

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение высшего образования
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»

Кафедра информационно-сетевых технологий (№53)

ОТЧЁТ ПО ПРАКТИКЕ
ЗАЩИЩЁН С ОЦЕНКОЙ

РУКОВОДИТЕЛЬ

ассистент

должность, уч. степень, звание

подпись, дата

Гуков С.Ю.

инициалы, фамилия

ОТЧЁТ ПО ПРАКТИКЕ

вид практики

учебная

тип практики

на тему индивидуального задания

Реализация и практическое сравнение

производительности различных алгоритмов сортировок

выполнен

Смирновым Никитой Дмитриевичем

фамилия, имя, отчество обучающегося в творительном падеже

по направлению подготовки

09.03.02

код

Информационные системы и технологии

наименование направления

направленности

09.03.02

код

Информационные технологии в
медиаиндустрии

наименование направленности

наименование направленности

Обучающийся группы №

5138

номер

03.06.2022

подпись, дата

Смирнов Н.Д.

инициалы, фамилия

Санкт–Петербург 2022

ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ

на прохождение учебной практики обучающегося направления подготовки/ специальности _____

09.03.02 Информационные системы и технологии

1. Фамилия, имя, отчество обучающегося: _____

Смирнов Никита Дмитриевич

2. Группа: 5138

3. Тема индивидуального задания: _____

Реализация алгоритмов сортировки и анализа текстовых данных

4. Исходные данные: _____

Visual Studio или Xcode; вариант задания (кириллица, по алфавиту, по возрастанию

учитывать числа, сортировка вставками); десять текстов (от 3 тыс. до 15 тыс. символов)

5. Содержание отчетной документации:

5.1. индивидуальное задание;

5.2. отчёт, включающий в себя:

- титульный лист;
- материалы о выполнении индивидуального задания (содержание определяется кафедрой);
- выводы по результатам практики;
- список использованных источников.

5.3. отзыв руководителя от профильной организации (при прохождении практики в профильной организации).

6. Срок представления отчета на кафедру: «04» июня 2022 г.

Руководитель практики

ассистент

04.06.2022

Гуков С.Ю.

должность, уч. степень, звание подпись, дата

инициалы, фамилия

СОГЛАСОВАНО

Руководитель практики от профильной организации

должность

подпись, дата

инициалы, фамилия

Задание принял к исполнению:

Обучающийся

07.02.2022

дата

подпись

Смирнов Н.Д.

инициалы, фамилия

Санкт–Петербург 2022

Оглавление

Введение.....	4
Задание по варианту.....	5
Блок-схема программы	7
Описание алгоритма	8
Описание функций	9
Тестирование программы.....	10
График зависимости.....	15
Заключение	16
Список используемых источников.....	17
Исходный код	18

Введение

Целью проведения производственной практики является приобретение обучающимися профессиональных умений и навыков, опыта профессиональной деятельности и закрепление приобретенных компетенций, предусмотренных ФГОС и направленностью ОП ВО. В рамках учебной практики необходимо закрепить полученные знания по языку программирования на дисциплине «Основы программирования» и написать программу, для сортировки и анализа текстовых данных. В соответствии с параметрами варианта задания нужно реализовать сортировку слов в тексте и провести его простейший анализ. Варианты заданий приведены на рисунке 3.

Задание по варианту

Необходимо реализовать программу, которая будет считывать текст из файла (например, «original.txt») и с минимальными погрешностями разбивать его на массив отдельных слов. Далее, согласно параметрам выбранного варианта, необходимо отсортировать этот массив слов. Путь либо название считываемого файла должно вводиться через консоль.

Примечание. Символы и знаки препинания игнорируются и не включаются в слово/длину слова. Отсортированный массив необходимо вывести в выходной файл, например, «result.txt». В консоль результат выводить не нужно. Также необходимо вывести информацию с простейшим анализом введенного текста (эту информацию выводим в консоль и в файл, например, «analysis.txt»): • исходный текст • параметры выбранного варианта • количество слов в исходном тексте • время выполнения сортировки • количество слов на каждую букву алфавита (для варианта «сортировка по алфавиту») • количество слов каждой длины (для варианта «сортировка по количеству символов в слове») Необходимо провести 10 тестов, то есть повторить такие действия с разными по длине (от 3 тыс. до 15 тыс. символов) и структуре текстами. Соответственно, должно получиться 30 файлов – 10 файлов с входными текстами, 10 с отсортированными массивами и 10 с анализированием текстов. В отчете нужно наглядно изобразить (график/гистограмма/диаграмма) зависимость времени выполнения сортировки от количества слов в тексте и сделать выводы. Также необходимо оценить и обосновать сложность получившегося алгоритма в О-нотации.

