



# 项目说明文档

## 数据结构课程设计

### ——关键字检索系统

作者姓名：\_\_\_\_\_陆诚彬\_\_\_\_\_

学号：\_\_\_\_\_2254321\_\_\_\_\_

指导教师：\_\_\_\_\_张颖\_\_\_\_\_

学院、专业：\_\_\_\_\_软件学院 软件工程\_\_\_\_\_

## 目录

|                         |   |
|-------------------------|---|
| 1 项目背景 .....            | 3 |
| 2 项目需求分析 .....          | 3 |
| 2.1 功能需求 .....          | 3 |
| 2.2 非功能需求 .....         | 3 |
| 2.3 项目输入输出需求 .....      | 3 |
| 2.3.1 输入格式 .....        | 3 |
| 2.3.2 输出格式 .....        | 4 |
| 2.3.3 项目示例 .....        | 4 |
| 3 项目设计 .....            | 4 |
| 3.1 数据结构设计 .....        | 4 |
| 3.1.1 文件处理 .....        | 4 |
| 3.1.2 字符串处理 .....       | 4 |
| 4 项目实施 .....            | 5 |
| 4.1 文件创建与写入实现 .....     | 5 |
| 4.1.1 文件创建与写入功能简介 ..... | 5 |
| 4.1.2 文件创建与写入核心代码 ..... | 5 |
| 4.2 关键字检索实现 .....       | 5 |
| 4.2.1 关键字检索功能简介 .....   | 5 |
| 4.2.2 关键字检索核心代码 .....   | 6 |
| 4.4 系统总体功能流程图 .....     | 7 |
| 5 设计小结 .....            | 8 |
| 5.1 功能与非功能需求: .....     | 8 |
| 5.2 数据结构与算法设计: .....    | 8 |
| 5.3 代码实现: .....         | 8 |
| 6 软件测试 .....            | 9 |
| 6.1 输入文件名及英文段落测试 .....  | 9 |
| 6.2 关键词查找测试 .....       | 9 |

# 1 项目背景

在当今信息时代，数据的获取和处理变得日益重要。文本数据，作为信息传递的一种重要形式，拥有广泛的应用领域，例如文献检索、数据挖掘、信息检索等。对文本数据进行有效的检索和分析，对于提高工作效率、加强信息管理和处理具有重要意义。关键字检索系统是一种常见的文本数据处理工具，能够快速从大量文本中找到特定的信息，大大节省了查找信息的时间和精力。

本项目旨在开发一个基础的关键字检索系统，用户能够创建一个文本文件，并输入一个关键词，系统将自动统计出这个关键词在文本文件中出现的次数。这种系统的设计和实现对于初学者理解基本的文件操作和字符串处理是非常有益的。

## 2 项目需求分析

### 2.1 功能需求

- 1) **文件创建与命名：**系统应允许用户通过键盘输入来创建一个新的文本文件。用户应能为文件自定义名称，且文件名不应包含空格。
- 2) **关键字输入：**用户输入一个不包含空格的关键字。系统需对用户输入进行检查，确保其符合格式要求。
- 3) **关键字检索与统计：**系统应对文本文件进行扫描，识别并统计关键字出现的次数。输出关键字在文本中出现的总次数。

### 2.2 非功能需求

- 1) **错误处理：**当用户输入不合规范的文件名或关键字时，系统应给出明确的错误提示，并引导用户重新输入。系统应能妥善处理文件读写中可能出现的异常。
- 2) **性能要求：**关键字的检索过程应尽可能快速，以提高用户体验。对于较大的文本文件，系统应保持稳定的运行性能。
- 3) **用户界面：**界面友好，操作简便，用户易于上手。明确的指示和提示信息，以指导用户进行操作。

### 2.3 项目输入输出需求

#### 2.3.1 输入格式

输入英文文件名、英文文字段落、英文短语。

### 2.3.2 输出格式

输出英文文字段落、搜索结果。

### 2.3.3 项目示例

```
关键字检索系统
请输入文件名: src.txt

请输入一段英文:
With over two billion pages and more being added daily, the Web is a massive col-
lection of interrelated pages. With so much available information, locating the
precise information you need can be difficult. Fortunately, a number of organiza-
tions called search services or search providers can help you locate the informa-
tion you need. They maintain huge databases relating to information provided on
the Web and the Internet. The information stored at the databases includes addre-
sses, content descriptions or classifications, and keywords appearing on Web pag-
es and other Internet informational resources.^
本段文本已保存在文本文件"src.txt"中。

请输入要检索的关键词: Web
显示源文件"src.txt" :

With over two billion pages and more being added daily, the Web is a massive col-
lection of interrelated pages. With so much available information, locating the
precise information you need can be difficult. Fortunately, a number of organiza-
tions called search services or search providers can help you locate the informa-
tion you need. They maintain huge databases relating to information provided on
the Web and the Internet. The information stored at the databases includes addre-
sses, content descriptions or classifications, and keywords appearing on Web pag-
es and other Internet informational resources.^

在源文件中共检索到: 3 个关键字 "Web"
Press any key to continue
```

