

****

信 息 学 院

《社交网络技术与应用》

作业四 实验报告

**实验题目 文本挖掘**

**组 员 学号37220222203661姓名 李林江**

# 文本分类实验

## 实验介绍

### 简介

本次实验我们将继续对前端问答窗口的功能进行升级改造。这一个小节的实验将带领大家动手实践文本分类，并在问答前端实现用户上传数据集压缩包进行文本分类，并对结果进行可视化的提示语句响应。

### 实验目的

掌握常用的文本分类算法

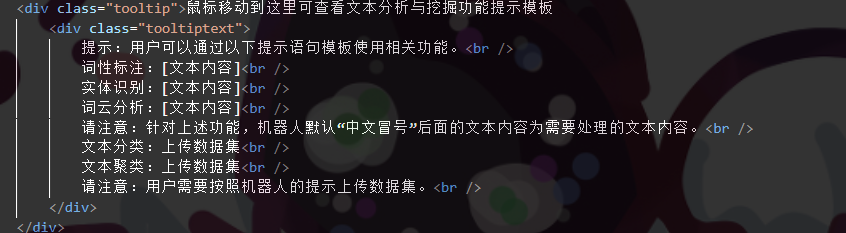
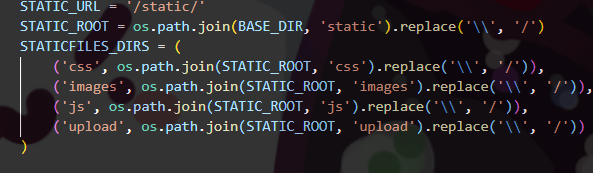
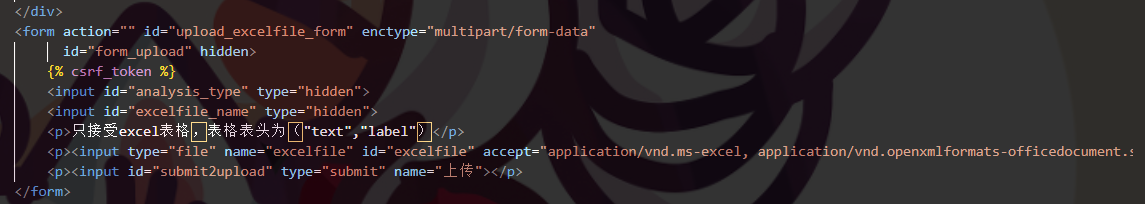
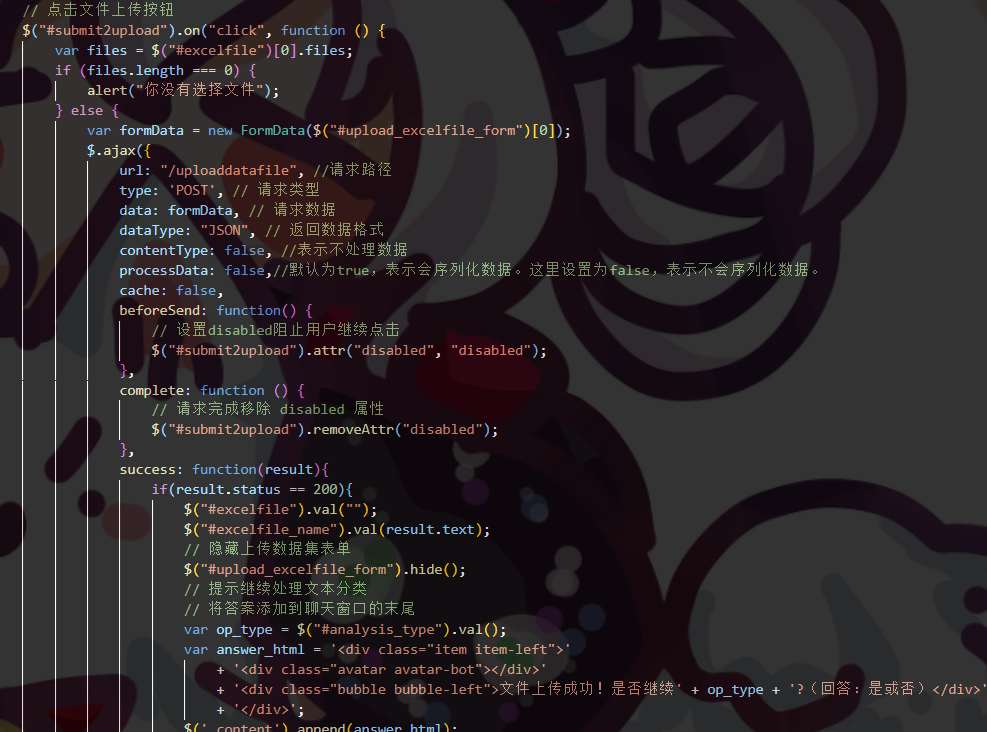
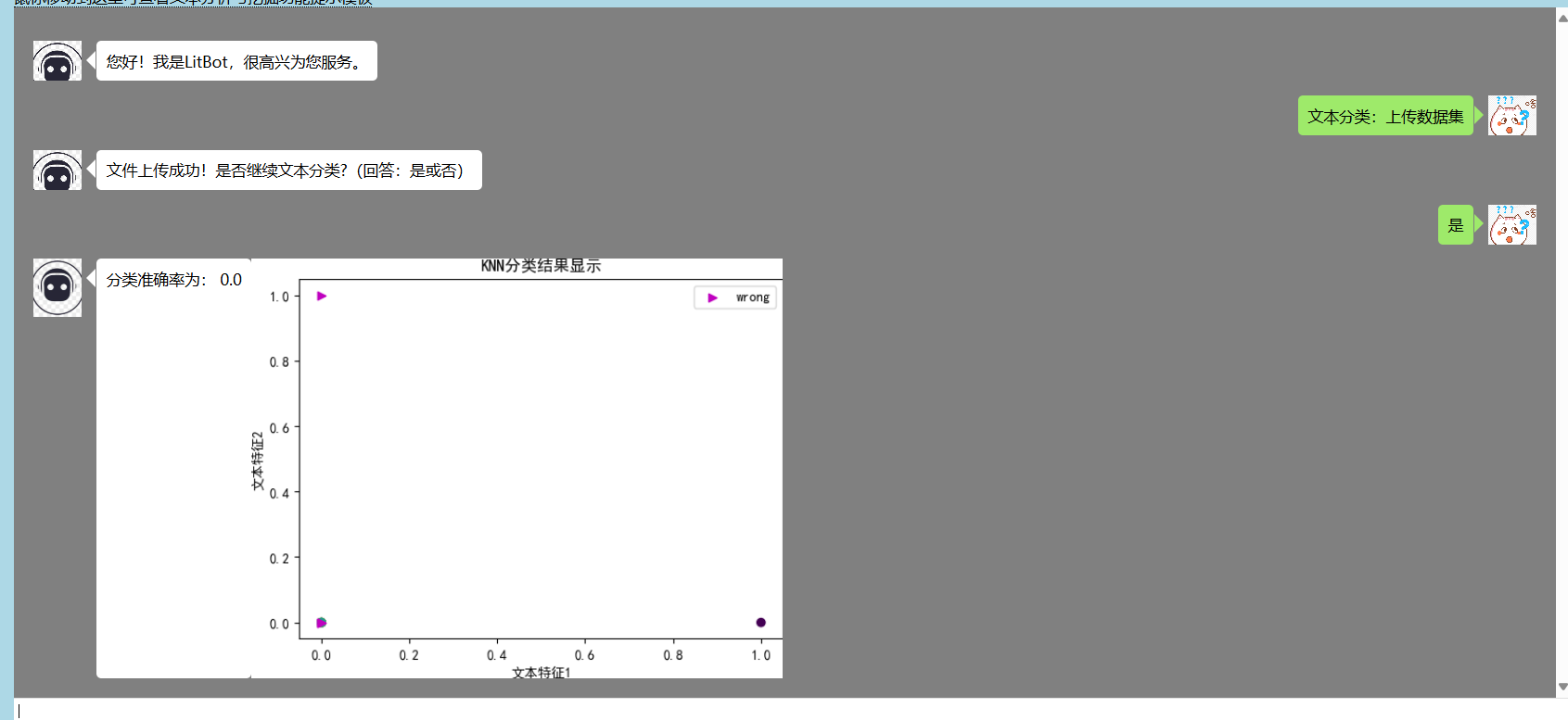
掌握常用的可视化工具

