编译原理报告

姓名	学号	职务
郑书桦	2019611077	组长
郑梦宜	2019611076	组员
张菊敏	2019611072	组员
郭清扬	2019611021	组员
张国润	2019611071	组员

1. 目标

①构造一个能够解析 SRT 文件的字幕分析器,包括以下内容:

- 分析字幕的词法和语法,分别写出词法和语法的文法
- 在文法的基础上设计词法分析器和语法分析器
- 建立内存数据结构,缓冲字幕
- 成功读取 srt 文件并成功打印文件内容
- 实现整体字幕的前进和后退
- 实现单条字幕的前进和后退 (辐射影响时间在其之后的字幕)
- 实现按照字幕序号 Sid 查找某条字幕
- 实现按照字幕时间点查找某条字幕
- 实现保存和更新字幕文件功能
- 输入和运行异常时会报相应的错误
- 该字幕分析器界面清晰、简洁,操作方便友好

②通过小组合作完成该项目学习和提高以下的能力:

- 深入了解了编译原理中词法和语法分析的具体原理
- 通过团队配合和磨合,提高团队合作能力
- 提高编写 java 代码的能力

- 提高调试和测试程序的能力
- 提高搜索资料的能力
- 提高自身的抗压能力
- 提高解决问题的能力

2. 内容

2.1 字幕的词法和语法的分析

2.1.1 字幕的词法分析

词法分析的定义:编译器扫描源文件的字符流,过滤掉字符流中的空格、注释等,并将其分割为一个个的词。

> 字幕元素格式:

srt 字幕文件的格式为:

0

00:00:01,000 --> 00:00:25,000

English subtitle by: Eduun

1

00:00:36,700 --> 00:00:38,700

The late fourth century A.D. the

Roman Empire began to crumble.

字幕元素

字幕 Sid: 通过字幕 Sid 可唯一确定该字幕,字幕 Sid 是顺序的,可通过 Sid 知道字幕总

数

开始时间(小时:分钟:秒钟,毫秒数)--> 结束时间(小时:分钟:秒钟,毫秒数)

字幕内容(有的是一行有的是多行)

空行 (用于区分字幕)

> 词法分析思路:

- 1) 利用空行切分字幕文件为单条字幕,并存储在 ArrayList 中 ([sid, time, content])。
- 2) 针对字幕时间格式,将字幕时间进行切分,切分为:时/分/秒/毫秒。
- 3) 将处理好的数据存放进字幕节点 SrtNode 中 {sid, startTime, endTime, content}。

2.1.2 字幕的语法分析

语法分析的定义:语法分析是编译过程的一个逻辑阶段。语法分析的任务是在词法分析的基础上将单词序列组合成各类语法短语,如"程序","语句","表达式"等等。语法分析程序判断源程序在结构上是否正确,源程序的结构由上下文无关文法描述。

> 语法分析思路:

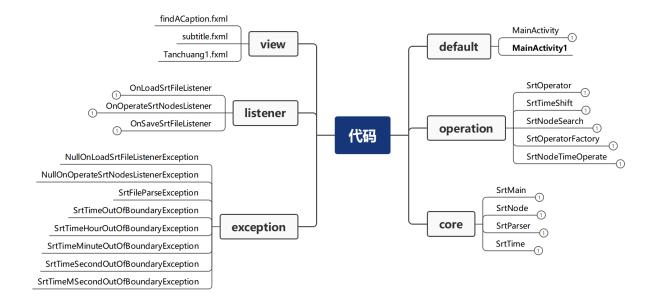
本项目的语法分析主要用于将字幕节点格式的数据(即一个个的"词")重新拼接转换成字幕 文件可以识别的"句子"。

- 1)针对字幕时间格式,对字幕节点中的 `startTime` 和 `endTime` 进行处理,处理成 "时:分:秒,毫秒" --> "时:分:秒,毫秒" 的形式,即 srt 格式的时间。
- 2) 将字幕节点中的 `sid` 和 `content` 与 上一点得到的时间组合起来,得到一条完整的 srt 文件可识别的字幕。

2.2 代码结构

代码说明: 代码主要分成 6 个部分,default 包放的是主类,operation 包里面存放的是这个字幕分析器功能的操作类,core 包存放的是 srt 文件结构类以及 srt 文件的解析类,listener 包放的是监听器的设置,exception 包放的是程序异常报错设置,第二个版本的字幕分析器加了个 view 包,然后这个 view 包放置的是图形用户界面的设计。