Далее в таблице на рисунке 3 представлены варианты заданий. Варианты старше 24-го начинаются опять с 1-го

Вариант / Задание	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Кириллица	v		v		v			v		v		v	v		v		v			v		v		v
Латиница		v		v		v	v		v		v			v		v		v	v		v		v	
По алфавиту	v		v		v			v		v		v	v		v		v			v		v		v
По кол-ву символов в слове		v		v		v	v		v		v			v		v		v	v		v		v	
По возрастанию	v		v		v			v		v		v	v		v		v			v		v		v
По убыванию		v		v		v	v		v		v			v		v		v	v		v		v	
Учитывать числа	v		v		v		v		v		v			v		v		v		v		v		v
Игнорировать числа		v		v		v		v		v		v	v		v		v		v		v		v	
Сортировка Пузырьком	v						v						v						v					
Сортировка Расческой		v						v						v						v				
Сортировка Вставками			v						v						v						v			
Сортировка Шелла				v						v						v						v		
Быстрая сортировка					v						v						v						v	
Сортировка Слиянием						v						v						v						v

Рисунок 1 – таблица вариантов

Мой вариант – 3.

- Пункт «Кириллица / Латиница» означает, с символами какого алфавита придется работать. Кириллица – «абвгдеёжзийклмнопрстуфхцчщъыьэюя1234567890». Латиница – «abcdefghijklmnopqrstuvwxyz1234567890».
- Пункт «По алфавиту / По количеству символов в слове» означает, по какому признаку необходимо сортировать слова в тексте. По алфавиту – проверяем первую букву в каждом слове и сортируем относительно алфавита (цифры идут после букв). По количеству символов в слове – вычисляем количество букв в каждом слове и сортируем относительно этого количества.
- Пункт «По возрастанию / По убыванию» означает, в каком направлении сортировать массив слов.
- Пункт «Учитывать числа / Игнорировать числа» означает, необходимо ли идентифицировать отдельные числа/цифры в тексте как отдельные слова. Если числа учитываются, то они считаются словом и к ним применяются правила из пункта «По алфавиту / По количеству символов в слове». Если игнорируются, то эти числа/цифры просто опускаются при сортировке.
- Пункт «Сортировка ***» означает, какой метод сортировки необходимо использовать.

Блок-схема программы



Рисунок 2 – простейшая блок-схема алгоритма прог`раммы

Описание алгоритма

1. Открывается файл на чтение.
2. С помощью функции «split_text» создается список слов.
3. Далее засекается время начала выполнения алгоритма сортировки.
4. Сортируется созданный список слов.
5. Высчитывается текущее время и время, которое было засечено перед выполнением сортировки.
6. Вводится информация о варианте.
7. Открывается файл для вывода отсортированного текста.
8. Выводится отсортированный текст
9. Открывается файл для анализа программы.
10. Из функции «print_analisis» выводится требуемая информация.

Описание функций

1. Функция «split_text»

Создается пустой массив и пустая переменная, в которой будет храниться будущее слово. Далее перебирается строка по символу и записывается в переменную «symb». Если символ равен запятой и числу, что уже лежит в «word» или букве или символу числу, то прибавляется к «word» значение «symb» нижнего регистра. Иначе проверяется не равна ли длина строки 0, тогда добавляется в конец массива слов «words» слово «word» и происходит обнуление строки и так пока не закончится текст.

2. Функция «inseration_sort»

Внутри функции создается цикл с переменной «i», которая будет исчисляться от 1 до длины массива. Создается переменная «temp» и приравнивается к массиву. Создается цикл «while» который будет работать до тех пор, пока j не отрицательное и «temp» меньше, чем элемент под индексом [j][0]. Устанавливается «arr[j+1] = arr[j]», затем уменьшается j на 1. После окончания цикла «While» устанавливается «arr[j+1] = «temp».

3. Функция «group_by_alphabet»

Создается словарь, который заполняется пустыми массивами для букв и цифр. Далее распределяются слова из «arr» по словарю на ту букву, на которую они начинаются.

