

同濟大學

TONGJI UNIVERSITY

《WEB 技术》

实验报告（大作业）

实验名称

万维网运行原理分析

小组成员

徐海航(2153484)

学院（系）

电子与信息工程学院

专 业

计算机科学与技术专业

任课教师

郭玉臣

日 期

2023 年 3 月 12 日

1 网页选择

选择网站为: <https://1.tongji.edu.cn/workbench>

工作台

通知公告 [更多](#)

- 2022-2023学年第二学期本科生重修选课通知 2023-02-27 10:02:07
- 2022-2023学年第二学期结业生（含补授学位）选课通知 2023-02-27 10:02:02
- 关于2023届毕业设计（论文）近期工作及时间节点的通知 2023-02-21 10:54:04
- 关于同济大学《大学美育》学习指南的通知 2023-02-17 09:40:27
- 2022-2023学年第二学期本研互选通知 2023-02-16 09:15:12
- 关于2022-2023学年第二学期第四轮选课安排（包括跨学科选课）及选课... 2023-02-16 09:15:08
- 2023年第一学期公共课线上教学平台上线支持选课事宜通知 2023-02-13 10:22:40

查看课表

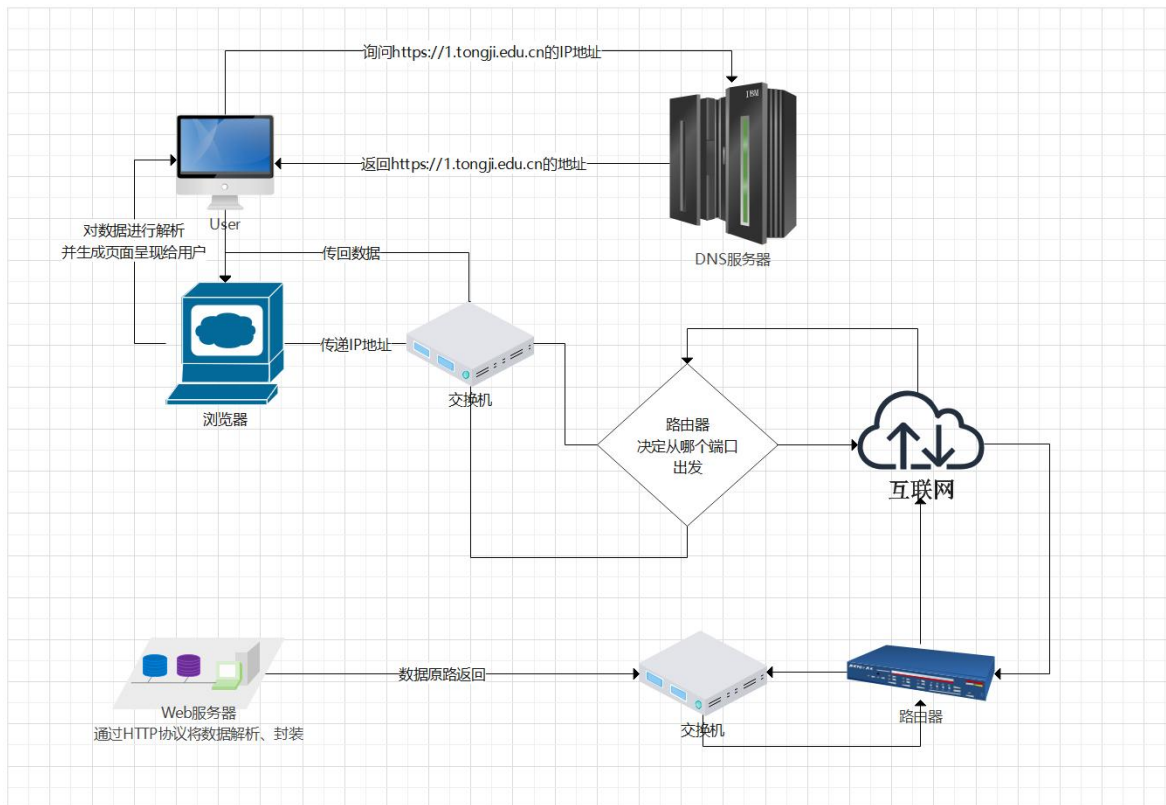
《 < 2023年 3月 > 》

日	一	二	三	四	五	六
26	27	28	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	1
2	3	4	5	6	7	8

暂无数据...

