

Approaching Information Systems Challenges with NLP

Rule-based Transformation of regulatory passive sentences into active voice

Xena Striebel

30.01.2024



01 Introduction

02 Project Overview

03 Basic Pipeline

04 Evaluation

05 Code Demo

06 Limitations &
Outlook

Introduction

Topic C:

Transformation from passive to active voice of regulatory sentences using the rule-based approach

Input

“These details, which shall be regulated by a federal law, require consent of the Bundestag.”

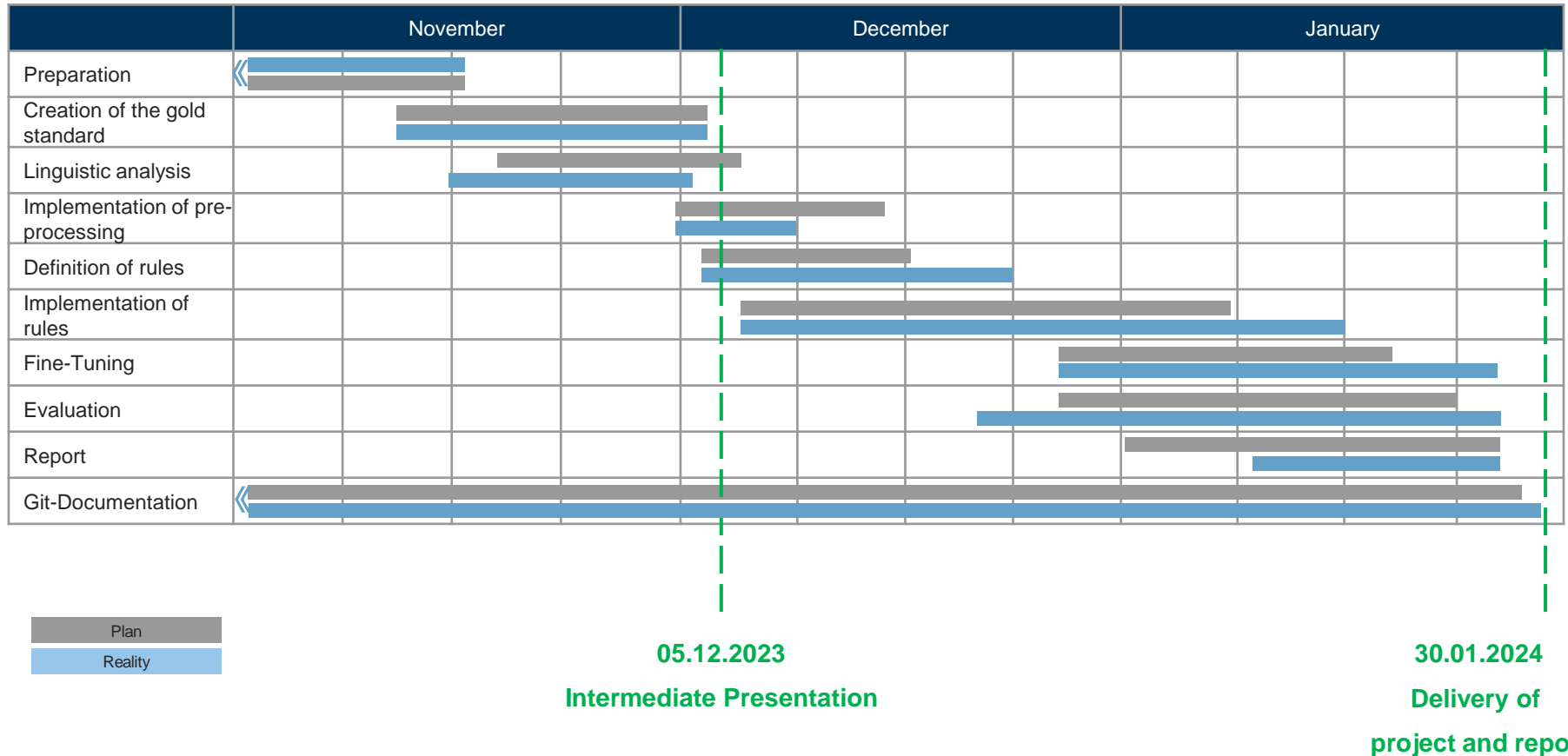


Output

“These details, which a federal law shall regulate, require consent of the Bundestag.”

