实验五

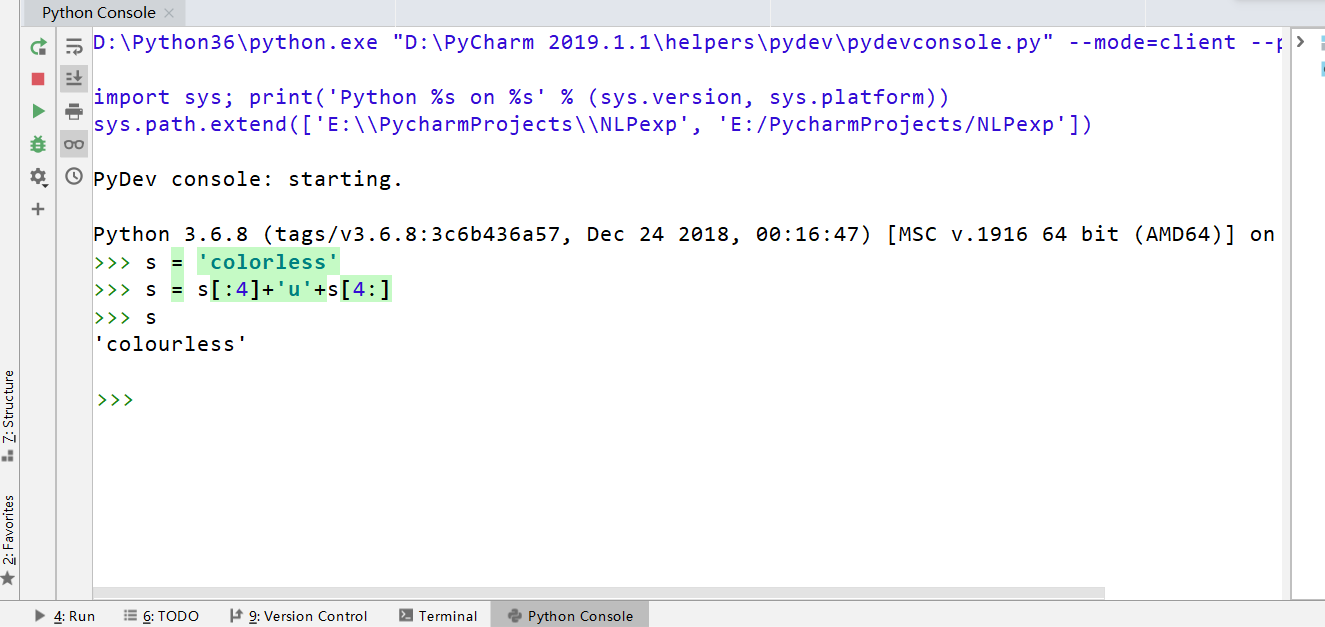
目的：会初步使用nltk进行文本预处理、实现一个简单的文本分类器

内容：

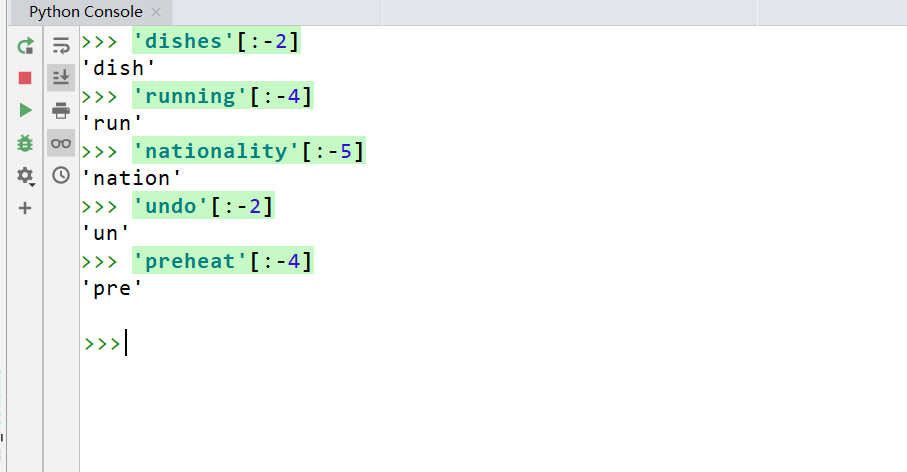
1. 第三章中使用示例演示一遍，并把结果复制在文件中



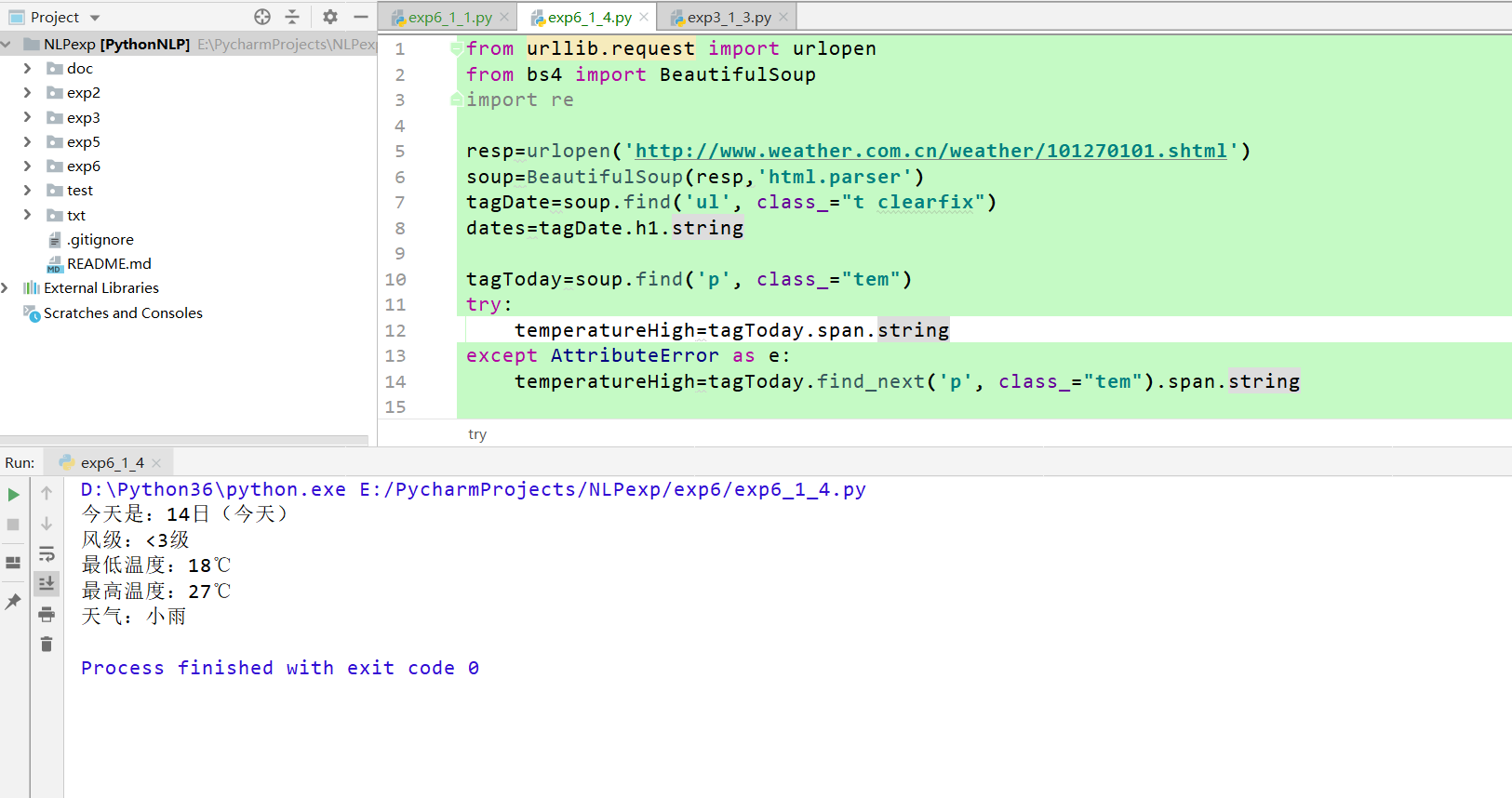
1. 从3.12中练习选做第1、2、4、20

T1、

T2、



T4、



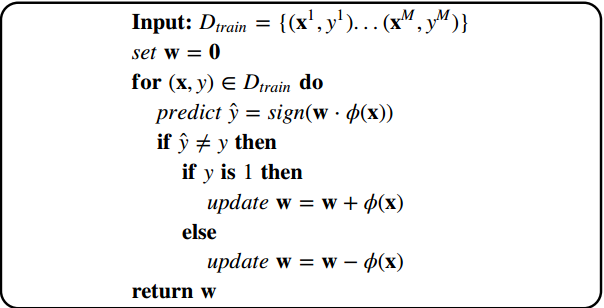
（3）采用感知机算法实现文本分类

步骤：

A: 语料处理：（见附件中的中文语料）

1. 从xml中提取文本
2. 把语料分成两部分：训练集80%,测试集：20%
3. 分词、去掉停用词。

B: 实现算法：



C： 用模型去测试测试集，计算准确度

