向量数据库实习设计

本次实习设计的目的是借助自然语言处理的典型任务，体验一个非结构化数据的典型处理流程。大概处理过程如下：

1. 建立小说表 yttlj ( phaseId, phaseText);（平台提供）
2. 建立人物表 person(personId, personName );（平台提供）
3. 建立共现表 coCurrence(personId1, personId2, counts);

（这部分请同学们用SQL实现，一个思路是在person表中增加一个类似位向量的字段phaseContains，每个位对应一个段落，如果某段落包含了该person，则将该位设置为1，这样通过两个将person的phaseContains进行位与运算，即可得到他们的共现次数。另一种做法是生成人物对(person1,person2), 针对每对人物，看看有多少段落同时包含了这两个人物）；

1. 通过矩阵分解，得到词向量表示，这部分可以通过平台提供的Python代码实现。
2. 将上面的词向量存入pgvector，计算人物之间位向量相似度，找到每个人和他最相似的另一个人。

张三丰

张翠山

殷素素

俞岱岩

俞莲舟

殷梨亭

纪晓芙

张无忌

周芷若

赵敏

殷离

小昭

黛绮丝

谢逊

朱长龄

朱九真

杨逍

范遥

灭绝师太

何足道

胡铁牛