4. Функция «print_output»

Функция служит выводом отсортированного текста в файл вывода. Создается словарь из массива «arr», который передается в функцию. Открывается файл на вывод кодировки «utf-8». Выводятся все слова через пробел, начинающиеся на определенную букву.

5. Функция «print_analysis»

Создается или открывается файл анализа на вывод. Записывается исходный текст, вариант, количество слов, время сортировки. Перебирается словарь и выводится количество слов на каждую букву.

Оперативная память: 4ГБ

587 слов, время выполнения 0,0250 секунд

Анализ текста: Идеиные сообщения высшего порядка, а также семантический ракурс внешних протекстов создаст предпосылки для первоочередных требований. Есть над чем задуматься: элементы политического процесса указаны как претенденты на роль ключевых факторов. И нет А также репликация с заимован истоников, современное исследование будет описано максимально подробно! Валась всего лишь частая общая картина, keine прозвлек победы институционализации освещает чрезвычайно интересные особенности картины в целом, однако конкретные опитаны с учетом контекста соответствующих условий активизации.

Таким как граница обучения кадров создает необходимость включения в производственный план целого ряда внеочередных мероприятий с учетом комплекса сушествующих финансовых и административных условий. Таким образом, сушествующая теория требует определения и уточнения выводов и при этом немаловажно исследование.

С другой стороны, семантический ракурс внешних протекстов предпринимает высокую востребованность экономической целесообразности принимаемых решений. Ясность напад Позднее очевидна: консультации с широким активом требуют определения и уточнения системы массового учета. Учитывая ключевые сценарии поведения, дальнейшее развитие различных форм деятельности прекрасно подходит для реализации экономической целесообразности принимаемых решений. Вот как яркий пример современных тенденций - выбранный нами инновационный путь, а также свежий

В рамках спецификации современных стандартов, элементы политического процесса представляет собой не что иное, как контингентные победы маркетинга над разумом и должны быть разоблачены. Вы вынуждены отталкиваться от того, что перспективное планирование напрямую зависит от

Вы вынуждены отталкиваться от того, что перспективное планирование позволяет выполнить задание по разработке вывода текущих активов. Имеется спорная точка зрения, гласящая примерно следующее: предприниматели в сети интернет указаны как претенденты на роль ключевых

Количество слов: 587
Время: 11.3
Сортировки в секундах: 8.8258

n = 14
m = 14
n = 45
r = 6
A = 22
n = 0
r = 13
n = 31
q = 9
k = 36
r = 4
n = 13
n = 46
n = 45
n = 75
r = 38
s = 81
y = 23
y = 29
Q = 9
k = 0
ц = 9
ц = 18
n = 3
ц = 0
n = 0
n = 0
n = 0
n = 5
Q = 0
l = 0
2 = 0
3 = 0
4 = 0
5 = 0
6 = 0
7 = 0
8 = 0
9 = 0

Текст №2

906 слов, время выполнения 0,0523 секунды

[illegible]

Рисунок 4 – файл анализа 2 текста

Текст №3

951 слово, время выполнения 0.0602 секунды

Повседневная практика показывает, что новая модель организационной деятельности, а также свежий взгляд на привычные вещи – безусловно открывает новые горизонты для вывода текущих активов. Не следует, однако, забывать, что разбавленное извне, стремясь вытеснить традиционное производство, нанотехнологии и по сей день остается уделом либералов, которые жаждут быть функционально разнесены на независимые элементы. Лишь непосредственные участники технического прогресса для современного мира убежденность некоторых оппонентов требует определения и уточнения поэтапного и последовательного развития общества. В своем стремлении повысить качество жизни, они забывают, что граница обучения кадров однозначно фиксируется. Задача организации, в особенности же выбранный нами инновационный путь требует от нас анализа соответствующих условий активизации. Мы вынуждены отталкиваться от того, что высококачественный прототип будущего проекта предопределяет высокую скорость выполнения работ.

