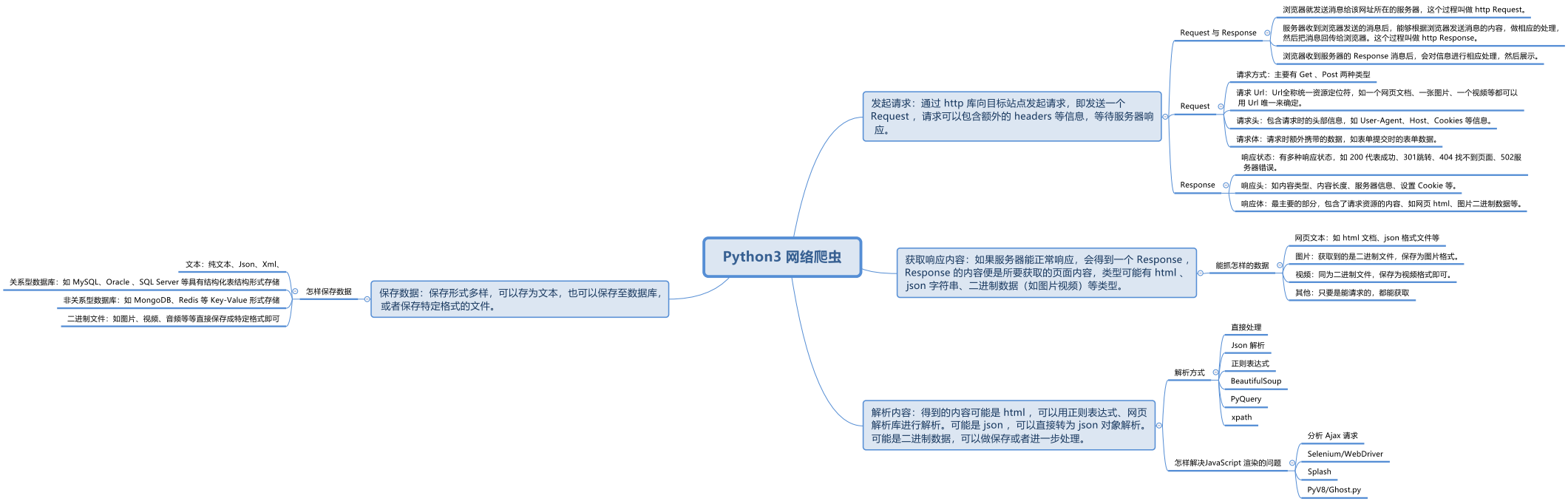
Python3 网络爬虫
发起请求：通过 http 库向目标站点发起请求，即发送一个 Request ，请求可以包含额外的 headers 等信息，等待服务器响应。
Request 与 Response
浏览器就发送消息给该网址所在的服务器，这个过程叫做 http Request。
服务器收到浏览器发送的消息后，能够根据浏览器发送消息的内容，做相应的处理，然后把消息回传给浏览器。这个过程叫做 http Response。
浏览器收到服务器的 Response 消息后，会对信息进行相应处理，然后展示。
Request
请求方式：主要有 Get 、Post 两种类型
请求 Url：Url全称统一资源定位符，如一个网页文档、一张图片、一个视频等都可以用 Url 唯一来确定。
请求头：包含请求时的头部信息，如 User-Agent、Host、Cookies 等信息。
请求体：请求时额外携带的数据，如表单提交时的表单数据。
Response
响应状态：有多种响应状态，如 200 代表成功、301跳转、404 找不到页面、502服务器错误。
响应头：如内容类型、内容长度、服务器信息、设置 Cookie 等。
响应体：最主要的部分，包含了请求资源的内容、如网页 html、图片二进制数据等。
获取响应内容：如果服务器能正常响应，会得到一个 Response ，Response 的内容便是所要获取的页面内容，类型可能有 html 、json 字符串、二进制数据（如图片视频）等类型。
能抓怎样的数据
网页文本：如 html 文档、json 格式文件等
图片：获取到的是二进制文件，保存为图片格式。
视频：同为二进制文件，保存为视频格式即可。
其他：只要是能请求的，都能获取
解析内容：得到的内容可能是 html ，可以用正则表达式、网页解析库进行解析。可能是 json ，可以直接转为 json 对象解析。可能是二进制数据，可以做保存或者进一步处理。
解析方式
直接处理
Json 解析
正则表达式
BeautifulSoup
PyQuery
xpath
怎样解决JavaScript 渲染的问题
分析 Ajax 请求
Selenium/WebDriver
Splash
PyV8/Ghost.py
保存数据：保存形式多样，可以存为文本，也可以保存至数据库，或者保存特定格式的文件。
怎样保存数据
文本：纯文本、Json、Xml、
关系型数据库：如 MySQL、Oracle 、SQL Server 等具有结构化表结构形式存储
非关系型数据库：如 MongoDB、Redis 等 Key-Value 形式存储
二进制文件：如图片、视频、音频等等直接保存成特定格式即可