МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «ВГТУ»)

Факультет информационных технологий и компьютерной безопасности

Кафедра Систем управления и информационных технологий в строительстве

КУРСОВОЙ ПРОЕКТ

по дисциплине Основы программирования и алгоритмизации Тема: Разработка программы для работы с файловой базой данных

«Аналитика данных».

Расчетно-пояснительная записка

Разработал студент В.А.Шелеповский

Подпись, дата Инициалы, фамилия

Руководитель О.В. Минакова

Подпись, дата Инициалы, фамилия

Нормоконтролер О.В. Минакова

Подпись, дата Инициалы, фамилия

Защищена Оценка

дата

2023

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВА- ТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «ВОРОНЕЖ- СКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «ВГТУ»)

Кафедра Систем управления и информационных технологий в строительстве

ЗАДАНИЕ

на курсовой проект

по дисциплине Основы программирования и алгоритмизации

Тема: Разработка программы для работы с файловой базой данных «Аналитика данных»

Студент группы бИСТ-231 Шелеповский Виталий Андреевич

Фамилия, имя, отчество

База данных «Образовательные онлайн-курсы», Признак: Название курса, Вариант сортировки: по рейтингу (возрастание),

Технические условия Windows 10, MicrosoftVisualStudio2022, язык про- граммирования C

Содержание и объем проекта (графические работы, расчеты и прочее):

Сроки выполнения этапов анализ и постановка задачи (10.9-19.10.23); разработка пошаговой детализации программы (19.10 -11.11.23); реализация программы (11.11-5,12.23); тестирование программы (6.12-11.12.23); оформление пояснительной записки (11.12-14.12.23).

Срок защиты курсового проекта

Руководитель О.В. Минакова

Подпись, дата Инициалы, фамилия

Задание принял студент В.А.Шелеповский

Подпись, дата Инициалы, фамилия

СОДЕРЖАНИЕ

Введение……………………3

Раздел 1……………………..4

Приложение 1………………9

ВВЕДЕНИЕ

Цель: Создать базу данных «Аналитика данных» с использованием языка «си» с заданным функционалом.

Задачи:

1. Реализовать загрузку данных из файла для дальнейшей работы с ним.
2. Реализовать сортировку по Рейтингу.
3. Реализовать поиск по Версии и Разработчику.
4. Реализовать запись данных в файл.
5. Создать меню выбора опций.

РАЗДЕЛ 1

В наши дни базы данных являются неотъемлемой частью нашей жизни.

Необходимо дать определение, что такое база данных.  База данных — это набор упорядоченных и структурированных данных, которые хранятся на определённом компьютере. По сути, это большая таблица с заполненными по образцу значениями. Самые большие базы данных могут занимать миллионы гигабайт памяти.

Базы данных помогают собрать и сохранить информацию, необходимую для следования требованиям законодательства, регуляторных органов и стандартов безопасности. Они также облегчают выполнение бизнес-процессов и управление ресурсами. Фирмы могут использовать базы данных для управления продажами, запасами, производством и другими аспектами бизнеса.

Рассмотрим преимущества баз данных.

Базы данных и их системы управления заточены на работу с большим объёмом данных и от лица большого числа пользователей.

**Скорость** — ещё одно преимущество базы данных. База данных устроена так, что она легко и быстро находит, записывает, переписывает и снова находит данные. Всё потому, что СУБД всегда знает, что где лежит и по какому критерию искать. Там не будет случайных данных в случайном месте.

Скорость важна ещё и потому, что СУБД обычно обслуживает сразу много потоков: одновременно ей могут пользоваться десятки и сотни тысяч человек.

Сложность. Базы данных нужны в числе прочего для хранения сложно структурированных данных. Мы привыкли думать, что база данных — это такая таблица, где есть строки и столбцы. Но база данных при правильной организации может намного больше:

1. Связывать одну единицу данных с множеством других. Например, если один человек совершил много заказов со множеством товаров внутри каждого, база данных способна хранить и обрабатывать такие связи.
2. База может хранить дерево данных.
3. В базах могут жить ссылки на другие фрагменты и отделы базы.

