数据结构实践课报告

题目：文本格式化

班级：17052317 姓名：於文卓 学号17061833 完成日期：2018年11月14日

1. **需求分析**

试写一个文本格式化程序可以处理一个输入文件，它按照用户指定的下列版面规格参数重排版面：实现页内调整、分段、分页等文本处理功能，排版结果存入输出文本文件中。

（1）输出文件中字与字之间只留一个空格符，即实现多余空格符的压缩。

（2）输出文件中，任何完整的字不能分割在两行，行尾可以不齐，但行首要对齐。

（3）页底空行书不少于3行则将页号印在页底空行中第2行的中间位置上。否则不印

（4）自行设计测试用例

(5) 参数包括

1.每页内文字的行数

2.每页内文字所占最大字符数

3.每页文字前的固定空格数

4.每页页顶所空行数

5.每页页底所空行数

6.起始页号

**概要设计**

1. **主要算法介绍：**
2. 去掉多余的空格：
3. 标点符号后的空格
4. 页脚
5. 左对齐
6. **部分算法实现代码**

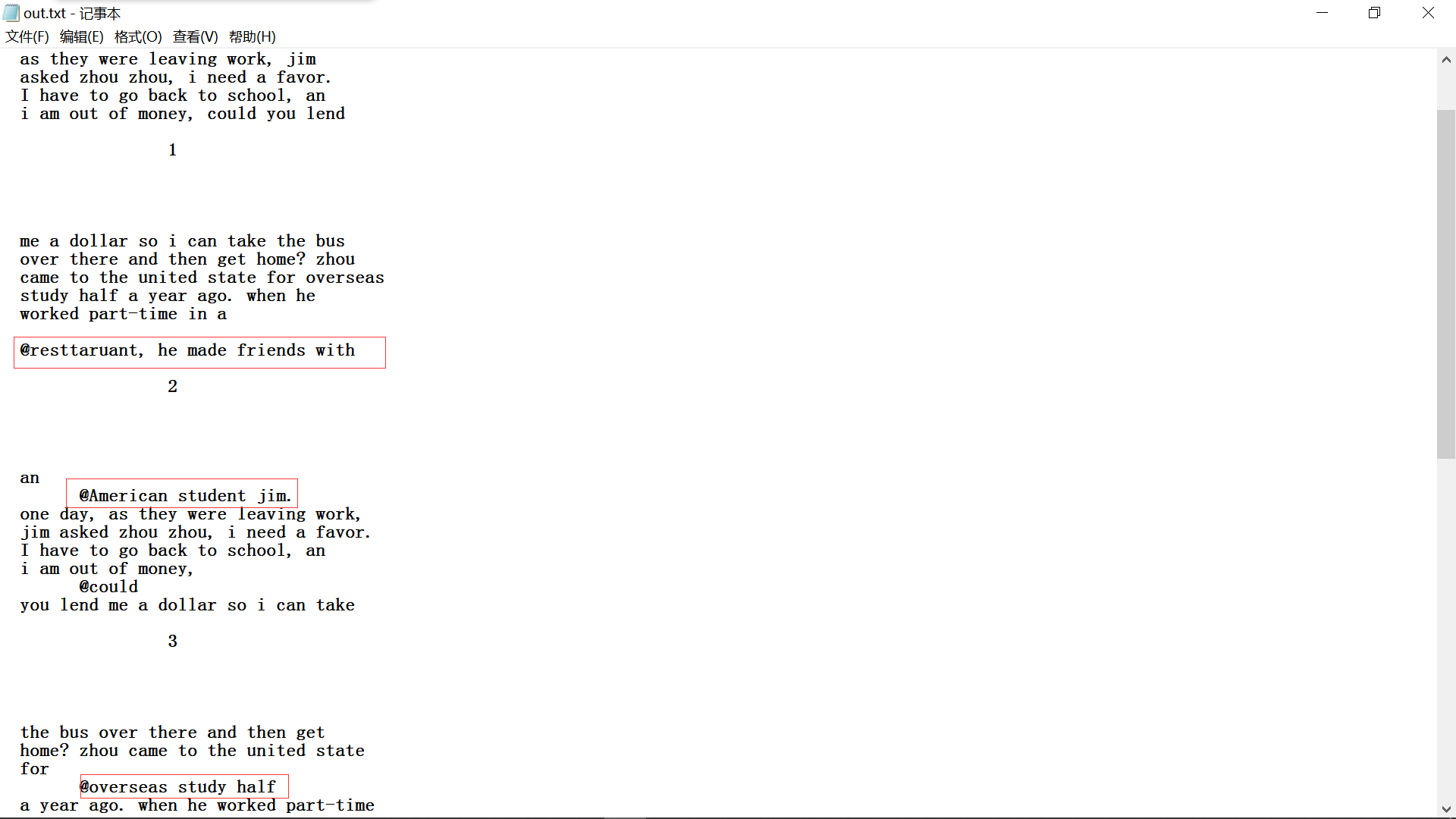
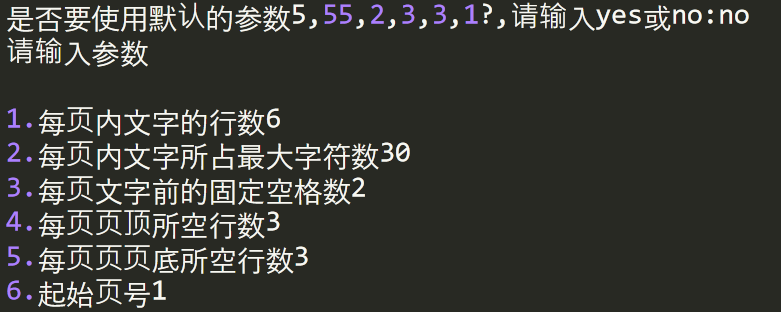
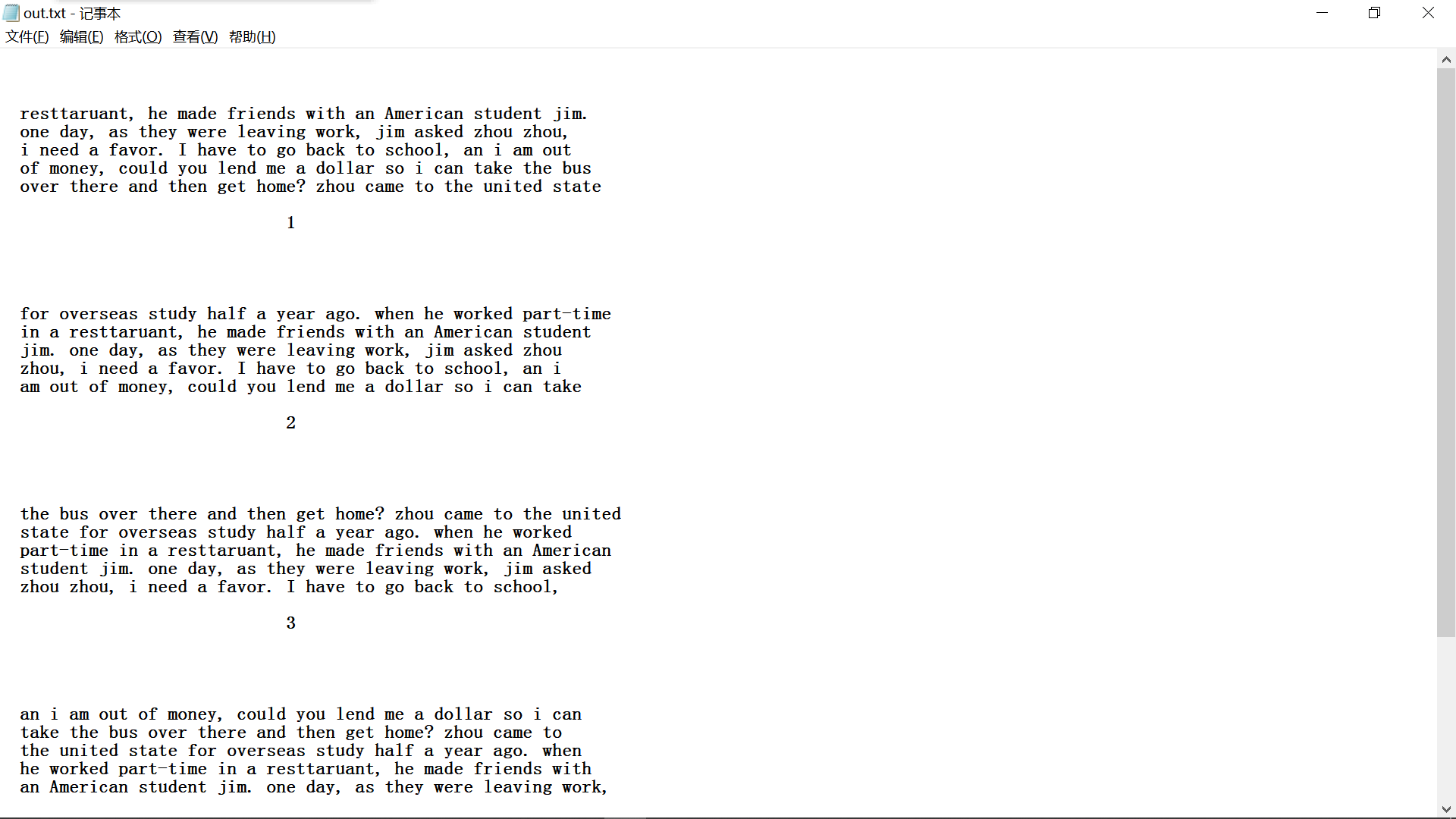
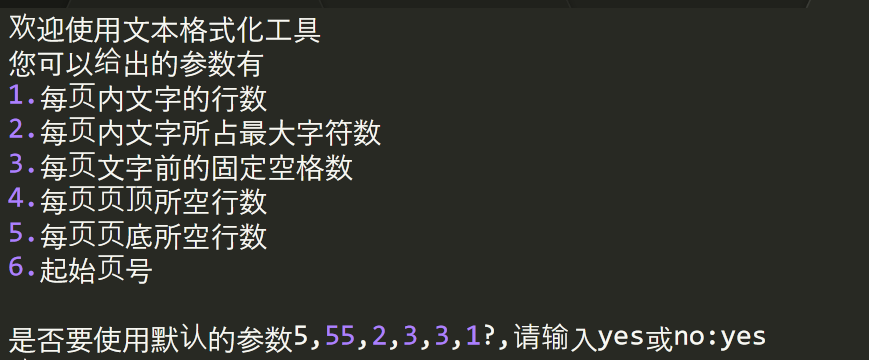
|  |
| --- |
| def preprocess(article):  article = article.strip()  article = article.replace(",", ", ")  article = article.replace(" ,", ",")  article = article.replace(".", ". ")  article = article.replace(" .", ".")  article = article.replace("?", "? ")  article = article.replace(" ?", "?")  return article |
| |  | | --- | | word\_list = article.split()#分割每个单词，形成列表    str\_info = " ".join(word\_list)#目的是去掉连续的空格 |  |  | | --- | | for i in range(footing\_len):#底部空格  if footing\_len >=3 :  if i==1:  print(" "\*((word\_in\_line+margin)//2),str(start\_page\_num),end = "",file = f)  print("\n",end = "",file = f)    start\_page\_num+=1 | |

