МИНОБРНАУКИ РОССИИ САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ «ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА) Кафедра МО ЭВМ

КУРСОВАЯ РАБОТА

по дисциплине «Программирование»

Тема: Обработка текстовых данных

Студент гр. 1304	 Мусаев А. И.
Преподаватель	Чайка К. В.

Санкт-Петербург 2021

ЗАДАНИЕ НА КУРСОВУЮ РАБОТУ

Студент: Мусаев А. И.				
Группа: 1304				
Тема работы: обработка текстовых данных				
Исходные данные:				
консольное приложение на языке Си, обработка введённого пользователем				
текста согласно введённым кодам действий, пока пользователь сам не				
завершит работу в программе; использовать структуры и широкие символы				
Содержание пояснительной записки:				
«Содержание», «Введение», «Разработка программного кода», «Сборка				
программы», «Тестирование», «Пользовательская инструкция»,				
«Заключение», «Список использованных источников», «Приложение А.				
Программный код»				
Предполагаемый объем пояснительной записки:				
Не менее 12 страниц.				
Дата выдачи задания: 15.10.2021				
Дата сдачи реферата: 21.12.2021				
Дата защиты реферата: 22.12.2021				
Студент Мусаев А. И.				
Преподаватель Чайка К. В.				

АННОТАЦИЯ

Разработано диалоговое консольное приложение, демонстрирующее пользователю возможности обработки текста в Си. Хранение текста организовано в виде структур в динамической памяти. Вводимый текст поддерживает кириллические (широкие) символы. При выполнении работы использовались функции стандартной библиотеки языка Си.

SUMMARY

A dialog console application has been developed that demonstrates to the user the possibilities of text processing in C. Text storage is organized in the form of structures in dynamic memory. The input text supports Cyrillic (wide) characters. When performing the work, the functions of the C standard library were used.

```
Введите текст:
Мистер и миссис Дурсль проживали в доме ноmer четыре по Тисовой улице и всегда с гордостью заявляли, что они, слава богу, абсолютlу нормальные люди. Уж от кого-кого, а от них никак нельзя было ожидать, чтобы они попали в какую-нибудь странную или загадочную ситуацию. Мистер и миссис Дурсль проживали в доме ноmeр четыре по Тисовой улице и всегда с гордостью заявляли, что они, слава богу, абсолютно нормальные люди. Мистер дурсль возглавляла фирму под назв анием «Граннингс», которая специализировалась на производстве дрелей. Это был полный мужчина с очень пышными усами и очень короткой шеей. Что же касается я миссыз Дурсль, она была тощей блондинкой с шеей почти вдвое длиннее, чем положено при ее росте. Однако этот недостаток пришелся ей весьма кстати, поск ольку большую часть времени миссис Дурсль следила за соседями и подслушивала их разговоры. А с такой шеей, как у нее, было очень удобно заглядывать за ч ужие заборы. Это был полный мужчина с очень пышными усами и очень короткой шеей. Что же касается миссис Дурсль, она была тощей блондинкой с шеей почти в двое длиннее, чем положено при ее росте. Шалаш, шалаш, мадам!

Выберите желаемое действие:

Введите 4, если хотите для каждого предложения вывести "Палиндромы на месте!" если все слова являются палиндромами, "Кто-то потерял все палиндромы." есл и ни одного слово не является палиндромом, в остальных случаях "чего-то не хватает."

Введите 2, если хотите заменить все специальные символы (включая .,;) в тексте на подстроку ">special symbol<".

Введите 3, если хотите эзменить все специальные символы (включая .,;) в тексте на подстроку ">special symbol<".

Введите 4, если хотите отсортировать предложения в которых в одном слове встречаются латинские и кириллические буквы.
Введите 6, если хотите вывести текст.

Введите 6, если хотите завершить работу программы.
```

