课前思考：  
1.已经处理过的数据类型，主要有：（有、无符号）整型数、（有、无符号）浮点数、布尔型、字符、字符串、C语言中的“指针”等  
从另一个分类来说，包括非结构化数据，如，文本、图片、语音；以及结构化数据。  
2.数据库管理系统，是一种操纵和管理数据库的大型软件，用于建立、使用和维护数据库，对数据库进行统一的管理和控制，保证数据的安全性和完整性。同时提供了对数据的录入、修改、查询操作，具有数据定义，数据操作、数据存储与管理、数据维护、通信等功能。  
3.仅粗略了解过SQL  
4.不能。我认为，一篇文本如果作为一个记录，那么他的“正文”字段将会非常臃肿。在“查询”时，也造成资源的浪费和效率的低下。

课堂思考：  
1.一个信息检索应该包括几个部分？需要完成什么？  
1）信息获取（爬取）——爬虫  
2）文本处理——对爬取到的信息进行预处理，得到文章特征  
3）匹配——当用户输入一定关键词，与文章特征进行匹配，对结果进行排序  
4）推送——将排序结果返回给用户  
  
2.怎么找到满足“条件”的戏剧？  
逐个operator进行查找，并从所有文档中不断筛选满足要求的文档（戏剧） 如，先找所有含Brutus的，得到集合1，再从集合1这些文档中找包含了Caesar的文档，得到集合2，最后从集合2中找到不含Calpurnia的文档，得到集合3.