

Readme da API IDP

Visão Geral

A API Intelligent Diacritic Placer (IDP API) é uma aplicação baseada em Python e Flask, projetada para converter textos em inglês para o formato Inglês com Acentos (IcA). Esta API serve como uma ferramenta para pesquisa linguística e fins educacionais, especialmente no contexto do Projeto PI.

Recursos

- Converte textos em inglês padrão (IP) para o formato IcA.
- Suporta diferentes sotaques ingleses para conversão fonética.
- Utiliza um processo único de conversão fonética-grafêmica com base no mapeamento grafema-fonema do SDPI.

Instalação

1. Clone o repositório.
2. Instale as dependências necessárias: `pip install -r requirements.txt`.

Uso

Inicie a API com `python app.py`. A API fornece dois endpoints principais:

1. `/api/v1/get-phonetized-text`: Aceita texto em IP e um sotaque inglês, retornando a versão fonetizada.
2. `/api/v1/convert-en-to-ewd`: Converte texto em IP para IcA, considerando o sotaque inglês especificado.

Endpoints

GET Phonetized Text

- URL: `/api/v1/get-phonetized-text`
- Método: `POST`
- Parâmetros do Corpo:

- `text`: Texto em IP a ser fonetizado.
- `accent`: Sotaque inglês para conversão fonética (padrão é `en-us`).

Convert EN to IcA

- URL: `/api/v1/convert-en-to-ewd`
- Método: `POST`
- Parâmetros do Corpo:
 - `text`: Texto em IP a ser convertido para IcA.
 - `accent`: Sotaque inglês para conversão (padrão é `en-us`).

Processo de Conversão

A conversão de IP para IcA envolve um processo de transformação fonética-grafêmica em duas etapas:

1. **Conversão Fonética**: O texto em IP é primeiramente convertido em sua representação fonética (fonetizada) com base no sotaque inglês especificado. Este processo utiliza a biblioteca `phonemizer` para alcançar transcrições fonéticas precisas.
2. **Mapeamento Grafêmico (Mapamento Grafema-Fonema do SDPI)**: Os dados fonéticos são então mapeados para sua representação correspondente em IcA usando o Mapamento Grafema-Fonema do SDPI. Este mapeamento envolve tabelas detalhadas de conversão que pareiam fonemas específicos com grafemas correspondentes (letras com diacríticos), transformando efetivamente a sequência fonética em IcA.

Documentação

Documentação Swagger está disponível em `/swagger` para uso detalhado dos endpoints.

Testes

Testes unitários são fornecidos para vários componentes. Execute os testes usando `pytest`.

Arquitetura

A IDP API é construída como uma arquitetura monolítica com foco em simplicidade e facilidade de implantação. Utiliza SQLite para gestão de dados leve e Flask como framework web.

Desenvolvimento e Implantação

- **Abordagem TDD:** Garante qualidade e minimiza bugs.
- **CI/CD:** Configuração simplificada usando GitHub Actions.
- **Monitoramento e Logging:** Monitoramento básico com Grafana e logging interno do Flask.

Segurança

- **Autenticação:** Implementação simplificada do OAuth 2.0.
- **Segurança dos Dados:** TLS básico para transmissão de dados.

Expansão para Outros Idiomas e Sotaques

Possibilidades de Expansão

A IDP API, atualmente focada na conversão do Inglês Padrão para Inglês com Acentos, foi projetada com uma arquitetura que permite expansão para outros idiomas e uma variedade de sotaques. Este potencial de expansão abre novas frentes para pesquisa linguística e aplicações educacionais em um contexto global.

Suporte a Diferentes Idiomas

A incorporação de novos idiomas na IDP API pode ser realizada através da adaptação do processo de fonetização e do mapeamento grafema-fonema para cada língua específica. Isso inclui a integração de bibliotecas fonéticas e bases de dados que suportem a transcrição fonética precisa desses idiomas.

Sotaques Variados

Assim como já é feito com o inglês, a IDP API pode ser adaptada para suportar diferentes sotaques dentro de um mesmo idioma. Isso é essencial para capturar as nuances fonéticas que caracterizam os sotaques regionais ou nacionais, enriquecendo a experiência de aprendizado e a precisão da pesquisa linguística.

Desafios e Considerações

A expansão para outros idiomas e sotaques envolve desafios como a necessidade de um conhecimento aprofundado das características fonéticas de cada língua e a disponibilidade de recursos fonéticos e lexicais detalhados. Além disso, é crucial considerar as particularidades culturais e linguísticas para garantir uma representação precisa e respeitosa dos idiomas.

Futuro da IDP API

A visão futura da IDP API inclui a criação de uma plataforma robusta e versátil que possa servir como uma ferramenta valiosa para a educação e pesquisa linguística em diversos idiomas e sotaques, promovendo assim uma compreensão mais ampla e inclusiva das diversas formas de expressão humana ao redor do mundo.

Contribuições

Como este é um projeto privado atualmente em desenvolvimento para a pesquisa do Projeto PI, as contribuições são limitadas apenas a membros da equipe autorizados. Para consultas sobre envolvimento ou colaboração, por favor, envie uma mensagem de e-mail para contact@thipimethod.org.

Contato

Para consultas ou assistência, entre em contato com contact@thipimethod.org.

Licença

Este projeto é licenciado sob a Licença Internacional Creative Commons Attribution 4.0. Para visualizar uma cópia desta licença, visite <http://creativecommons.org/licenses/by/4.0/> ou envie uma carta para Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

© Danilo Florentino Maia, 2024