**分析HTTP请求和响应数据**

打开淘宝和京东主页，分别在搜索关键字输入“耳机”，点击搜索按钮，分析浏览器发送给服务器的请求内容和响应结果，完成下面的报告。

**1.分析搜索逻辑是由哪个服务链接或者页面完成，如果有多个请分别列出，并且将发送到服务器的数据以列表方式进行说明；**

**淘宝：**

**搜索逻辑是在https://www/taobao.com跳转s.taobao.com到用“GET”方法请求发出search请求****。目标资源的URL ： /search**

**URL参数为**q=%E8%80%B3%E6%9C%BA&imgfile=&commend=all&ssid=s5- e&search\_type=item&sourceId=tb.index&spm=a21bo.2017.201856-taobao-item.1&ie=utf8&initiative\_id=tbindexz\_20170306

1. 将URL参数解码可获得以下参数：

q: 耳机 =========>我们搜索的关键词为“耳机”

Imagfile： =========>图片文件为空，这里应该是用于图片查找

commend：all ========>查找全部商品

ssid：s5-e ========> SSID是服务集标识。SSID技术可以将一个无线局域网分为几个需要不同身份验证的子网络，每一个子网络都需要独立的身份验证，只有通过身份验证的用户才可以进入相应的子网络，防止未被授权的用户进入本网络。

search\_type: item ==========>查找类型为item

sourceId: tb.index ==========>资源Id

spm: a21bo.2017.201856-taobao-item.1 ==========>淘宝提供的SPM是淘宝社区电商业务（xTao）为外部合作伙伴（外站）提供的一套跟踪引导成交效果数据的解决方案。

ie: utf8 =========>IE浏览器使用utf8字符集

initiative\_id: tbindexz\_20170306

1. **scheme:**https 页面跳转协议为https
2. **accept：**

text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,\*/\*;q=0.8,application/signed-exchange;v=b3

1. **accept-encoding:** gzip, deflate, br  =======>gzip是一种数据格式；deflate是一种压缩算法,是huffman编码的一种加强。
2. **accept-language:**  zh-CN,zh;q=0.9 ========>支持的语言
3. **referer:**  https://www.taobao.com/ ========>请求的来源
4. **sec-fetch-mode:**  navigate 与数据挖掘有关（不确定，粗略查了一下）
5. **sec-fetch-site:**  same-site
6. **sec-fetch-user:**  ?1
7. **cookie: 这里省略了**
8. **upgrade-insecure-requests:**  1
9. **user-agent:**  Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/76.0.3809.132 Safari/537.36

——用户的浏览器： Mozilla/5.0、AppleWebKit/537.36 、Chrome/76.0.3809.132 Safari/537.36

============================================================================================================================================

**京东：**

搜索服务是在https://www.jd.com/下跳转至search.jd.com使用GET方法请求/Search

**URL参数为** keyword=%E8%80%B3%E6%9C%BA&enc=utf-8&pvid=99b71f1721ec49ab9869eb76bfca4906

1. **将URL解码得到以下：**

**keyword:**  耳机 =====> 查找关键词“耳机”

**enc:** utf-8 =======>字符集为‘utf-8’

**pvid:**  99b71f1721ec49ab9869eb76bfca4906 =======>Port-base Vlan ID，也就是端口的虚拟局域网ID号，关系到端口收发数据帧时的VLAN TAG 标记

1. **Accept:**

text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,\*/\*;q=0.8,application/signed-exchange;v=b3

1. **Accept-Encoding:** gzip, deflate, br
2. **Accept-Language:** zh-CN,zh;q=0.9
3. **Connection:** keep-alive =========>希望保持长连接
4. **accept：**

text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,\*/\*;q=0.8,application/signed-exchange;v=b3

1. **accept-encoding:** gzip, deflate, br
2. **accept-language:**  zh-CN,zh;q=0.9 ========>支持的语言
3. **referer:**  https://www.taobao.com/ ======>请求的来源
4. **sec-fetch-mode:**  navigate =========>FetchMode为navigate
5. **sec-fetch-site:**  same-site
6. **sec-fetch-user:**  ?1
7. **Cookie: 这里省略**
8. **upgrade-insecure-requests:**  1
9. **user-agent:**  Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/76.0.3809.132 Safari/537.36