|  |
| --- |
| **print(" "\*margin,end="",file = f)#每一行开头的空格**  **print(line,file = f)#打印一页** |

1. **调试分析**
2. **考虑到python对字符串操作的方便，这里选用python来实现功能**
3. **最后剩下一行字符串如果上一页没有满，则在上一页中，如果上一页满了，就起新的一页**
4. **用户手册**
5. **本程序运行环境为win10操作系统，执行文件为文本格式化.py**
6. **用户界面及测试数据**

**原文为**

|  |
| --- |
| **resttaruant, he made friends with an American student jim.one day ,as they were leaving work,jim asked zhou zhou ,i need a favor . I have to go back to school,an i am out of money ,could you lend me a dollar so i can take the bus over there and then get home ?zhou came to the united state for overseas study half a year ago .when he worked part-time in a @resttaruant, he made friends with an @American student jim.one day ,as they were leaving work,jim asked zhou zhou ,i need a favor . I have to go back to school,an i am out of money ,@could you lend me a dollar so i can take the bus over there and then get home ?zhou came to the united state for @overseas study half a year ago .when he worked part-time in a resttaruant, he made friends with an American student jim.one day ,as they were leaving work,jim asked zhou zhou ,i need a favor . I have to go back to school,an i am out of money ,could you lend me a dollar so i can take the bus over there and then get home ?zhou came to the united state for overseas study half a year ago .when he worked part-time in a resttaruant, he made friends with an American student jim.one day ,as they were leaving work,jim asked zhou zhou ,i need a favor . I have to go back to school,an i am out of money ,could you lend me a dollar so i can take the bus over there and then get home ?** |

****

1. **附件源代码**

*# -\*- coding: utf-8 -\*-*

*# @Author: TD21forever*

*# @Date: 2018-11-14 15:41:57*

*# @Last Modified by: TD21forever*

*# @Last Modified time: 2018-11-15 01:57:20*

file = open('input.txt','r')*#读取文件*

*#预处理 传入字符串*

**def** preprocess(article):

article = article.strip()

article = article.replace(",", ", ")

article = article.replace(" ,", ",")

article = article.replace(".", ". ")

article = article.replace(" .", ".")

article = article.replace("?", "? ")

article = article.replace(" ?", "?")