2.2.1 代码结构思维导图



2.2.2 代码包内相应类内函数说明

core 包		
类名	类内部的实现和类内部函数说明	
对 srt 文件进行解析以及做各种操作的统领类 SrtMain.java		

	③loadSrtFileFail():加载 srt 文件失败报错
	④getSrtFilePath():返回当前 srt 文件路径
	⑤setSrtFilePath(srtFilePath):设置当前 srt 文件路径
	⑥getOnLoadSrtFileListener():返回当前的 srt 文件监听器
	⑦setOnLoadSrtFileListener():设置当前的 srt 文件监听器
	⑧getSrtNodeListSize():返回链表长度
	⑨printSrtNodeList():打印 srt 文件所有结点
	⑩printSrtNode(node):打印 srt 文件某个结点
	①convertString(node):将单个结点转成字符型
	②convertString(list):将全部结点转成字符型
	③operateSrtNodes(opName,parameter,listener):设置操作
	④saveSrtFile(filePath,listener):将当前状态的 srt 文件内容保存到原
	文件中
	⑤loadSrtFile(path):根据路径导入 srt 文件
	字幕节点结构的定义以及相关的函数操作
6.00.1.	①定义字幕结点(SrtNode)结构: sid, begin(SrtTime),
SrtNode.java	end(SrtTime), content
	②对 sid,begin,end,content 变量提供 set()和 get()方法
	字幕时间结构的定义以及相关的函数的操作
	①定义字幕时间(SrtTime)结构:hour、minute、second、
SrtTime.java	msecond
Si trime.java	②设定了可输入的时间范围,例如 MAX_HOUR=99
	③对 hour、minute、second、msecond 变量提供 set()和 get()方
	法

	④提供将 SrtTime 转换成 String 的 toString()方法 ⑤提供将 msecond 转换成 hour、minute、second 的方法 asMsecondInt()
SrtParser.java	解析字幕文件:将 srt 文件转换成链表存储,相当于语义分析,先将整个字幕文件划分成一条条独立的字幕,再对每一条字幕进行 id、时间、内容上的切分,并将其存储在结点中。 ①void execute(file,listener):读取 srt 文件,用 ArrayList 将每一条字幕进行存放,每一条字幕(sid,times,content)对应一个结点 node,将 String 转换成 SrtNode (parseStrNode(node)),然后将所有结点串成链表。 ②SrtNode parseStyToSrtNode(strNode):如果结点长度大于 2 则: node.setSid(sid):将字幕 id 放在结点的 sid 里面 node.setBegin(begin):将字幕开始时间放在结点的 begin 里面 node.setEnd(end):将字幕结束时间放在结点的 end 里面 node.setContent(content):将字幕具体文本内容放在结点的 content 里面 ③其中时间分隔步骤为: ">" 分隔开始时间和结束时间 ":" 分隔时、分、秒 "," 分隔秒和毫秒
	operation 包
SrtOperator.java	①SrtOperator(parameters):this.parameters=parameters ②execute(srtNodes,onOperateSrtNodesListener)

	③setParameter(name,value):this.parameters.put(name,value) ④getParameter(name,value):this.parameters.get(name,value)
SrtTimeShift.java	实现时间轴平移 ①SrtTimeShift(parameters):如果 parameters 获取 shiftType、srtMsecond、srartNode 中存在不空,则将不空的相应的值赋给当前类生命的 shiftType、shiftMsecond、startNode。 ②execute(srtNodes,onOperateSrtNodesListener):doTimeShift() ③doTimeShift(srtNodes,onOperateSrtNodesListener):实现所有字幕平移 ④doTimeShift(srtNodes,startNode,onOperateSrtNodesListener) :实现单个字幕的平移
SrtNodeSearch.jav a	根据 sid 或者时间查找字幕 ①SrtNodeSearch(parameters): 如果 sid 不空,获取 sid; 如果 hour 不空,获取 hour; 如果 minute 不空,获取 minute; 如果 second 不空,获取 second; 如果 msecond 不空,获取 second; ②execute(srtNodes,onOperateSrtNodesListener): 如果 sid 不为空则遍历链表查找该字幕结点,如果没有找到则报错; sid 为空则按照输入的时间进行查找,没有找到则报错。

