QUESTION

Input: plain text question

Types:
- Who is the [director] of Y?
- Who [directed] Y?
- What [movies] were directed by Y?

- How long is Y?
- Where was Y [born/filmed]?

- ???

Has a direct effect on

Detect question type

Perform linguistic analysis:

HOW?

Use:
- Part of speech
- Dependency relations
- Entity detection
- Key words

Internal variable, string

Has a direct effect on

Find property and entity

Position of property and entity depend on question type / linguistic analysis

Type: strings

Text

Order of wd:, wdt: and variable can vary according to question type, e.g. different with passive questions

Construct Wikidata query

Since we know we are searching in the movie domain: leverage this?

HOW?

Type: string

Select correct answer(s)

Since multiple things in real life (e.g. movies, people or roles) can have the same, we should select the answer(s) given

Type: JSON object

Format should be compatible with testing

Write answers to file/terminal

Other problems:

- Construct our own test set for evaluation performance
- Add more question types: which appear in the Test Questions?
- What is the desired answering format?
- Do we use any external libraries (etc.), instead of only SpaCy?