**2.分析服务器的响应结果，数据的格式、数据的类型以及呈现方式；**

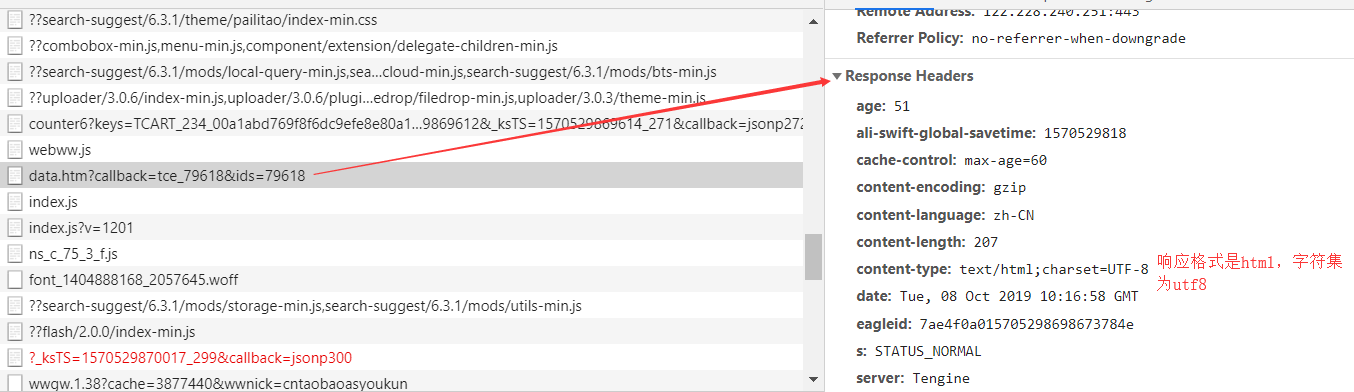
**淘宝：**

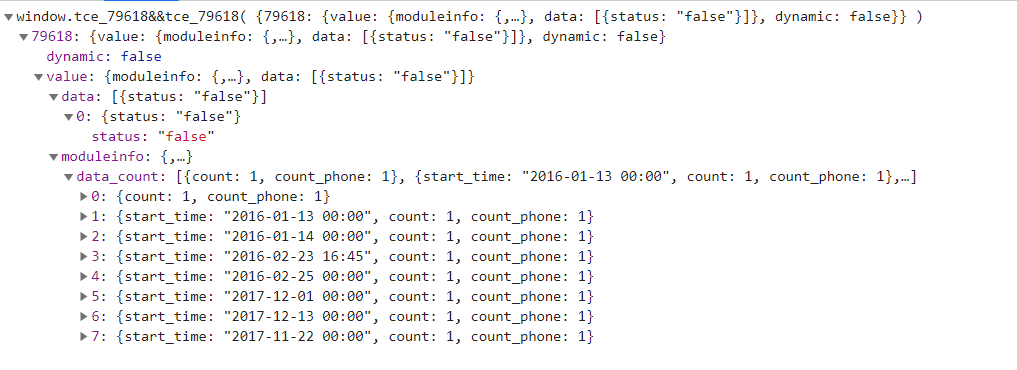
响应结果类型是DOC类型

数据格式为json格式

**content-encoding:** gzip ——内容编码

1. **content-language:** zh-CN ——语言
2. **content-type: text/html;charset=UTF-8 ——媒体格式是HTML，字符集为UTF8**
3. **date:** Sat, 07 Sep 2019 02:12:11 GMT ======>时间
4. **eagleeye-traceid:** 0b8395d715678223310263890e8419 =====>用于日志关联上下文的ID
5. **server:** Tengine/Aserver ——服务，（查了一下似乎是定制版的Nginx服务器）
6. **set-cookie:** JSESSIONID=C34647F3038811239B1B208377A4E6B7; Path=/; HttpOnly
7. **set-cookie:** alitrackid=www.taobao.com; Path=/
8. **set-cookie:** lastalitrackid=www.taobao.com; Path=/
9. **status:** 200 ======>状态码200表示请求成功
10. **strict-transport-security:** max-age=31536000 =====>是一个安全功能，它告诉浏览器只能通过HTTPS访问当前资源，最大时间31536000
11. **timing-allow-origin:** \*
12. **vary:** Accept-Encoding ======>告诉代理服务器缓存两种版本的资源：压缩和非压缩，这有助于避免一些公共代理不能正确地检测Content-Encoding标头的问题





淘宝使用的是JSONP，该协议的一个要点就是允许用户传递一个callback参数给服务端，然后服务端返回数据时会将这个callback参数作为函数名来包裹住JSON数据，这样客户端就可以随意定制自己的函数来自动处理返回数据了。（具体信息我这边不知道怎么回事看不到）

============================================================================================================================================