Project Overview

My plan to approach the problem consisted of 10 subparts and three milestones. Overall, the main parts were followed as planned with some small deviations.



Preparation: Goldstandard

I identified and extracted passive sentences from three regulatory documents and defined the active voice for each sentence.

Documents

1. **Basic Law**
2. **GDPR**
3. **Regulation on medical devices**

Criteria

1. **Divers length**
2. **Divers complexity and structure**
3. **Different tenses**
4. **Sentences with relative clauses and subclauses**
5. **Some sentences with negoations**



Σ 160 Sentences to form the Goldstandard

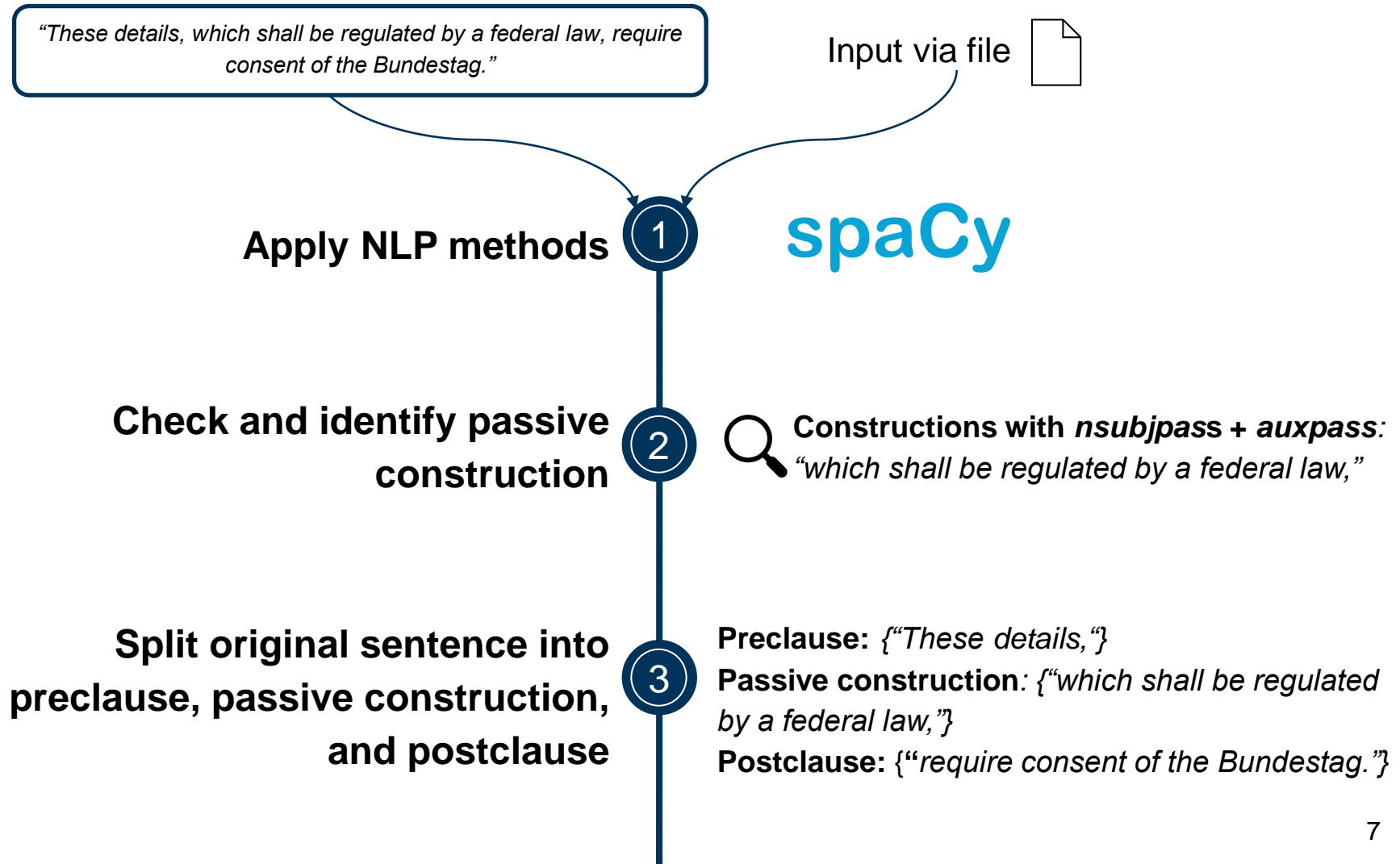
Preparation: Literature

There have already been a few basic approaches on how to transform passive voice into active voice or vice versa. However, they are all very limited to predefined, not complex sentences as input.

Related Work:

- Fahad and Beenish (2020): An Approach towards Implementation of Active and Passive voice using LL(1) Parsing
- Ilukkumbura and Rupasinghe (2021): Sinhala Active Voice into Passive Voice Converter using Rule Based Approach with Grammar Error Correction
- Madaan et al. (2016): A novel approach to paraphrase english sentences using natural language processing
- Pawale et al. (2015): sentence correction for english language using Grammar rules and syntax parsing

Basic Pipeline



Analyse each passive construction

4

Analysis of POS tags and dependencies:

{ subpass: „which“

aux: „shall“, „be“

verb: „regulated

agent: „a federal law“ }

Conjugate active verb with library pattern

5

{activeVerb: „shall regulate“ }

Adapt components and compose them and the active verb to get active construction

6

Active construction:

{“which a federal law shall regulate,“}

Assemble everything to get final active sentence

7

Preclause + Active construction + Postclause:

{“These details,“} + {“which a federal law shall regulate,“} + {“require consent of the Bundestag.”}

In the case of more than one passive construction in the original sentence:

Combine all sentences with one active construction to get final sentence with only active constructions

8

Original:

*It **shall be tested** and **it shall be called** the configurable device.*

First construction:

*One **shall test it** and it shall be called the configurable device.*

Second construction:

*one **shall call it** the configurable device.*

Final:

*One **shall test it** and **one shall call it** the configurable device.*

Return final sentence

9

These details, which a federal law shall regulate, require constant of the Bundestag.

Evaluation

For precision and recall both the ability to identify passive constructions and correctness of the passive-to-active conversion were considered. To compare the output with the goldstandard sentence semantically the SBERT method was applied.

Defined Metrics

- **True Positive (TP):** correctly identified as passive and transformed correctly (SBERT-Score > 0.95)
- **False Positive (FP):** wrongly identified as passive and attempted to transform
- **True Negative (TN):** correctly identified as active and thus not transformed
- **False Negative (FN):** (wrongly identified as active and not transformed) PLUS (correctly identified as passive and transformed wrongly (SBERT-Score \leq 0.95))

TP: 150 FP: 0 TN: 15 FN: 10



Recall: 0.9375

Precision: 1.0

Accuracy Rate by human evaluation:
93.75%

Code Demo

Limitations & Outlook

Three main limitations were identified:

1

LIMITATIONS:

Some sentences having more than one passive construction:

- Usually more complex and nested
- Composition of transformed components to get final sentence difficult



POSSIBLE APPROACHES:

Use LLM or ML to compose components correctly
Check output automatically for grammar and punctuation

2

Parsing difficulties and ambiguities with longer and more complex sentences.



Test other parsers like NLTK or Stanford Parser
Use of context-free grammar and syntax parsing

3

If no agent is provided, the default agent “one” was used.



Using LLM or ML could be possible to derive the agent from the context.

Thank you for your attention!

Any questions?