## 3 项目设计

本项目的主要功能是创建文本文件、写入用户输入的文本，然后读取该文件以统计指定关键字的出现次数。基于此，项目的数据结构设计主要围绕文件处理和字符串处理展开。

### 3.1 数据结构设计

#### 3.1.1 文件处理

文件对象：

1. `ofstream outfile`: 用于写入用户输入的文本到文件。
2. `ifstream infile`: 用于读取文件内容，以便进行关键字统计。

这些文件对象对应 C++ 标准库中的文件流处理类，允许进行文件的打开、读取、写入和关闭操作。

#### 3.1.2 字符串处理

字符串变量：

1. `string filename`: 存储用户输入的文件名。
2. `string text`: 存储用户输入的文本内容。
3. `string keyword`: 存储用户输入的用于检索的关键字。
4. `string word`: 临时存储从文件中读取的单词，用于与关键字比较。

这些字符串变量是实现文本输入、存储和检索的基础。

## 4 项目实施

### 4.1 文件创建与写入实现

#### 4.1.1 文件创建与写入功能简介

##### 1) 获取文件名:

程序首先提示用户输入文件名，使用 `getline(cin, filename)` 从标准输入获取。

##### 2) 创建文件:

使用 `ofstream outfile(filename)` 创建一个输出文件流对象 `outfile`，并以用户指定的文件名打开文件。如果文件创建失败，程序将输出错误信息并返回。

##### 3) 写入用户文本:

程序接着提示用户输入一段文本。使用 `getline(cin, text)` 获取用户输入的文本，并将其写入到文件中。

#### 4.1.2 文件创建与写入核心代码

```
1.  cout << "请输入文件名: ";
2.  getline(cin, filename);
3.  ofstream outfile(filename);
4.  if (!outfile.is_open()) {
5.      // Error message if the file fails to create
6.      cerr << "文件创建失败" << endl;
7.      return 1;
8.  }
9.  cout << "请输入一段英文: ";
10. string text;
11. getline(cin, text);
12. outfile << text;
13. outfile.close();
14. cout << "本段文本已保存在文本文件" << "\"" << filename << "\"中。" << endl;
```

