САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ ФАКУЛЬТЕТ ИНФОКОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

Отчет по лабораторной работе №1 по курсу «Алгоритмы и структуры данных» Тема: Сортировка слиянием, Метод декомпозиции Вариант 22

Выполнил: Чан Тхи Лиен К3140

Проверил: Афанасьев А.В.

Санкт-Петербург 2024 г.

Содержание отчета

| Содержание отчета | 2 |
|---|----|
| Задачи по варианту | |
| Задание 1. Сортировка слиянием | 3 |
| Задание 5. Представитель большинства | 6 |
| Задание 9. Метод Штрассена для умножения матриц | 9 |
| Дополнительные задачи | |
| Задание 3. Число инверсий | 13 |
| Задание 4. Бинарный поиск | 15 |
| Задание 8. Умножение многочленов | 18 |
| Вывод | 20 |

Задачи по варианту

Задание 1. Сортировка слиянием

- 1. Используя *псевдокод* процедур Merge и Merge-sort из презентации к Лекции 2 (страницы 6-7), напишите программу сортировки слиянием на Python и проверьте сортировку, создав несколько рандомных массивов, подходящих под параметры:
 - Формат входного файла (input.txt). В первой строке входного файла содержится число n ($1 \le n \le 2 \cdot 10^4$) число элементов в массиве. Во второй строке находятся n различных целых чисел, по модулю не превосходящих 10^9 .
 - Формат выходного файла (output.txt). Одна строка выходного файла с отсортированным массивом. Между любыми двумя числами должен стоять ровно один пробел.
 - Ограничение по времени. 2сек.
 - Ограничение по памяти. 256 мб.
- 2. Для проверки можно выбрать наихудший случай, когда сортируется массив размера $1000,\ 10^4,\ 10^5$ чисел порядка $10^9,\$ отсортированных в обратном порядке; наилучший, когда массив уже отсортирован, и средний. Сравните, например, с сортировкой вставкой на этих же данных.
- 3. Перепишите процедуру Merge так, чтобы в ней не использовались сигнальные значения. Сигналом к остановке должен служить тот факт, что все элементы массива L или R скопированы обратно в массив A, после чего в этот массив копируются элементы, оставшиеся в непустом массиве.
- *или* перепишите процедуру Merge (и, соответственно, Merge-sort) так, чтобы в ней не использовались значения границ и середины p, r и q.

```
# Merge Sort in python
from time import perf_counter
def mergeSort(array): 5 usages * lien
if len(array) == 1:
    return array
else:

r = len(array)//2
L = array[:r]
M = array[r:]

mergeSort(L)
mergeSort(M)
```

```
i = j = k = 0
while i < len(L) and j < len(M):
    if L[i] < M[j]:</pre>
        array[k] = L[i]
        i += 1
    else:
        array[k] = M[j]
        j += 1
    k += 1
while i < len(L):
    array[k] = L[i]
    i += 1
    k += 1
while j < len(M):
    array[k] = M[j]
    k += 1
return array
```

- 1. Импортировать библиотеку time.
- 2. Использую **def** для создания функции.
- проверить количество элементов массива и рекурсивно разделить массив на две части.
- сравнить элементы двух подмассивов, упорядочить и объединить два массива.
- 3. Использовать **with open** чтобы открыть файл и **readline** для чтения каждой строки в файле.

```
t1_start = perf_counter()

if __name__ == '__main__':

with open('input.txt', 'r') as infile:

n = int(infile.readline())

data = str(infile.readline())

data = data.split(' ')

arr = [int(i) for i in data]

mergeSort(arr)
result = ' '.join(str(i) for i in arr)

with open('output.txt', 'w') as outfile:

outfile.write(result)

t1_stop = perf_counter()

print('Время работы: %s секунд '% (t1_stop - t1_start))

t3
```