Рис. 1.

```
Мистер и миссис Дурсль проживали в доме ноmer четыре по Тисовой улице и всегда с гордостью заявляли, что они, слава богу, абсолютlу нормальные люди. Чего-то не хватает.

Уж от кого-кого, а от них никак нельзя было ожидать, чтобы они попали в какую-нибудь странную или загадочную ситуацию. Чего-то не хватает.

Мистер и миссис Дурсль проживали в доме номер четыре по Тисовой улице и всегда с гордостью заявляли, что они, слава богу, абсолютно нормальные люди. Чего-то не хватает.

Мистер Дурсль возглавлял фирму под названием «Граннингс», которая специализировалась на производстве дрелей. Кто-то потерял все палиндромы.

Это был полный мужчина с очень пышными усами и очень короткой шеей. Чего-то не хватает.

Что же касается миссія Дурсль, она была тощей блондинкой с шеей почти вдвое длиннее, чем положено при ее росте. Чего-то не хватает.

Однако этот недостаток пришелся ей весьма кстати, поскольку большую часть времени миссис Дурсль следила за соседями и подслушивала их разговоры. Чего-то не хватает.

А с такой шеей, как у нее, было очень удобно заглядывать за чужие заборы. Чего-то не хватает.

Что же касается миссис Дурсль, она была тощей блондинкой с шеей почти вдвое длиннее, чем положено при ее росте. Чего-то не хватает.

Что же касается миссис Дурсль, она была тощей блондинкой с шеей почти вдвое длиннее, чем положено при ее росте. Чего-то не хватает.

Что же касается миссис Дурсль, она была тощей блондинкой с шеей почти вдвое длиннее, чем положено при ее росте. Чего-то не хватает.
```

Рис 2.

Micropy-special symbol<arbonal-psecial symbol<arbonal-psecial-pse

Рис. 3.

3

Мистер и миссис Дурсль проживали в доме ноmer четыре по Тисовой улице и всегда с гордостью заявляли, что они, слава богу, абсолюtly нормальные люди. Что же касается мисsis Дурсль, она была тощей блондинкой с шеей почти вдвое длиннее, чем положено при ее росте. Уж от кого-кого, а от них никак нельзя было ожидать, чтобы они попали в какую-нибудь странную или загадочную ситуацию. Мистер и миссис Дурсль проживали в доме номер четыре по Тисовой улице и всег да с гордостью заявляли, что они, слава богу, абсолютно нормальные люди. Мистер Дурсль возглавлял фирму под названием «Граннингс», которая специализиров алась на производстве дрелей. Это был полный мужчина с очень пышными усами и очень короткой шеей. Однако этот недостаток пришелся ей весьма кстати, поск ольку большую часть времени миссис Дурсль следила за соседями и подслушивала их разговоры. А с такой шеей, как у нее, было очень удобно заглядывать за ч ужие заборы. Что же касается миссис Дурсль, она была тощей блондинкой с шеей почти вдвое длиннее, чем положено при ее росте. Шалаш, шалаш, мадам!

Рис. 4.

4

Уж от кого-кого, а от них никак нельзя было ожидать, чтобы они попали в какую-нибудь странную или загадочную ситуацию. Мистер и миссис Дурсль проживали в доме номер четыре по Тисовой улице и всегда с гордостью заявляли, что они, слава богу, абсолютно нормальные люди. Мистер Дурсль возглавлял фирму под н азванием «Граннингс», которая специализировалась на производстве дрелей. Это был полный мужчина с очень пышными усами и очень короткой шеей. Однако этот недостаток пришелся ей весьма кстати, поскольку большую часть времени миссис Дурсль следила за соседями и подслушивала их разговоры. А с такой шеей, ка к у нее, было очень удобно заглядывать за чужие заборы. Что же касается миссис Дурсль, она была тощей блондинкой с шеей почти вдвое длиннее, чем положен опри ее росте. Шалаш, шалаш, мадам!

Рис. 5.

СОДЕРЖАНИЕ

	Введение	6
1.	Разработка программного кода	7
1.1.	Техническое задание	7
1.2.	Разбиение по задачам	8
1.3.	Запись и хранение текста	9
1.4.	Общие функции обработки текста	10
1.5.	Функции пользовательской обработки текста	11
2.	Сборка программы	12
3.	Тестирование	13
4.	Пользовательская инструкция	14
	Заключение	15
	Список использованных источников	15
	Приложение А. Программный кол	16

ВВЕДЕНИЕ

Целью данной работы является изучение встроенных средств языка Си для работы с текстом и разработка пользовательского приложения с использованием этих средств.

