

Run GPT translation on terms

Read in a dataframe with terms and translate them into multiple languages

Install OpenAI Library

In [1]:

```
import sys # installing this version of openai should automatically install the CLI
!{sys.executable} -m pip install openai
```

Collecting openai

Using cached openai-0.28.0-py3-none-any.whl (76 kB)

Requirement already satisfied: typing-extensions in /artifacts/virtualenv/python3/lib/python3.7/site-packages (from openai) (4.4.0)

Requirement already satisfied: tqdm in /artifacts/virtualenv/python3/lib/python3.7/site-packages (from openai) (4.50.2)

Requirement already satisfied: requests>=2.20 in /artifacts/virtualenv/python3/lib/python3.7/site-packages (from openai) (2.24.0)

Collecting aiohttp

Using cached aiohttp-3.8.5-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (976 kB)

Requirement already satisfied: certifi>=2017.4.17 in /artifacts/virtualenv/python3/lib/python3.7/site-packages (from requests>=2.20->openai) (2020.6.20)

Requirement already satisfied: idna<3,>=2.5 in /artifacts/virtualenv/python3/lib/python3.7/site-packages (from requests>=2.20->openai) (2.10)

Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /artifacts/virtualenv/python3/lib/python3.7/site-packages (from requests>=2.20->openai) (1.25.10)

Requirement already satisfied: chardet<4,>=3.0.2 in /artifacts/virtualenv/python3/lib/python3.7/site-packages (from requests>=2.20->openai) (3.0.4)

Requirement already satisfied: charset-normalizer<4.0,>=2.0 in /artifacts/virtualenv/python3/lib/python3.7/site-packages (from aiohttp->openai) (2.0.12)

Collecting multidict<7.0,>=4.5

Using cached multidict-6.0.4-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (94 kB)

Collecting yarl<2.0,>=1.0

Using cached yarl-1.9.2-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (236 kB)

Collecting aiosignal>=1.1.2

Using cached aiosignal-1.3.1-py3-none-any.whl (7.6 kB)

Collecting frozenlist>=1.1.1

Using cached frozenlist-1.3.3-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl (148 kB)

Requirement already satisfied: attrs>=17.3.0 in /artifacts/virtualenv/python3/lib/python3.7/site-packages (from aiohttp->openai) (20.2.0)

Collecting async-timeout<5.0,>=4.0.0a3

Using cached async_timeout-4.0.3-py3-none-any.whl (5.7 kB)

Collecting asyncctest==0.13.0

```
Using cached asyncctest-0.13.0-py3-none-any.whl (26 kB)
Installing collected packages: multidict, frozenlist, yarl, asyncctest, async-timeout, aiosignal, aiohttp, openai
Successfully installed aiohttp-3.8.5 aiosignal-1.3.1 async-timeout-4.0.3 asyncctest-0.13.0 frozenlist-1.3.3 multidict-6.0.4 openai-0.28.0 yarl-1.9.2
WARNING: You are using pip version 21.3.1; however, version 23.2.1 is available.
You should consider upgrading via the '/artifacts/virtualenv/python3/bin/python -m pip install --upgrade pip' command.
```

Load dataframe with cleaned/escaped text extracted from HTML for each URL in sample

```
In [2]: import pandas
import openai
openai.api_key = "YOUR_API_KEY"
df = pd.read_csv('mydata/terms_sample.csv')
```

```
In [3]: df.shape
```

```
Out[3]: (12, 3)
```

```
In [4]: languages = ['Spanish (Spain)', 'Spanish (LatAm)', 'German', 'Italian', 'Japanese', 'Chinese (S)', 'French', 'E
```

