**爬虫基本原理**

1. 爬虫是**请求**网站并**提取**数据的自动化程序
2. （爬虫）访问网站的基本流程

发起请求：通过http库向目标站点（服务器）发送请求Request，等待服务器响应

获取响应内容：若服务器正常响应，返回一个Response

解析内容

保存数据

1. Request包含？

请求方式：GET(信息在url中，直接通过url访问)，POST（信息在form data中，需提交表单才可访问），其他（）

请求URL

请求头：包含请求的头部信息，如User-Agent,cookies等

请求体：请求额外携带的数据，如form data

1. Response包含？

响应状态：status

响应头：内容类型、长度、cookies等

响应体：请求资源的内容，如网页HTML等