WEB的三大基础：**URI**、http和html。由此可见URI的重要性。

当前的URI几乎等价于URL，即统一资源定位符。互联网上任何资源都是通过URI来定位的。

**URL组成**

<scheme>://<username>:<password>@<host>:<port>/<path>;<param>?<query>#frag

* scheme：协议
* username:password：用户访问时输入的用户名和密码，可选，若不填，则为匿名访问
* host：port：主机名和端口，其中主机名可用ip地址代替
* path：资源的本地路径，其根路径不是/，而是自定义的
* param：某些协议会用这个组件来指定输入参数。参数为键/值对。URL可以包含多个参数字段，它们相互之间以及与路径的其余部分之间用分号(；)隔开。
* query：某些协议会用这个组件传递参数以激活应用程序（比如数据库，搜索引擎以及其他因特网网关）
* frag：一小片或一部分资源的名字。这个字段是在客户端内部使用的。不会传递给服务器的。

**URL转换**

Html文档中可以使用相对URL路径，但在http会话中必须是URL绝对路径，因此，需要做URL路径转换。

**URL字符集**

考虑到以下情况：

* URL必须统一的定义互联网上的资源，并且**所有的互联网协议**都能够通过URL进行资源传送。不同的互联网协议的工作机制千差万别，因此，某些特殊字符必须用%进行转义。
* URL必须是人类友好的，因此如果要使用不可打印字符，必须用%进行转义。