[切换旧版](#)

2 绘制分析网络拓扑和数据流向



3 分析网页组成

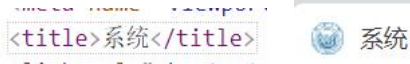
网页组成部分可以分为首部和主体

3.1 首部

```
<head>
  <meta charset="utf-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
  <meta name="renderer" content="webkit">
  <meta name="viewport" content="width=device-width,initial-scale=1,maximum-scale=1,user-scalable=1">
  <title>同济大学-教学管理信息系统</title>
  <link rel="shortcut icon" type="images/x-icon" href="/static/images/logo.ico">
  <style type="text/css">...</style>
  <link href="/static/css/app.3cfcfb2...css" rel="stylesheet">
  <link rel="stylesheet" type="text/css" href="/static/css/165.7c294ba...css">
  <script charset="utf-8" src="/static/js/165.ce90f6a...js"></script>
  <script charset="utf-8" src="/static/js/36.37792ba...js"></script>
  <link rel="stylesheet" type="text/css" href="/static/css/289.8619caa...css">
  <script charset="utf-8" src="/static/js/289.082eef3...js"></script>
</head>
```

首部包含文档的标题(title), 以及系统用来标识文档的一些其他信息。标题相当于文件名。

<head>部分中的 <title>同济大学-教学管理信息系统</title>部分 即为网页的标题，更改<title>标签里面的部分可以发现网页的标题得到了更改



3.2 主体

是 HTML 文档最主要的部分，往往又由若干个更小的元素组成，如段落 (paragraph)、表格(table)和列表(list)等

```
...<body class style="overflow: hidden;" == $0
  <script type="text/javascript" src="/static/js/explorerCheck.js"></script>
  <script type="text/javascript" src="/static/js/vue.min.js"></script>
  <script type="text/javascript" src="/static/js/lodash.min.js"></script>
  <script type="text/javascript" src="/static/js/moment.min.js"></script>
  <script type="text/javascript" src="/static/js/echarts.min.js"></script>
  <script type="text/javascript" src="/static/js/esdk-obs-browserjs-without-polyfill.3.22.3.min.js"></script>
  <script type="text/javascript" src="/static/tinymce4.7.5/tinymce.min.js"></script>
  <script src="/static/jspdf.1.5.3/jspdf.min.js"></script>
  ><div id="app">...</div>
  <script type="text/javascript" src="/static/js/app.bef70ba...js"></script>
  ><div style="position: static; display: none; width: 0px; height: 0px; border: none; padding: 0px; margin: 0px;">...</div>
</body>
```