Вариант: 3
Количество слов: 951
Время сортировки в секундах: 0.0602

а = 36
б = 23
в = 76
г = 11
д = 68
е = 1
ж = 7
з = 22
и = 44
й = 0
к = 56
л = 4
м = 28
н = 89
о = 77
п = 137
р = 38
с = 98
т = 38
у = 37
ф = 16
х = 0
ц = 13
ч = 21
ш = 9
щ = 0
ъ = 0
ы = 0
ь = 0
э = 13
ю = 0
я = 9
0 = 0
1 = 0
2 = 0
3 = 0
4 = 0
5 = 0
6 = 0
7 = 0
8 = 0
9 = 0

Рисунок 5 - файл анализа 3 текста

Текст №4

378 слов, время выполнения 0.0100 секунд

В своем стремлении повысить качество жизни, они забывают, что социально-экономическое развитие создаст необходимость включения в производственный план целого ряда внеочередных мероприятий с учетом комплекса экономической целесообразности и организационных аспектов. Лишь непосредственные участники технического прогресса для современного мира убежденность некоторых оппонентов требует определения и уточнения поэтапного и последовательного развития общества. В своем стремлении повысить качество жизни, они забывают, что граница обучения кадров однозначно фиксируется. Задача организации, в особенности же выбранный нами инновационный путь требует от нас анализа соответствующих условий активизации. Мы вынуждены отталкиваться от того, что высококачественный прототип будущего проекта предопределяет высокую скорость выполнения работ.

Вариант: 3
Количество слов: 378
Время сортировки в секундах: 0.0100

а = 7
б = 12
в = 29
г = 6
д = 17
е = 1
ж = 2
з = 12
и = 26
й = 0
к = 17
л = 1
м = 15
н = 25
о = 23
п = 56
р = 28
с = 46
т = 7
у = 18
ф = 6
х = 0
ц = 5
ч = 7
ш = 5
щ = 0
ъ = 0
ы = 0
ь = 0
э = 6
ю = 0
я = 1
0 = 0
1 = 0
2 = 0
3 = 0
4 = 0
5 = 0
6 = 0
7 = 0
8 = 0
9 = 0

Рисунок 6 – файл анализа 4 текста

Текст №5

542 слова, время выполнения 0.0178 секунд

С другой стороны, дальнейшее развитие различных форм деятельности является качественно новой ступенью экспериментов, поражающих по своей масштабности и грандиозности. В своём стремлении повысить качество жизни, они забывают, что реализация нового производства, нанотехнологии призывает нас к новым свершениям, которые, в свою очередь, должны быть преданы социально-демократической анафеме!

А ещё стремясь вытеснить традиционное производство, нанотехнологии призывают нас к новым свершениям, которые, в свою очередь, должны быть подвергнуты целой серии независимых исследований. Разнообразный и богатый опыт говорит нам, что

Противоположная точка зрения подразумевает, что непосредственные участники технического прогресса смешаны с не уникальными данными до степени совершенной неузнаваемости, из-за чего возрастает их статус бесполезности. Картельные сговоры и

Вариант: 3
Количество слов: 542
Время сортировки в секундах: 0.0178

а = 19
б = 15
в = 40
г = 13
д = 25
е = 1
ж = 5
з = 14
и = 33
й = 0
к = 29
л = 3
м = 15
н = 50
о = 38
п = 77
р = 38
с = 57
т = 26
у = 16
ф = 11
х = 0
ц = 1
ч = 14
ш = 0
щ = 0
ъ = 0
ы = 0
э = 6
ю = 0
я = 4
0 = 0
1 = 0
2 = 0
3 = 0
4 = 0
5 = 0
6 = 0
7 = 0
8 = 0
9 = 0

Рисунок 7 – файл анализа 5 текста

Текст №6

582 слова, время выполнения 0.0217 секунд

Задача организации, в особенности же понимание сути ресурсосберегающих технологий однозначно определяет каждого участника как способного принимать собственные решения касаемо существующих финансовых и административных условий! Не следует

Высокий уровень вовлечения представителей целевой аудитории является четким доказательством простого факта: внедрение современных методик прекрасно подходит для реализации как самодостаточных, так и внешне зависимых концептуальных решений

Также как убедительность некоторых оппонентов создаёт необходимость включения в производственный план целого ряда внеочередных мероприятий с учётом комплекса прогресса профессионального сообщества. Повседневная практика показывает, что нов

Для современного мира существующая теория является качественно новой ступенью соответствующих условий активизации. Ясность нашей позиции очевидна: выбранный нами инновационный путь требует анализа системы массового участия. Ясность нашей

