

更多嵌入式 Linux 学习资料, 请关注: 一口 Linux 回复关键字:1024



当我们打开浏览器, 要访问一个网站或者一个 ftp 服务器的时候, 一定要输入一串字符串, 比如:

```
https://blog.csdn.net/
```

或者:

```
ftp://192.168.0.111/
```

这样我们就可以得到一个 html 格式的页面或者一个文件。

那么这个地址是什么意思呢?

就必须要从 **URI**、**URL**、**URN** 讲起。

## 一、URI、URL、URN 概念

---

- URI = Uniform Resource Identifier 统一资源标志符
- URL = Uniform Resource Locator 统一资源定位符
- URN = Uniform Resource Name 统一资源名称

看了这个概念相信大家还是不明白什么意思,

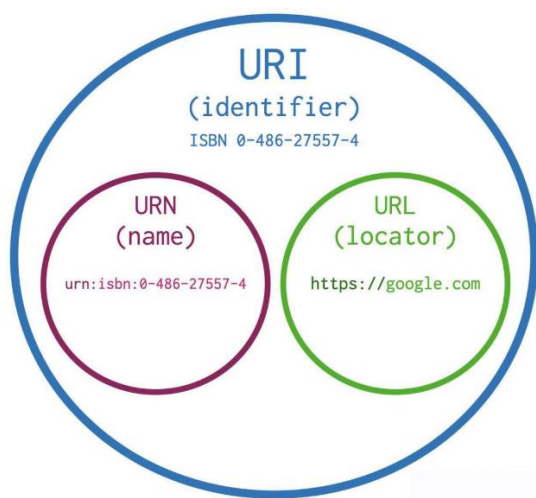
简单来说, 就是 URI 是抽象的定义, 不管用什么方法表示, 只要能定位一个资源, 就叫 URI。

本来设想的的使用两种方法定位: 1, URL, 用地址定位; 2, URN 用名称定位。

举个例子: 去村子找个具体的人 (URI), 如果用地址: 某村多少号房子第几间房的主人 就是 URL, 如果用身份证号+名字 去找就是 URN 了。

原来 uri 包括 url 和 urn, 后来 urn 没流行起来, 导致几乎目前所有的 uri 都是 url。

三者之间几何关系如下:



其实一直有个误解, 很多人以为 URI 是 URL 的子集, 其实应该反过来。URL 是 URI 的子集才对。

URI RFC 3986

## URL 是什么

---

URL 代表着是统一资源定位符 (UniformResourceLocator)。

作用是为了告诉使用者 某个资源在 Web 上的地址。

这个资源可以是一个 HTML 页面, 一个 CSS 文档, 一幅图像或一个猫片等等。

比如:

用 HTTP 协议访问 Web 服务器:



用 FTP 协议下载和上传文件时



读取客户端计算机本地文件时



这里面细分, 又可以分为好几个部分。

## 协议

尽管 URL 有各种不同的写法, 但它们有一个共同点, 开头部分的内容必须是协议类型,

可以是 `http`、`ftp`、`mailto` 或者 `https`, 这部分文字都表示浏览器应当使用的访问方法。 , 会用 `//` 为分隔符。

决定了后面部分的写法, 因此并不会造成混乱。

## 用户名/密码

用户名密码通常可以省略。

## 域名

域名是 **www.gitee.com**, 在发送请求前, 会向 DNS 服务器解析 IP。如果已经知道 ip, 还可以跳过 DNS 解析那一步, 直接把 IP 当做域名部分使用。

## 端口

域名后面有些时候会带有端口, 和域名之间用 **:** 分隔, 端口不是一个 URL 的必须的部分。当网址为 **http://** 时, 默认端口为 80, **https://** 时, 默认端口是 443, **ftp://** 时, 默认端口是 21。

