**学习内容：**

**对项目步骤的总结：**

1. **数据库**
   1. 数据库连接
   2. 构建表格（11个 字段 ，id和五个全文关键字和 文章序号 ，标题分词五个）
   3. 构建读取每个文章，存入数据库中的算法：

采用jieba库中IF-IDF算法进行关键字抽取，因为从title中得到关键字的数量可能不足5个需要按情况进行分类，构建函数进行分类。

* 1. 调用1.3算法，根据目录将目录下的所有txt文件保存在数据库中。
  2. 关闭数据库的连接

1. **UI界面**

2.1 构建窗口

2.2 获得输入在检索框的检索值并保存

2.3 调用检索函数获得检索值返回到显示结果框

**3. SimCSE检索部分**

3.1 将从UI界面获得的检索值通过pycorrector改错

3.2 对检索值进行分词（关键字抽取俩个值）

3.3 连接数据库

3.4 从数据库中读取数据保存到字典中（{10个关键字:文章名}）

3.5 关闭数据库

3.6 构建函数，获得新字典，{每个元素的文章名：检索值与该10个关键字得到的余弦相似度}：

调用SimCSE模型，这俩个检索值分别对单个元素10关键字进行余弦相似度搜索，将最大的余弦相似度保存进行相加，则可以得到检索值中抽取的俩个值与这个文章的最优余弦相似度。

3.7 对新字典进行排序，得到前三个最优的结果

3.8 根据文章名和目录找到文章返回到UI界面