

极客学院  
jikexueyuan.com

# 定向爬虫 简单的模拟登录

## 定向爬虫：简单的模拟登录 — 课程概要

- 抓包利器Fiddler
- Cookies与保持登录
- Cookies模拟登录
- 模拟登录新浪微博
- 实战——追女神助手v0.1

■ 定向爬虫：简单的模拟登录

# 抓包利器Fiddler

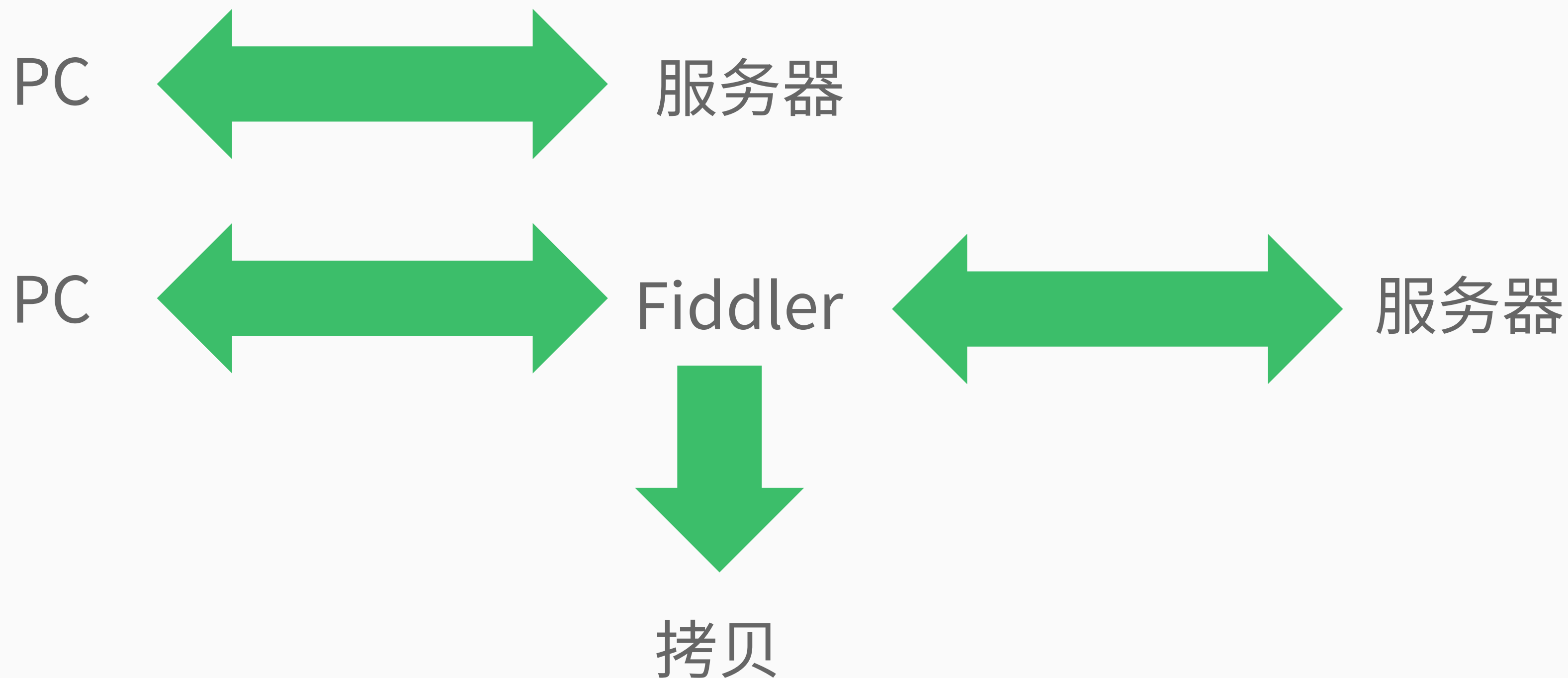
# 抓包利器Fiddler

- Fiddler介绍
- Fiddler抓包

## 抓包利器Fiddler — Fiddler介绍

电脑与互联网之间的通信是通过不同的数据包收发来实现的。Fiddler可以从中间对数据进行拦截，拷贝一份数据以后再将数据发送给目的端。

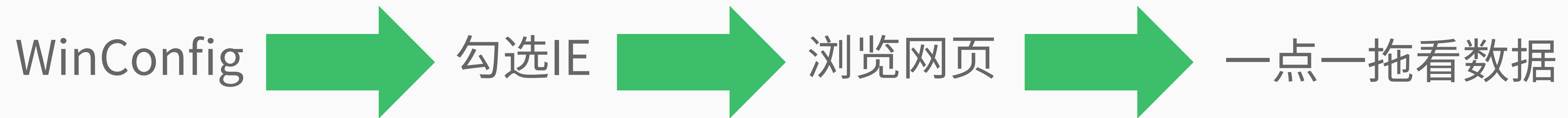
同类的还有WireShark。



## 抓包利器Fiddler — **Fiddler**抓包

安装：一路下一步。

配置：



# Cookies与保持登录

# Cookies与保持登录

- Cookies介绍
- 保持登录机制



## Cookies与保持登录 — Cookies介绍

Cookie（复数形态Cookies），中文名称为小型文本文件或小甜饼，指某些网站为了辨别用户身份而储存在用户本地终端（Client Side）上的数据（通常经过加密）。

