

**Πανεπιστήμιο Πειραιώς  
Τμήμα Πληροφορικής  
Έτος: 2021 - 2022**



**Μάθημα:**  
**«ΕΠΕΞΕΡΓΑΣΙΑ ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ»**  
**Εργασία: 1η Προαιρετική εργασία**  
**Εξάμηνο: 5ο**

**Στοιχεία:**

**Όνομα:** Αιμιλιανός Κουρπάς-Δανάς

**ΑΜ :** Π 20100

**Ημερομηνία παραδοσης :** 13.07.2022

## **ΕΚΦΩΝΗΣΗ ΕΡΓΑΣΙΑΣ**

**Εργασία 5 :** Ανάπτυξη & Τεκμηρίωση Chatbot (μπορούν να χρησιμοποιηθούν υποδείγματα από τα folders στο Teams– πρέπει να γίνει αναφορά σε αυτά μέσα στην εργασία)

### **Σημείωση**

Τα προγράμματα γράφτηκαν στο προγραμματιστικό περιβάλλον VScode.

Στον κώδικα υπάρχουν περιεκτικά σχόλια ουσίας για την καλύτερη επεξήγησή του.(προστέθηκαν μετά τα screenshots)

## ΤΕΚΜΗΡΙΩΣΗ ΚΩΔΙΚΑ

Γλώσσα Υλοποίησης: Python

### main.py

```
import numpy as np
import random
from data import data
from utils import removeStopwords
from mlmodel import MLmodelClass
from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer
from keras.utils import to_categorical

# Collect all data and prepare to train
categories = []
xTrain = []
yTrain = []

for category in data:
    categories.append(category["tag"])
    for sentence in category["patterns"]:
        xTrain.append(removeStopwords(sentence))
        yTrain.append(category["tag"])

yTrain = to_categorical(np.array([categories.index(y) for y in yTrain]).reshape(-1, 1)))

# Transform sentences
countVectorizer = CountVectorizer()
xTrainCounts = countVectorizer.fit_transform(xTrain)

tfidfTransformer = TfidfTransformer()
xTrainTfidf = tfidfTransformer.fit_transform(xTrainCounts).toarray()

# Create and train the model
model = MLmodelClass(len(xTrainTfidf[0]), len(yTrain[0]))
model.train(xTrainTfidf, yTrain)

# Interaction with the user
print("Hello, how can I help you?")
while True:
    question = input()
    result = removeStopwords(question)
    xNewCounts = countVectorizer.transform([result])
    xNewTfidf = tfidfTransformer.transform(xNewCounts).toarray()

    pred = model.predict(xNewTfidf)
    responses = [resp for resp in data if resp["tag"] == categories[pred]]
    print(random.choice(responses[0]["responses"]))
    if categories[pred] == "goodbye":
        break
```

Αυτός ο κώδικας ξεκινάει με την εισαγωγή των απαραίτητων βιβλιοθηκών. Ο κώδικας προετοιμάζει τα δεδομένα εκπαίδευσης εξάγοντας κατηγορίες, προτάσεις από τα δεδομένα(data.py). Οι προτάσεις υπόκεινται σε προεπεξεργασία με την αφαίρεση των συνηθισμένων λέξεων. Οι ετικέτες μετατρέπονται σε μορφή μίας κωδικοποίησης one-hot

για την εκπαίδευση. Στη συνέχεια, τα κείμενα μετατρέπονται σε αριθμητικά χαρακτηριστικά χρησιμοποιώντας τον CountVectorizer και τον TfidfTransformer από το scikit-learn.

Ο κώδικας εισέρχεται σε έναν βρόχο, όπου ζητά από τον χρήστη να εισάγει δεδομένα, επεξεργάζεται την είσοδο αφαιρώντας τις συνηθισμένες λέξεις, μετατρέπει την επεξεργασμένη είσοδο σε αριθμητικά χαρακτηριστικά και χρησιμοποιεί το εκπαιδευμένο μοντέλο για να προβλέψει την κατηγορία. Με βάση την προβλεπόμενη κατηγορία, επιλέγεται τυχαία μια απόκριση και εμφανίζεται στο χρήστη. Ο βρόχος συνεχίζεται μέχρι η προβλεπόμενη κατηγορία να είναι "αντίο", στο οποίο σημείο το πρόγραμμα τερματίζει.

