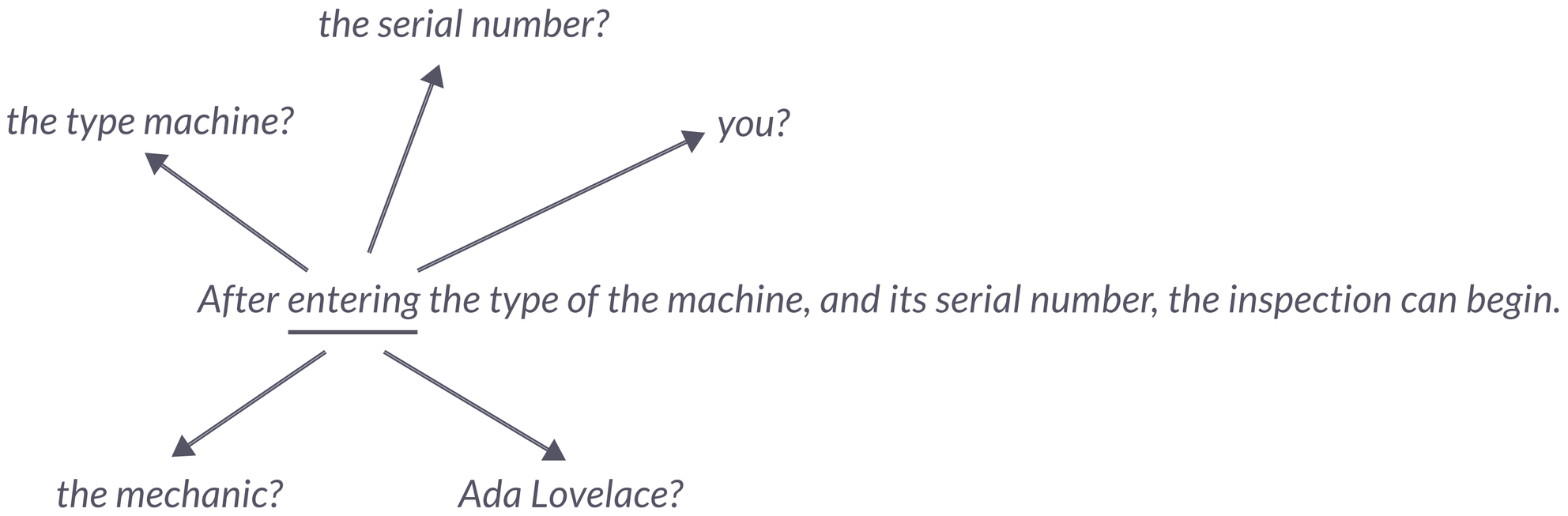
Detection and Insertion of Implicit Subject

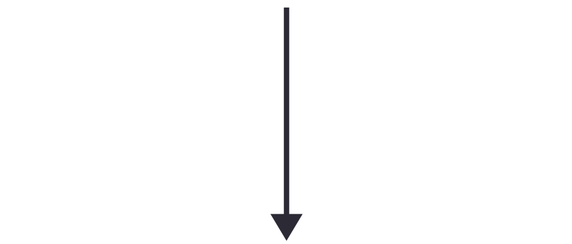
Approaching Information System Challenges with Natural Language Processing

Lukas Rossi


the serial number?
the type machine?
you?
After entering the type of the machine, and its serial number, the inspection can begin.
the mechanic?
Ada Lovelace?

Inspected Text

"After entering the type of the machine, and its serial number, the inspection can begin."



- Passive Detector
- Gerund Detector
- Nominalized Gerund Detector
- Imperative Detector

1. Implicit Subject Detection

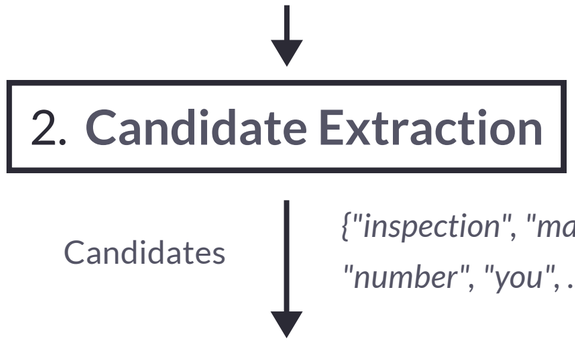
{"entering"}

Detections

Context

"Inspection of an Energy Drink Bottling Machine

You develop an application that helps you with the inspection of a machine. After entering the type of the machine, and its serial number, the inspection can begin:
..."