Дли достижения поставленной цели требуется решить следующие задачи:

- изучение теоретических материалов (структуры, функции стандартной библиотеки, цветной вывод текста, широкие символы);
 - разработка программного кода в соответствии с тех. заданием;
- сборка программы с использованием заголовочных файлов и мэйкфайла;
 - тестирование приложения и отладка;
 - создание пользовательской инструкции.

1. РАЗРАБОТКА ПРОГРАММНОГО КОДА

1.1. Техническое задание

Вариант 10

Программе на вход подается текст (текст представляет собой предложения, разделенные точкой. Предложения - набор слов, разделенные пробелом или запятой, слова - набор латинских или кириллических букв, цифр и других символов кроме точки, пробела или запятой) Длина текста и каждого предложения заранее не известна.

Для хранения предложения и для хранения текста требуется реализовать структуры Sentence и Text

Программа должна сохранить (считать) текст в виде динамического массива предложений и оперировать далее только с ним. Функции обработки также должны принимать на вход либо текст (Text), либо предложение (Sentence).

Программа должна найти и удалить все повторно встречающиеся предложения (сравнивать их следует посимвольно, но без учета регистра).

Далее, программа должна запрашивать у пользователя одно из следующих доступных действий (программа должна печатать для этого подсказку. Также следует предусмотреть возможность выхода из программы):

- 1) Для каждого предложения вывести "Палиндромы на месте!" если все слова являются палиндромами, "Кто-то потерял все палиндромы." если ни одного слово не является палиндромом, в остальных случаях "Чего-то не хватает.".
- 2) Заменить все специальные символы (включая .,;) в тексте на подстроку ">special symbol<".
- 3) Отсортировать предложения по уменьшению количества латинских букв.
- 4) Удалить все предложения в которых в одном слове встречаются латинские и кириллические буквы.

Все сортировки и операции со строками должны осуществляться с использованием функций стандартной библиотеки. Использование собственных функций, при наличии аналога среди функций стандартной библиотеки, запрещается.

Каждую подзадачу следует вынести в отдельную функцию, функции сгруппировать в несколько файлов (например, функции обработки текста в один, функции ввода/вывода в другой). Также, должен быть написан Makefile.

1.2. Разбиение по задачам

Таким образом, код разбивается на следующие задачи:

- 1) ввести текст и организовать его удобное расположение в памяти;
- 2) сравнить предложения между собой без учёта регистра и удалить повторяющиеся;
- 3) найти палиндромы в каждом предложении и определить на какую часть предложение состоит из них;
- 4) написать функцию-компаратор для встроенной функции qsort();
- 5) написать функцию удаления предложений по условию;

Исходя из характера задач, целесообразно разделить их по группам:

- 1) функции ввода-вывода;
- 2) функции обработки текста;
- 3) действия, реализуемые непосредственно в функции main().

1.3. Запись и хранение текста

Размер текста до компиляции не определён и может быть очень большим. Используем динамическую память.

Текст должен быть структурирован удобно для дальнейшей обработки. Разобьём его на предложения, а предложения — на слова.

Реализуем для каждой составной части структурный тип данных: Text и Sentence.

Sentence — массив слов, он будет иметь переменную size, чтобы знать размер предложения.

Text — массив предложений и его size.

Функция readText() считывает текст и сохраняет его в структуру Text. Функция output() отвечает за вывод, она выводит текст с пробелами между предложениями.

1.4. Общие функции обработки текста

Функция remove_repetes() сравнивает попарно предложения в тексте. Если находятся два совпадающих предложения, остается только то, которое стоит в предложении раньше.

1.5. Функции пользовательской обработки текста

В функции main() устанавливается кириллическая локализация (встроенная функция setlocale()).

