МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Пензенский государственный технологический университет»

(ПензГТУ)

Факультет автоматизированных информационных технологий

Кафедра «Информационные технологии и системы»

Дисциплина «Алгоритмизация и программирования»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ № 1

на тему «Вычисление энтропии по Шеннону»

Вариант №3

Выполнил: студент гр. 19ИС1бп Ведюшкин А.Е.

Проверил: ст. преподаватель каф. ИТС Володин К.И.

Работа защищена с оценкой: \_\_\_\_\_\_\_\_\_\_\_\_\_

Пенза 2021

Задание

1. Реализуйте алгоритм расчета энтропии по Шеннону указанных файлов с заданным расширением;
2. Предложите оптимизированную версию при условии больших входных данных;
3. Реализуйте задачу параллельной обработки больших входных данных с помощью модуля *multiprocessing* и *functools*;
4. Проведите исследования изменения времени вычисления от:

а. размера буфера чтения из файла;

б. количества потоков обработки.

1. Алгоритмы реализовать в Python 3.6+.

Для реализации на *Python* достаточно стандартной библиотеки *Python*.

Код реализации

import time

import math

from multiprocessing import Pool

from collections import Counter

def uniq\_count(message):

return list(Counter(message).values())

def entrop(message):

return -sum([e/len(message)\*math.log2(e/len(message)) for e in uniq\_count(message)])

start = time.time()

if \_\_name\_\_ == '\_\_main\_\_':

file\_name = 'text.txt'

a = 10

buffer1 = []

n = 1

G=a

g = open(file\_name, 'rb')

kolvo = g.read()

while len(kolvo)>G:

n+=1

G+=a

g.close()

with open(file\_name, 'rb') as f:

for i in range(n):

print("nomer buffer - ", i+1)

buffer = f.read(a)

buffer1.append(buffer)

print(buffer1, '\n')

del buffer

with Pool(2) as p:

print("entropy = ", (sum(p.map(entrop,buffer1)))/len(buffer1))

file = open('enttime.txt', 'a+')

file.write(str(time.time() -start) + '\n')

print("%s seconds" % (time.time() -start))

Исследование

Рисунок 1 – Зависимость времени обработки от кол-ва потоков

Рисунок 2 – Зависимость времени обработки от длины буффера