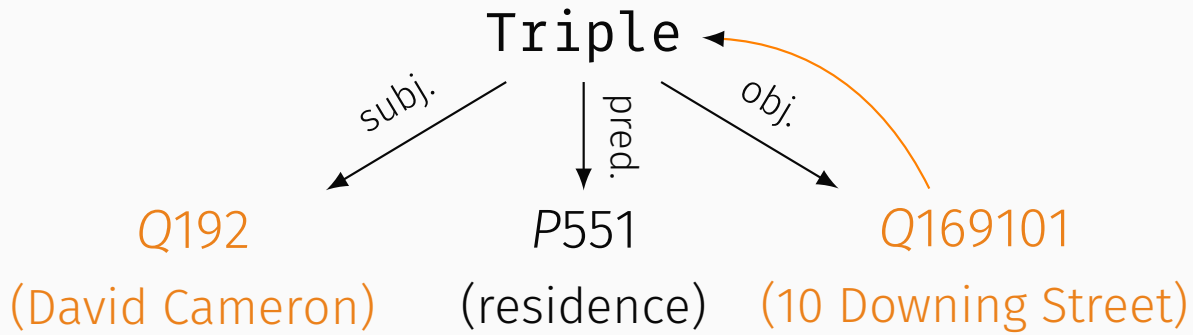
Answer extraction: Module work



Answer extraction: Module work



Answer extraction: Module work



Answer extraction: Module work

*Q*169101
(10 Downing Street)

Answer extraction: Final output

Ask me anything alpha version

Where does the prime minister of the United Kingdom live?

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. Wikipedia



Simple research

Ask me anything alpha version

10 Downing Street

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; The building contains approximately one hundred rooms. There is a private residence on the third floor and a kitchen in the basement. Wikipedia

CAS

Computer Algebra System

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

Computer Algebra System

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

Two syntaxes:

Computer Algebra System

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

Two syntaxes:

- Mathematica

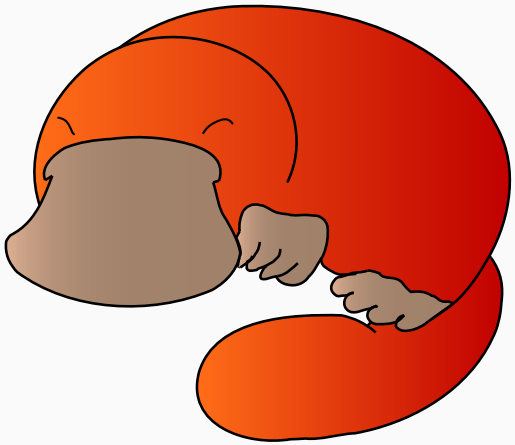
Computer Algebra System

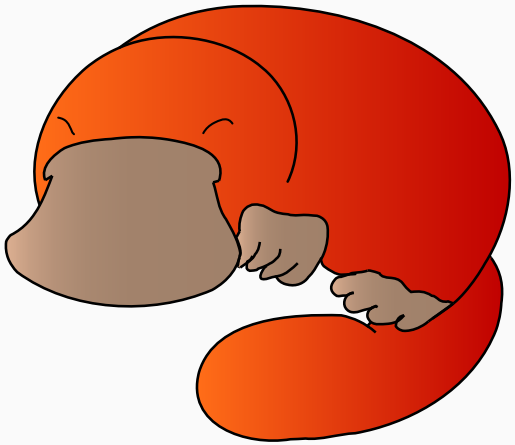
Based on Sympy (math) and PLY (parsers).

Two syntaxes:

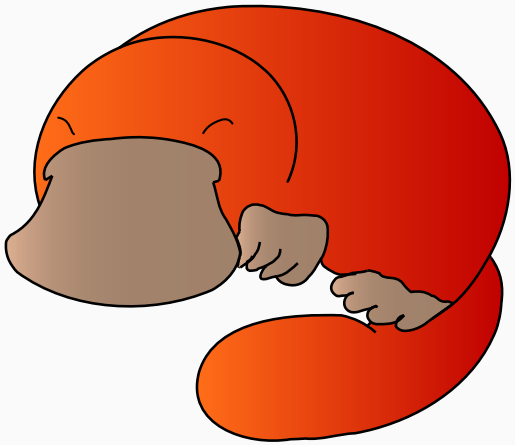
- Mathematica
- Intuitive syntax with permissive notations

LET’S SUM UP





VS



VS

Wolfram Alpha®

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?”

→ Improve Wikidata?

→ 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

meteo

literature

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	LaTeX	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/

https://twitter.com/ProjetPP

https://github.com/ProjetPP

contact@askplatyp.us