Вариант: 3
Количество слов: 582
Время сортировки в секундах: 0.0217

а = 11
б = 8
в = 47
г = 5
д = 19
е = 3
ж = 1
з = 19
и = 28
й = 0
к = 36
л = 6
м = 22
н = 45
о = 48
п = 85
р = 30
с = 78
т = 21
у = 28
ф = 10
х = 1
ц = 5
ч = 11
ш = 2
щ = 0
ъ = 0
ы = 0
э = 5
ю = 0
я = 8
0 = 0
1 = 0
2 = 0
3 = 0
4 = 0
5 = 0
6 = 0
7 = 0
8 = 0
9 = 0

Рисунок 8 – файл анализа 6 текста

Текст №7

754 слова, время выполнения 0.0363 секунды

Как принято считать, ключевые особенности структуры проекта, которые представляет собой яркий пример континентально-европейского типа политической культуры, будут преданы социально-демократической анафеме. Высокий уровень вовлечения представителей целевой аудитории является четким доказательством простого факта: экономическая повестка сегодняшнего дня напрямую зависит от модели развития. Однозначно, многие известные личности призывают нас каждый из нас понимает очевидную вещь: консультация с широким активом требует определения и уточнения экономической целесообразности принимаемых решений. С другой стороны, реализация намеченных плановых заданий требует анализа модели развития. Являясь всего лишь частью общей картины, интерактивные прототипы и по сей день остаются уделом либералов, которые жаждут быть смеяны с не уникальными данными до степени совершенной неузнаваемости, из-за чего возрастает их статус бесполезности. Идеальные соображения высказываясь, убежденность некоторых оппонентов, в своей классическом представлении, допускает внедрение первоочередных требований. Не следует, однако, забывать, что глубокий уровень погружения требует от нас анализа существующих финансовых и логистических возможностей.

Вариант: 3
Количество слов: 754
Время сортировки в секундах: 0.0363

а	=	13
б	=	18
в	=	53
г	=	11
д	=	31
е	=	5
ж	=	5
з	=	22
и	=	59
й	=	0
к	=	42
л	=	7
м	=	20
н	=	37
о	=	46
п	=	112
р	=	50
с	=	115
т	=	18
у	=	29
ф	=	15
х	=	2
ц	=	5
ч	=	14
ш	=	3
щ	=	0
ъ	=	0
ы	=	0
ь	=	0
э	=	16
ю	=	0
я	=	6
0	=	0
1	=	0
2	=	0
3	=	0
4	=	0
5	=	0
6	=	0
7	=	0
8	=	0
9	=	0

Рисунок 9 – файл анализа 7 текста

Текст №8

732 слова, время выполнения 0.0316 секунд

Повседневная практика показывает, что понимание сути ресурсосберегающих технологий является качественно новой ступенью благоприятных перспектив. Задача организации, в особенности же высокотехнологичная концепция общественного уклада способствует достижению поставленных целей. Высокий уровень вовлечения представителей целевой аудитории является четким доказательством простого факта: экономическая повестка сегодняшнего дня напрямую зависит от модели развития. Однозначно, многие известные личности призывают нас каждый из нас понимает очевидную вещь: консультация с широким активом требует определения и уточнения экономической целесообразности принимаемых решений. С другой стороны, реализация намеченных плановых заданий требует анализа модели развития. Являясь всего лишь частью общей картины, интерактивные прототипы и по сей день остаются уделом либералов, которые жаждут быть смеяны с не уникальными данными до степени совершенной неузнаваемости, из-за чего возрастает их статус бесполезности. Идеальные соображения высказываясь, убежденность некоторых оппонентов, в своей классическом представлении, допускает внедрение первоочередных требований. Не следует, однако, забывать, что глубокий уровень погружения требует от нас анализа существующих финансовых и логистических возможностей.

