

Linux操作系统下的离线浏览器介绍 (1) PDF转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/144/2021\\_2022\\_Linux\\_E6\\_93\\_8D\\_E4\\_BD\\_c103\\_144570.htm](https://www.100test.com/kao_ti2020/144/2021_2022_Linux_E6_93_8D_E4_BD_c103_144570.htm) 本文将介绍如何实现一个离线浏览器，以下载并浏览网上资源。 镜像目录结构 离线浏览下载到本地的网页时，需要解决的一个关键性问题就是“如何通过某一网页中的超级链接正确地定位其他网页”。比较简便的方法是在用户指定的本地目录下建立一个目标网站的完整或部分镜像。也就是按照文件在服务器端的目录结构保存下载的文件(参见下图)。这样一来，如果网页中的超级链接是以相对路径形式给出的，那么浏览程序就可以直接通过此相对路径访问到本地文件系统中的网页；如果网页中的超级链接是以绝对的URL形式给出的，那么必须在保存网页之前将这些URL转换为本地绝对路径。在网络中，一个有效的URL应该只有唯一的网络文件与之对应。因此，只要将网络上由URL所确定的层次关系，转化为本地文件系统中由目录路径所确定的层次关系，就可以建立网站在本地的完全或部分镜像。下面讨论建立镜像的具体方法。 镜像路径算法 首先，将下载网页时生成的URL拆分成协议类名(protocol)、IP地址(ipaddr)、目录名(directory)和文件名(file)。KDE环境提供了一个用于解析URL的类KURL，只需要定义一个对象KURL u((const char\*)URL)，就可以利用该类提供的成员函数将URL拆解为所需的部分。但是，此类未提供对ASP定位语句的支持，所以读者可以在KURL的基础上编写自己的拆解函数，以完善程序功能。需要注意的是，在同一网络文件的URL中，网址部分可能是以域名地址形式给出的，也可能

是以IP地址形式给出的。为了避免将同一文件镜像到不同目录下，如果网址是域名形式的，应该使用socket函数gethostbyname()将其转换为IP地址。其次，确定网络文件在本地的镜像路径。假设用户指定的本地目录存放在字符数组LDir中，则代码如下：  
QString LocalDir = LDir “ / ”  
protocol “ \_ ” ipaddr directory. QString LocalPath = LocalDir file.  
这样一来，如果一个网络文件的URL是http://11.171.38.32/webfile/relax/index.html，而用户指定的本地目录是/home/yangjx/web，则此网页文件对应的镜像路径为/home/yangjx/web/http\_11.171.38.32/webfile/relax/index.html。

100Test 下载频道开通，各类考试题目直接下载。详细请访问  
[www.100test.com](http://www.100test.com)