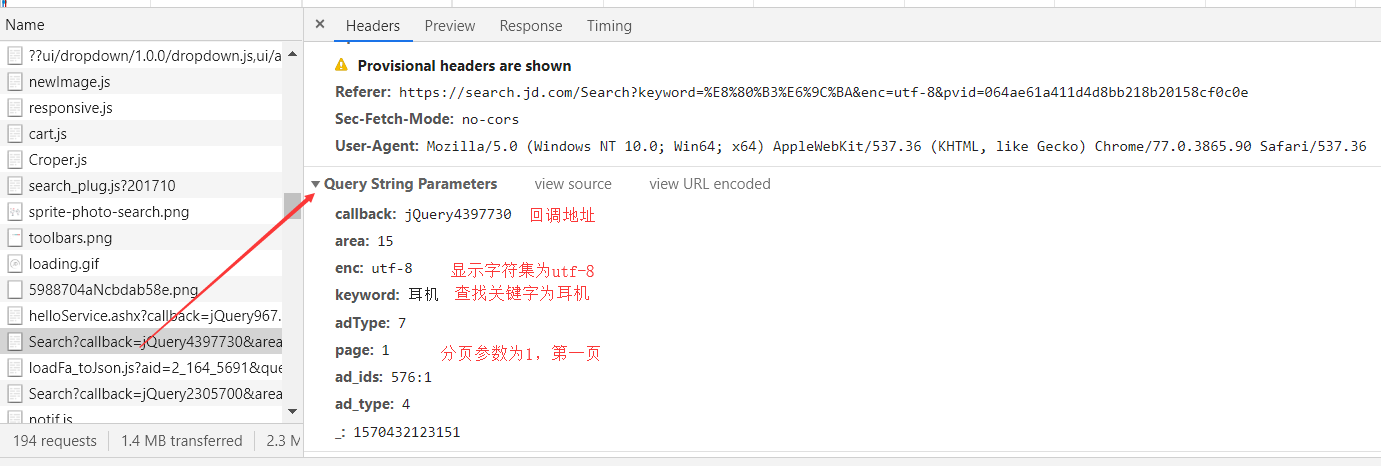
**京东：**

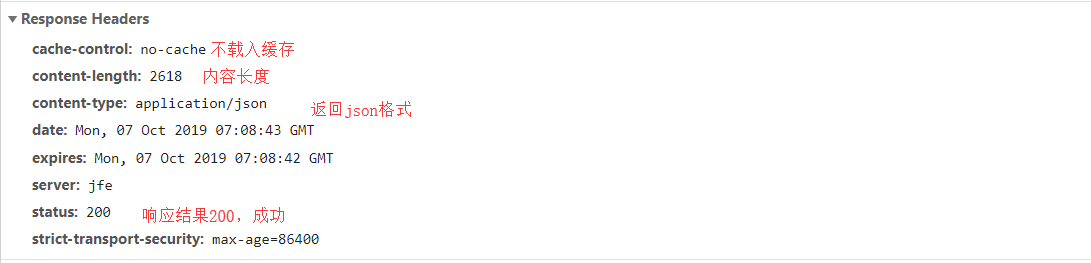
服务器响应结果为DOC类型

响应的数据格式为json格式

1. **HTTP/1.1** 200 OK =======>状态码为200即响应成功
2. **Date:** Sat, 07 Sep 2019 02:31:44 GMT =======>时间
3. **Content-Type:** text/html ======>媒体格式为HTML
4. **Transfer-Encoding:** chunked=====>传输编码
5. **Connection:** close
   ========>这里服务端希望断开长连接
6. **Vary:** Accept-Encoding
7. **Set-Cookie:** xtest=8095.cf6b6759; expires=Mon, 07-Oct-2019 02:31:44 GMT; Max-Age=2592000; domain=search.jd.com
8. **Content-Encoding:** gzip ======>内容用gzip编码，指定“Vary:Accept-Encoding”标头，网站需要启用**gzip**
9. **Server:** jfe
10. **Strict-Transport-Security:** max-age=7776000 =======>是一个安全功能，它告诉浏览器只能通过HTTPS访问当前资源，最大时间 7776000

之后展现出商品页面，点开其中一个search请求，如下图





查看具体响应参数，是json格式，内部包含了查询的商品的信息，以及需要加载资源的url



**3.对比淘宝和京东搜索功能的数据结构，请说明它们的区别和特点；**

|  |  |  |
| --- | --- | --- |
| 对比 | 淘宝 | 京东 |
| 请求参数 | **q:**  **imgfile:**  **commend:**  **ssid:**  **search\_type:**  **sourceId:**  **spm:**  **ie:**  **initiative\_id:** | **keyword:**  **enc:**  **pvid:** |
| 总请求数目 | 106 | 122 |
| 总资源大小 | 2.8MB | 1.9MB |
| DOM加载时间 | 981ms | 1.03s |

1. 淘宝和京东在发出搜索请求的时候淘宝的请求参数要多于京东
2. 淘宝使用JSONP方式处理跨域请求问题。
3. 淘宝采用了eagle eye进行监控，每次服务端响应的时候会返回一个与上下文关联的ID
4. 淘宝发送请求后服务端响应信息中 **timing-allow-origin:** \*指的是用于指定特定站点，以允许其访问Resources Timing API提供的相关信息，否则这些信息会由于跨源限制将被报告为零，\*是通配符从而允许所有域都具有访问定时信息的权限。
5. 京东的请求参数比淘宝要少一些，并且在响应内容中**Connection**为**close，但是请求的时候请求头中Connection为keep-alive**

 在http1.1中，client和server都是默认对方支持长链接的， 如果client使用http1.1协议，但又不希望使用长链接，则需要在header中指明connection的值为close；如果server方也不想支持长链接，则在response中也需要明确说明connection的值为close。 那么在这里也就是说，客户端希望保持连接，但是服务端将客户端的请求完成后希望断开连接。