实验3内容：文本单词统计

基本内容：

统计某篇英文小说中某些单词的出现次数和位置。英文小说存于一个文本文件中。待统计的词汇集合要一次输入完毕，即统计工作必须在程序的一次运行之后就全部完成。程序的输出结果是每个词的出现次数和出现位置所在行的行号，将输出结果保存在一个文本文件中。

测试数据：

以自己的源程序模拟英文小说，程序语言保留字集作为待统计的词汇集。实现提示：

设小说中的词汇一律不跨行。这样，每读入一行，就统计每个词在这行中的出现次数。出现位置所在行的行号可以用链表存储。若某行中出现了不止一次，不必存多个相同的行号。

算法描述：定义一个数组存单词和出现次数，分隔符分开单词，转为小写，当读取到单词时，检查是否在数组中出现，若没有，则将次数设为1，若存在，则次数+1重复直到没有单词

代码：

#include <stdio.h>

#include <stdlib.h>

#include <string.h>

#include <ctype.h>

struct Word {

char word[100]; // 单词

int count; // 出现次数

};

int main() {

char delimiters[] = " \t\n,.!?:;\"(){}[]<>";

char input\_filename[100], output\_filename[100];

printf("请输入输入文件名：");

scanf("%s", input\_filename);

printf("请输入输出文件名：");

scanf("%s", output\_filename);

FILE\* input\_file = fopen(input\_filename, "r");

FILE\* output\_file = fopen(output\_filename, "w");

if (input\_file == NULL) {

printf("错误打开文件");

exit(1);

}

struct Word words[100];

int num\_words = 0; // 存储单词的数量

char line[1000]; // 存储读入的每一行

while (fgets(line, 1000, input\_file) != NULL) {

char\* token = strtok(line, delimiters);

while (token != NULL) {

// 将单词转换为小写形式

for (int i = 0; token[i]; i++) {

token[i] = tolower(token[i]);

}

// 查找单词是否出现过

int word\_index = -1;

for (int i = 0; i < num\_words; i++) {

if (strcmp(words[i].word, token) == 0) {

word\_index = i;

break;

}

}

if (word\_index == -1) {

// 如果单词是第一次出现，则加入到 words 数组中

strncpy(words[num\_words].word, token, 100);

words[num\_words].count = 1;

num\_words++;

}

else {

// 否则增加该单词的出现次数

words[word\_index].count++;

}

token = strtok(NULL, delimiters);

}

}

// 将单词及其出现次数写入输出文件

for (int i = 0; i < num\_words; i++) {

fprintf(output\_file, "%s\t%d\n", words[i].word, words[i].count);

}

// 关闭文件

fclose(input\_file);

fclose(output\_file);

return 0;

}

输入：请输入输入文件名：data\_3input.txt（Economic development can out of energy consumption.

Nowadays,the most common sources of energy,including oil,natural gas and coal,are renewable.

That is to say the total amount of these resources is limited and can't be replenish.

So how do we maintain sustainable development?

On the one hand,renewable energy,such as wind and solar energy,needs to be exploited.

On the other hand,we should conserve energy,for example making polices to encourage to drive new energy car,reducing dependence on cars,riding or walking to work and gradually learning to live a sustainably lifestyle）

请输入输出文件名：data\_3out.txt

输出在文件中：economic 1

development 2

can 1

out 1

of 3

energy 6

consumption 1

nowadays 1

the 4

most 1

common 1

sources 1

including 1

oil 1

natural 1

gas 1

and 4

coal 1

are 1

renewable 2

that 1

is 2

to 6

say 1

total 1

amount 1

these 1

resources 1

limited 1

can't 1

be 2

replenish 1

so 1

how 1

do 1

we 2

maintain 1

sustainable 1

on 3

one 1

hand 2

such 1

as 1

wind 1

solar 1

needs 1

exploited 1

other 1

should 1

conserve 1

for 1

example 1

making 1

polices 1

encourage 1

drive 1

new 1

car 1

reducing 1

dependence 1

cars 1

riding 1

or 1

walking 1

work 1

gradually 1

learning 1

live 1

a 1

sustainably 1

lifestyle 1

学习反思：

以下都是不太懂，查资料懂的

1. strncpy复制字符串的应用
2. strtok分隔字符串的应用
3. 文件操作的函数的基础应用

一开始想不到用什么分割和什么转换为小写字母，后来一查就知道了。