# 十六、分析网页该如何抓取

不同网站

不同的程序员

不同的代码结构

不同的展现形式

**sitemap 完全无法复用**

**遇到新网站，该如何下手？**

# 一、判断选择器

**单个信息**

文字 —— Text 选择器（**必选**）

链接 —— Link 选择器

图片 —— Image 选择器

表格 —— Table 选择器

**多个信息**

**Element 选择器 + 其他**

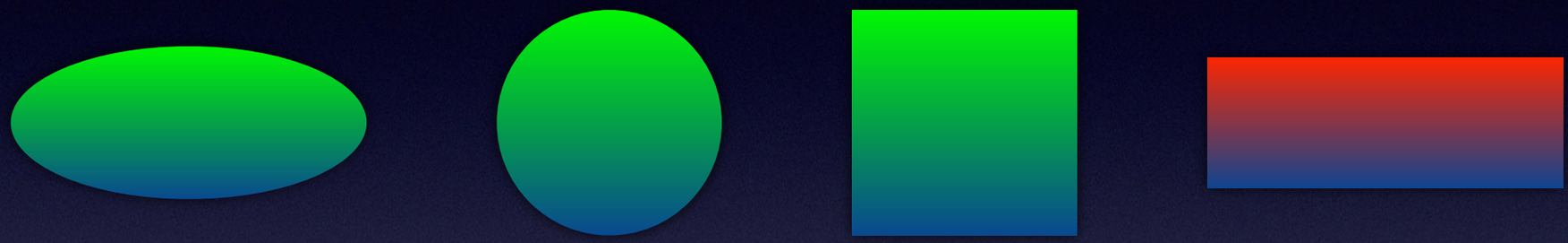
**(建议每次都使用！！)**

# 二、分析同类型信息的区域

方法：多次点击尝试

1、凭常识、感觉点击 —— 形状
2、进一步分析所需信息的类型，再次点击
3、勾选“强制”按钮，再次点击（**不建议用！！**）

# 网页中的类型



A　　B　　C　　D

**最小优先原则** ——→ 多点几次，就会有惊喜

# 抓取多页的分析方法

1、分析 URL —— 规律分页（第8节）
2、鼠标下拉加载更多 —— Element scroll down（第9节）
3、点击加载更多 —— Element click （第10节）
4、URL无规律 && 点击“页码”翻页 —— Element click（第11节）
5、页码的 URL = 网页真实链接 —— 循环分页（第12节）
6、其他 —— 看看第 19 节的奇技淫巧
7、再其他 —— 私信明白（我微信你有吧）

**先data preview确保单页数据正确，再抓多页**

# 私信明白的原则

1、上面的步骤，你 **全部** 都尝试过了

2、红包得有吧 🙂