

GECko

Software project, work update 2

Eduardo Calò, Léo Jacqmin, Thibo Rosemplatt

NLP M2
IDMC, Université de Lorraine

October 6th, 2020

Table of Contents

Pre-processing

Back-end

Front-end

Pipeline

Action plan

Table of Contents

Pre-processing

Back-end

Front-end

Pipeline

Action plan

File format

■ Incorrect-correct parallel dataset:

Source: The rich people will buy a car but the poor people always need to use a bus or taxi

.

Target: Rich people will buy a car , but poor people always need to use a bus or taxi .

```
$STARTSEPL||SEPR$KEEP TheSEPL||SEPR$DELETE richSEPL||SEPR$TRANSFORM_CASE_CAPITAL peopleSEPL||SEPR$KEEP willSEPL||SEPR$KEEP buySEPL||-  
SEPR$KEEP aSEPL||SEPR$KEEP carSEPL||SEPR$APPEND_, butSEPL||SEPR$KEEP theSEPL||SEPR$DELETE poorSEPL||SEPR$KEEP peopleSEPL||SEPR$KEEP  
alwaysSEPL||SEPR$KEEP needSEPL||SEPR$KEEP toSEPL||SEPR$KEEP useSEPL||SEPR$KEEP aSEPL||SEPR$KEEP busSEPL||SEPR$KEEP orSEPL||SEPR$KEEP  
taxiSEPL||SEPR$KEEP .SEPL||SEPR$KEEP
```

Figure: Output of *preprocess_data.py*, input for the model.

■ 5000 tags (KEEP, DELETE, APPEND, CASE, MERGE...)

Table of Contents

Pre-processing

Back-end

Front-end

Pipeline

Action plan

Overview

Project Workflow



Table of Contents

Pre-processing

Back-end

Front-end

Pipeline

Action plan

Preview

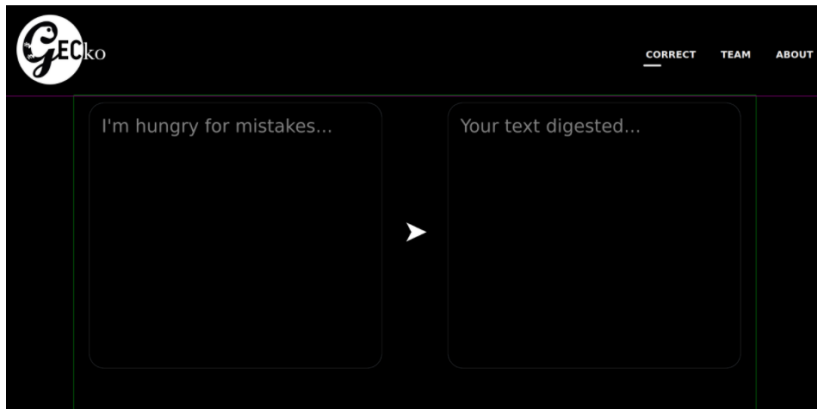


Table of Contents

Pre-processing

Back-end

Front-end

Pipeline

Action plan

Pipeline

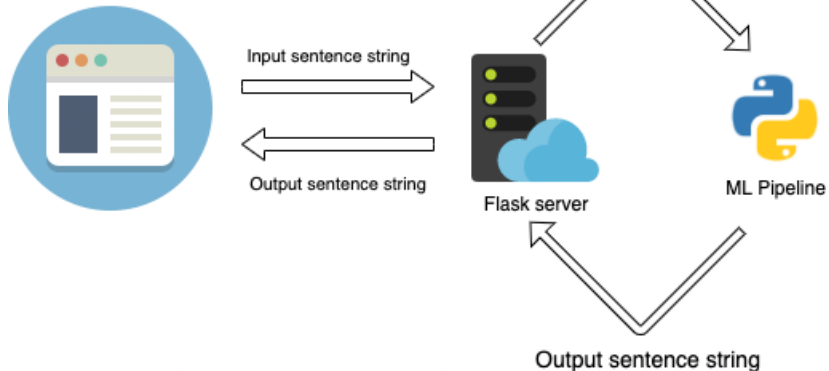


Table of Contents

Pre-processing

Back-end

Front-end

Pipeline

Action plan

Milestones (revised)



