文本类型数据是非结构化的数据，无论是管理类应用还是内容类应用，都只能对结构化的数据进行处理。因此，在文本挖掘中非常重要的技术环节就是数据的结构化过程。基于结构化的数据形式，可以进行统计建模分析，并在计算机中进行自动化处理。