- 4. Meтод **split()** разделить строку на массив.
- 5. Метод **join()** собирает строку из элементов списка.
- 6. Использовать **with open** чтобы открыть файл и **write** для запись результаты в файле.

```
1 7
2 19 34 5 17 9 43 11
1 5 9 11 17 19 34 43
```

Время работы: 0.0015571000003546942 секунд

7. Записать файл для тест.

```
import unittest
import random
from Ja6_2.Task1.Task1 import mergeSort

class TestMergeSort(unittest.TestCase): *lien*
def test_sort_large_array(self): *lien*
large_array = [random.randint(1,10000) for _ in range(10000)]
sorted_array = sorted(large_array)
mergeSort(large_array)
self.assertEqual(large_array, sorted_array)

if __name__ == '__main__':
unittest.main()
```

Задание 5. Представитель большинства

Правило большинства - это когда выбирается элемент, имеющий больше половины голосов. Допустим, есть последовательность A элементов $a_1, a_2, ... a_n$, и нужно проверить, содержит ли она элемент, который появляется больше, чем n/2 раз. Наивный метод это сделать:

Очевидно, время выполнения этого алгоритма квадратично. Ваша цель - использовать метод "Разделяй и властвуй" для разработки алгоритма проверки, содержится ли во входной последовательности элемент, который встречается больше половины раз, за время $O(n \log n)$.

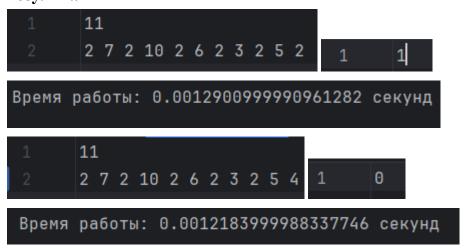
- Формат входного файла (input.txt). В первой строке входного файла содержится число n ($1 \le n \le 10^5$) число элементов в массиве. Во второй строке находятся n положительных целых чисел, по модулю не превосходящих 10^9 , $0 \le a_i \le 10^9$.
- **Формат выходного файла (output.txt).** Выведите 1, если во входной последовательности есть элемент, который встречается строго больше половины раз; в противном случае 0.
- Ограничение по времени. 2сек.
- Ограничение по памяти. 256 мб.

```
from time import perf_counter
def find_mode(arr, left, right): 3 usages ≛ lien
    if left > right:
    if left == right:
        return arr[left], 1
   mid = (left + right) // 2
   left_mode, left_count = find_mode(arr, left, mid)
    right_mode, right_count = find_mode(arr, mid + 1, right)
    total_left = count_mode(arr, left_mode, left, right)
    total_right = count_mode(arr, right_mode, left, right)
   if total_left > total_right:
        return left_mode, total_left
   else:
        return right_mode, total_right
def count_mode(arr, element, left, right): 2 usages ≗ lien
    count = 0
    for i in range(left, right + 1):
        if arr[i] == element:
            count += 1
    return count
```

- 1. Импортировать библиотеку **time.**
- 2. Создать 2 функции:
- функция "**find_mode**":
- Рекурсивно разделить массив на две половины, чтобы найти моду в каждой половине.
- использовать функцию "count_mode" подсчета количества появления моды.
- Сравнить количество вхождений моды в каждой половине и найдите моду массива.
- функция "count_mode": Подсчитать количество вхождений элемента

```
if n>len(array)//2:
        return str(1)
        return str(0)
t1_start = perf_counter()
if __name__ == "__main__":
    with open('input.txt', 'r') as infile:
       num = int(infile.readline())
        data = str(infile.readline())
    data = data.split(' ')
    array = [int(i) for i in data]
    result = find_mode(array, left: 0, len(array)-1)
    n = result[1]
    with open('output.txt', 'w') as outfile:
        outfile.write(print_result(array, n))
t1_stop = perf_counter()
print('Время работы: %s секунд '% (t1_stop - t1_start))
```

- 3. Сравнить количество вхождений моды с количеством элементов в массиве.
- если количество вхождений больше половины, печать 1.
- если нет, печать **0**.
- 4. Использовать **with open** чтобы открыть файл и **readline** для чтения каждой строки в файле.
- 5. Метод **split()** разделить строку на массив.
- 6. Метод join() собирает строку из элементов списка.
- 7. Использовать **with open** чтобы открыть файл и **write** для запись результаты в файле.