——Wikipedia

## Cookies与保持登录 — 保持登录机制

当登录一个网站时，网站往往会请求用户输入用户名和密码，并且用户可以勾选“下次自动登录”。如果勾选了，那么下次访问同一网站时，用户会发现没输入用户名和密码就已经登录了。这正是因为前一次登录时，服务器发送了包含登录凭据（用户名加密码的某种加密形式）的Cookie到用户的硬盘上。第二次登录时，（如果该Cookie尚未到期）浏览器会发送该Cookie，服务器验证凭据，于是不必输入用户名和密码就让用户登录了。

——wikipedia

■ 定向爬虫：简单的模拟登录

# Cookies模拟登录

# Cookies模拟登录

- Fiddler获取Cookies
- Requests提交Cookies

# Cookies模拟登录 — Fiddler获取Cookies

- 抓取登录数据包
- 获取Cookies

# Cookies模拟登录 — Requests提交Cookies

```
cookie = {"Cookie": "XXXXXXXXX"}  
Html   = requests.get(url, cookies=cookie)
```

# 模拟登录新浪微博

# 模拟登录新浪微博

- 分析新浪微博登录机制
- 分析需要提交的数据
- Requests提交数据



## 模拟登录新浪微博 — 分析新浪微博登录机制

- 输入用户名密码
- 发送给服务器
- 服务器验证成功
- 返回Cookies与正常页面

## 模拟登录新浪微博 — 分析需要提交的数据

- Mobile (用户名)
- Password (密码)
- remember (是否保持登录)
- backURL (登录以后返回的地址)
- backTitle (登录以后返回的标题)
- tryCount (尝试次数)
- vk (一个简单的验证码)
- submit (登录)
- **action (URL参数)**

## 模拟登录新浪微博 — Requests提交数据

```
data = {"mobile" : "XXXXXXXXX",  
        password : "XXXX",  
        .....  
        }  
  
Html  = requests.post(url, data=data)
```

■ 定向爬虫：简单的模拟登录

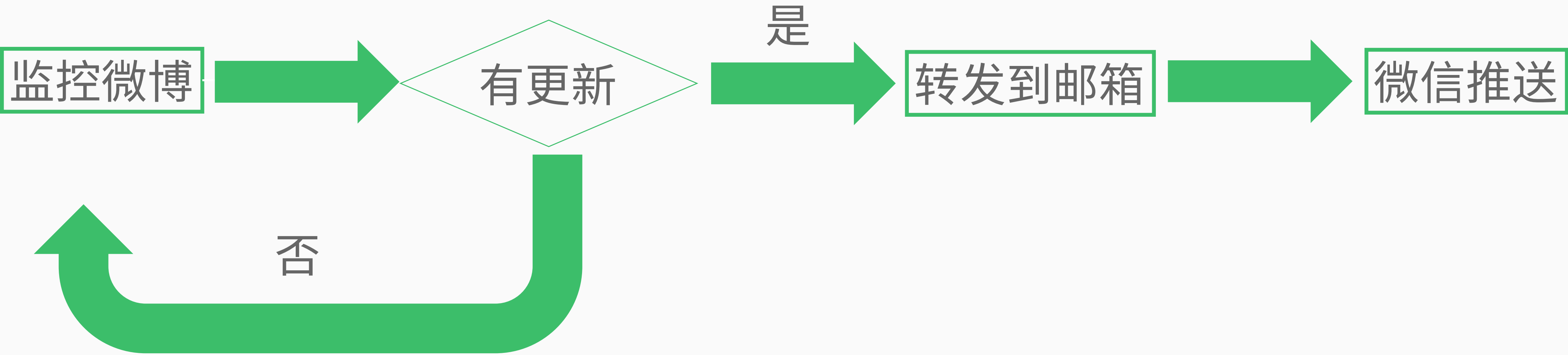
# 实战——追女神助手v0.1

# ■ 实战——追女神助手v0.1

目标网站：新浪微博

目标内容：微博内容

目标流程：



## 实战——追女神助手v0.1 — 更多功能

为什么是v0.1而不是v1.0?

因为功能太简单了。如果你还能：

- 爬取她的关注列表，从而分析她的社交关系网络
- 爬取她全部的微博内容，然后分析她的性格与喜好
- 她微博里面经常提到某个明星？
- 别忘了我们讲过如何爬取百度贴吧~
- 她微博里面经常爆照？
- 别忘了我们也会爬取照片~

爬不是目的，爬是为了实现目的~

Come On 少年，把学到的东西用起来~

## 定向爬虫：简单的模拟登录

本套课程中我们学习了爬虫的模拟登录功能，你应当掌握以下知识：

- 通过Cookies登录新浪微博
- 通过提交表单登录新浪微博

你可以使用这些技巧来制作一个监控性的爬虫，例如本课实战项目中我们制作的追女神助手v0.1。从下一课开始，我们将讲解Scrapy，如果想继续提高，你可以继续在极客学院学习《定向爬虫入门》课程。

# 极客学院

jikexueyuan.com

中国最大的IT职业在线教育平台

