

1 LibCurl 简介

LibCurl 是免费的客户端 URL 传输库，支持 FTP,FTPS, HTTP, HTTPS, SCP, SFTP, TFTP, TELNET, DICT, FILE , LDAP 等协议，其主页是 <http://curl.haxx.se/>。Libcurl 具备线程安全、IPv6兼容、易于使用的特点。本文档主要介绍 LibCurl 在 http 协议方面的应用。

1.1HTTP 协议格式

说明：本节主要介绍 http 协议，若熟悉 http 协议者可不看。

Http（超文本传输协议）是分布式双向超媒体信息系统应用层协议，主要应用于 WWW。通常 HTTP 消息包括客户机向服务器的请求消息和服务器向客户机的响应消息。http 消息（请求或者响应）消息的通用格式实质相同，这两种类型的消息由一个起始行，一个或者多个头域，一个只是头域结束的空行和可选的消息体组成。HTTP 的头域包括通用头，请求头，响应头和实体头四个部分。

起始行：请求消息中的的起始行称为请求行，由3个字段组成，它们定义请求的类型、URL 和 http 版本，最后是回车和换行符。请求类型包括 get、head、post、put、move 等。响应消息中的的起始行称为，也由三个部分组成，http 版本、状态码和状态短语，最后是回车和换行符。所有的 Http 消息题头

http 头域：HTTP 的头域按其所属性质包括通用头，请求头，响应头和实体头四个部分。通用头域允许出现在请求或者响应消息中，包含 Cache-Control、 Connection、 Date、Pragma、Transfer-Encoding、Upgrade、Via。请求头域只允许出现在请求消息中，响应头域只允许出现在响应消息中，实体头部分提供有消息文档主体信息，主要在响应消息中发送；但是请求消息（如 post 和 put 方法）也可以使用实体题头。

每个头域由一个域名，冒号（:）和域值三部分组成。域名是大小写无关的，域值前可以添加任何数量的空格符，头域可以被扩展为多行，在每行开始处，使用至少一个空格或制表符。下表为一个典型的请求消息，下面介绍常用头域：

```
GET http://download.microtool.de/somedata.exe HTTP/1.1
```

```
Host: download.microtool.de
```

```
Accept: */*
```

```
Pragma: no-cache
```

```
Cache-Control: no-cache
Referer: http://download.microtool.de/
User-Agent: Mozilla/4.04[en](Win95;I;Nav)
Range: bytes=554554-
```

消息的第一行是请求行，“GET”表示我们所使用的 HTTP 动作，其他可能的还有“POST”等，GET 的消息没有消息体，而 POST 消息是有消息体的，消息体的内容就是要 POST 的数据。后面 `http://download.microtool.de:80/somedata.exe` 就是我们要请求的对象，之后 HTTP1.1 表示使用的是 HTTP1.1 协议。

从第2行开始进入 http 头域，本例中共包括 **Host, Accept, Pragma, Cache-Control, Referer, User-Agent, Range 域**。

Host 域：表示我们所请求的主机和端口

Accept 域：表示我们所用的浏览器能接受的 Content-type（一般包括 image/gif, image/x-xbitmap, image/jpeg, image/pjpeg, application/x-shockwave-flash, application/vnd.ms-excel, application/vnd.ms-powerpoint, application/msword），

Pragma 域：Pragma 头域用来包含实现特定的指令，最常用的是 Pragma:no-cache。在 HTTP/1.1 协议中，它的含义和 Cache-Control:no-cache 相同。

Cache-Control 域：Cache-Control 指定请求和响应遵循的缓存机制。

Referer：Referer 头域允许客户端指定请求 ui 的原资源地址，这可以允许服务器生成回退链表，可用来登陆、优化 cache 等。他也允许废除的或错误的连接由于维护的目的被追踪。如果请求的 uri 没有自己的 uri 地址，Referer 不能被发送。如果指定的是部分 uri 地址，则此地址应该是一个相对地址。

User-Agent 头域：User-Agent 头域的内容包含发出请求的用户信息。

Range 头域：Range 头域可以请求实体的一个或者多个子范围，**跟断点续传相关**。例如，

表示头500个字节：bytes=0-499

表示第二个500字节：bytes=500-999

表示最后500个字节：bytes=-500

表示500字节以后的范围: bytes=500-

第一个和最后一个字节: bytes=0-0,-1

同时指定几个范围: bytes=500-600,601-999

但是服务器可以忽略此请求头, 如果无条件 GET 包含 Range, 请求头, 响应会以状态码 206 (PartialContent) 返回而不是以200 (OK)。

一个典型的响应消息格式如下:

```
HTTP/1.0200OK
Date:Mon,31Dec200104:25:57GMT
Server:Apache/1.3.14(Unix)
Content-type:text/html
Last-modified:Tue,17Apr200106:46:28GMT
Etag:"a030f020ac7c01:1e9f"
Content-length:39725426
Content-range:bytes554554-40279979/40279980
```

消息第一行为响应行, 起格式为如下, 包括 http 版本、状态码和文本描述。

```
HTTP-Version SPStatus-CodeSP Reason-Phrase CRLF
```

HTTP -Version 表示支持的 HTTP 版本, 例如为 HTTP/1.1。Status- Code 是一个三个数字的结果代码。Reason-Phrase 给 Status-Code 提供一个简单的文本描述。Status-Code 主要用于机器自动识别, Reason-Phrase 主要用于帮助用户理解。Status-Code 的第一个数字定义响应的类别, 后两个数字没有分类的作用。第一个数字可能取5个不同的值:

1xx:信息响应类, 表示接收到请求并且继续处理

2xx:处理成功响应类, 表示动作被成功接收、理解和接受

3xx:重定向响应类, 为了完成指定的动作, 必须接受进一步处理

4xx:客户端错误, 客户请求包含语法错误或者是不能正确执行

5xx:服务端错误, 服务器不能正确执行一个正确的请求

响应头域允许服务器传递不能放在状态行的附加信息，这些域主要描述服务器的信息和 Request-URI 进一步的信息。响应头域包含 Age、Location、Proxy-Authenticate、Public、Retry- After、Server、Vary、Warning、WWW-Authenticate。对响应头域的扩展要求通讯双方都支持，如果存在不支持的响应头域，一般将会作为实体头域处理。

在上表中，从第二行开始进入 http 头域。包括 **Date、Server、Content-type、Last-modified、Etag、Content-length、Content-range** 域。一些比较重要的头域包括：

Location 域：Location 响应头用于重定向接收者到一个新 URI 地址。

Content-Type：属于实体头，用于向接收方指示实体的介质类型，指定 HEAD 方法送到接收方的实体介质类型，或 GET 方法发送的请求介质类型 Content-Range 实体头

Content-length 域：属于实体头，指明实体（文档）长度，表示实际传送的字节数。

Content-Range：属于实体头，用于指定整个实体中的一部分的插入位置，他也指示了整个实体的长度。在服务器向客户返回一个部分响应，它必须描述响应覆盖的范围和整个实体长度。一般格式：

```
Content-Range:bytes-unitSP first-byte-pos-last-byte-pos/entity-length
```

例如，传送头500个字节次字段的形式：Content-Range:bytes 0-499/1234如果一个 http 消息包含此节（例如，对范围请求的响应或对一系列范围的重叠请求），Content-Range 表示传送的范围