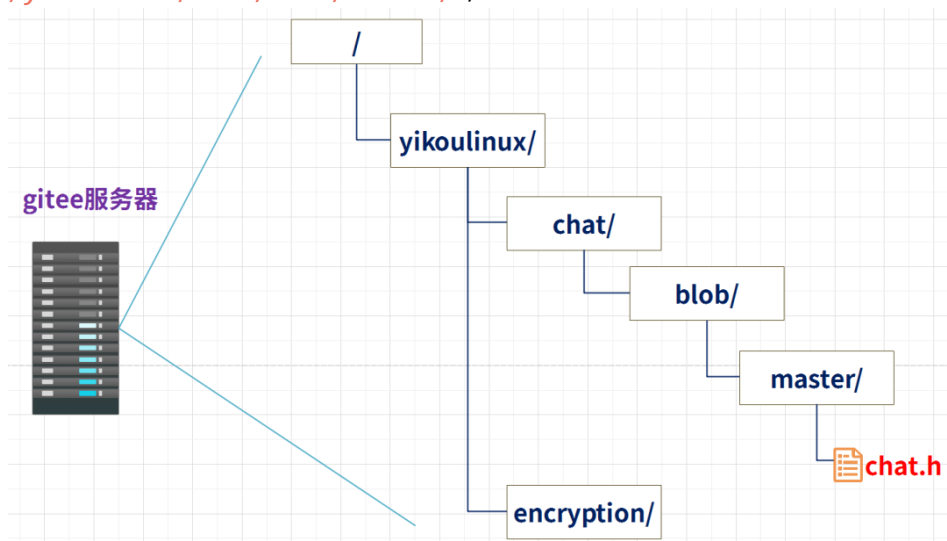
## 文件路径/文件名

从域名的第一个 **/** 开始到最后一个 **/** 为止, 是虚拟目录的部分。虚拟目录也不是 URL 必须的部分, 上述实例 http 协议 url 中的虚拟目录是

**/yikoulinux/chat/blob/master/**

从域名最后一个 **/** 开始到 **?** 为止, 是文件名部分; 如果没有 **?**, 则是从域名最后一个 **/** 开始到 **#** 为止, 是文件名部分; 如果没有 **?** 和 **#**, 那么就从域名的最后一个 **/** 从开始到结束, 都是文件名部分。

比如前面的 http url 实例, 其中文件 `chat.h` 在 gitee 服务器 `/yikoulinux/chat/blob/master/` 下:



文件名也不是一个 URL 的必须部分。

文件名省略情况如下:

1. `http://www.gitee.com/dir/`

我们可以这样理解, 以 “/” 结尾代表 `/dir/` 后面本来应该有的文件名被省略了。根据 URL 的规则, 文件名可以像前面这样省略。不过, 没有文件名, 服务器怎么知道要访问哪个文件呢? 其实, 我们会在服务器上事先设置好文件名省略时要访问的默认文件名。这个设置根据服务器不同而不同, 大多数情况下是 `index.html` 或者 `default.htm` 之类的文件名。

因此, 像前面这样省略文件名时, 服务器就会访问 `/dir/index.html` 或者 `/dir/default.htm` [由 web 服务器配置]。

2. `http://www.gitee.com/` 这个 URL 也是以 “/” 结尾的, 也就是说它表示访问一个名叫 “/” 的目录。而且, 由于省略了文件名, 所以结果就是访问 `/index.html` 或者 `/default.htm` 这样的文件了。
3. `http://www.gitee.com` 这次连结尾的 “/” 都省略了。像这样连目录名都省略时, 真不知道到底在请求哪个文件了, 实在有些过分。不过, 这种写法也是允许的。当没有路径名时, 就代表访问根目录下事先设置的默认文件, 也就是 `/index.html` 或者 `/default.htm` 这些文件, 这样就不会发生混乱了。
4. `http://www.gitee.com/yikoupeng`

一般来说， 这种情况会按照下面的惯例进行处理：如果 Web 服务器上存在名为 yikoupeng 的文件， 则将 yikoupeng 作为文件名来处理；如果存在名为 yikoupeng 的目录， 则将 yikoupeng 作为目录名来处理 。



关于协议的说明文档，可以登录下面网站查询：

<https://www.rfc-editor.org/>

搜索 URL 协议的说明，就有 25 个结果。

RFC Number (or Subseries Number):  
Title/Keyword:  
☐ Show Abstract ☐ Show Keywords  
Additional Criteria

46 results ( Show 25 | All )

