信息处理技术 作业3

From: 梁鑫宇 3160104494

题目：英文自动分词，要求能处理标点符号

思路分析：

1. 语言选择

根据上课时课件中给出的算法，分析使用指针可以完成任务且较为直观，故选择c语言来完成。

2. 思路分析

分词需要辨别英文单词与各种标点符号，其中包含属于单词结构的连接符，缩写词的’ 号等以及分隔单词的分隔符。

考虑英文输入标准要求分隔单词要有空格。又有可能段尾无空格直接换行，故考虑以空格和换行作为分隔单词的标准。

由于各种标点符号作用不同，且数量较多。若依次判断太过繁琐，反向考虑通过筛选读入只保留有意义的英文单词和连接符等，其他符号直接未读入存储数组。由是则只要合理地输出所保存的数组即可。

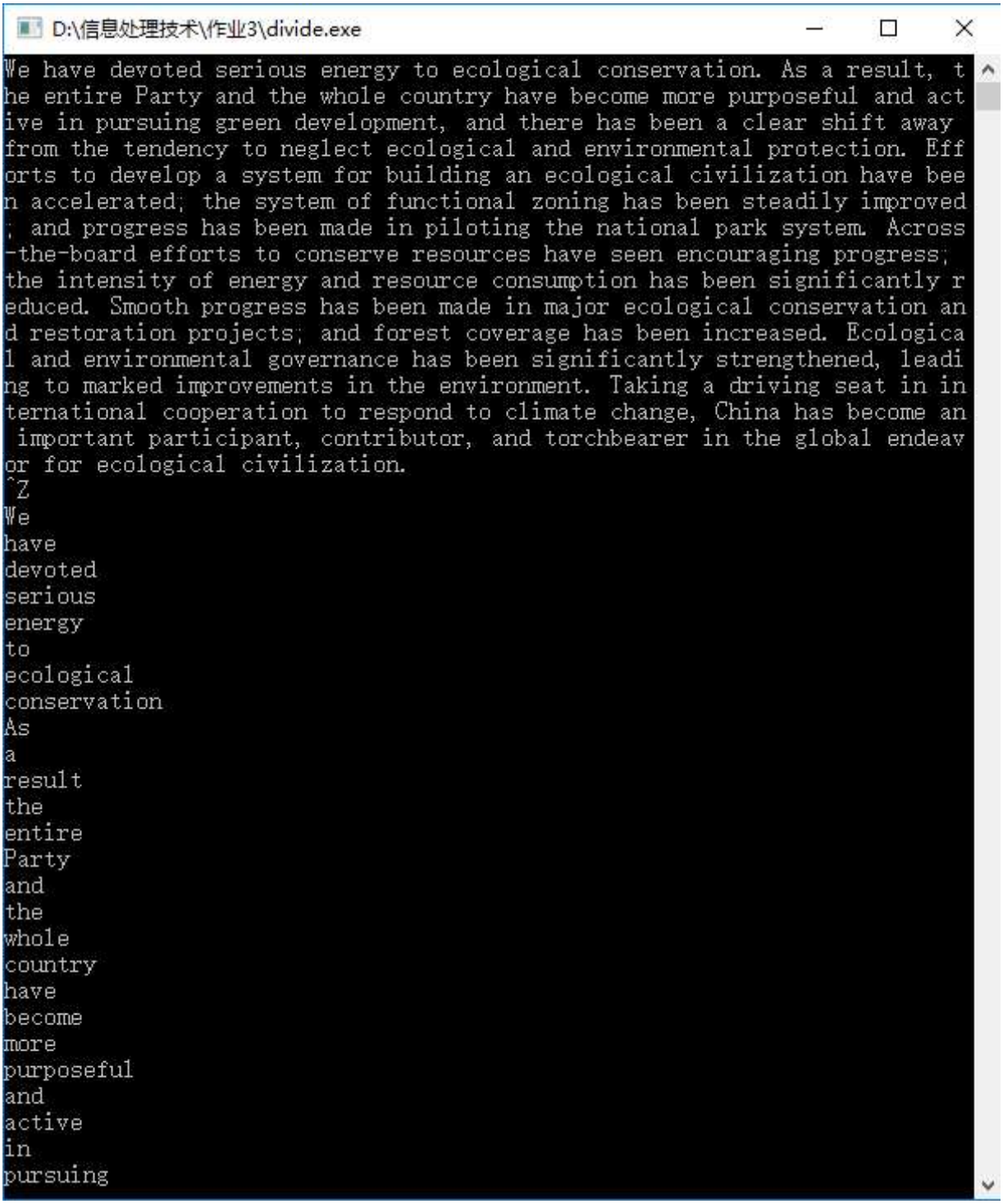
源代码：

```
#include <stdio.h>
#include <stdlib.h>

#define DEFAULTSIZE 100

void show(char* words,int size);
char* morespace(char* words,int* maxsize);

int main() {

    //通过malloc函数获取一个指针形式的字符数组
    char* words = (char*)malloc(sizeof(char)*DEFAULTSIZE);

    int size = 0;//记录字符数组目前存储了多少字符

    int maxsize = DEFAULTSIZE;//记录目前字符数组的大小

    char ch;

    //当用户没有使用EOF结束输入时进入如下循环
    while(scanf("%c",&ch) != -1){
        //由于任务目标是英文分词，此处只将英文字符，可能出现的连接符以及空格和换行读入字符数组
        if(ch == ' ' || (ch>=65&&ch<=122) || ch == '-' || ch == '\'' || ch == '\n'){
            words[size] = ch;
            ++size;
        }
        //每读入一个字符，判断一下数组是否满了。如果已经满了，调用morespace函数
```

```
        if(size == maxsize){
            words = morespace(words,&maxsize);
        }
    }

    show(words,size);//输出结果

    free(words);//释放内存

    return 0;
}

//输出分词结果，每个单词一行
void show(char* words,int size){

    int new_line = 0; //记录是否已经换过行，可处理掉输入中多余的空行

    for(int i = 0;i<size-1;++i){
        if(words[i] == ' ' || words[i] == '\n'){
            //如果此前输出字符还没有换过行，那么输出换行符分隔此单词
            if(new_line == 0){
                printf("\n");
                new_line = 1;//并将此值标记为1
            }
        }
        //若不是分隔符空格或换行，则为单词的组成部分，正常输出
        else{
            printf("%c",words[i]);
            new_line = 0;//此时为一个单词的字符逐个输出，标记换行次数为0
        }
    }
}

//获取更大空间的字符数组
char* morespace(char* words,int* maxsize){

    //通过malloc函数获得一个更大的字符数组
    char* new_words = (char*)malloc(sizeof(char)*(*maxsize+DEFAULTSIZE));

    //利用指针操作修改maxsize的值
    *maxsize = *maxsize + DEFAULTSIZE;

    //将原来字符数组里的字符逐个赋到新的数组中
    for(int i = 0;i<*maxsize;++i){
        new_words[i] = words[i];
    }

    //返回新的字符数组
    return new_words;
}
```

