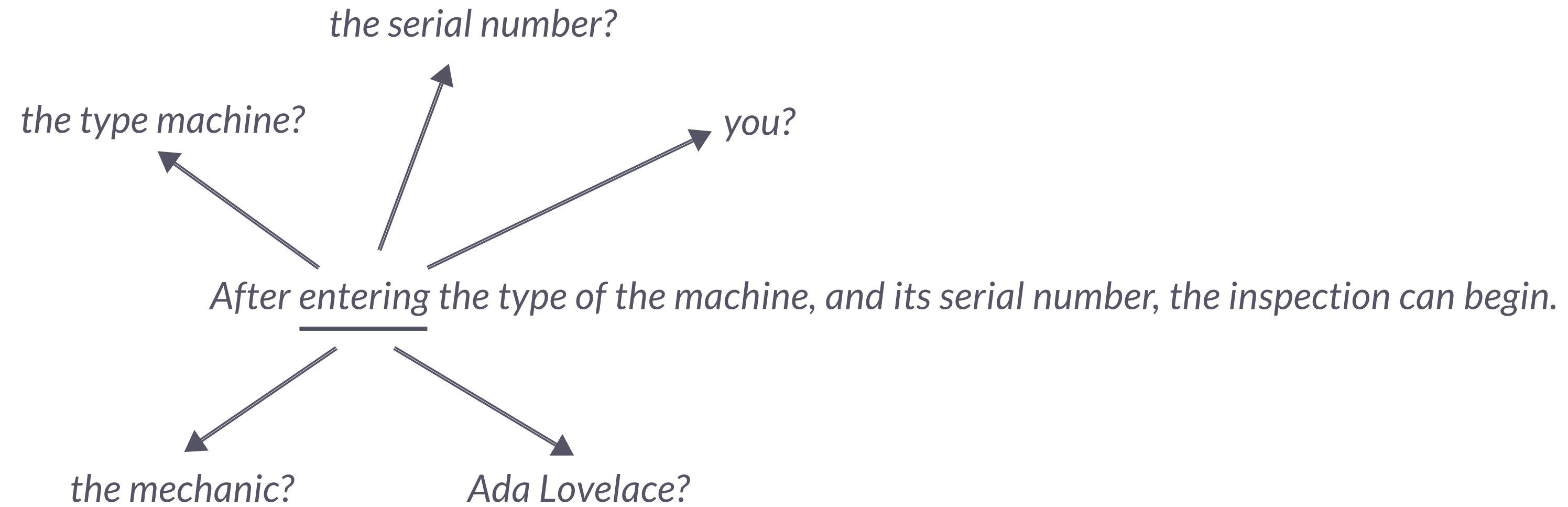


Detection and Insertion of Implicit Subject

Approaching Information System Challenges with Natural Language Processing

Lukas Rossi





Inspected Text

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



1. Implicit Subject Detection

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



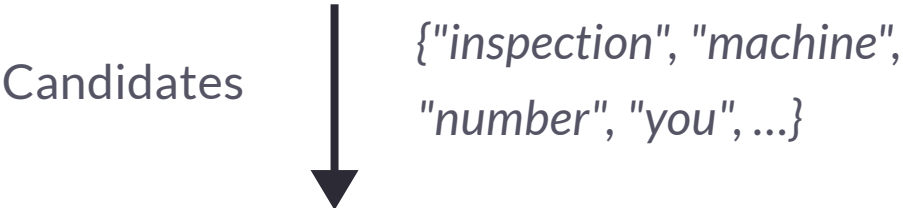
Context

"Inspection of an Energy Drink Bottling Machine"

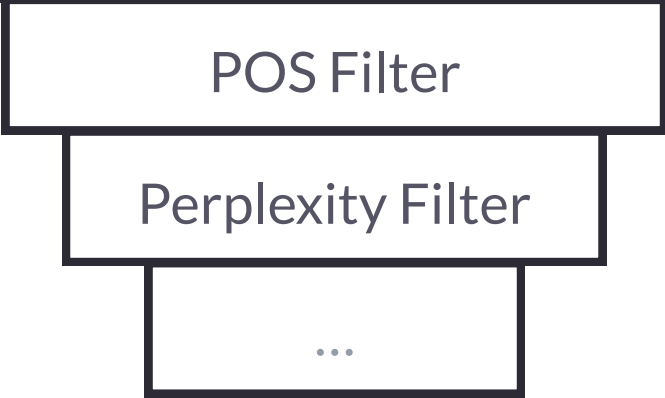
You develop an application that helps you with the inspection of a machine. After entering the type of the machine, and its serial number, the inspection can begin: ..."



