**作业1要求**

**作业内容：**请编写一个网络爬虫程序，能够从豆瓣网（或新浪新闻等网站）上抓取10万个以上评论页（或新闻页面），作为接下来Hadoop存储与计算实验的原始数据。

**编程语言：**任选，可用已有爬虫框架。