Задание 9. Метод Штрассена для умножения матриц

Умножение матриц. Простой метод. Если есть квадратные матрицы $X = (x_{ij})$ и $Y = (y_{ij})$, то их произведение $Z = X \cdot Y \Rightarrow z_{ij} = \sum_{k=1}^{n} x_{ik} \cdot y_{kj}$. Нужно вычислить n^2 элементов матрицы, каждый из которых представляет собой сумму n значений.

```
Matrix_Multiply(X, Y)::
    n = X.rows
    Z - квадратная матрица размера n
    for i = 1 to n:
        for j = 1 to n:
        z[i,j] = 0
        for k = 1 to n:
        z[i,j] = z[i,j] + x[i,k]*y[k,j]
    return Z
```

- **Цель**. Применить метод Штрассена для умножения матриц и сравнить его с простым методом. *Найти размер матриц п, при котором метод Штрассена работает существенно быстрее простого метода*.
- Формат входа. Стандартный ввод или input.txt. Первая строка размер квадратных матриц n для умножения. Следующие строки соответсвенно сами значения матриц A и B.
- Формат выхода. Стандартный вывод или output.txt. Матрица $C = A \cdot B$.

```
from time import perf_counter
from math import *

def add_matrices(A, B): 12 usages *lien

return [[A[i][j] + B[i][j] for j in range(len(A[0]))] for i in range(len(A))]

def subtract_matrices(A, B): 6 usages *lien

return [[A[i][j] - B[i][j] for j in range(len(A[0]))] for i in range(len(A))]
```

- 1. Импортировать библиотеку **time**, **math**.
- 2. Создать две функции для умножения и вычитания двух матриц.

```
def strassen(A, B): 8 usages ≛ lien
   n = len(A)
   if n == 1:
       return [[A[0][0] * B[0][0]]]
   mid = n // 2
   A11 = [row[:mid] for row in A[:mid]]
   A12 = [row[mid:] for row in A[:mid]]
   A21 = [row[:mid] for row in A[mid:]]
   A22 = [row[mid:] for row in A[mid:]]
   B11 = [row[:mid] for row in B[:mid]]
   B12 = [row[mid:] for row in B[:mid]]
   B21 = [row[:mid] for row in B[mid:]]
   B22 = [row[mid:] for row in B[mid:]]
   P1 = strassen(A11, subtract_matrices(B12, B22))
   P2 = strassen(add_matrices(A11, A12), B22)
   P3 = strassen(add_matrices(A21, A22), B11)
   P4 = strassen(A22, subtract_matrices(B21, B11))
   P5 = strassen(add_matrices(A11, A22), add_matrices(B11, B22))
   P6 = strassen(subtract_matrices(A12, A22), add_matrices(B21, B22))
   P7 = strassen(subtract_matrices(A11, A21), add_matrices(B11, B12))
    C11 = add_matrices(subtract_matrices(add_matrices(P5, P4), P2), P6)
    C12 = add_matrices(P1, P2)
    C21 = add_matrices(P3, P4)
    C22 = subtract_matrices(add_matrices(P5, P1), add_matrices(P3, P7))
    C = []
    for i in range(mid):
        C.append(C11[i] + C12[i])
    for i in range(mid):
        C.append(C21[i] + C22[i])
    return C
```

- 3. Создать функцию "strassen" для умножения двух матриц:
- Рекурсивно разделить и умножить матриц и вычислить Р1 Р7.
- использовать P1 P7 для создания результата матрицы C.

```
def next_size(n): 1 usage *lien
    num = floor(log2(n))
if n == pow(2, num):
    size = n
    return size
else:
    size = int(pow(2, num+1))
    return size

def pad_matrix(A, size): 2 usages *lien
    padded = [[0] * size for _ in range(size)]
for i in range(len(A)):
    for j in range(len(A[i])):
        padded[i][j] = A[i][j]
return padded

t1_start = perf_counter()
```