2. Candidate Extraction



3. Candidate Filtering



4. Candidate Insertion



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*

Candidate Filter



you?

~~the serial number?~~

~~Ada Lovelace?~~

~~the type of machine?~~

~~the mechanic?~~



Tie breaking

Techniques: Context topic analysis, asking generative AI,

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

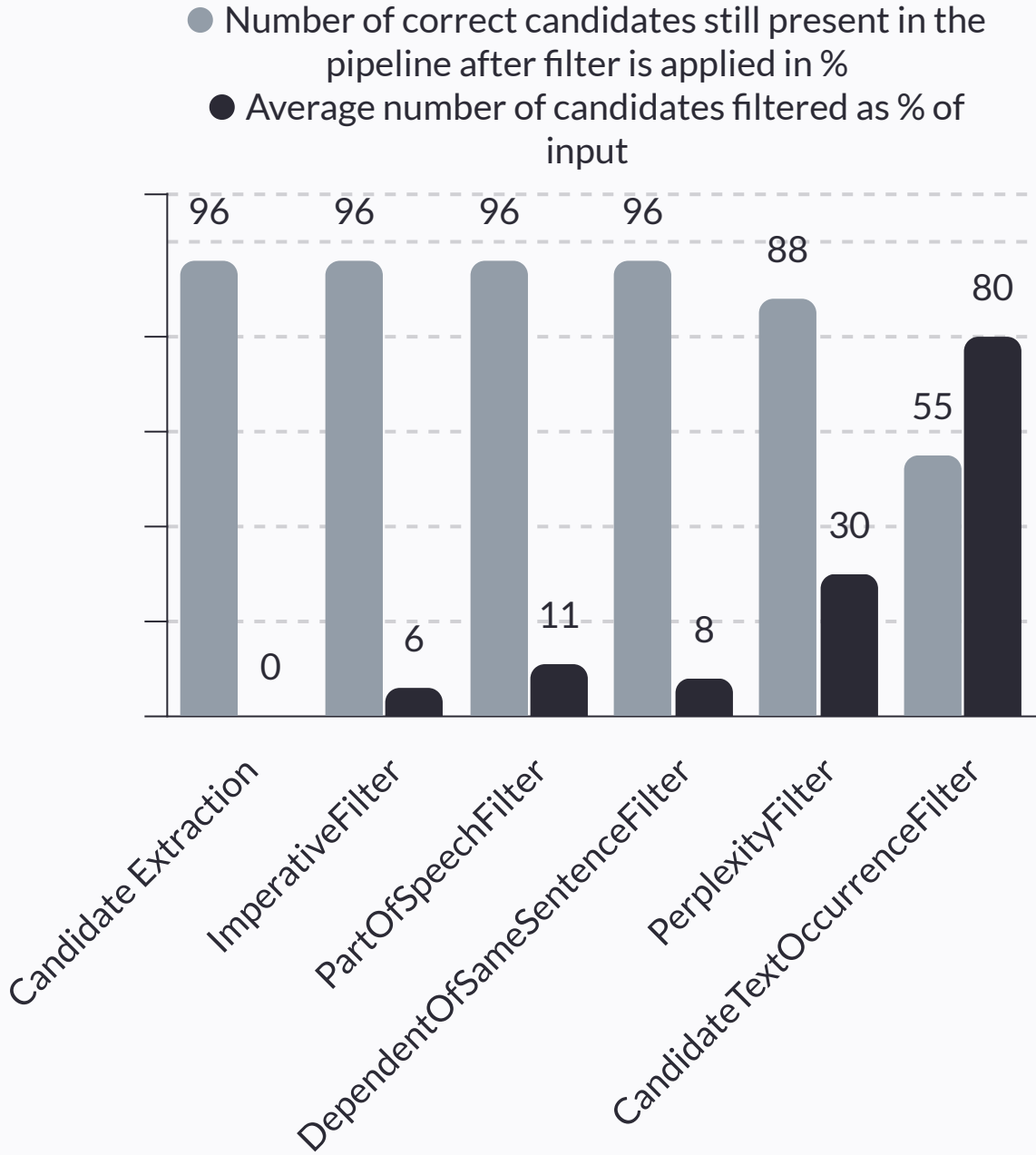
Real

Detection Results

Recall:
83%

Precision:
52%

Filtering Results



E2E Results

27% of GS entries
matched

Time for some Code