### mlmodel.py

```
import numpy as np
from keras.models import Sequential
from keras.layers import Dense

class MLmodelClass:
    def __init__(self, inputLen, outputLen):
        self.model = Sequential()
        self.model.add(Dense(50, activation='relu', input_dim=inputLen))
        self.model.add(Dense(20, activation='relu'))
        self.model.add(Dense(30, activation='relu'))
        self.model.add(Dense(outputLen, activation='softmax'))

        self.model.compile(optimizer='adam',
                           loss='categorical_crossentropy',
                           metrics=['accuracy'])

    def train(self, x, y, epochs=500, verbose=1):
        self.model.fit(x, y, epochs=epochs, verbose=verbose)

    def predict(self, x):
        pred = self.model.predict(x)
        return np.argmax(pred)
```

Αυτός ο κώδικας υλοποιεί ένα μοντέλο μηχανικής μάθησης με χρήση του Keras. Το μοντέλο περιλαμβάνει πλήρως συνδεδεμένα στρώματα και χρησιμοποιεί την 'relu' συνάρτηση ενεργοποίησης. Εκπαιδεύεται με τα δεδομένα εισόδου και προβλέπει την κατηγορία εξόδου.

## data.py

```
data = [
    {
        "tag": "greetings",
        "patterns": ["hello", "hi", "whats up", "how are you"],
        "responses": ["Hello! How can I help you?"]
    },
    {
        "tag": "goodbye",
        "patterns": ["thanks", "thank you", "bye", "goodbye", "good night"],
        "responses": ["Goodbye!", "It was nice talking to you! Bye", "Have a nice day"]
    },
    {
        "tag": "who",
        "patterns": ["who are you", "what are you", "why do you exist"],
        "responses": ["I am a virtual assistant and I am here to give you some information. How can I help you?"]
    },
    {
        "tag": "opening_hours",
        "patterns": ["what are the library opening hours?", "when does the library open?", "what time does the library close?", "when can I visit the library?"],
        "responses": ["The library is open from 9:00am to 6:00pm on weekdays and from 10:00am to 4:00pm on weekends."]
    },
    {
        "tag": "services",
        "patterns": ["what services does the library offer?", "can I borrow books from the library?", "are there study rooms available?", "do you provide printing and photocopying services?"],
        "responses": ["The library offers book borrowing services, study rooms for students, as well as printing and photocopying facilities."]
    },
    {
        "tag": "membership",
        "patterns": ["how can I become a library member?", "what are the requirements to join the library?", "is library membership free?", "do I need to provide any documents to register?"],
        "responses": ["To become a library member, you can visit the library's membership desk and fill out a registration form. Membership is free, and you may be required to provide identification and proof of address."]
    },
    {
        "tag": "events",
        "patterns": ["are there any upcoming library events?", "do you host author talks or book clubs?", "are there any workshops or seminars happening?", "what events are scheduled for this month?"],
        "responses": ["The library regularly hosts various events, including author talks, book clubs, workshops, and seminars. You can check the library's website or social media pages for information about upcoming events."]
    },
    {
        "tag": "location",
        "patterns": ["where is the library located?", "what is the address of the library?", "how can I find the library?", "which street is the library on?"],
        "responses": ["The library is located at 123 Main Street, in the heart of the city. It is easily accessible and situated near the central bus station."]
    }
]
```

Τα παρεχόμενα δεδομένα αναπαριστούν ένα σύνολο συνομιλιακών προτύπων και των αντίστοιχων απαντήσεων για ένα σύστημα chatbot που έχει σχεδιαστεί ειδικά για να προσομοιώνει αλληλεπιδράσεις που σχετίζονται με μια **βιβλιοθήκη**. Οι διάφορες κατηγορίες μέσα στα δεδομένα καλύπτουν διάφορες πτυχές μιας βιβλιοθήκης, όπως πληροφορίες για τις ώρες λειτουργίας της βιβλιοθήκης, οι υπηρεσίες που προσφέρονται, τα στοιχεία μέλητης, τα γεγονότα που διοργανώνονται και η τοποθεσία της βιβλιοθήκης. Αυτά τα δεδομένα θα χρησιμοποιηθούν για την εκπαίδευση του chatbot για να παρέχει σχετικές και ενημερωτικές απαντήσεις στο πλαίσιο μιας βιβλιοθήκης.