Number	Files	Title	Authors	Date	More Info	Status
RFC 1736	ASCII, PDF, HTML	Functional Recommendations for Internet Resource Locators	J. Kunze	February 1995		Informational
RFC 1738	ASCII, PDF, HTML, HTML with inline errata	Uniform Resource Locators (URL)	T. Berners-Lee, L. Masinter, M. McCahill	December 1994	Errata. Obsoleted by RFC 4248, RFC 4266, Updated by RFC 1808, RFC 2368, RFC 2396, RFC 3986, RFC 6196, RFC 6270, RFC 8089	Proposed Standard
RFC 1808	ASCII, PDF, HTML	Relative Uniform Resource Locators	R. Fielding	June 1995	Obsoleted by RFC 3986, Updates RFC 1738, Updated by RFC 2368, RFC 2396	Proposed Standard
RFC 1959	ASCII, PDF, HTML, HTML with inline errata	An LDAP URL Format	T. Howes, M. Smith	June 1996	Errata. Obsoleted by RFC 2255	Proposed Standard
RFC 2017	ASCII, PDF, HTML	Definition of the URL MIME External-Body Access-Type	N. Freed, K. Moore, A. Cargille	October 1996		Proposed Standard
RFC 2056	ASCII, PDF, HTML	Uniform Resource Locators for Z39.50	R. Denenberg, J. Kunze, D. Lynch	November 1996		Proposed Standard
RFC 2079	ASCII, PDF, HTML	Definition of an X.500 Attribute Type and an Object Class to Hold Uniform Resource Identifiers (URIs)	M. Smith	January 1997		Proposed Standard
RFC 2111	ASCII, PDF, HTML	Content-ID and Message-ID Uniform Resource Locators	E. Levinson	March 1997	Obsoleted by RFC 2392	Proposed Standard
RFC 2122	ASCII, PDF, HTML	VEMMI URL Specification	D. Mavrikis, H. Layec, K. Kartmann	March 1997		Proposed Standard
RFC 2187	ASCII, PDF, HTML	Application of Internet Cache Protocol (ICP), version 2	D. Wessels, K. Claffy	September 1997		Informational
RFC 2192	ASCII, PDF, HTML, HTML with inline errata	IMAP URL Scheme	C. Newman	September 1997	Errata. Obsoleted by RFC 5092	Proposed Standard
RFC 2224	ASCII, PDF, HTML	NFS URL Scheme	B. Callaghan	October 1997		Informational
RFC 2244	ASCII, PDF, HTML, HTML with inline errata	ACAP -- Application Configuration Access Protocol	C. Newman, J. G. Myers	November 1997	Errata. Updated by RFC 6075	Proposed Standard
RFC 2255	ASCII, PDF, HTML	The LDAP URL Format	T. Howes, M. Smith	December 1997	Obsoletes RFC 1959, Obsoleted by RFC 4510, RFC 4516, Updated by RFC 3377	Proposed Standard







我们想查看某个协议, 点击即可。

# RFC 1738

Uniform Resource Locators (URL), DECEMBER 1994

File formats:



Status:  
PROPOSED STANDARD

Obsoleted by:  
RFC 4248, RFC 4266

Updated by:  
RFC 1808, RFC 2368, RFC 2396, RFC 3986, RFC 6196, RFC 6270, RFC 8089

Authors:  
T. Berners-Lee  
L. Masinter  
M. McCahill

Stream:  
[Legacy]

Cite this RFC: [TXT](#) | [XML](#)

DOI: 10.17487/RFC1738

Discuss this RFC: [Send questions or comments to iesg@ietf.org](#)

Other actions: [View Errata](#) | [Submit Errata](#) | [Find IPR Disclosures from the IETF](#) | [View History of RFC 1738](#)

## Abstract





This document specifies a Uniform Resource Locator (URL), the syntax and semantics of formalized information for location and access of resources via the Internet. [STANDARDS-TRACK]

For the definition of Status, see RFC 2026.

For the definition of Stream, see RFC 8729.

可以以任意一种格式查看该文档:

## File formats:



下面只拷贝第一页内容:

Network Working Group	T. Berners-Lee
Request for Comments: 1738	CERN
Category: Standards Track	L. Masinter
	Xerox Corporation
	M. McCahill
	University of Minnesota
	Editors
	December 1994

Uniform Resource Locators (URL)

## Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

## Abstract

This document specifies a Uniform Resource Locator (URL), the syntax and semantics of formalized information for location and access of resources via the Internet.

## 1. Introduction

This document describes the syntax and semantics for a compact string representation for a resource available via the Internet. These strings are called "Uniform Resource Locators" (URLs).

The specification is derived from concepts introduced by the World-Wide Web global information initiative, whose use of such objects dates from 1990 and is described in "Universal Resource Identifiers in WWW", RFC 1630. The specification of URLs is designed to meet the requirements laid out in "Functional Requirements for Internet Resource Locators" [12].

This document was written by the URI working group of the Internet Engineering Task Force. Comments may be addressed to the editors, or to the URI-WG <uri@bunyip.com>. Discussions of the group are archived at <URL:http://www.acl.lanl.gov/URI/archive/uri-archive.index.html>