Существует много разных типов баз данных. Наиболее популярные типы:

1. Реляционные базы данных
2. Key-value базы данных
3. Документно-ориентированные базы данных
4. Графовые базы данных
5. Колоночные базы данных.

Рассмотрим некоторые инструменты для аналитики данных:

-SQL.

Это специальный язык, который формирует запросы к базе данных по определённой структуре. С его помощью аналитик «достаёт» данные и затем обрабатывает их с помощью другого, более гибкого языка программирования, например Python. У SQL есть разновидности. Например, система управления базами данных MySQL, в которой можно хранить любые данные: контакты клиентов, карточки товаров, информацию о дате публикации материалов и т. д. PostgreSQL — более сложная система, которая подходит для управления большими базами данных и обработки сложных запросов, например в финансовой сфере, промышленности, крупном ретейле. SQL и его разновидности — это инструменты с открытым исходным кодом, поэтому доступны бизнесу любой сферы и формата.

-Python.

Это язык программирования и универсальный инструмент для работы с данными. У него простой синтаксис, поэтому писать код на нём получается быстрее, чем на других языках, например на Java. У Python есть встроенный интерпретатор, с помощью которого аналитики кодят и проверяют гипотезы в режиме реального времени. Чтобы пользователям было проще работать с этим инструментом, разработчики создали библиотеки — файлы с шаблонами. Они нужны, чтобы не набирать каждый раз один и тот же код, а вводить свои данные и получать результаты. Вот самые популярные библиотеки:

1. Pandas. Самая простая библиотека, которая помогает анализировать неполные и неупорядоченные данные — именно такие и встречаются чаще всего. В библиотеке Pandas есть готовые методы для фильтрации, группировки и объединения данных.
2. Matplotlib. Библиотека для визуализации, которая позволяет строить разные графики: от простых до гистограмм и круговых диаграмм.
3. NumPy. Библиотека для обработки больших массивов данных и облегчения математических операций с массивами.

-DBeaver.

Универсальный инструмент для работы с различными базами данных (MySQL, PostgreSQL, Oracle и др). Программа для аналитики DBeaver умеет:

1. администрировать базы данных: анализировать, экспортировать и импортировать данные, работать с пользовательскими сессиями;
2. визуализировать данные;
3. просматривать, редактировать, сортировать и фильтровать данные.

-Tableau, Power BI.

Информацию, которую аналитик получает в результате анализа данных, нужно наглядно представить — в виде графиков, диаграмм, отчётов. Чтобы не делать это вручную, используют инструменты визуализации. Например, с помощью Tableau можно импортировать данные из любой базы данных и представить их в нужном виде. Tableau — дорогое удовольствие и доступно не каждой компании.

-Excel.

Аналитика с помощью таблиц Microsoft всё ещё популярна в маленьких компаниях — для сбора и анализа небольшого объёма данных. А вот как инструмент визуализации Excel используют часто. Удобно экспортировать сюда результаты аналитики и строить здесь даже сложные графики и диаграммы.

В своем проекте я буду создавать базу данных записанную в файл.

В итоге должно получиться как в файле «База данных» (рис.1.3).

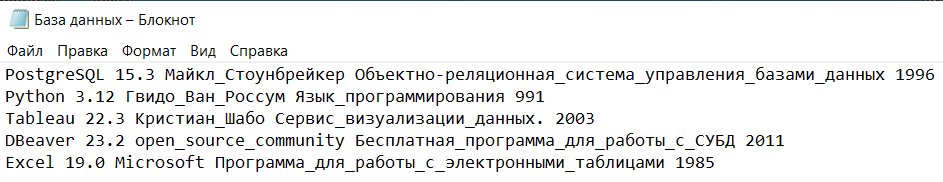


Рисунок 1.3 – Пример базы данных

Первое поле – Название инструмента, второе – Версия, третье – Разработчик, четвертое – Описание, пятое – Год создания.

