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: «Энтропия»

Выполнил: студент гр. 19ИС1бп Жалдыбин А.Е

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

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

Пенза 2021

Содержание

[Задача 1. Энтропия 3](#_Toc62666294)

[Код реализации 4](#_Toc62666295)

[Исследование 5](#_Toc62666296)

[Вывод 6](#_Toc62666297)

# Задача 1. Энтропия

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:
   1. размера буфера чтения из файла;
   2. количества потоков обработки.
5. 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

st = time.time()

def uniq\_count(message):

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

def ent(message):

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

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

file\_name = 'newfile.txt'

bs = 10000

bs2 = 1

full\_ent = 0

buffer\_list = []

pool = Pool(processes=4)

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

buffer = f.read(bs)

buffer\_list.append(buffer)

del buffer

print(pool.map(ent,buffer\_list))

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

# Исследование

|  |  |
| --- | --- |
| Исходный файл – 1 Гб | Время обработки |
| Размер буффера – 1 Гб | 0,6188924312591553 |
| Кол-во потоков - 2 |

|  |  |
| --- | --- |
| Исходный файл – 1 Гб | Время обработки |
| Размер буффера – 1 Гб | 0,2719101905822754 |
| Кол-во потоков - 4 |

# Вывод

В данной работе было разработана программа расчета энтропии по Шеннону.

Также была реализована параллельная обработка входных данных, были проведены исследования измерения времени от размеров буфера и числа обработки данных.