**Một số lưu ý khi sử dụng bộ chuyển đổi âm tiết sang âm vị dành cho tiếng Việt**

1. **Cách sử dụng ứng dụng**

Có hai cách sử dụng dành cho người dung:

* *Cách 1: Sử dụng trên nền tảng website*

Truy cập địa chỉ website : <http://172.16.75.74/vnp/>

Nhập văn bản cần phân tích vào ô textbox rồi sau đó click chuột vào nút Phân tích văn bản, kết quả sẽ được hiển thị ở ô textbox bên dưới.

* *Cách 2: Sử dụng bằng câu lệnh trên terminal console*

Có thể chạy ứng dụng trên cửa sổ terminal trên hệ điều hành Window hoặc Linux bằng cách sử dụng câu lệnh dưới đây:

java -jar MICAPhonetizer.jar path/to/xlsx\_File path/to/file\_input path/to/file\_output

Trong đó:

+ java -jar MICAPhonetizer.jar : cú pháp để chạy file thực thi jar (MICAPhonetizer.jar ) trong môi trường java

+ path/to/xlsx\_File : đường dẫn đến tệp cơ sở dữ liệu ngữ âm được lưu trữ dưới dạng tệp excel (.xlsx)

+ path/to/file\_input: đường dẫn đến tệp văn bản đầu vào được lưu trữ theo chuẩn UTF-8

+ path/to/file\_output: đường dẫn đến tệp văn bản chưa kết quả phân tích được lưu theo chuẩn UTF-8

Dưới đây là ví dụ cụ thể cho một câu lệnh hoàn chỉnh:

java -jar MICAPhonetizer.jar D:/test/sources/phonemes.xlsx D:/test/sources/input.txt D:/test/sources/output.txt

1. **Một số quy ước trong tệp cơ sở dữ liệu ngữ âm**

Tệp cơ sở dữ liệu ngữ âm được lưu trữ dưới dạng tệp excel bao gồm 4 sheets theo thứ tư chính xác là: phụ âm đầu, âm cuối, nguyên âm và âm đệm.

Trong mỗi sheet sẽ có một bảng gồm 3 cột: cột đầu tiên là biểu diễn về mặt chữ cái của âm vị, cột thứ hai là biểu diễn âm vị theo bảng phiên âm IPA, cột thứ ba là biểu diễn theo chuẩn XSAMPA.

Khi nhà phát triển muốn cập nhật cơ sở dữ liệu vào bảng, xin vui lòng giữ đúng thứ tự của 4 sheets từ trái qua phải lần lượt là : phụ âm đầu, âm cuối, nguyên âm và âm đệm. Trong mỗi sheet, xin vui lòng giữ nguyên thứ tự lần lượt của các cột lần lượt trong bảng là: biểu diễn, IPA và XSAMPA

|  |  |  |
| --- | --- | --- |
| Biểu diễn | IPA | XSAMPA |
| b | ɓ | b |
| đ | ɗ | d |
| x | s | s |
| s | s | s |
| gh | ɣ | G |

1. **Một số lưu ý dành cho nhà phát triển**

Mã nguồn chương trình được viết bằng ngôn ngữ java

Một số hàm chính trong mã nguồn chương trình:

* **void** readPhonemeDatabaseExcell(String filename)

Là hàm đọc cơ sở dữ liệu từ tệp cơ sở dữ liệu được lưu trữ dưới dạng tệp excel (đuôi mở rộng là .xlsx)

* ArrayList<String> convertSentenceToPhonemesWithTone(String sentence)

Là hàm chuyển đổi câu hoặc đoạn văn bản thành các âm vị không kèm theo thanh điệu.

+ Đầu vào của hàm là biến String biểu diễn một câu hoặc đoạn văn bản

+ Đầu ra của hàm là một ArrayList gồm các âm vị được phân tách ra từ câu theo cả 2 bảng mã phiên âm là IPA và XSAMPA

* ArrayList<String> convertSyllableToPhonemes(String syllable)

Là hàm chuyển đổi âm tiết thành các âm vị không kèm theo thanh điệu.

+ Đầu vào của hàm là biến String biểu diễn một âm tiết

+ Đầu ra của hàm là một ArrayList gồm các âm vị được phân tách ra từ âm tiết theo cả 2 bảng mã phiên âm là IPA và XSAMPA

* ArrayList<String> convertSyllableToPhonemesWithTone(String syllable)

Là hàm chuyển đổi âm tiết thành các âm vị kèm theo thanh điệu.

+ Đầu vào của hàm là biến String biểu diễn một âm tiết

+ Đầu ra của hàm là một ArrayList gồm các âm vị được phân tách ra từ âm tiết theo cả 2 bảng mã phiên âm là IPA và XSAMPA có kèm theo thông tin thanh điệu