	获取对应 srt 文件操作功能		
SrtOperatorFactor	getSrtOperator(name,parameters):		
y.java	如果 name= "SrtTimeShift",则时间轴平移;		
	如果 name= "SrtNodeSearch",则进行结点搜索,否则报错。		
	时间轴平移的操作函数		
	①plusBoth(node,time):所有字幕向前移进 time 时间		
	②plusBoth(node,mSecond):单条字幕向前移进 mSecond		
SrtNodeTimeOper ate.java	③minusBoth(node,time):所有字幕回退 time 时间		
ate.java	④minusBoth(node,mSecond):单条字幕后退 mSecond		
	⑤isGreaterThan(SrtTime1,SrtTime2):判断 SrtTime1 > SrtTime2?		
	⑥isIn(begin,end,SrtTime):判断输入移动的时间是否在范围内		
	listener 包		
	定义 OnLoadSrtFileListener 接口,该接口包含的抽象方法有:		
OnLoadSrtFileListe	①onLoadSrtFileStart(): 开始读取 srt 文件		
ner.java	②onLoadSrtFileSuccess(list):成功读取 srt 文件		
	③onLoadSrtFileFail(Exception):读取 srt 文件失败		
	定义 OnOperateSrtNodesListener 接口,该接口包含的抽象方法		
	有:		
OnOperateSrtNod	①onOperationStart():开始操作		
esListener.java	②onOperationSuccess(node):对某个结点操作成功		
	③onOperationSuccess(SrtNode):对所有结点操作成功		
	④onOperationSuccess(Exception):操作失败		

OnSaveSrtFileListe ner.java

定义 OnSaveSrtFileListener 接口,该接口包含的抽象方法有:

- ①onSaveSrtFileSuccess():保存 srt 文件成功
- ②onSaveSrtFileFail(Exception):保存 srt 文件失败

exception 包

- ①NullOnLoadSrtFileListenerException:没有成功导入 srt 文件监听器时会报此错误
- ②NullOnOperateSrtNodesListenerException: 没有成功导入 srt 结点监听器时会报此错误
- ③SrtFileParseException: srt 文件解析异常时会报此错误

(1) main():

- ④SrtTimeOutOfBoundaryException:输入时间异常时会报此错误
- ⑤SrtTimeHourOutOfBoundaryException:输入的小时不在范围时会报此错误
- ⑥SrtTimeMinuteOutOfBoundaryException:輸入的分钟不在范围时会报此错误
- ⑦SrtTimeSecondOutOfBoundaryException:输入的秒不在范围时会报此错误
- ⑧SrtTimeMSecondOutOfBoundaryException: 输入的毫秒不在范围时会报此错误

default 包

MainActivity.java (第一版:不含前端 界面)

1.srtMain.setOnLoadSrtFileListener{

onloadSrtFileSuccess(list):提示读取 srt 文件成功并打印字幕总数; onloadSrtFileFail(Exception):提示读取 srt 文件失败。 }

- 2.srtMain.loadSrtFile(scanInFilePath());
- ②doLoop(srtMain):
- 1.整体字幕进行移动 2.查找某一条的字幕 3.更改读取字幕文件
- 4.保存字幕文件修改 5.打印当前字幕内容 6.退出当前所在程序

	否则报错输入非法。		
	③打印一些提示信息的函数:		
	printAvailableOperator();		
	printAvailableSingleNodeOperator();		
	printAvailableSearchOperator();		
	scanInFilePath();		
	scanInNodeSearchSid();		
	scanInNodeSearchSrtTime();		
	scanInTimeShiftParam();		
	1 句今第一版的 Main Activity iava 包由低有功能		
	1、包含第一版的 MainActivity.java 包中所有功能		
	2、前端监听的方法:		
	(1) scanInFilePath():接受前端输入的文件路径,调用后端的打印 方法。		
	クi法。 (2) saveFile(): 保存整个字幕文件。		
MainActivity1.java	(3) tanChuang1(): 弹出整体字幕时间轴平移的界面。		
(第二版:包含前端	(4) moveOverall(): 对字幕时间轴平移界面输入的监听。		
界面)	(5) findACaption(): 弹出查找某一条字幕的界面。		
	(6) findBySid(): 获取输入的 SID, 查找字幕并打印结果。		
	(7) findByTime(): 获取输入的时间,查找字幕并打印结果。		
	(8) moveACaption(): 获取移动某一条字幕的时间。		
	(9) goHome(): 退回主界面。		
	(10) exitAndClose(): 退出程序,关闭窗口。		
	view 包		

view 包

subtitle.fxml	该页面是字幕分析器的主页面,有相应的功能按键。
Tanchuang1.fxml	该页面是点击整体字幕移动按钮时出现的弹窗页面设计。
findACaption.fxml	该页面是点击查找某条字幕按钮后出现的页面。