Название инструмента, Разработчик и Описание – это будет тип данных указатель «char», Версия – тип данных «float», Цена – тип данных «int», т.к. для слов, состоящих из нескольких символов, требуется такой тип данных, позволяющий хранить несколько символов, ведь просто «char» позволяет хранить только один символ. Тип «float» используется для Версии, т.к. все значения дробные. Цена – это всегда целые числа. Сначала мы реализуем в программе ввод данных, представленных на рис.1.3 , в пустой файл с названием «База данных». Важно отметить, что в полях «Название инструмента», «Разработчик» и «Описание» (первый, третий и четвертый столбец файла) используется символ «\_», т.к. пробел будет использоваться в качестве разделителя данных столбцов. Это необходимо чтобы элементы поля «Преподаватель», а также поля «Наименование» считывались как единое целое.

Для хранения базы данных будем использовать структуру (struct). Поля таблицы будут полями структуры. Полю Название инструмента будет соответсвовать instrument\_name, полю Версия – version, полю Разработчик – developer, полю Описание – Description, полю Год создания – year в структуре с названием DataBase. Для того, чтобы запускать эти функции, будем записывать их в функцию main. Исходный код прототипа представлен на рисунке 1 и рисунке 2 в приложении А.

Описание модулей программы Таблица 1

|  |  |  |
| --- | --- | --- |
| **Наименование** | **Назначение** | **Примечание** |
| main | Главная функция | Выводит информацию о программе и вызывает другие функции |
| Search | Поиск данных по полям  Версия и Разработчик | Ищет нужную строчку из базы данных |
| Sort | Сортировка данных по Рейтингу | Сортирует строки базы данных |
| Check | Вывод диагностических сообщений в ходе проверки корректности ввода | Возвращает 1, если ввод данных верный, 0 – если неверный. |
| Output | Загрузка базы данных в программу из файла | Загружает базу данных в массив |
| Input | Ввод данных из программы в файл-базу данных | Загружает данные из программы в базу данных. |