在所选网站中主体部分包含了对外部 js 文件的引入，以及两个 div 的划分。这两个 div 内部又有详细的划分和功能实现

比如：

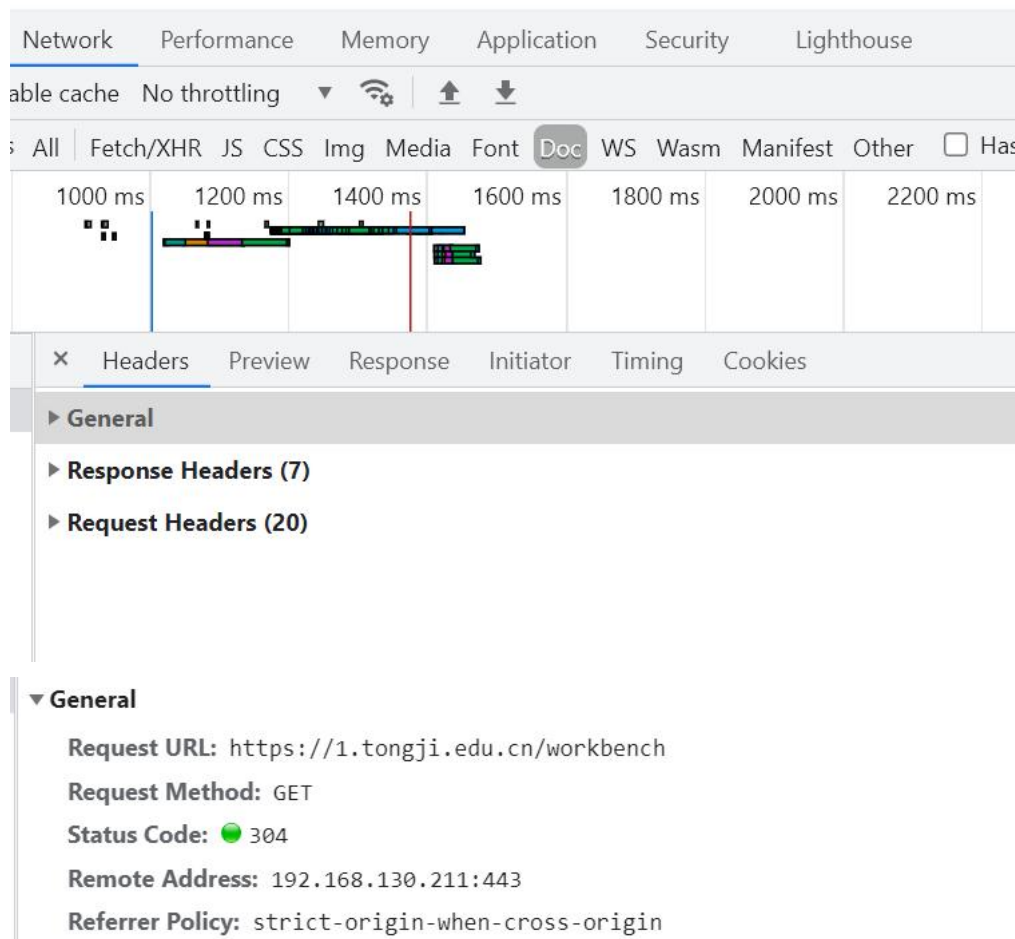
```

▼<div id="app">
  ▼<div data-v-f9858724 style="display: inline;">
    ▼<div data-v-38de9b68 data-v-f9858724 style="height: 100%;">
      ▼<div data-v-0f62e21e data-v-38de9b68 class="edu-header">
        ▼<div data-v-0f62e21e class="edu-header-left">
          ▼<div data-v-0f62e21e class="float-left">
            <div data-v-0f62e21e class="show-menus-btn close"></div>
          </div>
          <div data-v-0f62e21e class="edu-header-logo"></div>
        </div>
        ▼<div data-v-0f62e21e class="edu-header-right"> (flex)
          ▼<div data-v-5db3a1d4 data-v-0f62e21e class="benchHeader"> (flex)
            ▼<div data-v-5db3a1d4 class="userInfo">
              
              ▼<div data-v-5db3a1d4 class="infoContent"> (flex)
                ▼<div data-v-5db3a1d4>
                  <b data-v-5db3a1d4 class="user-name">Hi ~ 徐海航</b>
                  <span data-v-5db3a1d4 class="user-dept">电子与信息工程学院</span>
                </div>
                ▼<div data-v-5db3a1d4 class="calender">
                  <span data-v-5db3a1d4 class="el-tag el-tag--small el-tag--light">2022-2023学年度第2学期</span>
                  <span data-v-5db3a1d4 class="el-tag el-tag--small el-tag--light">第4周 </span>
                </div>
              </div>
            </div>
            ▼<div data-v-5db3a1d4 class="loginInfo">
              ▼<li data-v-5db3a1d4 style="margin-bottom: 5px;">
                " 上次登录: "
                <span data-v-5db3a1d4>2023.03.16 09:56:29</span>
              </li>
            </div>
          </div>
        ▼<div data-v-0f62e21e class="edu-item">
          ▼<div data-v-30bc5512 data-v-0f62e21e class="edu-msgContent"> (flex) == $0
            ▼<div data-v-30bc5512 class="el-badge bellbadge">
              <div data-v-30bc5512 class="edu-msg"></div>
              <sup class="el-badge__content is-fixed" style="display: none;">0</sup>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>

```

该划分采用”app”的标签，内部也包含了不同划分以及划分内的样式

4 分析网页 HTTP 协议



一些通用信息的报文会显示在此处

网址

Request URL: `https://1.tongji.edu.cn/workbench`

请求方法

Request Method: GET

HTTP 状态码，304 表示 Not Modified

如果客户端发送了一个带条件的 GET 请求且该请求已被允许，而文档的内容（自上次访问以来或者根据请求的条件）并没有改变，则服务器应当返回这个状态码。

Status Code: 304

代表的是当前 HTTP 请求的远程地址，即 HTTP 请求的源地址。

Remote Address: 192.168.130.211:443

对于同源的请求，会发送完整的 URL 作为引用地址；在同等安全级别的情况下，发送文件的源作为引用地址(HTTPS->HTTPS)；在降级的情况下不发送此首部 (HTTPS->HTTP)。

