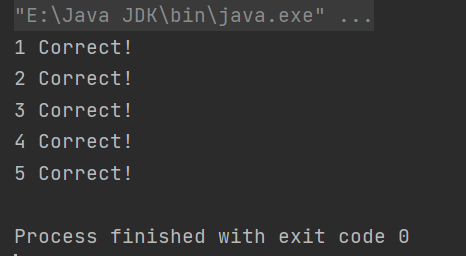
HW5

段易非 2018010303

1. DataChooser



【代码实现】

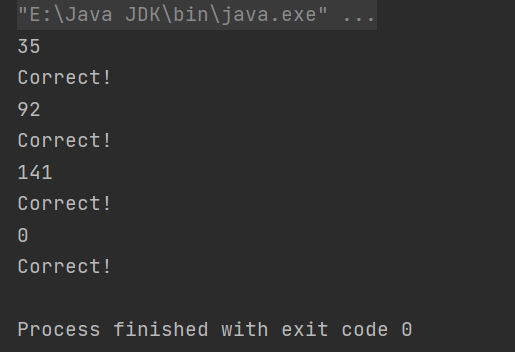
代码主要分为两个模块，全部在构造函数DataChooser中完成。第一个是将data.csv文件读入并提取出第四列Department的数据，第二个则是将这列数据写入一个输出的文件depatment.csv当中并保存。

对于第一个模块，实现的思路是先用BufferedReader嵌套FileReader来将输入文件的字节流转换为字符流并进行缓冲，再循环使用BufferedReader的readLine函数来逐行读取csv中的数据，并且每次读取之后立即根据csv格式特点用“,”来split改行字符串，获取第4个即我们所需的院系信息，每次都将这行的院系信息存入一个数组便于后续处理。

在第二个模块中，首先用File创建指定的输出csv文件，再用BufferedWriter、OutputStreamWriter和OutputStream依次嵌套这一文件来实现字符流缓冲区处理，随即将之前数组中储存的院系信息依次利用BufferedWriter的append函数写入其中即可。

【问题解决】

1. 由于测试文件可能长度不定，因此对于用来存储院系信息的数组，要用ArrayList类实现为可变长数组；
2. BufferedWriter在最后要用其close函数来关闭，否则无法完成写文件的过程，导致输出department.csv为空。
3. FileSearcher



【代码实现】

代码分为两个主要模块，分别对应两个函数。构造函数FileSearcher用于读取dialog文件，search函数用于处理文件内容并完成单词查找计数的过程。

在读取文件时除了要将其转化为字符缓冲流之外，还要创建一个File类对象来存储数据信息，便于后续在处理读取顺序时明确文件总长度。

在search函数中，利用循环结构配合readLine函数来逐行读取dialog文件，每读一行都随即进行搜索处理。具体处理方式是：首先将“’”“-”都替换为空格，方便之后在split时将由之连接的单词拆分开，接着将“,”“.”“!”“?”都替换为“”，从而确保split截取出的单词为纯字母组合。接下来利用“ ”来对改行文本进行split操作，然后遍历获得的一系列单词，与待搜索单词进行比对，一致即对count加1，dialog文件全部处理完成后返回最终得到的count值即可。

【问题解决】

1. Count在search函数开头要置零，否则在变更单词查询时会有问题；
2. br读完后一次后需要回到开头重新读才能搜寻另外的单词，利用BufferedWriter的mark和reset两个函数来实现；
3. replaceAll中对于有特殊含义的“.” “?”要用“\\”来标注转义，否则会报错。