Приложение А

**Прототип программы**

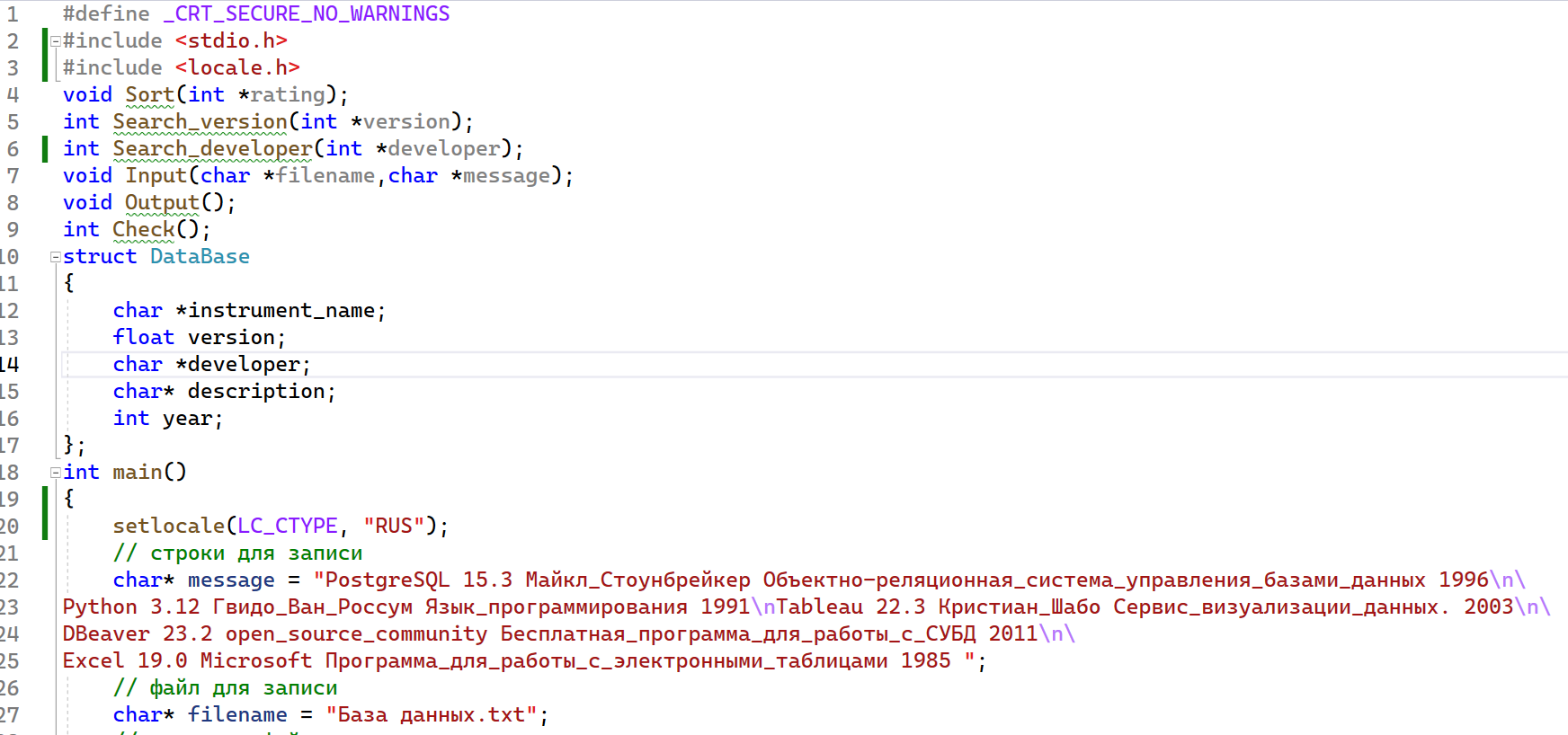


Рисунок А.1 – Прототип программы

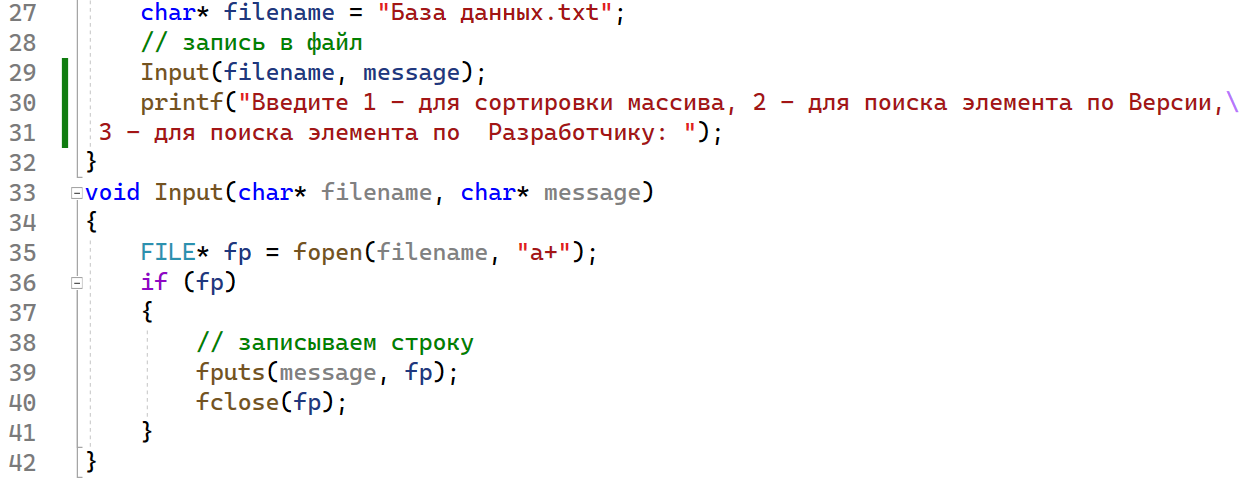


Рисунок А.2 – Прототип программы (продолжение)