### 4.2 关键字检索实现

#### 4.2.1 关键字检索功能简介

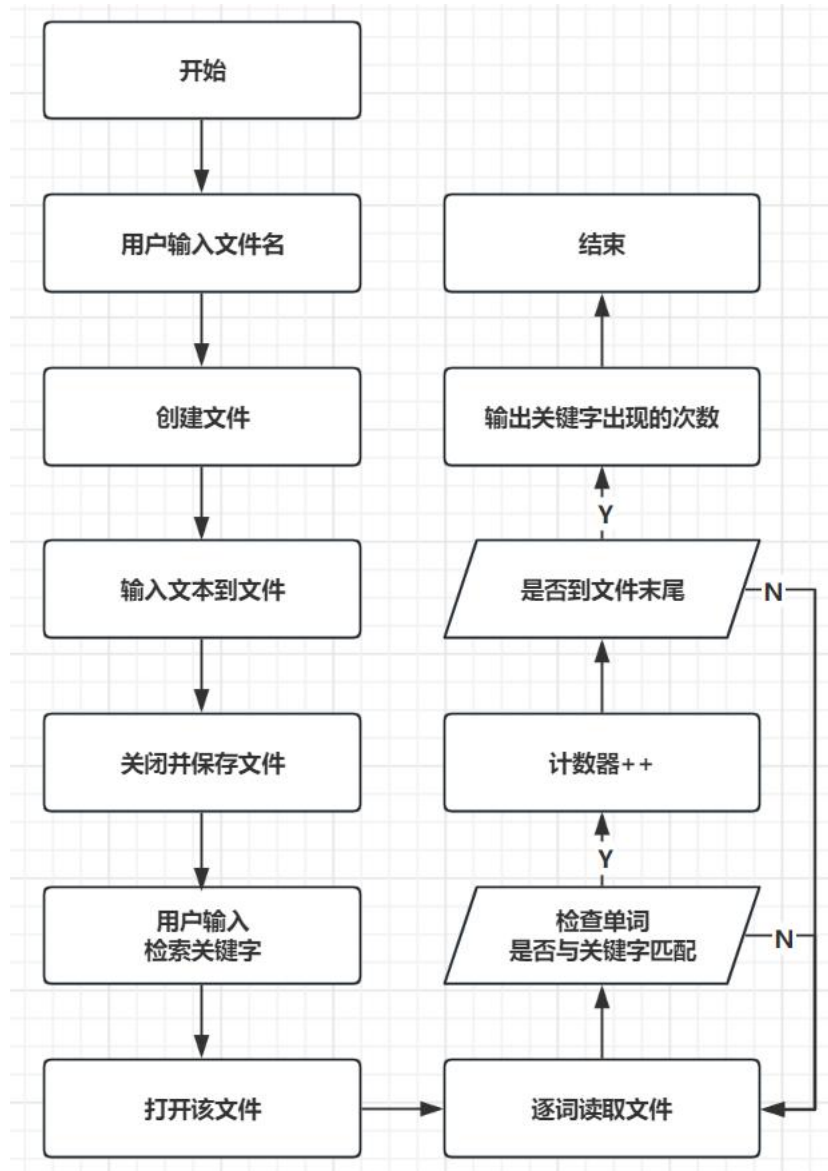
使用 `ifstream infile(filename)` 创建输入文件流对象 `infile`，并以相同的文件名打开文件。如果文件打开失败，程序将输出错误信息并返回。

**逐词检索:** 读取文件中的每个单词，并与关键字（`keyword`）进行比较。如果找到匹配的单词，计数器 `count` 递增。

#### 4.2.2 关键字检索核心代码

```
1.  ifstream infile(filename);
2.  if (!infile.is_open()) {
3.      cerr << "文件读取失败" << endl;
4.      return 1;
5.  }
6.  else {
7.      cout << text << endl << endl;
8.  }
9.  int count = 0;
10. string word;
11. while (infile >> word) {
12.     if (word == keyword) {
13.         ++count;
14.     }
15. }
16. cout << " 在源文件中共检索到 " << count << " 个关键  
    字 " << "\" " << keyword << "\" " << "。" << endl;
```

#### 4.4 系统总体功能流程图



## 5 设计小结

本项目针对的是信息时代文本数据处理的需求，特别是关键字检索方面。它解决了在大量文本中快速定位特定信息的需求，对工作效率和优化信息管理具有显著意义。

### 5.1 功能与非功能需求：

从功能需求角度看，系统包括文件创建、关键字输入和检索统计等功能。非功能需求方面，系统注重错误处理、性能优化和用户界面的友好性。这些需求确保了系统的有效性、稳定性和易用性，使之适应不同用户群体的需求。

### 5.2 数据结构与算法设计：

项目的数据结构主要围绕文件和字符串处理。使用标准库中的文件流类处理文件读写，以及通过字符串变量实现文本输入、存储和检索。算法设计上，注重高效的关键字匹配和统计，这是项目实现关键字检索功能的核心。

### 5.3 代码实现：

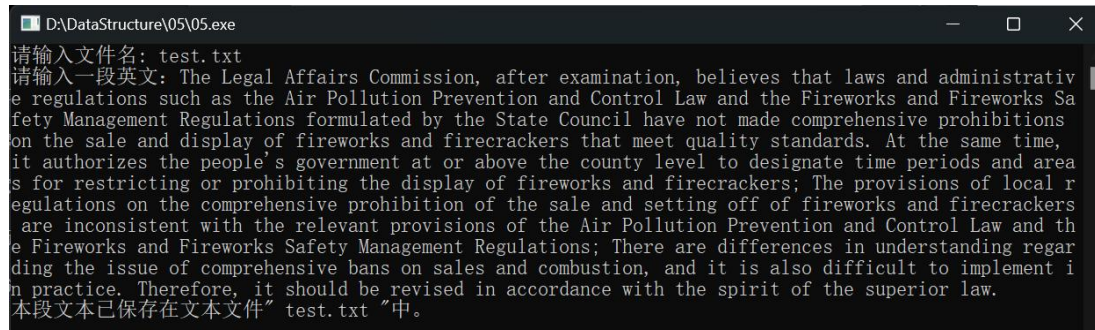
项目的代码实现分为文件创建与写入、关键字检索两部分。详细的步骤和核心代码展示了如何获取用户输入、创建和读取文件，以及如何实现关键字的有效检索。这些实现细节不仅体现了功能需求的满足，也显示了项目对错误处理和性能优化的考虑。

