URI和URL及URN的区别

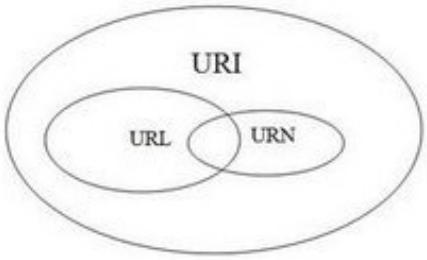
来源: biaodianfu 发布时间: 2011-02-14 22:58 阅读: 2139 次 推荐: 1 原文链接 [收藏]

对于URL,大家都比较熟悉，其他两个词就比较陌生了。URI、URL和URN是识别、定位和命名互联网上的资源的标准途径。1989年Tim Berners-Lee发明了互联网（World Wide Web）。WWW被认为是全球互连的实际的和抽象的资源的集合–它按需求提供信息实体–通过互联网访问。实际的资源的范围从文件到人，抽象的资源包括数据库查询。

因为要通过多样的方式识别资源（人的名字可能相同，然而计算机文件只能通过唯一的路径名称组合访问），所以需要标准的识别WWW资源的途径。为了满足这种需要，Tim Berners-Lee引入了标准的识别、定位和命名的途径：URI、URL和URN。

- URI：Uniform Resource Identifier，统一资源标识符；
- URL：Uniform Resource Locator，统一资源定位符；
- URN：Uniform Resource Name，统一资源名称。

在这个体系中的URI、URL和URN是彼此关联的。URI的范畴位于体系的顶层，URL和URN的范畴位于体系的底层。这种排列显示URL和URN都是URI的子范畴。



三者中，其中URL和URI特别容易混淆。

URL是Internet上用来描述信息资源的字符串，主要用在各种WWW客户程序和服务器程序上。采用URL可以用一种统一的格式来描述各种信息资源，包括文件、服务器的地址和目录等。

URL的格式由下列三部分组成：

1. 协议（或称为服务方式）；
2. 存有该资源的主机IP地址（有时也包括端口号）；
3. 主机资源的具体地址。如目录和文件名等。

第一部分和第二部分之间用”：//”符号隔开，第二部分和第三部分用”/”符号隔开。第一部分和第二部分是不可缺少的，第三部分有时可以省略。

目前最大的缺点是当信息资源的存放地点发生变化时，必须对URL作相应的改变。因此人们正在研究新的信息资源表示方法。

URI是以某种统一的（标准化的）方式标识资源的简单字符串，一般由三部分组成：

1. 访问资源的命名机制。
2. 存放资源的主机名。
3. 资源自身的名称，由路径表示。

典型情况下，这种字符串以scheme开头，语法如下：

[scheme:] scheme-specific-part

http://www.google.com，其中http是scheme，//www.google.com是 scheme-specific-part，并且它的scheme与scheme-specific-part被冒号分开了。

有的URI指向一个资源的内部。这种URI以”#”结束，并跟着一个anchor标志符（称为片断标志符）。

相对URI不包含任何命名规范信息。它的路径通常指同一台机器上的资源。相对URI可能含有相对路径（如：“..”表示上一层路径），还可以包含片断标志符。

URI的常见问题

- 难以输入，URI不必要的冗长。
- 莫明其妙的大写字母。
- 不常见的标点符号。
- 在纸介质上显示很困难，一些字符在纸上打印出来不容易辨认。
- 主机和端口的问题除了 scheme-specific 部分，domain 和port 也可能给用户带来困惑。

设计**URI**应该遵循的规则（具体还可以参考上一篇：优秀的**URI**不会改变）

URI 是网站UI的一部分，因此，可用的网站应该满足这些URL 要求

- 简单，好记的域名
- 简短（short）的URI
- 容易录入的URI
- URI 能反应站点的结构
- URI 是可以被用户猜测和hack的（也鼓励用户如此）
- 永久链接，Cool URI don’t change

聪明的选择**URI**

一定要短 为了URI能被方便的录入，写下，拼写和记忆，URI 要尽可能的短，根据w3c 提供的参考数据，一个URI 的长度最好不要超过80个字节（这并非一个技术限制，经验和统计提供的数据），包括

schema 和host,port 等。

大小写策略 URI的大小写策略要适当，要么全部小写，要么首字母大写，应避免混乱的大小写组合，在Unix 世界，文件路径队大小写是敏感的，而在Windows 世界，则不对大小写敏感。

允许**URI**管理 URI映射 管理员可以重新组织服务器上的文件系统结构，而无需改动URI，这就需要URI和真实的服务器文件系统结构之间有一个映射机制。，而不是生硬的对应。这种映射机制可以通过如下技术手段实现：

- Aliases ，别名，Apache 上的目录别名，IIS 上的虚拟目录
- Symbolic links ，符号链接，Unix 世界的符号链接
- Table or database of mappings ，数据库映射，URI 和文件系统结构的对应关系存储在数据库中。

标准的重定向 管理员可以简单的通过修改HTTP 状态代码来实现服务器文件系统结构变更之后的URI兼容，可以利用的HTTP Status Code 有：

- 301 Moved Permanently ([RFC2616] section 10.3.2)
- 302 Found (undefined redirect scheme, [RFC2616] Section 10.3.3)
- Temporary Redirect ([RFC2616] Section 10.3.8)

用独立的**URI**

技术无关的URI

- 提供动态内容服务时，应使用技术无关的URI。即URI不暴露服务器端使用的脚本语言，平台引擎，而这些语言，平台，引擎的变化也不会导致URI的变更。因此，sevelet,cgi-bin之类的单词不应该出现在URI 中。
- 提供静态内容服务时，应当隐去文件的扩展名取而代之的技术是content-negotiation, proxy, 和 URI mapping

身份标志和Session 机制

- 使用标准的身份认证机制，而不是每个用户一个特定的URI
- 使用标准的Session 机制，而不是把Session ID 放在URI 中使用。

内容变更时使用标准转向

- 对变更的内容使用标准的重定向
- 对删除的资源使用 HTTP410

提供索引代理

索引策略

- Content-Location
- Content-MD5

提供适当的缓存信息

- 缓存相关的HTTP头
- 缓存策略
- 缓存生成内容 HTTP HEAD和HTTP GET

总结

- URI 是Web UI 的一部分，应当像对待网站Logo 和公司品牌一样对待它
- URI 是网站和普通用户之间的唯一接口，应当像对待你的商务电话号码一样对待它

读懂并记住上面两句话，你下次设计URI 的时候就会给它应有的重视了。

- URL 应当是用户友好的
- URI 应当是可读的
- URI 应当是可预测的
- URI 应当是统一的

读懂和记住上面四句话，你就知道应该设计什么样的URI了。