Requests 模块

基于网络请求的两个模块：urllib（古老）/ requests（简洁、高效）

requests 模块：py中原生的一款基于网络请求的模块，

功能强大，简单便捷，效率极高。

作用：模拟浏览器发请求

如何使用：① 指定url

② 发起请求

③ 获取响应数据

④ 持久化存储

编码流程

实战编码：

- 需求：爬取搜狗首页的页面数据

实战巩固

- 需求1：爬取搜狗指定词条对应的搜索结果页面（建议网页采集器）
  - UA伪装（User-Agent）

-UA检测

-需求2：破解百度翻译

- post请求（携带了参数param）
- 相应数据是一组json数据

-需求3：爬取豆瓣电影分类排行榜中的电影详情数据

-需求4：爬取国家药监区中化妆品生产许可证相关数据

-动态加载数据（Ajax）

-首页中对应的企业信息数据是通过ajax动态请求到的

-通过对企业详情页url的观察发现：

-url的域名都是一样的，只有携带的参数（id）不一样

-id值可以从首页对应的ajax请求到的json串中获取

-域名和id值拼接出一个完整的企业对应的详情页的url

-详情页的企业详情数据也是动态加载出来的

-观察后发现，所有的post请求的url都是一样的，只有参数id值是不一样的

-如果我们可以批量获取多家企业的id后，就可以将id值和url形成一个完整的详情页对应的url

数据解析：

- 聚焦爬虫
- 正则
- bs4
- xpath