

3rd task

TASK:
Create chatbot in Arabic

Name :atheer al-mutairi

Steps:

-first I need to import the libraries by using the code:

```
import pickle
```

```
from os import listdir
```

```
from tqdm import tqdm
```

```
import numpy as np
```

```
np.random.seed(0)
```

```
users_files=listdir("users")
```

```
chars_c=set('abcdefghijklmnopqrstuvwxyzابتثجحذرزسشصضطظعغفقكلمنهوى')
```

```
data=[]
```

```
for file in users_files:
```

```
    try:
```

```
        with open('users/'+file,'rb') as fp:
```

```
            data.extend(pickle.load(fp))
```

```
    except EOFError:
```

```
        pass
```

it has both Arabic and English alphabet

- Second clean data :

```
qq=dict()
```

```
questions=[]
```

```
answers=[]
```

```
from collections import Counter
```

```
import re
```

```
chars=Counter()
```

```
for (q,a) in tqdm(data):
```

```
    if len(chars_c-set(q))!=len(chars_c) and 1<len(a)<280 and 4<len(q)<280 and  
len(re.findall('https?:/(?:[-\w.]|(?:%[\da-fA-F]{2}))+', q+' '+a))==0:
```

```
if q not in qq:
```

```
    qq[q]=None
```

```
    q=q.lower().replace("<br>","\n").replace("<br/>","\n").replace("</br>","\n")
```

```
    a=a.lower().replace("<br>","\n").replace("<br/>","\n").replace("</br>","\n")
```

```
    questions.append(q)
```

```
    chars.update(q)
```

```
    answers.append(a)
```

```
    chars.update(a)
```

- Helping ML model to understand how to begin and end:

```
is_mask=False
```

```
n_chars=100
```

```
n_chars=min(n_chars,len(chars))
```

```
all_chars=[i[0] for i in chars.most_common(n_chars)]
```

```
indexes=np.arange(n_chars)+2+is_mask
```

```
indexes=indexes.tolist()
```

```
n_chars_p=n_chars+2+is_mask
```

```
all_chars_map=dict(list(zip(all_chars,indexes)) + list(zip(indexes,all_chars)))
```

```
with open("all_chars_map.pkl",'wb') as fp:
```

```
    pickle.dump(all_chars_map,fp)
```

- Finally encode the charterers into integers :

```
questions=[[all_chars_map[char] for char in s if char in all_chars_map] for s in questions]
```

```
answers=[[all_chars_map[char] for char in s if char in all_chars_map] for s in answers]
```

```
indexss=[i for i,(q,a) in enumerate(zip(questions,answers)) if (1<len(q)) and (1<len(a))]
```

```
questions=[questions[i] for i in indexss]
```

```
answers=[answers[i] for i in indexss]
```

- After preprocessing I have to do the molding and and training so I get the chat reso

Resource:

- <https://hashim.id/creating-arabic-chatbot-using-keras-and-ask-fm/>
- https://pytorch.org/tutorials/beginner/chatbot_tutorial.html