В зависимости от введенного пользователем значения происходит следующие действия:

- 1) Запускается функция palindroms(), которая находит слова в каждом предложении и проверяет, являются ли они палиндромами. Для того чтобы определить на какую часть предложение состоит из палиндромов, есть два флага, один «поднимается», если находится хоть один палиндром, другой, если находится не палиндром. Так можно определить нужные данные. Затем для каждого предложения выводится найденная информация.
- 2) Происходит замена всех специальных символов в функции «replace» на строку ">special symbol<", это происходит при помощи нескольких вспомогательных массивов. Один из них для хранения подстроки, другой для пересобирания предложений и третий для подсчета новой длины предложения.
- 3) Сортировка происходит при помощи встроенной функции qsort(), функция компоратор funcmp(), подсчитывает количество латинских букв в предложении и сравнивает по их количеству два предложения.
- 4) Удаление происходит по средствам функции delete_lat_and_kir(), она находит все предложения, в которых в одном слове встречаются латинские и кириллические буквы и удаляет их.

2. СБОРКА ПРОГРАММЫ

Функции ввода-вывода readWord(), readSentence(), readText() и printText() объявлены в отдельном файле inputoutput.c, которому соответствует заголовочный файл inputoutput.h. Все необходимые для этих функций заголовочные файлы стандартной библиотеки и константные выражения прописаны в inputoutput.h, также в нём описаны структуры для хранения текста.

Функции обработки текста сгруппированы в файле processing.c: remove_repetes(), delete_lat_and_kir(), funcmp(), replace(), palindroms(), . Все необходимые заголовочные файлы стандартной библиотеки, прототипы самих функций и описание структур сохранены в processing.h.

Так как в функцию main() будут импортированы оба заголовочных файла, помимо обычной защиты от повторного включения требуется предусмотреть условную компиляцию для описания структур, зависящую от определения одного и того же выражения.

Сборка программы будет осуществляться утилитой make в соответствии с инструкциями в мэйкфайле Makefile.

3. ТЕСТИРОВАНИЕ

Результаты тестирования представлены в таблице 1. В целях удобства учитываются только значимые части ввода и вывода. Полностью детали интерактива продемонстрированы на рисунках 1-4 в аннотации.

Таблица 1 – Результаты тестирования

таоли	ца 1 – Результаты тестирования				
№ п/п	Входные данные	Выходные данные			
Удаление дубликатов предложений. Ввод по разделителям, вывод через пробелы.					
1.	Привет, как ты? Я нормально, а ты? ПРИВЕТ, КАК ТЫ? Я тоже нормально, надеюсь, всё будет хорошо. И я!	Привет, как ты? Я нормально, а ты? Я тоже нормально, надеюсь, всё будет хорошо. И я!			
	Действие 1: информация про палиндромы.				
2.	Привет, как ты? Я нормально, а ты? ПРИВЕТ, КАК ТЫ? Я тоже нормально, надеюсь, всё будет хорошо. И я!	Привет, как ты? Чего-то не хватает. Я нормально, а ты?			
		Чего-то не хватает.			
		Я тоже нормально, надеюсь, всё будет хорошо. Чего-то не хватает.			
		И я!			
		Палиндромы на месте!			
Действие 2: замена					
3.	Для тебя живу. Тебя, тебя люблю, ТЕБЯ. Наверно, тебячата. 2	Привет>special symbol<>special symbol<как>special symbol<ты? Я>special symbol<+ормально>special symbol<>special symbol <a>special symbol<ты? Я>special symbol<ты? Я>special symbol<тоже>special symbol<+ормально>special symbol<+ормально>special symbol<>special symbol<>special symbol<special symbol<br=""></special>symbol<special symbol<ol=""> symbol<special symbol<br="" symbol<special=""></special>sypecial symbol<будет>special symbol<будет>special</special>			
	Действие 3: сортиров	ка по количеству букв.			
5.	Я fine, а ты? Привет, как ты? Я тоже нормально, надеюсь, всё будет хорошо. И я!	Я fine, а ты? Привет, как ты? Я тоже нормально, надеюсь, всё будет хорошо. И я!			
	•	инскими и кириллическими			
6.	СЛОЕ Привет, как ты? Я finaльно, а ты? ПРИВЕТ, КАК ТЫ? Я тоже нормально, надеюсь, всё будет хорошо. И я!	вами. Привет, как ты? Я тоже нормально, надеюсь, всё будет хорошо. И я!			

