Университет Информационных Технологий, Механики и Оптики

Факультет ПИиКТ

Лабораторная работа №1

По дисциплине

«Прикладная математика»

Студент: Базарова А. А.

Группа: P3300

Санкт-Петербург

2019 г.

1. Постановка задачи.

Реализовать процедуру вычисления энтропии для текстового файла.

В процедуре необходимо подсчитывать частоты появления символов (прописные и заглавные буквы не отличаются, знаки препинания рассматриваются как один символ, пробел является самостоятельным символом), которые можно использовать как оценки вероятностей появления символов.

Вычислить величину энтропии. Точность вычисления -- 4 знака после запятой. Обязательно предусмотреть возможность ввода имени файла, для которого будет вычисляться энтропия.

Проверить запрограммированную процедуру на нескольких файлах

Вычислить значение энтропии для тех же файлов, но с использованием частот вхождений пар символов

2. Решение поставленной задачи.

import math,sys

p=dict()#1 letter probs

h=dict()#1 letter entrop

coupleP=dict()#2 letters probs

count=0

print("Write name of the file, please")

name=input()

print()

try:

file=open(name,'r')

except FileNotFoundError:

print("There is no such a file :0")

sys.exit(0)

for c in file.read():#here file.read is a string that contains all the symbols of the file

if c.isalpha() or c.isdigit():

sym=c.upper()

elif c==" ":

sym=" "

else:

sym="."

p.setdefault(sym,1)#this function looks for sym in the dict, and if there is no such a thing, it adds this to dict with the 1

p[sym]+=1

if count>0:

coupleP.setdefault(prev+sym,1)

coupleP[prev+sym]+=1

prev=sym

count+=1

#now in the probs dicts there are quantities, not probs

#lets fix it

H=0 #file entropy

for x in p:

p[x]=p[x]/count

h[x]=math.log2(1/p[x])

H+=h[x]\*p[x]

print("Character: ", f'{x}',", Probability: ",f'{p[x]:.4f}',", Entropy: ",f'{h[x]:.4f}')

print()

print("H: ",f'{H}')

print()

#now we need to count H\*

coupleH=0 #couples entropy

for x in coupleP:

coupleP[x]=coupleP[x]/(count-1)#couple count=count-1

coupleH-=coupleP[x]\*p[x[1]]\*math.log2(coupleP[x])

print("H\*: ", f'{coupleH:.4f}')

3. Описание входных данных программы.

На вход программе подается строка, название файла

Программа выводит данные для заполнения таблиц 1.1 и 1.2

4. Тестирование программы.

Для тестирования было составлено три файла, содержащие различные отрывки из произведений Уильяма Шекспира.

Соответствующие объемы 40 кБ, 20 кБ и 60 кБ.

5. Анализ результатов работы программы.

6.Заключение.