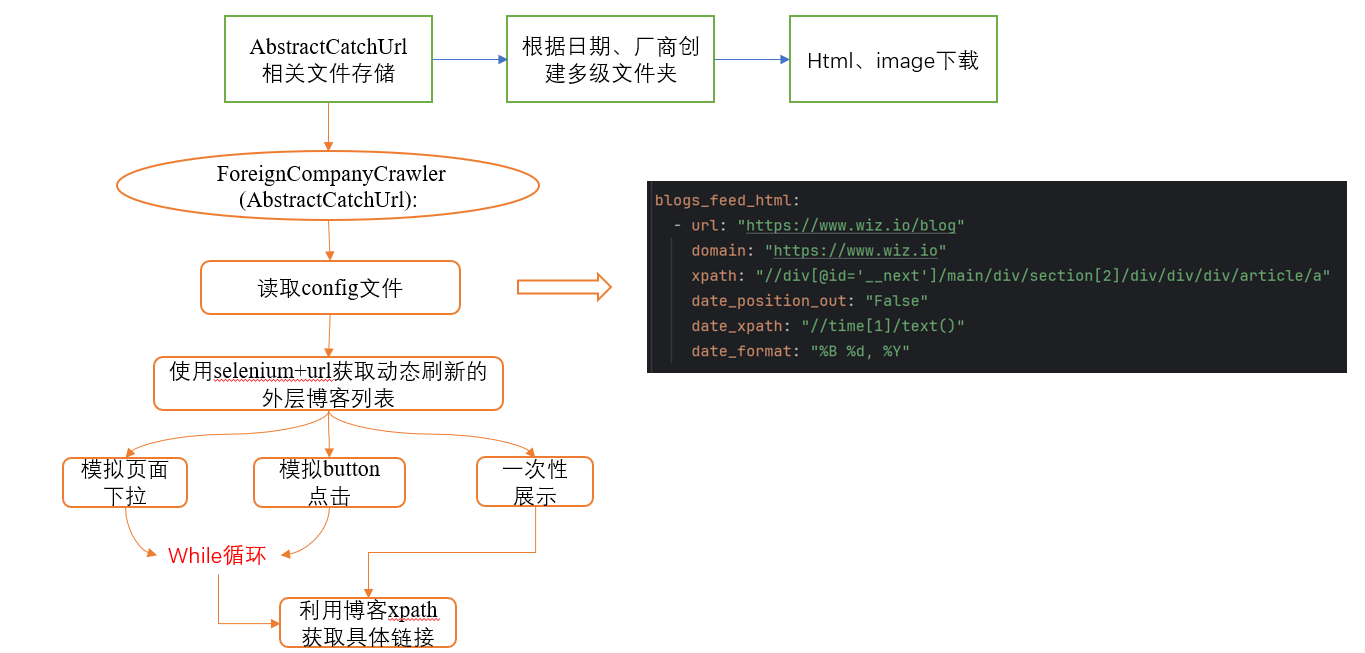
赵映雅 – 搭建三类博客的爬取框架

总体流程：



图片下载利用tree.xpath('//img/@src')，结果：

1. 内嵌的base64编码
2. 可以直接访问的图片url

处理逻辑：

* 利用“base64”字段区分两类图片
* 利用正则或者split函数提取图片扩展名称
* 使用base64decode解码Ⅰ类图片；使用request请求Ⅱ类图片