	Произвольный код действия.				
7.	7	Введите число от 1 до 6.			
8.	p				
	Действие 5: выход из программы.				
9.	Привет, как ты? Я тоже нормально, надеюсь, всё будет хорошо. И я!				

4. ПОЛЬЗОВАТЕЛЬСКАЯ ИНСТРУКЦИЯ

«Программа позволяет производить заданные действия над введённым текстом. Какое именно действие из набора применить, определяете Вы.

Шаг первый — запуск приложения

Запустите терминал из папки с файлами приложения. Введите команду make && ./a.out.

Шаг второй — ввод текста

Введите в терминал текст и нажмите клавишу «Enter».

Текст вводится в одну строку как последовательность предложений, разделённых точкой с пробелом.

Слова в предложении разделяются пробелом или пробелом и запятой.

Шаг третий — выберите действие

Введите номер действия из представленных на экране, нажмите клавишу «Enter».

Шаг четвёртый — продолжаем работу

Вы можете продолжать применять действия друг за другом к новому тексту.

Шаг пятый — выход из приложения

Чтобы завершить работу, введите номер действия «6».

На каждом шаге вы можете выводить новый текст.

ЗАКЛЮЧЕНИЕ

Изучены новые средства работы с текстом на языке Си. Разработано интерактивное консольное приложение по техническому заданию. Программа собрана, протестирована и отлажена. Написана инструкция для пользователя.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1. Правила оформления пояснительной записки. // se.moevm.info URL: http://se.moevm.info/doku.php/courses:programming:report (дата обращения: 11.12.2021).
- 2. Функции стандартных библиотек. // wikipedia.org URL: https://ru.wikipedia.org/wiki/Wchar.h (дата обращения: 10.12.2021).
- 3. Функции стандартных библиотек и другие инструкции. // cplusplus.com URL: https://www.cplusplus.com/reference (дата обращения: 10.12.2021).

ПРИЛОЖЕНИЕ А ПРОГРАММНЫЙ КОД

Название файла «main.c»

```
#include "inputoutput.h"
#include "processing.h"
int main()
   setlocale(LC ALL,"");
   wprintf(L"Введите текст:\n");
   struct Text text=readText();
   wchar t k,c;
   remove repetes (&text);
   wprintf(L"Выберите желаемое действие:\n");
   wprintf(L"Введите 1, если хотите для каждого
предложения вывести "Палиндромы на месте!" если все слова
являются палиндромами, "Кто-то потерял все палиндромы."
если ни одного слово не является палиндромом, в остальных
случаях "Чего-то не хватает."\n");
   wprintf(L"Введите 2, если хотите заменить все
специальные символы (включая .,;) в тексте на подстроку
">special symbol<".\n");
   wprintf(L"Введите 3, если хотите отсортировать
предложения по уменьшению количества латинских букв.\n");
   wprintf(L"Введите 4, если хотите удалить все
предложения в которых в одном слове встречаются латинские
и кириллические буквы. \n");
   wprintf(L"Введите 5, если хотите вывести текст.\n");
   wprintf(L"Введите 6, если хотите завершить работу
программы. \n");
   k=getwchar();
   while (k!=L'6') {
       switch(k){
            case L'1':
                palindroms (&text);
                break;
```

```
case L'2':
                  replace(&text);
                  break;
             case L'3':
                  gsort(text.sents, text.size,
 sizeof(struct Sentence**), funcmp);
                  break;
             case L'4':
                  delete lat and kir(&text);
                  break;
             case L'5':
                  output(&text);
                  break;
             default:
                  wprintf(L"Введите число от 1 до 6\n");
                  break;
         }
         k=getwchar();
         c=qetwchar();
    }
    //for(i=0;i<text.size;i++) {</pre>
    // free(text.sents[i]->str);
    //}
    //for(i=0;i<text.size;i++) {</pre>
    // free(text.sents[i]);
    //}
 }
Название файла «inputouput.h»
#pragma once
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <locale.h>
#include <wchar.h>
#include <wctype.h>
#ifndef Structurs
```

```
#define Structurs
struct Sentence{
    wchar_t *str;
    int size;
};

struct Text{
    struct Sentence **sents;
    int size;
};

#endif

struct Text readText();

void output(struct Text* text);
```

```
Название файла « inputouput.c»
#include "inputoutput.h"
struct Text readText()
    int i, m, n;
    wchar t c;
    struct Text text;
    setlocale(LC ALL, "");
    m=0;
    text.sents=malloc(sizeof(struct Sentence*));
    if (text.sents!=NULL) {
         c=qetwchar();
         while (c!='\n') {
             n=0:
    text.sents=realloc(text.sents, (m+2) *sizeof(struct
Sentence));
             text.sents[m]=malloc(sizeof(struct Sentence));
             text.sents[m]->str=malloc(sizeof(wchar t));
             while ((c!='.') && (c!='!') && (c!='?')){
                  text.sents[m]->str=realloc(text.sents[m]-
>str,sizeof(wchar t)*(n+2));
                  text.sents[m]->str[n]=c;
                  text.sents[m]->str[n+1]='\0';
                  n++;
                  c=qetwchar();
             text.sents[m]->str=realloc(text.sents[m]-
>str,sizeof(wchar t)*(n+2));
             text.sents[m]->str[n]=c;
             text.sents[m]\rightarrowstr[n+1]='\0';
             text.sents[m]->size=n+1;
             c=getwchar();
             c=qetwchar();
             m++;
         }
    }
    else{
         wprintf(L"Недостаточно памяти\n");
    text.size=m;
    return text;
```