2. Candidate Extraction

Candidates

{"inspection", "machine", "number", "you", ...}

3. Candidate Filtering

- POS Filter
- Perplexity Filter
- ...

Detection + Selected Candidate + Original Sentence

"entering" + "you" + "After entering the type ..."

4. Candidate Insertion

"After you enter the type of the machine, and its serial number, the inspection can begin."

After entering the type of the machine, and its serial number, the inspection can begin.

you?
the serial number?
Ada Lovelace?
~~*the type of machine?*~~
the mechanic?

Filter syntactic improbabilities

Techniques: POS tag analysis, dependency analysis

you?
~~*the serial number?*~~
~~*Ada Lovelace?*~~
~~*the type of machine?*~~
the mechanic?

Filter semantic improbabilities

Techniques: Likelyhood estimates, semantic similarity estimates, analysis of previous relations

you?
~~*the serial number?*~~
~~*Ada Lovelace?*~~
~~*the type of machine?*~~
~~*the mechanic?*~~

Tie breaking

Techniques: Context topic analysis, asking generative AI,

Candidate Filter

Detection Example: Gerund Detector

After entering the type of the machine, and its serial number, the inspection can begin.

VBG

~~*Someone*~~ *entering the type of the machine begins the process.*

(nominal) subject

The ~~person~~ entering the type of the machine starts the process.

adjectival modifier/clause

Filter Example: Perplexity Filter

Semantic improbability filter

*"A job interview can be negotiated by **<UNK>**."*

~~*"A job interview can be negotiated by **[a job]**."*~~

*"A job interview can be negotiated by **[you]**."*

*"A job interview can be negotiated by **[reviews]**."*

~~*"A job interview can be negotiated by **[applications]**."*~~

$$\text{perplexity}(W) = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i|w_1 \ldots w_{i-1})}}$$

Gold Standard

Synthetic Business Process Descriptions

"You are planning a LAN party for 10 friends, so the first thing you have to do is to send invitations to these 10 friends. Next, you have to find out which games they want to play. As soon as you have received a list of games, you can appoint a date when the LAN party is going to..."

Simple

Synthetic

GDPR

"The controller shall communicate any rectification or erasure of personal data or restriction of processing carried out in accordance with Article 16, Article 17(1) and Article 18 to each recipient to whom the personal data have been disclosed, unless this proves impossible or..."

Complex

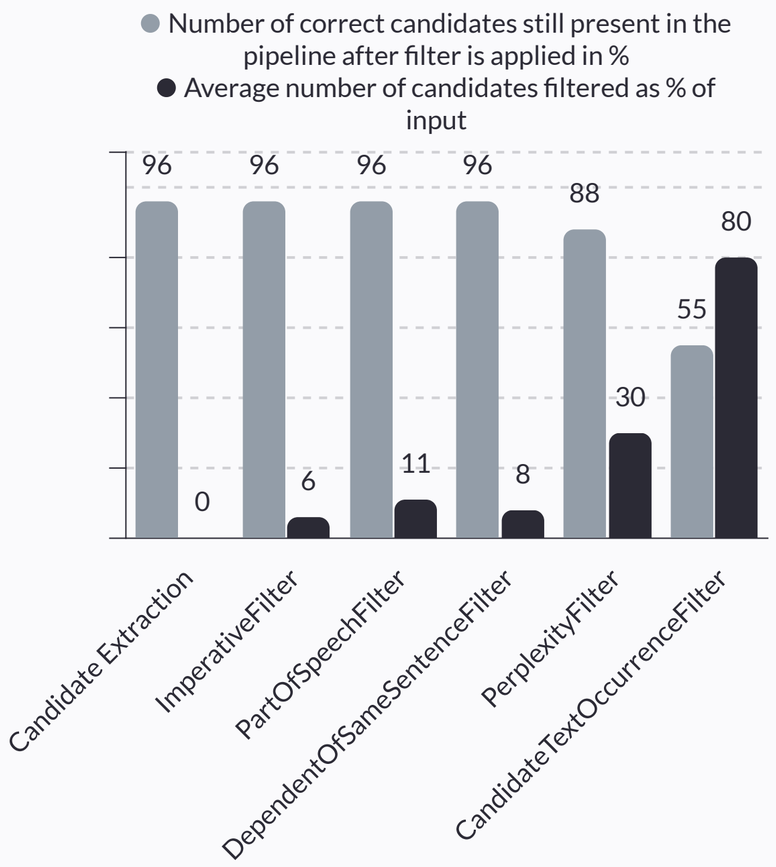
Real

Detection Results

Recall:
83%

Precision:
53%

Filtering Results



E2E Results

27% of GS entries matched

Time for some Code