

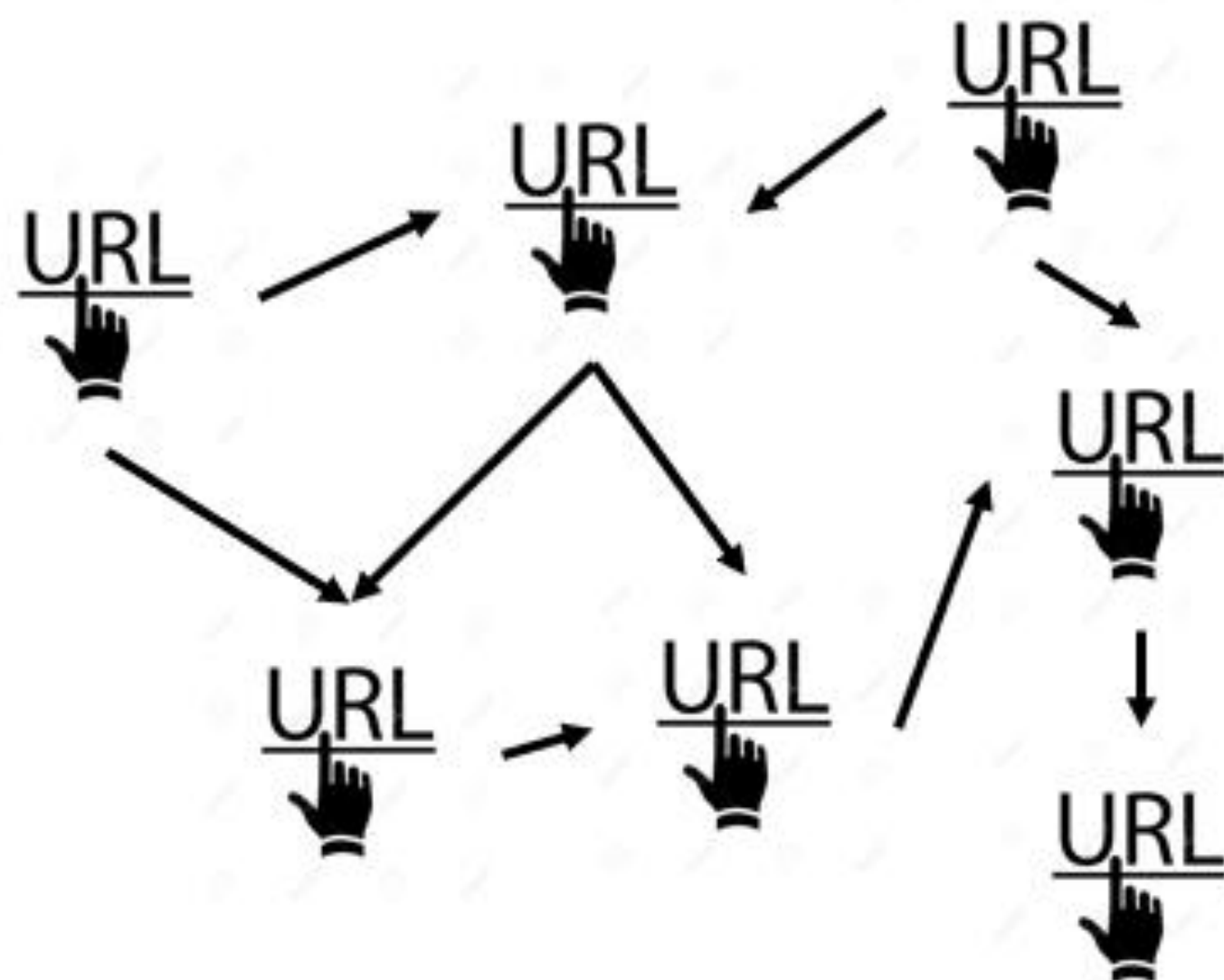
第三课 网络数据的获取与表示

第一节 爬虫简介

主 讲: Robin

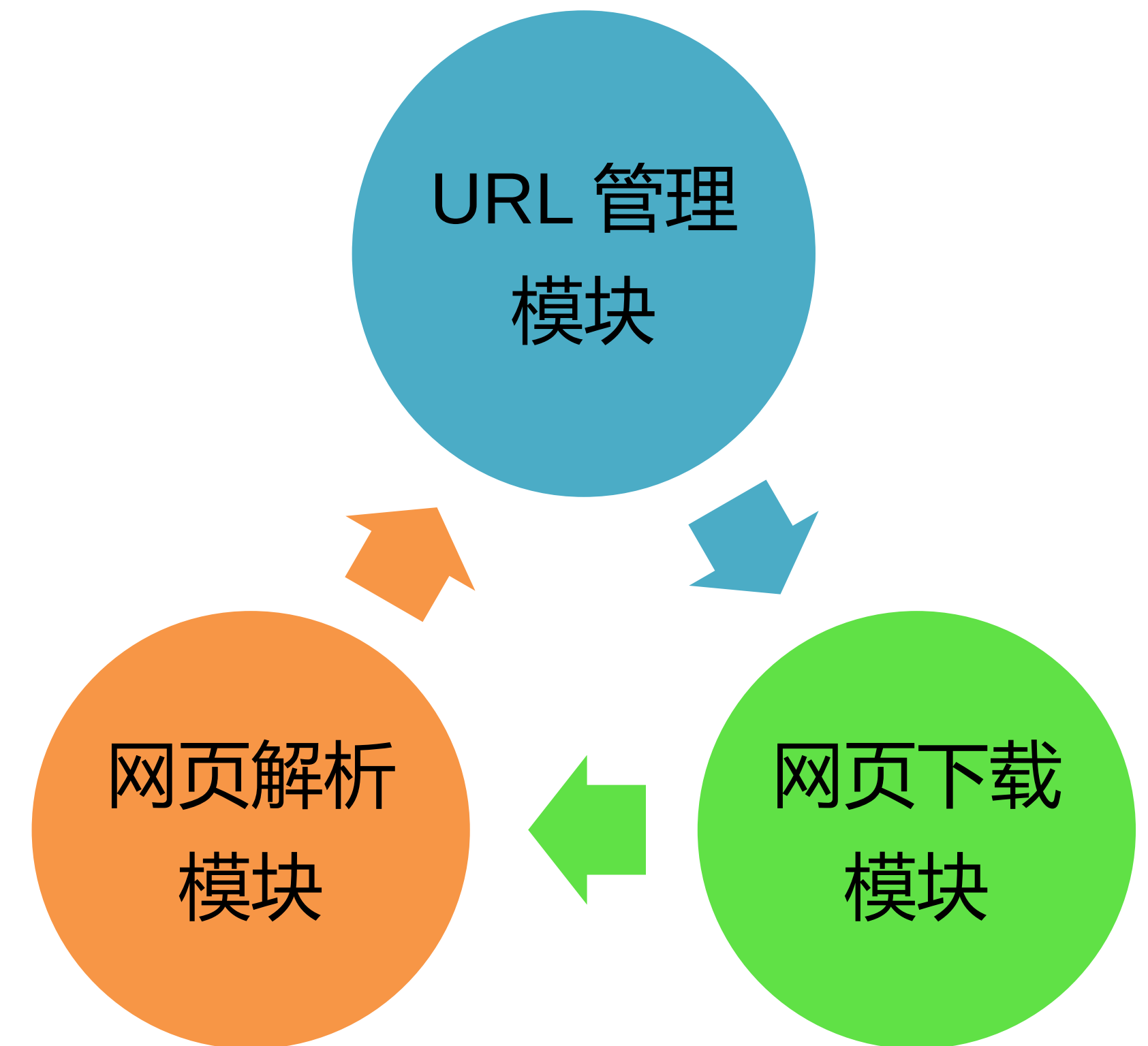
爬虫

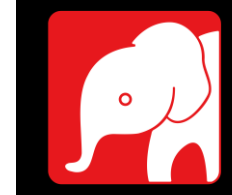
- 自动抓取互联网信息的程序
- 利用互联网数据进行分析、开发产品



基本架构

- URL 管理模块
 - 对计划爬取的或已经爬取的URL进行管理
- 网页下载模块
 - 将URL管理模块中指定的URL进行访问下载
- 网页解析模块
 - 解析网页下载模块中的URL，处理或保存数据
 - 如果解析到要继续爬取的URL，返回URL管理模块继续循环





只为遇见明天更优秀的你！