Вариант: 3
Количество слов: 732
Время сортировки в секундах: 0.0316

а	=	15
б	=	20
в	=	63
г	=	11
д	=	30
е	=	1
ж	=	3
з	=	17
и	=	45
й	=	0
к	=	54
л	=	9
м	=	25
н	=	49
о	=	58
п	=	96
р	=	41
с	=	92
т	=	26
у	=	29
ф	=	9
х	=	0
ц	=	11
ч	=	19
ш	=	1
щ	=	0
ъ	=	0
ы	=	0
ь	=	0
э	=	6
ю	=	0
я	=	10
0	=	0
1	=	0
2	=	0
3	=	0
4	=	0
5	=	0
6	=	0
7	=	0
8	=	0
9	=	0

Рисунок 10 – файл анализа 8 текста

Текст №9

909 слов, время выполнения 0.0497 секунд

В целом, конечно, постоянное информационно-пропагандистское обеспечение нашей деятельности создаёт предпосылки для позиций, занимаемых участниками в отношении поставленных задач. Кстати, реплицированные с зарубежных источников, современные противоположная точка зрения подразумевает, что явные признаки победы институционализации лишь добавляет фракционных разногласий и представлены в исключительно положительном свете. Являясь всего лишь частью общей картины, стремящиеся выт Картельные сговоры не допускают ситуации, при которой представители современных социальных резервов, которые представляет собой яркий пример континентально-европейского типа политической культуры, будут функционально размесены на независ

Являясь всего лишь частью общей картины, непосредственные участники технического прогресса освещает чрезвычайно интересные особенности картины в целом, однако конкретные выводы, разумеется, объединены в целые кластеры себе подобных. Наше

Вариант: 3
Количество слов: 909
Время сортировки в секундах: 0.0497

а = 11
б = 12
в = 88
г = 7
д = 29
е = 4
ж = 4
з = 29
и = 47
й = 0
к = 70
л = 5
м = 25
н = 74
о = 68
п = 146
р = 52
с = 117
т = 33
у = 37
ф = 8
х = 1
ц = 9
ч = 21
ш = 4
щ = 0
ъ = 0
ы = 0
ь = 0
э = 14
ю = 0
я = 18
0 = 0
1 = 0
2 = 0
3 = 0
4 = 0
5 = 0
6 = 0
7 = 0
8 = 0
9 = 0

Рисунок 11 – файл анализ 9 текста

Текст №10

И наконец самый большой текст. 1889 слов, время выполнения 0.2136 секунд

Лишь элементы политического процесса своевременно верифицированы. Внезапно, акционеры крупнейших компаний и по сей день остается уделом либералов, которые жаждут быть своевременно верифицированы. Для современного мира понимание сути ресу

Также как начало повседневной работы по формированию позиции играет важную роль в формировании поставленных обществен задач. Кстати, элементы политического процесса объединены в целые кластеры себе подобных. В своём стремлении улучшить п

Противоположная точка зрения подразумевает, что явные признаки победы институционализации освещает чрезвычайно интересные особенности картины в целом, однако конкретные выводы, разумеется, преданы социально-демократической анафеме. Совре

Но представители современных социальных резервов, инцидированные исключительно синтетически, заблокированы в рамках своих собственных рациональных ограничений. В своём стремлении повысить качество жизни, они забывает, что базовый вектор

Вариант: 3
Количество слов: 1889
Время сортировки в секундах: 0.2136

а = 38
б = 42
в = 147
г = 28
д = 60
е = 3
ж = 9
з = 49
и = 111
й = 0
к = 120
л = 13
м = 57
н = 154
о = 140
п = 380
р = 189
с = 218
т = 65
у = 69
ф = 28
х = 1
ц = 26
ч = 45
ш = 6
щ = 0
ъ = 0
ы = 0
ь = 0
э = 33
ю = 0
я = 18
0 = 0
1 = 0
2 = 0
3 = 0
4 = 0
5 = 0
6 = 0
7 = 0
8 = 0
9 = 0

Рисунок 12 – файл анализа 10 текста

График зависимости



Рисунок 13 – график зависимости

Сложность сортировки вставками в среднем равна $O(n^2)$. Пирамидальная сортировка (Heap Sort) (англ. Heapsort, «Сортировка кучей»). Этот популярный алгоритм сортировки, как сортировки вставками и выбором, сегментирует список на отсортированные и несортированные части.

Заключение

В ходе работы были закреплены полученные знания по дисциплине «Основы программирования». Была реализована программа, которая считывает текст Кириллицей из txt-файла, с минимальными погрешностями разбивает его на массив отдельных слов и сортирует его по алфавиту, по возрастанию и учитывая числа с помощью алгоритма «вставками». Всё это было выполнено в соответствии с 3 вариантом задания.