**return** article

**def** operate(line\_num=5,word\_in\_line=55,margin=2,heading\_len=3,footing\_len=3,start\_page\_num=1,file = file):

flag = 0

article = file.read()*#读到文件里的字符串*

file.close()

f = open ('out.txt','a')

article = preprocess(article)

word\_list = article.split()*#分割每个单词，形成列表*

str\_info = " ".join(word\_list)*#目的是去掉连续的空格*

str\_info = str\_info.replace("@", "**\n** @")

start = 0

end = word\_in\_line

**while** end < len(str\_info):

**for** i **in** range(heading\_len):*#顶部的空格*

print("**\n**",end="",file = f)

**for** one **in** range(line\_num):*#每一行*

line = str\_info[start:end]

temp = end

**if** end<=len(str\_info):

**if** str\_info[temp-1] != " " **or** str\_info[temp-1] **not** **in** word\_list:*#如果一行的最后一个不是空格说明那个单词被拆开了*

*# 另一个条件是防止出现as被分开的这种情况*

**while** str\_info[temp] != " ":*#下一行的第一个字母不是空格，就把这个字母加到上一行的末尾*

line = line + (str\_info[temp])

temp+=1

line = line + (str\_info[temp])*#temp最后移到空格，空格放在上一行的末尾*

end = temp+1

print(" "\*margin,end="",file = f)*#每一行开头的空格*

print(line,file = f)*#打印一页*

start = end

end+=word\_in\_line

**elif** one+1 < line\_num:*#如果每页的行数还没有得到要求*

line = str\_info[start:]

print(" "\*margin,end="",file=f)*#每一行开头的空格*

print(line,file = f)*#打印一页*

flag = 1

**for** i **in** range(footing\_len):*#底部空格*

**if** footing\_len >=3 :

**if** i==1:

print(" "\*((word\_in\_line+margin)//2),str(start\_page\_num),end = "",file = f)

print("**\n**",end = "",file = f)

**break**

**else**:*#如果这一页的行数已经达到了，那就另起一页*

**for** i **in** range(footing\_len):*#底部空格*

**if** footing\_len >=3 :

**if** i==1:

print(" "\*((word\_in\_line+margin)//2),str(start\_page\_num),end = "",file = f)

print("**\n**",end = "",file = f)

**for** i **in** range(heading\_len):*#顶部的空格*

print("**\n**",end="",file = f)

line = str\_info[start:]

print(" "\*margin,end="",file = f)*#每一行开头的空格*

print(line,file = f)*#打印一页*

**for** i **in** range(footing\_len):*#底部空格*

**if** footing\_len >=3 :

**if** i==1:

print(" "\*((word\_in\_line+margin)//2),str(start\_page\_num+1),end = "",file = f)

print("**\n**",end = "",file = f)

flag = 1

**if** flag == 1:

**break**

**for** i **in** range(footing\_len):*#底部空格*

**if** footing\_len >=3 :

**if** i==1:

print(" "\*((word\_in\_line+margin)//2),str(start\_page\_num),end = "",file = f)

print("**\n**",end = "",file = f)

start\_page\_num+=1

**if** \_\_name\_\_ == '\_\_main\_\_':

**while** **True**:

print("欢迎使用文本格式化工具")

print("您可以给出的参数有**\n**1.每页内文字的行数**\n**2.每页内文字所占最大字符数**\n**3.每页文字前的固定空格数**\n**4.每页页顶所空行数**\n**5.每页页底所空行数**\n**6.起始页号**\n**")

ans = "no"

ans = input("是否要使用默认的参数5,55,2,3,3,1?,请输入yes或no:")

**if** ans == 'yes':

operate()

**else**:

print("请输入参数**\n**")

a = int(input("1.每页内文字的行数"))

b = int(input("2.每页内文字所占最大字符数"))

**if** b>80:

b = int(input("每页内文字所占最大字符数小于80，请重新输入："))

c = int(input("3.每页文字前的固定空格数"))

d = int(input("4.每页页顶所空行数"))

e = int(input("5.每页页页底所空行数"))

ff = int(input("6.起始页号"))

operate(a,b,c,d,e,ff)

ff.close()

input.txt

out.txt