

Documentação do Modelo de Correção de Redações para a Competência 1 do ENEM

1. Introdução

Este projeto utiliza técnicas de **Sistemas Baseados em Conhecimento (SBS)** para corrigir redações com base na **Competência 1** do ENEM, que avalia o domínio da norma culta da língua escrita. O modelo identifica erros ortográficos, gramaticais e de concordância, fornecendo uma nota prevista para a redação em uma escala de 0 a 2.

2. Técnicas de Sistemas Baseados em Conhecimento (SBS) Utilizadas

2.1 Regras Baseadas em Conhecimento

O modelo utiliza duas abordagens principais para identificar erros nas redações:

- **Correção Ortográfica:** Utiliza a biblioteca **SymSpell** para aplicar regras ortográficas. A técnica faz uma busca de termos no dicionário e verifica a normalização das palavras, removendo acentuação para melhorar a detecção de erros.
- **Correção Gramatical e de Concordância:** Através da biblioteca **LanguageTool**, o modelo aplica regras de gramática para detectar erros como concordância nominal e verbal, uso inadequado de artigos e preposições, e erros de conjugação verbal.

Além disso, o modelo inclui uma verificação manual de **erros de concordância** com base em um arquivo de referência (`errosConcordancia.txt`), onde exemplos de erros são listados e constantemente atualizados conforme o modelo detecta novos erros durante a correção.

3. Funcionalidades do Código

3.1 Função de Extração de Features

A função `extrair_features()` coleta diversas características da redação, incluindo:

- Número de caracteres, palavras, parágrafos, e palavras únicas;
- Média de tamanho de palavras e parágrafos;
- Contagem de erros ortográficos, gramaticais e de concordância.

3.2 Treinamento do Modelo

O modelo utiliza **Regressão Linear** (via `LinearRegression` da biblioteca `sklearn`) para prever a nota da Competência 1 de acordo com os erros encontrados na redação.

3.3 Salvamento de Novos Erros de Concordância

Durante o treinamento, novos erros de concordância detectados são adicionados ao arquivo `errosConcordancia.txt`, tornando o sistema mais abrangente.

3.4 Avaliação de Redações

O modelo pode ser utilizado para avaliar redações inseridas pelo usuário. Ele retorna uma nota prevista para a Competência 1, além de listas de erros ortográficos, gramaticais e de concordância.

4. Como Utilizar o Código

4.1 Pré-requisitos

- Python 3.7+
- Instale as dependências necessárias, atualizando o arquivo `requirements.txt` com os seguintes pacotes:

```
symspellpy==6.7.0
language-tool-python==0.9.6
scikit-learn==1.2.0
joblib==1.3.0
numpy==1.23.0
```

4.2 Passos para Utilização

1. Treinamento do Modelo:

- O código `corretor.py` pode ser utilizado para carregar o dataset em formato XML e treinar o modelo. O modelo treinado será salvo no caminho especificado.

```
python corretor.py
```

2. Correção de Redações:

- Para corrigir uma redação, utilize o script `corrigirRedacao.py`, que solicitará o texto da redação e aplicará o modelo treinado para fornecer a nota e os erros detectados.

```
python corrigirRedacao.py
```

5. Estrutura do Código

- `corretor.py`: Responsável pelo treinamento e avaliação das redações.
- `processamento.py`: Contém funções para aplicação de regras de correção ortográfica, gramatical e de concordância.
- `errosConcordancia.txt`: Arquivo que armazena exemplos de erros de concordância e suas correções sugeridas.
- `modelo_treinado.pkl`: Modelo de regressão linear salvo após o treinamento.

6. Exemplo de Uso

Ao rodar o `corrigirRedacao.py`, o usuário insere o texto da redação e obtém a saída semelhante à seguinte:

Digite o texto da redação para correção:
Hoje as pessoas estão cada vez mais utilizando a tecnologia para tudo.

Nota Prevista para Competência 1: 1.5

Erros Ortográficos:

- Palavra incorreta: ...

Erros Gramaticais:

- Erro detectado: "Hoje as pessoas..."

Sugestão: "Hoje as pessoas"

7. Conclusão

Este sistema é uma poderosa ferramenta para corrigir redações automaticamente, utilizando um sistema baseado em conhecimento que aprende com os erros encontrados e se adapta às correções sugeridas.