- 4. Функция "**next_size**" используется для проверки того, является ли степень матрицы степенью 2 или нет.
- Если верно, не изменить степень матриц
- Если нет, увеличить степень матрицы до ближайшей степени по основанию 2.
- 5. Функция "**pad_matrix**" для добавления элемента 0 в недостающую позицию матриц A и B.

```
t1_start = perf_counter()
if __name__ == "__main__":
       n = int(infile.readline())
        data = infile.readline().split()
        A_values = list(map(int, data[:n*n]))
        A = [A_{values}[i*n:(i+1)*n] \text{ for } i \text{ in } range(n)]
        B_values = list(map(int, data[n*n:]))
        B = [B_values[i*n:(i+1)*n] for i in range(n)]
    size = next_size(n)
    A_padded = pad_matrix(A, size)
    B_padded = pad_matrix(B, size)
    C_padded = strassen(A_padded, B_padded)
    result = [row[:n] for row in C_padded[:n]]
    with open('output.txt', 'w') as outfile:
        for row in result:
            outfile.write(' '.join(map(str, row))+ '\n')
t1_stop = perf_counter()
print('Время работы: %s секунд '% (t1_stop - t1_start))
```

- 6. Использовать **with open** чтобы открыть файл и **write** для запись результаты в файле.
- 7. после получения результата мы удаляем 0 элементов, потому что мы добавили их в матрицы A и B



Дополнительные задачи

Задание 3. Число инверсий

Инверсией в последовательности чисел A называется такая ситуация, когда i < j, а $A_i > A_j$. Количество инверсий в последовательности в некотором роде определяет, насколько близка данная последовательность к отсортированной. Например, в сортированном массиве число инверсий равно 0, а в массиве, сортированном наоборот - каждые два элемента будут составлять инверсию (всего n(n-1)/2).

Дан массив целых чисел. Ваша задача — подсчитать число инверсий в нем. Подсказка: чтобы сделать это быстрее, можно воспользоваться модификацией сортировки слиянием.

- Формат входного файла (input.txt). В первой строке входного файла содержится число n ($1 \le n \le 10^5$) число элементов в массиве. Во второй строке находятся n различных целых чисел, по модулю не превосходящих 10^9 .
- Формат выходного файла (output.txt). В выходной файл надо вывести число инверсий в массиве.
- Ограничение по времени. 2сек.
- Ограничение по памяти. 256 мб.

```
# Bubble sort in Python

#function

from time import perf_counter

def bubbleSort(array): 1 usage * lien

if len(array)<=1:
    return 0

num = 0

for i in range(len(array)):

for j in range(0, len(array) - i - 1):

if array[j] > array[j + 1]:
    temp = array[j]
    array[j] = array[j+1]
    array[j+1] = temp
    num = num+1

return str(num)
```

- 1. Импортировать библиотеку time.
- 2. Использовать функцию "bubbleSort":
- Сравнит два соседних элемента.
- Если передний элемент больше заднего, поменять местами два. элемента и увеличить **num** на 1.
- Если нет, оставить эти два элемента на месте.
- печать результат

```
#result
t1_start = perf_counter()
with open('input.txt','r') as infile:
    n = int(infile.readline())
    data = str(infile.readline())

data = data.split(' ')
arr = [int(i) for i in data]

with open('output.txt', 'w') as outfile:
    outfile.write(bubbleSort(arr))
t1_stop = perf_counter()
print('Время работы: %s секунд '% (t1_stop - t1_start))
```

3. Использовать **with open** чтобы открыть файл для чтения данных и записи результат.

```
1 10
2 1821473236 1 17
```

Время работы: 0.001406800001859665 секунд

Задание 4. Бинарный поиск

В этой задаче вы реализуете алгоритм бинарного поиска, который позволяет очень эффективно искать (даже в огромных) списках при условии, что список отсортирован. Цель - реализация алгоритма двоичного (бинарного) поиска.

