1. 网络爬虫

通过HTTPClient的方式进行爬取页面 ：

**// 创建浏览器**CloseableHttpClient httpClient = HttpClients.*createDefault*();  
**// 创建请求连接**HttpGet get = new HttpGet("https://www.cnblogs.com/onepeace/p/4933567.html");  
**// 发送请求**CloseableHttpResponse response = httpClient.execute(get);  
**// 判断响应码**if(response.getStatusLine().getStatusCode() == 200){  
 HttpEntity entity = response.getEntity();  
 String s = EntityUtils.*toString*(entity);  
 System.*out*.println(s);  
}

通过工具Jsoup的方式进行进行爬取页面，可以针对 路径、文件、字符串进行解析：