2.2.3 关键代码以及注释

> 关键结构体

```
class SrtTime{
                    //字幕时间
int hour;
                     //小时
int minute;
                     //分钟
int second;
                     //秒
int msecond:
                     //毫秒
                     //字幕结点
class SrtNode{
int sid;
                     //字幕序号
SrtTime begin;
                     //字幕开始时间
                     //字幕结束时间
SrtTime end;
String content;
                    //字幕内容
}
> 对字幕进行语法分析
//将开始时间与结束时间分隔开
String[] timeStr = strNode.get(1).split("-->");
//将开始时间下的时、分、秒、毫秒分隔开
String[] beginStr = timeStr[0].split(":");
//将结束时间下的时、分、秒、毫秒分隔开
String[] endStr = timeStr[1].split(":");
begin.setHour(Integer.parseInt(beginStr[0].trim()));
begin.setMinute(Integer.parseInt(beginStr[1].trim()));
//分隔秒和毫秒
begin.setSecond(Integer.parseInt(beginStr[2].split(",")[0].trim()));
begin.setMsecond(Integer.parseInt(beginStr[2].split(",")[1].trim()));
end.setHour(Integer.parseInt(endStr[0].trim()));
end.setMinute(Integer.parseInt(endStr[1].trim()));
//分隔秒和毫秒
end.setSecond(Integer.parseInt(endStr[2].split(",")[0].trim()));
end.setMsecond(Integer.parseInt(endStr[2].split(",")[1].trim()));
node.setSid(Integer.parseInt(strNode.get(0).trim()));
node.setContent(content);
node.setBegin(begin);
node.setEnd(end);
```

▶ 读取 srt 文件,将每一条字幕用 ArrayList 形式存放,再转换成 SrtNode 形式

```
while (input.hasNextLine()){
   ArrayList<String> node = new ArrayList<>();
   String str = input.nextLine();
   //一条字幕(sid、times、content)对应一个 node
    while(!str.isEmpty()){
       node.add(str);
       if(input.hasNextLine()){
            str = input.nextLine();
         }else{
             break;
               }
     //将 String 转换成 SrtNode
     SrtNode srtNode = parseStrToSrtNode(node);
     //将所有字幕结点串成链表
     srtNodeList.add(srtNode);
}
```

2.4 运行结果和说明

2.4.1 初始版本运行结果和说明

> 打印组员信息

 - -	编译原	理+字幕分	 ↑析器	
	小	组成员		
 郑书桦、	郑梦宜、	张菊敏、	郭清扬、	张国润

输入 srt 文件路径并输出字幕总条数

请输入srt文件的绝对路径:

D:\Study\Junior\Junior_1\编译原理\三级项目\字幕分析器\subtitle.srt

读取srt文件成功提示: 该文件一共有 1309 条字幕

> 实现打印当前字幕内容的功能

```
1.整体字幕进行移动
                 2. 查找某一条的字幕
                 3. 更改读取字幕文件
                 4. 保存字幕文件修改
                 5.打印当前字幕内容
                 0. 退出当前所在系统
______
请输入你的功能选择:
0
00:00:00,800 --> 00:00:24,800
English subtitle by : Eduun
1
00:00:36,500 --> 00:00:38,500
The late fourth century A.D. the Roman Empire began to crumble.
2
00:00:38,500 --> 00:00:40,500
Alexandria, in the province of Egypt, still retained some of its lustre,...
00:00:40,500 --> 00:00:42,500
...possessed one of the Seven Wonders of the ancient world:...
00:00:42.500 --> 00:00:45.000
```

> 实现整体字幕时间点前进和后退的功能

-- | 该系统提供的功能 | -----1.整体字幕进行移动 2. 查找某一条的字幕 3. 更改读取字幕文件 4.保存字幕文件修改 5.打印当前字幕内容 0.退出当前所在系统 请输入你的功能选择: 请选择字幕移动的方向: (前进输"+",后退输入"-") 请选择字幕移动的时间长度: (整数,单位毫秒) 操作成功提示: 成功移动所有字幕的时间点 ---- | 该系统提供的功能 | -----1.整体字幕进行移动 2. 查找某一条的字幕 3. 更改读取字幕文件 4.保存字幕文件修改 5.打印当前字幕内容 0.退出当前所在系统 请输入你的功能选择: 1 请选择字幕移动的方向: (前进输"+",后退输入"-") 请选择字幕移动的时间长度: (整数,单位毫秒) 100 操作成功提示: 成功移动所有字幕的时间点