## utils.py

```
stopwords = {'a', 'an', 'the', 'in', 'on', 'is', 'and', 'I', 'you', 'he', 'she'} # stopwords to be removed

def removeStopwords(sentence):
    querywords = sentence.split()
    resultwords = [word for word in querywords if word.lower() not in stopwords]
    return ' '.join(resultwords)

sentence = "This is a sample sentence with stopwords."
processed_sentence = removeStopwords(sentence)
print(processed_sentence)
```

Αυτός ο κώδικας εκτελεί μια λειτουργία για την αφαίρεση συνηθισμένων λέξεων από μια πρόταση. Καταρχάς, ορίζεται ένα σύνολο stopwords που περιέχει συνηθισμένες λέξεις που πιθανώς να μην παρέχουν σημασία στην επεξεργαζόμενη πρόταση. Η συνάρτηση removeStopwords δέχεται μια πρόταση ως είσοδο και διαχειρίζεται την αφαίρεση των stopwords από την πρόταση. Αρχικά, η πρόταση χωρίζεται σε λέξεις. Στη συνέχεια, δημιουργείται μια νέα λίστα (resultwords) που περιέχει μόνο τις λέξεις που δεν ανήκουν στα stopwords. Τέλος, οι λέξεις συνενώνονται πάλι για να σχηματίσουν την επεξεργασμένη πρόταση. Στο παράδειγμα χρήσης, μια πρόταση δίνεται ως είσοδος στη συνάρτηση removeStopwords, και η επεξεργασμένη πρόταση εκτυπώνεται στην οθόνη.

## ΠΑΡΑΔΕΙΓΜΑΤΑ ΟΡΘΗΣ ΕΚΤΕΛΕΣΗΣ

Η εκπαίδευση

```
Epoch 471/500
2/2 [=====] - 0s 5ms/step - loss: 0.0037 - accuracy: 1.0000
Epoch 472/500
2/2 [=====] - 0s 5ms/step - loss: 0.0037 - accuracy: 1.0000
Epoch 473/500
2/2 [=====] - 0s 5ms/step - loss: 0.0037 - accuracy: 1.0000
Epoch 474/500
2/2 [=====] - 0s 5ms/step - loss: 0.0036 - accuracy: 1.0000
Epoch 475/500
2/2 [=====] - 0s 5ms/step - loss: 0.0036 - accuracy: 1.0000
Epoch 476/500
2/2 [=====] - 0s 9ms/step - loss: 0.0036 - accuracy: 1.0000
Epoch 477/500
2/2 [=====] - 0s 7ms/step - loss: 0.0036 - accuracy: 1.0000
Epoch 478/500
2/2 [=====] - 0s 14ms/step - loss: 0.0035 - accuracy: 1.0000
Epoch 479/500
2/2 [=====] - 0s 6ms/step - loss: 0.0035 - accuracy: 1.0000
Epoch 480/500
2/2 [=====] - 0s 6ms/step - loss: 0.0035 - accuracy: 1.0000
Epoch 481/500
2/2 [=====] - 0s 7ms/step - loss: 0.0035 - accuracy: 1.0000
Epoch 482/500
```

μία ενδεικτική συνομιλία :

```
1/1 [=====] - 0s 221ms/step
Hello!!.. How can I help you?
where is the library ?
1/1 [=====] - 0s 37ms/step
The library is located at 123 Main Street, in the heart of the city. It is easily accessible and situated near the
is it open ?
1/1 [=====] - 0s 37ms/step
The library is open from 9:00am to 6:00pm on weekdays and from 10:00am to 4:00pm on weekends.
is it free to get a book ?
1/1 [=====] - 0s 38ms/step
To become a library member, you can visit the library's membership desk and fill out a registration form. Members
thanks
1/1 [=====] - 0s 37ms/step
Have a nice day
```

## Βιβλιογραφία

- sklearn  
<https://scikit-learn.org/stable/install.html>
- keras  
<https://www.activestate.com/resources/quick-reads/how-to-install-keras-and-tensorflow/>
- numpy  
<https://www.edureka.co/blog/install-numpy/>

Τελος. Σας ευχαριστω πολυ.