测试样例

测试文本：十九大报告英文版节选


D:\信息处理技术\作业3\divide.exe
We have devoted serious energy to ecological conservation. As a result, t
he entire Party and the whole country have become more purposeful and act
ive in pursuing green development, and there has been a clear shift away
from the tendency to neglect ecological and environmental protection. Eff
orts to develop a system for building an ecological civilization have bee
n accelerated; the system of functional zoning has been steadily improved
; and progress has been made in piloting the national park system. Across
-the-board efforts to conserve resources have seen encouraging progress;
the intensity of energy and resource consumption has been significantly r
educed. Smooth progress has been made in major ecological conservation an
d restoration projects; and forest coverage has been increased. Ecologica
l and environmental governance has been significantly strengthened, leadi
ng to marked improvements in the environment. Taking a driving seat in in
ternational cooperation to respond to climate change, China has become an
important participant, contributor, and torchbearer in the global endeav
or for ecological civilization.
^Z
We
have
devoted
serious
energy
to
ecological
conservation
As
a
result
the
entire
Party
and
the
whole
country
have
become
more
purposeful
and
active
in
pursuing

包含连字符的单词


D:\信息处理技术\作业3\divide.exe
and
progress
has
been
made
in
piloting
the
national
park
system
Across-the-board
efforts
to
conserve
resources
have
seen
encouraging
progress
the
intensity
of
energy
and
resource
consumption
has
been
significantly
reduced
Smooth
progress
has
been
made
in
major
ecological
conservation
and
restoration
projects