- Формат входного файла (input.txt). В первой строке входного файла содержится число n ($1 \le n \le 10^5$) число элементов в массиве, и последовательность $a_0 < a_1 < ... < a_{n-1}$ из n различных положительных целых чисел в порядке возрастания, $1 \le a_i \le 10^9$ для всех $0 \le i < n$. Следующая строка содержит число k, $1 \le k \le 10^5$ и k положительных целых чисел $b_0, ... b_{k-1}, 1 \le b_i \le 10^9$ для всех $0 \le j < k$.
- Формат выходного файла (output.txt). Для всех i от 0 до k-1 вывести индекс $0 \le j \le n-1$, такой что $a_i = b_j$ или -1, если такого числа в массиве нет.
- Ограничение по времени. 2сек.

from time import perf_counter

def binary_search(arr, target): 1 usage * lien
left, right = 0, len(arr) - 1
while left <= right:
 mid = left + (right - left) // 2
 if arr[mid] == target:
 return mid
 elif arr[mid] < target:
 left = mid + 1
 else:
 right = mid - 1
return -1</pre>

- 1. Импортировать библиотеку time.
- 2. Создать функцию "binary_search" для поиска места элемента в массиве.
- разделить массив на две части.
- сравнить элемент, позицию которого нужно найти, со средним элементом массива.
- Если равно, печать эту позицию.
- Если больше, выполнить функцию со второй половиной массива.
- Если меньше, выполнить функцию с первой половиной массива.
- Если массив нет этого элемента, печать -1.

3. Создать функцию "**find_postions**" для поиска позиции каждого элемента

4. Использовать **with open** чтобы открыть файл для чтения данных и записи результат.

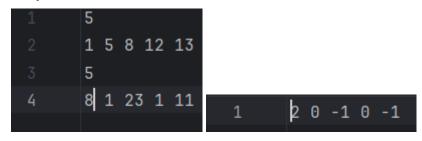
```
t1_start = perf_counter()
if __name__ == '__main__':

with open('input.txt', 'r') as infile:
    n = int(infile.readline())
    data1 = str(infile.readline()).split(' ')
    k = int(infile.readline())
    data2 = str(infile.readline()).split(' ')
    list_to_search = [int(i) for i in data1]
    list_to_find = [int(i) for i in data2]

array = find_positions(list_to_search, list_to_find)
    result = ' '.join(str(i) for i in array)

with open('output.txt', 'w') as outfile:
    outfile.write(result)

t1_stop = perf_counter()
print('Время работы: %s секунд '% (t1_stop - t1_start))
```



Время работы: 0.0011126000026706606 секунд

Задание 8. Умножение многочленов

Выдающийся немецкий математик Карл Фридрих Гаусс (1777—1855) заметил, что хотя формула для произведения двух комплексных чисел (a+bi)(c+di)=ac-bd+(bc+ad)i содержит *четыре* умножения вещественных чисел, можно обойтись и *тремя*: вычислим ac,bd и (a+b)(c+d) и воспользуемся тем, что bc+ad=(a+b)(c+d)-ac-bd.

Задача. Даны 2 многочлена порядка n-1: $a_{n-1}x^{n-1}+a_{n-2}x^{n-1}+\ldots+a_1x+a_0$ и $b_{n-1}x^{n-1}+b_{n-2}x^{n-1}+\ldots+b_1x+b_0$. Нужно получить произведение:

$$c_{2n-2}x^{2n-2}+c_{2n-3}x^{2n-3}+\ldots+c_1x+c_0$$
, где:

$$c_{2n-2} = a_{n-1}b_{n-1}$$

$$c_{2n-3} = a_{n-1}b_{n-2} + a_{n-2}b_{n-1}$$
...
$$c_2 = a_2b_0 + a_1b_1 + a_0b_2$$

$$c_1 = a_1b_0 + a_0b_1$$

$$c_0 = a_0b_0$$

Пример. Входные данные: n = 3, A = (3, 2, 5), B = (5, 1, 2)

$$A(x) = 3x^{2} + 2x + 5$$

$$B(x) = 5x^{2} + x + 2$$

$$A(x)B(x) = 15x^{4} + 13x^{3} + 33x^{2} + 9x + 10$$

Ответ: C = (15, 13, 33, 9, 10).