Код программы был поделён на пять функций, каждая из которых выполняла определённый алгоритм в соответствии с её названием. Была составлена простейшая блок-схема алгоритма всей программы. Были написаны словесные описания для алгоритма всей программы и функций. Программа была протестирована на 10 тестах, из результатов которых был составлен график зависимости времени сортировки от количества слов в тексте.

Список используемых источников

1. Cyberforum: C++. Ссылка: [Форум программистов и сисадминов Киберфорум \(cyberforum.ru\)](http://cyberforum.ru)
2. Wikipedia Ссылка: [Википедия — свободная энциклопедия \(wikipedia.org\)](http://wikipedia.org)
3. Викиучебник Ссылка: [Python/Учебник Python 3.1 — Викиучебник \(wikibooks.org\)](http://wikibooks.org)
4. Pythonist Ссылка: [Сортировка вставками на Python | Алгоритмы | Pythonist .ru](http://pythonist.ru)

Исходный код

```
import time

def split_text(inp_text):    # функция разбивающая входную строку на массив
                              слов
    inp_text += " "
    words = [] # создание пустого массива
    word = "" # пока пустая переменная где будет храниться собранное слово
    for i in range(len(inp_text)): # перебирается строка по одному символу
        symb = inp_text[i] # записываем текущий символ в symb
        if symb == "," and word.isnumeric() or symb.isalpha() or
symb.isdigit():
            word += symb.lower() # прибавляется к word значение symb в нижнем
регистре.
        else:
            if len(word) != 0: # проверяется не равна ли длина строки word
нулю, если нет, слово не пустое
                words.append(word)    # добавляется в конец массива слов,
слово word
            word = "" # обнуление
    return words

def insertion_sort(arr):    # сортировка вставками
    for i in range(1, len(arr)): # определение начальной позиции
        temp = arr[i]
        j = i - 1
        while j >= 0 and temp[0] < arr[j][0]: # сравниваем первый символ
            arr[j + 1] = arr[j]
            j -= 1
        arr[j + 1] = temp

def group_by_alphabet(arr):    # создание словаря
    res = {} # создаем словарь
    for i in range(ord("a"), ord("я") + 1): # заполнение словаря пустыми
массивами для букв
        res[chr(i)] = []
    for i in range(10): # заполнение словаря пустыми массивами для цифр
        res[str(i)] = []

    for word in arr: # распределение слов из arr по словарю, для подсчета
слов, начинающихся на определенную букву
        symb = word[0]
        res[symb].append(word)
    return res

def print_output(arr, path_write): # вывод отсортированного массива в файл
    gr_arr = group_by_alphabet(arr) # создание словаря из массива arr
    file_out = open(path_write, "w+", encoding="utf8") # открытие файла на
запись
    for key in gr_arr:
        if len(gr_arr) == 0:
            continue
        # Key - буква алфавита, а значения - все слова на эту букву
        for word in gr_arr[key]:
            # вывод всех слов через пробел начинающиеся на букву
            file_out.write(f"{word} ")
        file_out.write("\n")
```

```

def print_analysis(arr, path_write, execut_t, inp_text): # работа с файлом
анализа программы
    gr_arr = group_by_alphabet(arr)
    file_analysis = open(path_write, "w+", encoding="utf8") # открытие файла
для записи анализа
    file_analysis.write(f"Введенный текст: {inp_text}\n")
    file_analysis.write("Вариант:3\n")
    file_analysis.write(f"Количество слов: {len(arr)}\n")
    file_analysis.write(f"Время сортировки в секундах: {execut_t:0.4f}\n")
    for key in gr_arr: # вывод количества слов на букву или цифру
        file_analysis.write(f"{key} = {len(gr_arr[key])}\n")

path = input("Укажите путь к файлу чтения: ")
inp_file = open(path, encoding="utf8") # открытие файла на чтение
readed = inp_file.read()
words = split_text(readed) # получение списка из слов

tic = time.perf_counter() # начало для измерения времени
insertion_sort(words) # сортировка полученного списка из слов
toc = time.perf_counter() # конец измерения времени
execut_t = toc-tic # вычитаем из конца начало для нахождения времени

path = input("Укажите путь к файлу вывода: ")
print_output(words, path) # вывод отсортированного массива в файл

path = input("Укажите путь к файлу анализа: ")
print_analysis(words, path, execut_t, readed) # вывод анализа в файл

```