

### **\* Mô tả bài toán:**

#### **- Input:**

- + Một đoạn văn bản không giới hạn độ dài.
- + Đoạn văn bản ở bất kỳ ngôn ngữ nào.
- + Văn bản đầu vào có kí tự phù hợp với văn bản đó (VD: Tiếng Việt, Tiếng Anh --> Alphabet (bảng chữ cái Latin, Trung quốc --> Bảng chữ cái Hán tự (phồn thể hoặc giản thể))

#### **- Output:**

Tên gọi của loại ngôn ngữ đó bằng tiếng anh:

# Tiếng Việt -> Vietnamese

# Tiếng Nhật -> Japanese

# Tiếng Anh -> English

,...

### **\* Ý nghĩa:**

- Hiện nay, với sự bùng nổ và phát triển của khoa học và công nghệ, chúng ta dễ dàng tiếp cận hơn với những nguồn tài liệu, sách báo nước ngoài, và muốn dịch được văn bản này (thông qua công cụ dịch hoặc thuê người dịch...) thì trước tiên chúng ta cần biết được đây là ngôn ngữ gì. Đó là lí do mà ứng dụng này ra đời.

- Ứng dụng này có thể dùng để làm tiền xử lý cho các ứng dụng khác lớn hơn: Dịch thuật, phân loại văn bản theo ngôn ngữ,...

### **\* Mô tả API:**

- Nhận diện loại ngôn ngữ một cách tự động.
- Kết quả đầu ra theo tiêu chuẩn ISO 639.
- Thông qua thang đo thống kê N-grams.
- Danh sách ngôn ngữ có thể nhận diện chính xác là 180 ngôn ngữ khác nhau (phiên bản cũ là 60 ngôn ngữ).
- API cung cấp 20,000 request miễn phí mỗi tháng.
- Đơn giản và dễ sử dụng.
- API chỉ hoạt động khi câu từ rõ ràng, ít kí tự đặc biệt và không viết tắt.

### **\* Cách sử dụng:**

+ B1: Ta lưu URL của server API: "https://language-identification.p.rapidapi.com/lang-2.0"

+ B2: Tạo truy vấn chuỗi string.

Cấu trúc: {"type", "string", "threshold":"number", "verbose":"y/n"}

type: Kiểu lưu chuỗi (txt(UTF-8 encoding), doc, url).

string: Chuỗi kí tự bất kì.

number: Giá trị threshold cho biết ngưỡng chấp nhận kết quả đầu ra.

y/n: Nếu là 'y' sẽ cho biết thêm thông tin về ngôn ngữ đó, 'n' sẽ là thông tin tóm tắt của ngôn ngữ bao gồm:

iso639-2

iso639-3

language

name

relevance

,...

+ B3: Tạo header cho request (RapidAPI-key sẽ được cung cấp sau khi đăng ký sử dụng API)

+ B4: Tạo request gửi cho server và lưu lại reponse để lấy thông tin đầu ra mong muốn.

**\* Code ứng dụng demo: (file .ipynb đính kèm)**

```
pip install gradio
```

```
def Identify_laguage_string_txt(inputString):
```

```
    import requests
```

```
    url = "https://language-identification.p.rapidapi.com/lang-2.0"
```

```
    querystring = {"txt":inputString,"threshold":"100", "verbose":"y"}
```

```
    headers = {
```

```
        "Accept": "application/json",
```

```
"X-RapidAPI-Key": "779e73b7a8msh2bf647f53c47741p100badjsnc1416ee10f4e",  
"X-RapidAPI-Host": "language-identification.p.rapidapi.com"  
}
```

```
response = requests.request("GET", url, headers=headers, params=querystring)  
return response.json()['language_list'][0]['name']
```

```
import gradio as gr  
  
demo = gr.Interface(fn=Identify_language_string_txt, inputs="text", outputs="text")  
demo.launch()
```

**Demo:** <https://24469.gradio.app/>