Referrer Policy: strict-origin-when-cross-origin

请求报文

```
GET /workbench HTTP/2
Host: 1.tongji.edu.cn
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:105.0) Gecko/20100101 Firefox/105.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8
Accept-Language: zh-CN,zh;q=0.8,zh-TW;q=0.7,zh-HK;q=0.5,en-US;q=0.3,en;q=0.2
Accept-Encoding: gzip, deflate, br
Connection: keep-alive
Cookie: JSESSIONID=B64D3BD0A947FD7F058DFC382C46E3F2; sessionId=d7019732601c405086ff6b0188e44a63; language=cn
Upgrade-Insecure-Requests: 1
Sec-Fetch-Dest: document
Sec-Fetch-Mode: navigate
Sec-Fetch-Site: same-site
If-Modified-Since: Thu, 16 Mar 2023 11:29:51 GMT
If-None-Match: "6412fdaf-4e7"
```

请求行，包括 HTTP 方法和请求对应的 URL 地址以及协议名称和版本号

GET /workbench HTTP/2

报文头，格式为“属性名：属性值”

Host: 1.tongji.edu.cn

用户代理请求头包含一个特征串，其允许网络协议对等体，以确定请求软件的用户代理的应用程序类型，操作系统，软件供应商或软件版本。

User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:105.0) Gecko/20100101 Firefox/105.0

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8

Accept-Language: zh-CN,zh;q=0.8,zh-TW;q=0.7,zh-HK;q=0.5,en-US;q=0.3,en;q=0.2

Accept-Encoding: gzip, deflate, br

Connection: keep-alive

客户端的 Cookie 就是通过这个报文头属性传给服务端的。

Cookie:JSESSIONID=B64D3BD0A947FD7F058DFC382C46E3F2;sessionId=d7019732601c405086ff6b0188e44a63; language=cn

在 https 页面中，如果调用了 http 资源，那么浏览器就会抛出一些错误。因此会在 http 请求中加入 ‘Upgrade-Insecure-Requests: 1’，服务器收到请求后会返回 “Content-Security-Policy: upgrade-insecure-requests” 头，告诉浏览器，可以把所属本站的所有 http 连接升级为 https 连接。

Upgrade-Insecure-Requests: 1

Fetch 元数据请求标头指示请求的目标，即数据的来源以及如何使用这些获取到的数据

Document 表示目标是文档（HTML 或 XML），请求是用户发起的顶级导航的结果（例如，由用户单击链接产生）。

Sec-Fetch-Dest: document

该请求头表明了一个请求的模式。

Sec-Fetch-Mode: navigate

请求发起者的来源与目标资源来源之间的关系。

Sec-Fetch-Site: same-site

IMS，表示浏览器缓存记录中该文件的最后服务器修改时间，由客户端发送给服务器。

If-Modified-Since: Thu, 16 Mar 2023 11:29:51 GMT

HTTP 请求报头使得所述请求为条件

If-None-Match: "6412fdaf-4e7"

响应报文

```
HTTP/2 304 Not Modified
date: Sat, 18 Mar 2023 06:52:32 GMT
last-modified: Thu, 16 Mar 2023 11:29:51 GMT
etag: "6412fdaf-4e7"
server: elb
X-Firefox-Spdy: h2
```

响应行，包括报文协议及版本，状态码及状态描述，304 表示资源未修改，可读取本地缓存

HTTP/2 304 Not Modified

响应头

date: Sat, 18 Mar 2023 06:52:32 GMT

表示服务器本地记录中的文件最后修改时间，由服务器发送给客户端。

last-modified: Thu, 16 Mar 2023 11:29:51 GMT

一个代表响应服务端资源（如页面）版本的报文头属性，如果某个服务端资源发生了变化，这个 ETag 就会相应发生变化。

etag: "6412fdaf-4e7"

首部包含了处理请求的源头服务器所用到的软件相关信息。

server: elb

h2 字段表示其采用的是火狐框架的 SPDY

X-Firefox-Spdy: h2

5 实验总结

该实验让我认识到了如何去绘制一个网页的网络拓扑图和数据流向，同时也让我对网页即 HTML 的组成要素有了比较深入的了解，通过对所选网页的 HTTP 协议进行分析，可以通过请求报文和应答报文以及其中的关键点得到网页的部分信息以及访问状态等等。