> 实现根据字幕序号 Sid 查找某一条字幕的功能

	 - 该系统提供的功能
	1.整体字幕进行移动 2.查找某一条的字幕 3.更改读取字幕文件 4.保存字幕文件修改 5.打印当前字幕内容 0.退出当前所在系统
请输入你的功能选择: 2	
	 - 该页面提供的操作
	1.根据序号查找字幕
请选择你想要查找字幕的方式 1	=====================================
输入想要查找的序号Sid: 0 查找成功提示:该条字幕信息 0 00:00:00,900> 00:00 English subtitle by:	:24,900
	 该页面提供的操作

> 实现根据时间点查找某一条字幕的功能

	 该系统提供的功能		
 	1.整体字幕进行移动 2.查找某一条的字幕 3.更改读取字幕文件 4.保存字幕文件修改 5.打印当前字幕内容 0.退出当前所在系统		
请输入你的功能选择: 2	:======================================	=======================================	=
	该页面提供的操作		ı
	1.根据序号查找字幕 2.查找时间查找字幕		
请选择你想要查找字幕的 2	·===== 的方式:	=======================================	=
输入想要查找的字幕时间 00 00 01 800 输入的时间: 00:00:01 查找成功提示: 该条字和 0 00:00:00,800> 00 English subtitle by	幕信息如下所示→ 9:00:24,800	则输入00(enter)00(enter	r)00(enter)000(enter)
	该页面提供的操作		ı
	1.当前字幕时间移动 0.返回上一级的菜单		
1	:==========	=======================================	=
请选择字幕移动的方向:	(前进输"+",后退输入"-")		
	长度: (整数,单位毫秒)		
100 操作成功提示:成功移z	动该条字幕时间点		

> 实现保存字幕文件修改的功能

	 该系统提供的功能	
- 	1.整体字幕进行移动 2.查找某一条的字幕 3.更改读取字幕文件 4.保存字幕文件修改 5.打印当前字幕内容 0.退出当前所在系统	
请输入你的功能选择: 4 保存成功提示:保存路径为	D:\Study\Junior\Junior_1\编译原理\三级项	=== 目\字幕分析器\subtitle.srt
	 该系统提供的功能	
	1.整体字幕进行移动 2.查找某一条的字幕 3.更改读取字幕文件 4.保存字幕文件修改 5.打印当前字幕内容 0.退出当前所在系统	

请输入你的功能选择:

0

退出成功

2.4.2 优化后的版本说明

2.4.2.1 页面设计

> 主页面设计

	字幕分析器 请输入srt文件的路径 确认
	Manual Co. S. C. L. Haberley
移动整体字幕	
查找某条字幕	
保存	
打印	
退出	

> 整体字幕移动弹窗设计



> 查找字幕页面设计



2.4.2.2 操作说明

①首先运行该程序,在文本框输入 srt 文件的绝对路径后点击确认,下面就会自动打印出这个 srt 文件的所有内容,如下图所示:



②点击打印按钮, 右边就会将该 srt 文件的内容打印出来, 如下图所示:



③点击移动整体字幕按钮,就会跳出弹窗,在该弹窗输入要移动的毫秒数,点击前进或者后退按钮即可,具体显示如下图所示:

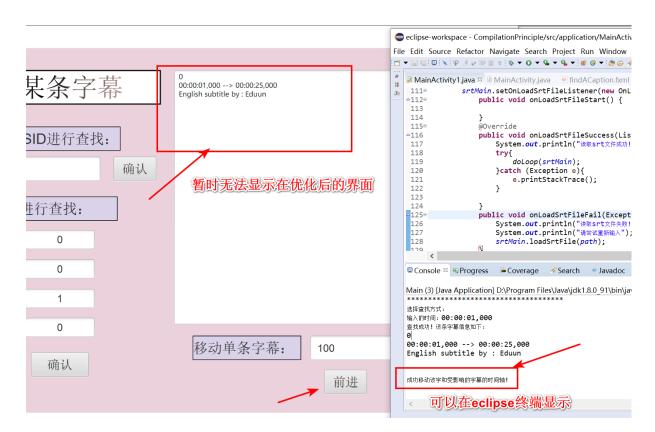


④点击查找某条字幕按钮,就会跳转去另一个页面,在这个页面,用户可以通过输入字幕序号查找字幕也可通过输入字幕时间点进行某条字幕的查找,只需要在相应的文本框输入相关信息,点击确认即可,查找到的某条字幕内容就会打印在右边白板上,具体如下图所示:





⑤在该页面还可以实现单条字幕时间点的移动,只需要在右下角输入要移动的毫秒数,点击前进或者后退就可以对该条字幕进行时间点移动,具体如下图所示:



⑥点击保存即可实现将字幕文件已经修改的内容保存到原文件上面,具体如下所示:



⑦点击退出即可退出程序。

2.4.2.3 问题说明

实现了第一个版本的字幕分析器之后,还想对这个版本进行界面优化,美美的页面都已经设计好了,但是由于我们对 javafx 的学习还处于初级阶段,很多操作还是不是很了解,所以在具

体实现的时候遇到了一些问题还无法解决,导致这个优化后的版本还存在这个问题: 就是明明已经成功移动了字幕的时间点,但是当打印的时候却还是原来的字幕内容。现在依旧还在继续搞这个优化后的版本,即使这个项目截止时间到了,但是我们依旧还会将这个版本进行优化,将其功能进行补齐和完善,因为我们不仅仅是为了完成项目而完成项目,还为了深入学习字幕分析器原理和实现一个完整的字幕分析器而继续学习和奋斗!

3. 个人感想

名字	个人感想
郑书桦	本次三级项目程序部分是用 Java 语言编写的,实践过程中感觉把编译原理(词法和语法分析)、数据结构与算法(链式存储)、Java 三门学科联系起来,把各个学科之间的知识联系起来共同参与解决问题。通过这次实践,我更加深刻的理解了什么是词法分析和语法分析,词法分析其实就是"切词",我们针对 srt 字幕的时间格式,构建了一个词法分析器,用于将"00:00:00,000 à 00:00:00,000"格式的字符串切成一个一个单独的值,即"时","分","秒","毫秒"。而语法分析,就相当于是拼图,将合适的"碎片"(或者说"词")拼接起来构成一张完整的"图",这个部分主要是用于将词法分析切出来的"词"重新拼接成 srt 文件可识别的语句。总的来说,这个项目不仅加深我对于词法分析、语法分析的理解,还重新复习了 Java 语言,锻炼了不少。
郑梦宜	一场简单的实践其实也会涉及许多的专业知识,也是一次知识体系的完整结合应用。通过这次实践,我理清了编译原理这门课程对于语言分析的本质工作。这个项目不仅加深我对于词法分析、语法分析的理解,还重新复习了 Java 语言,锻炼了不少。我在本次实践中主要负责了前端界面功能与后端代码接口的连接,前端界面的设计和功能实

	现是通过 javafx 平台实现的,因此本次课程实践也帮助我熟悉了 javafx 平台的使用。而在前后连接的过程中,我也发现了此次代码实现比较注重功能的实现,耦合度不高,这是此次时间成果有待提高的一个点。最后,在整个实践过程中,小组成员相处都十分融洽。很幸运能参与此次小组实践,期待下一次的进步。
郭清扬	只有对字幕进行词法分析和语法分析,才能实现这个字幕分析器,主要核心部分就在这两个点上,当然其他部分当然也很重要,当该项目分析器运行成功的时候,我感受到了很强的成就感,感觉从中真的有学到很多东西,感觉这个项目有一定的挑战,但同时觉得是一次很好的锻炼机会。
张国润	通过这次小组项目,让我对编译原理中词法分析和语法分析有一个更加深刻且清晰的认识。虽然在这个项目进展的时候遇到了很多的困难,例如如何用代码实现对字幕词法分析和语法分析,这一点我们虽然知道如何个逻辑分析,但是就是代码架构不知道怎么架构,最后还是参考了网上一些博主思路才成功写出来,说时候这个项目还是挺难的,但是我感觉我还是收获了很多,无论是知识上还是经验上。
张菊敏	这个小组的成员都是女生而且还是住在同一个书院的,感觉这次的合作就是天籁之和,无论是开会的时候还是交流任务的时候,我们队内都相处得非常融洽,一遇到问题和困难的时候,大家都会相互帮忙解决,所以在这个小组项目中,虽然项目不简单,但是过程却是很开心和充实的,当然在这个项目中也收获了很多知识和提升了自己的能力,这毋庸置疑,但是还是很开心得收获了她们的友谊,感谢项目!

4. 参考资料

srt 字幕解析 https://blog.csdn.net/piaozhiye/article/details/6661964

SRT 视频字幕的解析与同步 https://blog.csdn.net/hhbgk/article/details/22435723

视频显示字幕(解析 srt 文件) https://www.jianshu.com/p/94202c6a872d