Get list of page category labels

Run GPT Classification

```
In [5]: import openai
openai.api_base = "https://openai-proxy.shopify.ai/v1"
openai.api_key = "shopify-WyJ0ZWftIiwiaHR0cHM6Ly92YXVsdC5zaG9waWZ5LmlvL3RlYW1zLzEzNjEiLCJodHRwczovL3ZhdWx0LnNo"
```

```
In [6]: # will need to slow down rate at which we make requests—here we'll try an approach recommended by openai,
# wherein we can use the tenacity library to add a random wait time between requests
# source: https://github.com/openai/openai-cookbook/blob/main/examples/How_to_handle_rate_limits.ipynb
# from tenacity import (
#     retry,
#     stop_after_attempt,
```

```
# wait_random_exponential,  
# ) # for exponential backoff
```

```
In [9]: #@retry(wait=wait_random_exponential(min=1, max=60), stop=stop_after_attempt(2))  
  
def translate_terms(row, target_language):  
    """Translate the terms"""  
  
    user_message = (  
        f"Translate the following english term with the into {target_language}: '" + row['Term'] +  
        "'.\n Here is the definition of the term for more context: \n" + row['Definition'] +  
        "'.\n Here is the part of speech of the term for more context: \n" + row['Part of speech'] +  
        "'.\n Please only translate the term and not the definition. Do not provide any additional text besides  
    )  
  
    response = openai.ChatCompletion.create(  
        model="gpt-3.5-turbo",  
        messages=[  
            {"role": "system", "content": "You are a helpful translation assistant that translate terms into ot"},  
            {"role": "user", "content": user_message},  
        ],  
        max_tokens=50,  
        n=1,  
        temperature=0.1  
    )  
  
    return response['choices'][0]['message']['content']
```

Testing OpenAI connection with default code

```
In [10]: # Apply the function to the DataFrame  
for language in languages:  
    df[language+'_terms'] = df.apply(translate_terms, axis=1, args=(language,))
```

```
In [11]: df
```

Out[11]:

| | Term | Definition | Part of speech | Spanish (Spain)_terms | Spanish (LatAm)_terms | German_terms | Italian_terms | Japanese_terms | Chinese (S)_terms |
|---|-------------|--|----------------|-----------------------|-----------------------|-------------------|-----------------|----------------|-------------------|
| 0 | reseal | seal (something) again. | verb | volver a sellar. | volver a sellar | wieder versiegeln | richiudere | 再封する (さいふうする) | 重新密封 |
| 1 | circuitry | a circuit or system of circuits performing a particular function in an electronic device. | noun | circuitería | circuitería | Schaltkreis. | circuiteria | 回路 | 电路 |
| 2 | insecticide | a substance used for killing insects. | noun | insecticida | insecticida | Insektizid | insetticida | 殺虫剤 | 杀虫剂 |
| 3 | preferably | ideally; if possible. | adverb | preferiblemente | preferiblemente | vorzugsweise | preferibilmente | できれば | 最好 |
| 4 | bored | feeling weary because one is unoccupied or lacks interest in one's current activity. | adjective | aburrido | aburrido | gelangweilt | annoiato | 退屈 | 无聊 |
| 5 | conspiracy | a secret plan by a group to do something unlawful or harmful. | noun | conspiración | conspiración | Verschwörung. | conspirazione | 陰謀 | 阴谋 |
| 6 | addict | a person who is addicted to a particular substance, typically an illegal drug. | noun | adicto | adicto | Süchtiger. | dipendente. | 中毒者 | 癮君子 |
| 7 | heavy metal | a type of highly amplified harsh-sounding rock music with a strong beat, characteristically using violent or | noun | heavy metal | heavy metal | Schwermetall | heavy metal | ヘビーメタル | 重金属 |

| | Term | Definition | Part of speech | Spanish (Spain)_terms | Spanish (LatAm)_terms | German_terms | Italian_terms | Japanese_terms | Chinese (S)_terms |
|----|--------------|--|----------------|-----------------------|-----------------------|----------------|---------------|----------------|-------------------|
| | | fantastic imagery. | | | | | | | |
| 8 | make up | Compose or constitute a whole. | verb | maquillaje | maquillaje | zusammensetzen | truccarsi | メイクアップ | 化妆 |
| 9 | get together | Gather or assemble people or things in one place. | verb | reunirse | reunirse | sich treffen. | riunirsi | 集まる | 聚在一起 |
| 10 | season | Changes in the weather marked by shifting temperatures. These include summer, winter, spring and autumn. | noun | estación | estación. | Jahreszeit. | stagione | 季節 | 季节 |
| 11 | season | To enhance the flavor of a dish by using spices like basil, cumin, turmeric, paprika etc. | verb | condimentar | sazonar | würzen | condire | シーズン | 调味 |

In [12]:

```
df.to_csv('mydata/output_terms.csv')
```