## 实验步骤

本次实验基于实验四的问答系统继续完善文本挖掘功能。

实验任务介绍：我们将对数据集进行文本分类，并进行结果可视化的教学。

### 文本分类

## 实验小结

本实验针对问答窗口添加了文件上传、文本分类以及结果的可视化功能。

# 文本聚类实验

## 实验介绍

### 简介

本次实验我们将继续对前端问答窗口的功能进行升级改造。这一个小节的实验将带领大家动手实践文本聚类以及对聚类结果每类文本内容进行词云图分析。

### 实验目的

通过实践，掌握文本聚类和词云图分析的基本流程，并结合实践操作加深对文本聚类和可视化工具的理解和应用。

## 实验步骤

### 文本聚类与词云图可视化

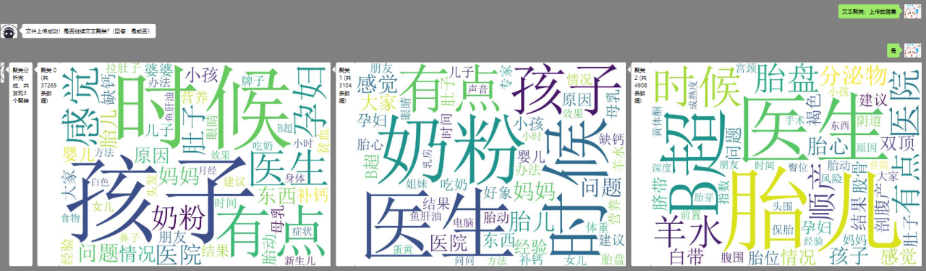
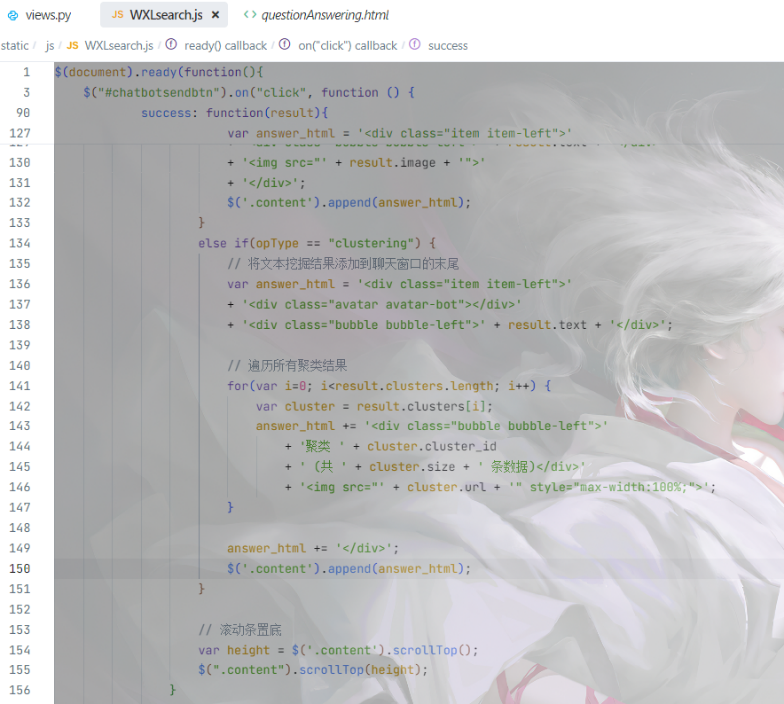
 