项目整体设计体现了从用户需求出发，通过合理的数据结构和算法设计，实现高效、稳定且用户友好的关键字检索系统。整个设计流程展示了软件开发的基本框架，包括需求分析、系统设计、代码实现和错误处理等关键环节。



## 6 软件测试

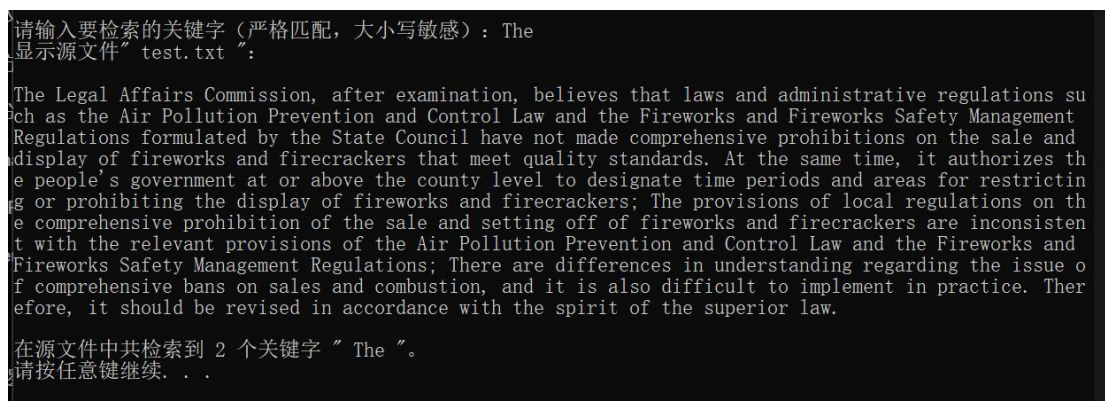
### 6.1 输入文件名及英文段落测试



```
D:\DataStructure\05\05.exe
请输入文件名: test.txt
请输入一段英文: The Legal Affairs Commission, after examination, believes that laws and administrative regulations such as the Air Pollution Prevention and Control Law and the Fireworks and Fireworks Safety Management Regulations formulated by the State Council have not made comprehensive prohibitions on the sale and display of fireworks and firecrackers that meet quality standards. At the same time, it authorizes the people's government at or above the county level to designate time periods and areas for restricting or prohibiting the display of fireworks and firecrackers; The provisions of local regulations on the comprehensive prohibition of the sale and setting off of fireworks and firecrackers are inconsistent with the relevant provisions of the Air Pollution Prevention and Control Law and the Fireworks and Fireworks Safety Management Regulations; There are differences in understanding regarding the issue of comprehensive bans on sales and combustion, and it is also difficult to implement in practice. Therefore, it should be revised in accordance with the spirit of the superior law.
本段文本已保存在文本文件" test.txt "中。
```

结论：符合输入逻辑判断

### 6.2 关键词查找测试



```
请输入要检索的关键词（严格匹配，大小写敏感）: The
显示源文件" test.txt ":
The Legal Affairs Commission, after examination, believes that laws and administrative regulations such as the Air Pollution Prevention and Control Law and the Fireworks and Fireworks Safety Management Regulations formulated by the State Council have not made comprehensive prohibitions on the sale and display of fireworks and firecrackers that meet quality standards. At the same time, it authorizes the people's government at or above the county level to designate time periods and areas for restricting or prohibiting the display of fireworks and firecrackers; The provisions of local regulations on the comprehensive prohibition of the sale and setting off of fireworks and firecrackers are inconsistent with the relevant provisions of the Air Pollution Prevention and Control Law and the Fireworks and Fireworks Safety Management Regulations; There are differences in understanding regarding the issue of comprehensive bans on sales and combustion, and it is also difficult to implement in practice. Therefore, it should be revised in accordance with the spirit of the superior law.

在源文件中共检索到 2 个关键字 " The "。
请按任意键继续. . .
```

结论：符合输出逻辑，且是所有存在的关键词。