- Формат входного файла (input.txt). В первой строке число n порядок многочленов A и B. Во второй строке коэффициенты многочлена A через пробел. В третьей строке коэффициенты многочлена B через пробел.
- Формат выходного файла (output.txt). Ответ одна строка, коэффициенты многочлена C(x) = A(x)B(x) через пробел.
- Нужно использовать метод "Разделяй и властвуй". Подсказка: любой многочлен A(x) можно разделить на 2 части, например, $A(x) = 4x^3 + 3x^2 + 2x + 1$ разделим на $A_1 = 4x + 3$ и $A_2 = 2x + 1$. И многочлен $B(x) = x^3 + 2x^2 + 3x + 4$ разделим на 2 части: $B_1 = x + 2$, $B_2 = 3x + 4$. Тогда произведение $C = A(x) * B(x) = (A_1B_1)x^n + (A_1B_2 + A_2B_1)x^{n/2} + A_2B_2$ требуется 4 произведения (проверьте правильность данной формулы). Можно использовать формулу Гаусса и обойтись всего тремя произведениями.

```
def add_polynomials(A, B): 2 usages # Hen

return [a + b for a, b in zip(A, B)] + A[len(B):] + B[len(A):]

def multiplication(A, B): 4 usages # Hen

n = len(A)

if n == 1:
    return [A[0] * B[0]]

m = n // 2

Al, A0 = A[:m], A[m:]
Bl, B0 = B[:m], B[m:]

Pl = multiplication(A1, B1)
P2 = multiplication(A0, B0)
P3 = multiplication(add_polynomials(A1, A0), add_polynomials(B1, B0))
```

```
21     result = [0] * (2 * n - 1)
22
23     for i in range(len(P1)):
24         result[i] += P1[i]
25
26     for i in range(len(P2)):
27         result[i + 2 * m] += P2[i]
28
29     for i in range(len(P3)):
30         result[i + m] += P3[i] - (P1[i] if i < len(P1) else 0) - (P2[i] if i < len(P2) else 0)
31
32     return result</pre>
```

- 1. Создать функцию "**add_polynomials**" для сложения двух многочленов.
- Использовать функцию **zip**, чтобы объединить коэффициенты одинаковой степени из многочленов A и B.
- Если многочлены разной длины, объединить оставшиеся коэффициенты из более длинного многочлена.
- 2. Создать функцию "multiplication" для умножения двух многочленов.
- Если длина многочлена равна 1, умножить два коэффициента многочлена вместе.
- Если больше 1, разделить каждый многочлен на две части.
- Рекурсивно выполнить три умножении.
- объединить результаты в список коэффициентов результирующего многочлена.

```
if __name__ == "__main__":

with open('input.txt', 'r') as infile:
    n = int(infile.readline())
    data_a = str(infile.readline())

data_b = str(infile.readline())

A = [int(i) for i in data_a.split()]

B = [int(i) for i in data_b.split()]

arr = multiplication(A, B)
    result = ' '.join(str(i) for i in arr)

with open('output.txt', 'w') as outfile:
    outfile.write(result)
```

3. Использовать **with open** чтобы открыть файл для чтения данных и записи результат.

```
      1
      4

      2
      3 2 4 1

      3
      2 7 1 6

      1
      6 25 25 50 23 25 6
```

Выхол

- Знать, как использовать метод "MergeSort" для сортировки ряда чисел.
- Использовать методы "декомпозиция" для более простого решения заданий.
- Знакомство с умножении многочленов и матриц.