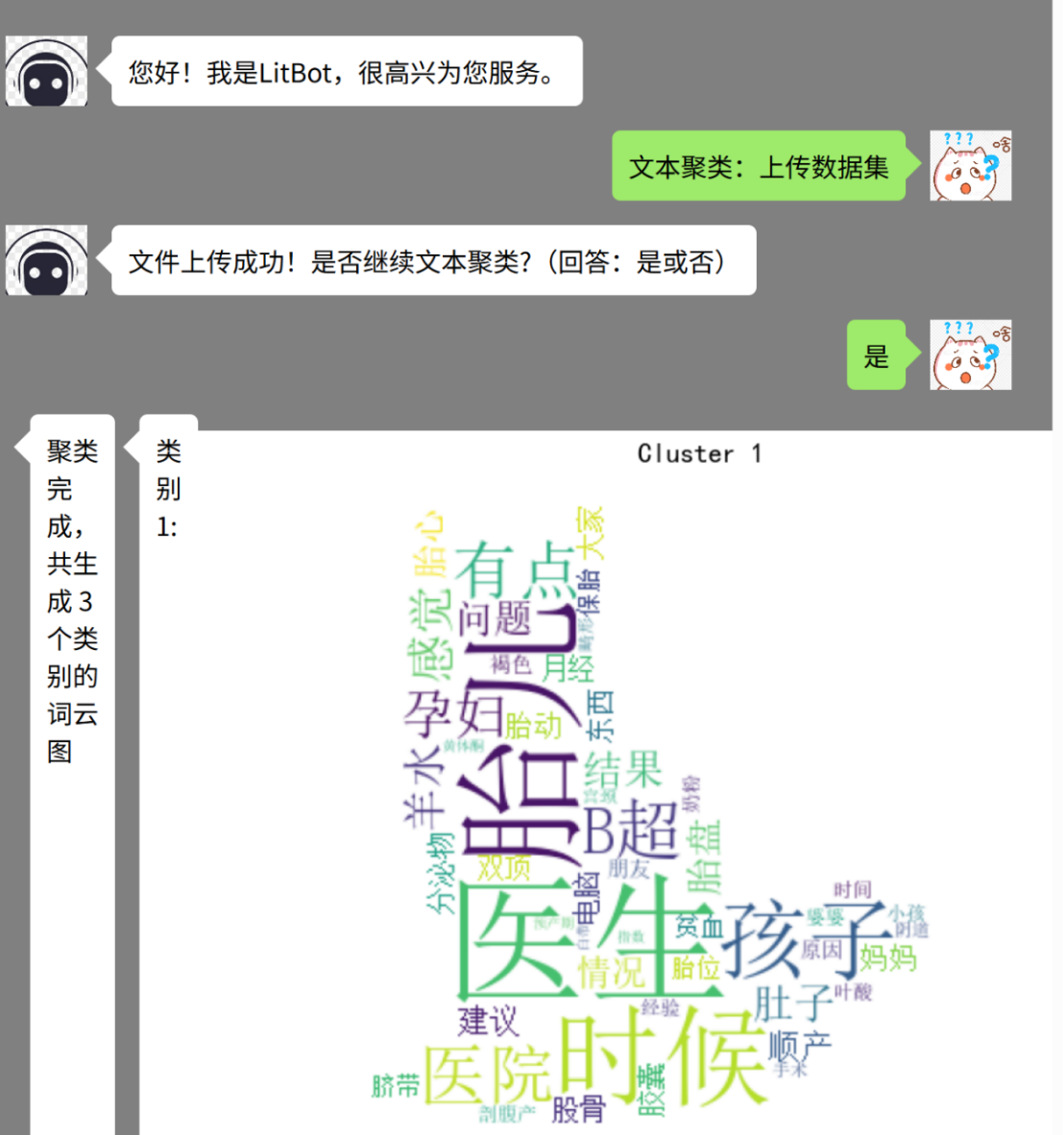
## 实验小结

本章主要自己动手实践文本聚类以及词云图分析的基本过程。

## 思考题

1. **本次作业请各小组使用实验二爬取的100000以上的问答数据集作为聚类的数据集，可以抽样3万条数据进行实验，请各个小组成员每人独立实现一种不同的聚类算法，并结合词云图结果分析聚类算法的效果。**

**陶婉婷：**

**李丹：**

**李林江：**

**张竣淇：**

**注意：本次实验4月30日23：59之前请把报告上传到课程平台（请每个人将”代码+实验报告”放入同一个命名为“姓名+学号+作业四”的文件夹并打包成” 姓名+学号+做业四.zip”）。**

**注意：请不要抄袭他人的结果（抄袭分数为0）。**