

实验一 分析HTTP请求和响应数据

学号：2017212212133

姓名：施莹丹

班级：计算机174

一、实验内容

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；
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；
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；

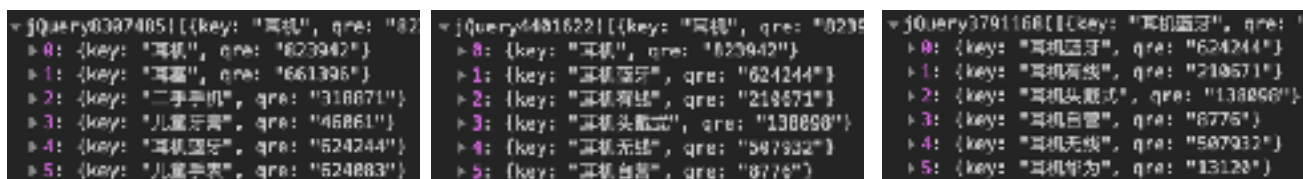
二、实验结果

1、

(1) 搜索逻辑

京东

- 输入变量：返回三份文件，均为耳机相关的模糊词汇。如下图。



```
< jQuery03074051[{"key": "耳机", "qre": "823942"}, {"key": "耳塞", "qre": "661396"}, {"key": "二手手机", "qre": "316671"}, {"key": "儿童牙膏", "qre": "46861"}, {"key": "耳机蓝牙", "qre": "624244"}, {"key": "儿童手表", "qre": "624883"}]
< jQuery44816221[{"key": "耳机", "qre": "823942"}, {"key": "耳机蓝牙", "qre": "624244"}, {"key": "耳机有线", "qre": "218671"}, {"key": "耳机头戴式", "qre": "138890"}, {"key": "耳机无线", "qre": "587932"}, {"key": "耳机台历", "qre": "8776"}]
< jQuery37911661[{"key": "耳机蓝牙", "qre": "624244"}, {"key": "耳机有线", "qre": "218671"}, {"key": "耳机头戴式", "qre": "138890"}, {"key": "耳机台历", "qre": "8776"}, {"key": "耳机无线", "qre": "587932"}, {"key": "耳机华为", "qre": "13128"}]
```

- 点击搜索：向<https://search.jd.com>发送参数，传递的参数分析见下方表格的参数分析。
- 返回结果：一个HTML界面

淘宝

- 输入变量：输入“er”就会出现8个热门的模糊词汇如“耳钉”、“耳环”，以及各词汇的后接热词，返回jsonp2267。如，耳钉及其热词“欧美”、“日韩”等。

举一个模糊词汇“耳机”内容为：

```
jsonp2267({"result": [{"耳机", "1090645.654299321"}, {"耳
环", "3578527.026986378"}, ... ],
"magic": [{"index": "1", "type": "tag_group", "data": [{"title": "耳塞式"}, {"title": "头戴式"}, {"title": "入耳式"}, {"title": "挂耳式", "type": "hot"}], [{"title": "重低音"}, {"title": "魔声"}, {"title": "Sony\\索尼"}, {"title": "森海塞尔"}, {"title": "无线"}, {"title": "漫步者", "type": "hot"}]}, ... ])
Index的下标表示第几个模糊词汇对应的热门后接词。
```

- 点击搜索：向<https://s.taobao.com>发送参数，传递的参数分析见下方表格的参数分析。
- 返回结果：一个HTML界面

(2) 参数

京东：

搜索链接：<https://search.jd.com>

参数：

参数名	参数内容	参数意义
keyword	耳机	搜索的变量名
enc	utf-8	变量的字符编码
pvid	b84fb3c2736048c78fa6e3 85f1238d75	Port-base Vlan ID，登陆端口的虚拟局域网的ID账号，和端口收发数据时的VLAN TAG有着直接的关系

淘宝：

搜索链接：<https://s.taobao.com>

未登录时在首页进行搜索的参数传输：

参数名	参数内容	参数意义
q	耳机	搜索的变量名
imgfile	''	淘宝有图片检索功能，此处为空
commend	all	推荐所有的“耳机”产品
ssid	s5-e	
search_type	item	搜索的类型为物品。淘宝共有宝贝、店铺两种搜索选择。
sourceId	tb.index	搜索的来源是tb.index这个页面
spm	a21bo.2017.201856- taobao-item.1	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。用来跟踪页面模块位置的编码，标准spm编码由4段组成，采a.b.c.d的格式。 <ul style="list-style-type: none"> • a代表站点类型，对于xTao合作伙伴(外站)； • b代表外站ID(即外站所使用的TOP appkey)； • c代表b站点上的频道ID，比如是外站某个团购频道，某个逛街频道，某个试用频道等； • d代表c频道上的页面ID，比如是某个团购详情页，某个宝贝详情页，某个试用详情页等； 完整的SPM四位编码能标识出某网站中某一个频道的某一个具体页面。

ie	utf8	变量的字符编码，采用utf-8的编码格式
initiative_id	staobaoz_20190927	该字段包含的数据有搜索的日期

未登录状态下在搜索界面进行搜索的参数：

参数名	参数内容	参数意义
q	耳机	搜索的变量名
imagefile		淘宝有图片检索功能，此处为空
js	1	增加筛选条件、改为搜索店铺、更改登录状态，该值均未变，因此无法分析
stats_click	search_radio_all:1	增加筛选条件、改为搜索店铺、更改登录状态，该值均未变，因此无法分析
initiateve_id	staobaoz_20190927	该字段包含的数据有搜索的日期
ie	utf8	变量的字符编码，采用utf-8的编码格式

登录后的搜索参数：

参数名	参数内容	参数意义
q	耳机	搜索的变量名
js	1	增加筛选条件、改为搜索店铺、更改登录状态，该值均未变，因此无法分析
initiative_id	staobaoz_20190927	该字段包含的数据有搜索的日期
ie	Utf-8	变量的字符编码，采用utf-8的编码格式

2、

	京东	淘宝
响应结果	200 OK	200 OK
数据格式	jquery	jsonp
数据类型	Doc	Doc、XHR
呈现方式	html页面	html页面

3、淘宝和京东搜索功能的数据结构的区别与特点

	京东	淘宝
传输数据量	少	多
数据搜索时间复杂	略快于淘宝 (6.40s)	(9.51s)