}

```
void output(struct Text* text)
{
    int i;
    for(i=0;i<text->size;i++) {
        wprintf(L"%ls ",text->sents[i]->str,text-
>sents[i]->size);
    }
    wprintf(L"\n");
}
```

Название файла «processing.h»

```
#pragma once
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <locale.h>
#include <wchar.h>
#include <wctype.h>
#ifndef Structurs
#define Structurs
struct Sentence{
    wchar t *str;
    int size;
};
struct Text{
    struct Sentence **sents;
    int size;
};
#endif
void delete lat and kir(struct Text * text);
int funcmp( const void * val1, const void * val2 );
void replace(struct Text *text);
void palindroms(struct Text* text);
void remove repetes(struct Text* text);
```

Название файла «processin.c»

```
#include "inputoutput.h"
struct Text readText()
    int i, m, n;
    wchar t c;
    struct Text text;
    setlocale(LC ALL, "");
    text.sents=malloc(sizeof(struct Sentence*));
    if (text.sents!=NULL) {
        c=getwchar();
        while (c!='\n') {
             n=0;
    text.sents=realloc(text.sents, (m+2) *sizeof(struct
Sentence));
             text.sents[m]=malloc(sizeof(struct Sentence));
             text.sents[m]->str=malloc(sizeof(wchar t));
             while ((c!='.') && (c!='!') && (c!='?')){
                 text.sents[m]->str=realloc(text.sents[m]-
>str,sizeof(wchar t)*(n+2));
                 text.sents[m]->str[n]=c;
                 text.sents[m]->str[n+1]='0';
                 n++;
                 c=getwchar();
             text.sents[m]->str=realloc(text.sents[m]-
>str,sizeof(wchar t)*(n+2));
             text.sents[m]->str[n]=c;
             text.sents[m]->str[n+1]='\0';
             text.sents[m]->size=n+1;
             c=getwchar();
             c=getwchar();
             m++;
         }
    }
    else{
        wprintf(L"Недостаточно памяти\n");
    text.size=m;
    return text;
```

```
}
void output(struct Text* text)
    int i;
    for(i=0;i<text->size;i++){
        wprintf(L"%ls ",text->sents[i]->str,text-
>sents[i]->size);
    wprintf(L"\n");
}
 Название файла «Makefile»
all: main.o inputoutput.o processing.o
    gcc main.o inputoutput.o processing.o
main.o: main.c inputoutput.h
    gcc -c main.c
inputoutput.o: inputoutput.c inputoutput.h
    gcc -c inputoutput.c
processing.o: processing.c processing.h
    gcc -c processing.c
clean:
    rm *.o
```