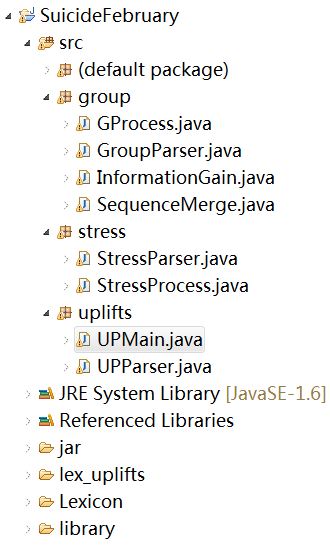
**Java Project**

【SuicideFebruary】



**UPMain.java //主文件**

**Function 1**

Depart() //Step1分离原始微博的文本与时间

// 输入路径String dirRead = "E:\\TEST\\suicide\\Depart\\suicide\\tweet\\";

// 输出路径String dirRead = "E:\\TEST\\suicide\\Depart\\suicide\\text\\";

// 输出路径String dirRead = "E:\\TEST\\suicide\\Depart\\suicide\\time\\";

**Function 2**

parseAllTeen(); //Step2为每条微博进行文本分析，得到14个文本相关特征 (for each teen)

// 输入路径 "E:\\TEST\\suicide\\Depart\\normal\\text\\";

// 输出路径 "E:\\TEST\\suicide\\Depart\\normal\_text.txt"

**其他功能**

**Function 0.1** InformationGain.java //计算每个feature的information gain

**Function 0.2** Proecess.java /prepareInputFCM() // weka|FCM| 输入数据归一化，格式整理

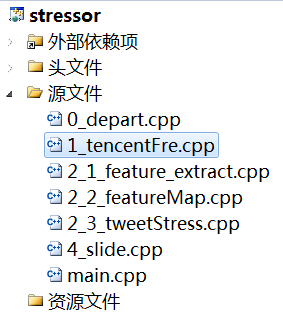
POSITIVE输出

【upTextFeature】positiveNum, degree, 6 categories, 3 rol/act/des

【upLevel】id, level[0-5]

**C++ Project**

【stressor】



**main.cpp //主文件**

**Function1 单条微博压力级别计算 //**step3

依次调用:

tencentFre();//输入输出为绝对路径，见代码

feature\_extract();//输入输出为绝对路径，见代码

featureMap();//输入输出为绝对路径，见代码

tweetStress();//输入输出为绝对路径，见代码

**Function2 划分压力区间，统计6个压力相关feature //**step4

**dayStaNew();//两个阈值设置变量为GAP, Coverage (这不是直接有confidence了)**

//1-daySta【DaySta】/【upDaySta】

**包括压力累计/天, 压力微博条数/天，微博数/天，日期+时间，微博起止ID**

timeResult1<<valStress<<" "<<numStress<<" "<<fre<<" "

<<vec\_year[i].second<<"-"<<vec\_month[i].second<<"-"<<vec\_day[i].second<<" "

<<vec\_hour[i].second<<" "<<i<<" "<<j-1<<endl;

//2-CateSta【CateSta】/【upCateSta】

**包括1-6种压力类型统计，rol/des/act统计，http[3项]统计，日期+时间**

for(int index=1;index<6;index++)

{

topicResult<<topicSta[index]<<" ";

}

for(int index=1;index<4;index++)

{

topicResult<<rolSta[index]<<" ";

}

for(int index=1;index<4;index++)

{

topicResult<<httpSta[index]<<" ";//http, reply, comment

}

topicResult<<vec\_year[i].second<<"-"<<vec\_month[i].second<<"-"<<vec\_day[i].second<<" "<<vec\_hour[i].second<<endl;

//3-Slide【Slide】/【upSlide】包括 ID（1），起止天ID（2），起止微博ID（2），起止日期（8）

slideProOut<<slideNum<<" "<<earlyID<<" "<<endID//slideID, earlyDayID, endDayID,startTwID, endTwID

<<" "<<vec\_daySE[earlyID].first<<" "<<vec\_daySE[endID].second<<" ";

slideProOut<<vec\_year[t1].second<<"-"<<vec\_month[t1].second<<"-"<<vec\_day[t1].second

<<" "<<vec\_hour[t1].second<<" ";

slideProOut<<vec\_year[t2].second<<"-"<<vec\_month[t2].second<<"-"<<vec\_day[t2].second

<<" "<<vec\_hour[t2].second<<endl;

e.g.,

//4-SlideValue【SlideValue】/【upSlideValue】

包括ID(1),peak,peakDay,span,cover,avgLevel,slope,RMS(7),cate(5),rol/act/des(3)

slideValueOut<<slideNum<<" "<<cur\_peak<<" "<<cur\_peak\_day<<" "<<j-i+1<<" "<<cur\_slide\_cover

<<" "<<cur\_slide\_avgLevel<<" "<<cur\_slide\_slope<<" "<<cur\_RMS;

for(int k=1;k<=6;k++)

{

slideValueOut<<" "<<cur\_cat[k];

}

for(int k=1;k<=3;k++){

slideValueOut<<" "<<cur\_rol[k];

}

slideValueOut<<endl;

**Function3统计正面每天统计量+区间划分**

dayStaUp()

-与dayStaNew()的差别在于：从5种类型变成6种类型

Stress topic 顺序

"Lexicon/eve\_topic\_school.txt", //0

"Lexicon/eve\_topic\_romantic.txt",//1

"Lexicon/eve\_topic\_peer.txt",//2

"Lexicon/eve\_topic\_healthy.txt",//3

"Lexicon/eve\_topic\_family.txt", //4

Positive topic 顺序

"lex\_uplifts/lexp\_study.txt", //0

"lex\_uplifts/lexp\_romantic.txt",//1

"lex\_uplifts/lexp\_friends.txt",//2

"lex\_uplifts/lexp\_self.txt",//3

"lex\_uplifts/lexp\_family.txt", //4

"lex\_uplifts/lexp\_enter.txt",//entertainment-new lexicon 5 !!!!