

# 2021年03月31日进度汇报

屈原斌  
首都师范大学  
ybqu@cnu.edu.cn

## 1 今日进度

1. Baseline复现（未完成）
2. 成语古诗文检错改进方案

## 2 工作详情

### 2.1 Baseline复现

- 完成特征的提取，共七个特征：
  1. Random Indexing: 使用S-Space package在Gigaword数据集上训练词向量
  2. N-grams: 提取1000个unigram, bigram和trigram特征
  3. Thesis Clarity Keywords
  4. Prompt Adherence Keywords
  5. LDA Topics
  6. Manually Annotated LDA Topics
  7. Predicted Thesis Clarity Errors

### 2.2 成语古诗文检错方案

- 方案流程图见图1：

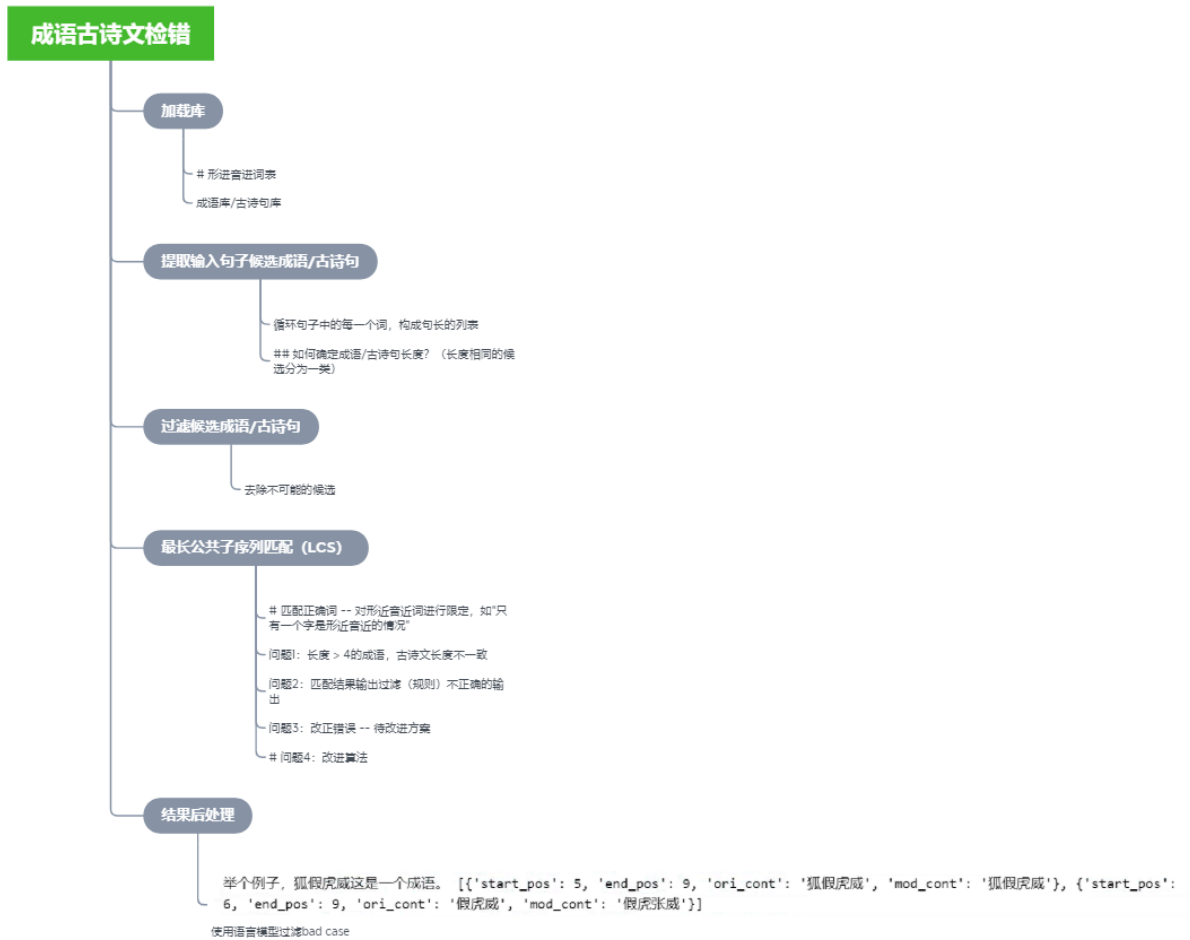


Figure 1: 成语古诗文检错方案流程图