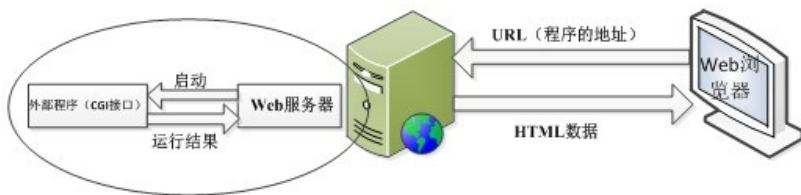


## CGI接口原理及实现 ( 2012-12-7 Over )



### 1.CGI定义：

CGI(CommonGateway Interface)是HTTP服务器与你的或其它机器上的程序进行“交谈”的一种工具，其**程序须运行在网络服务器上**。

### 2.CGI功能：

绝大多数的CGI程序被用来解释处理来自表单的输入信息，并在服务器产生相应的处理，或将相应的信息反馈给浏览器。CGI程序使**网页具有交互功能**。

### 3.CGI运行环境：

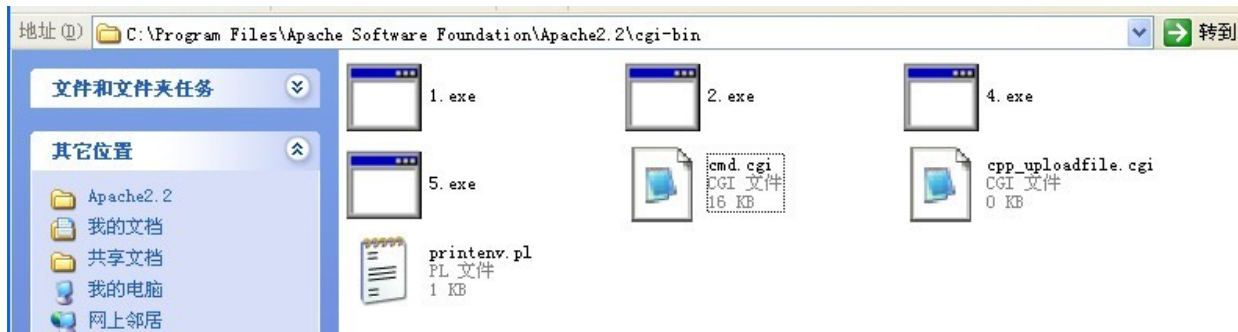
CGI程序在UNIX操作系统上CERN或NCSA格式的服务器上运行。在其它操作系统（如：windows NT及windows95等）的服务器上 也广泛地使用CGI程序，同时它也适用于各种类型机器。

### 4.CGI处理步骤：

- (1)通过Internet把用户请求送到服务器。
- (2)服务器接收用户请求并交给CGI程序处理。
- (3)CGI程序把处理结果传送给服务器。
- (4)服务器把结果送回到用户。

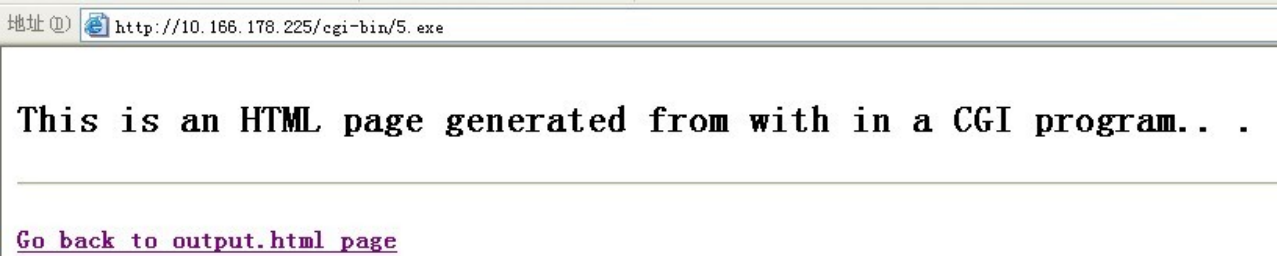
### 5.CGI服务器配置：

在许多服务器cgi-bin是仅能够**放置CGI脚本的目录**。



在Windows平台上将C或C++写好的程序的Debug或Release版本的.exe程序拷贝到cgi-bin的目录下（如上图所示），**将.exe改为.cgi也可同样运行**，如下2个图。





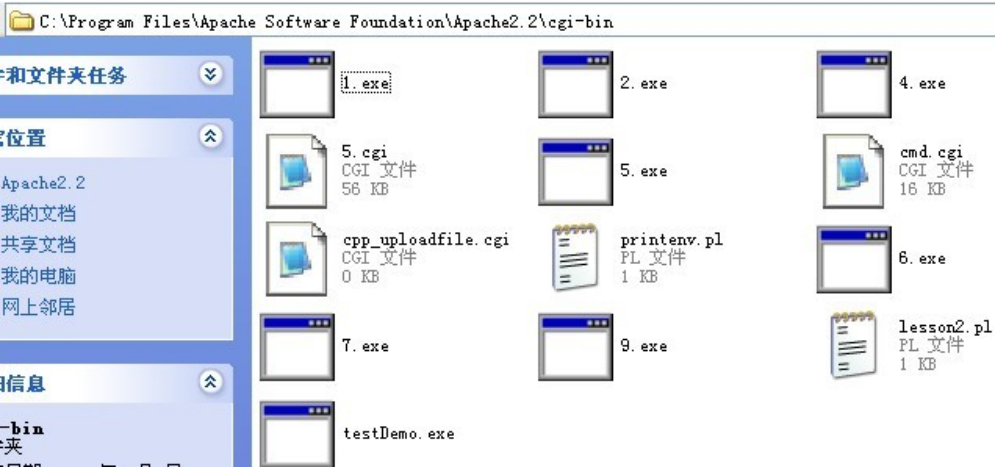
**cgi-bin目录**是存放CGI脚本的地方。这些脚本使WWW服务器和浏览器能运行外部程序，而无需启动另一个程序。它是运行在Web服务器上的一个程序，并由来自于浏览者的输入触发。

CGI程序不是放在服务器上就能顺利运行，如果要想使其在服务器上顺利的运行并准确的处理用户的请求，则须对所**使用的服务器**进行必要的设置。

**配置：**根据所使用的服务器类型以及它的设置把CGI程序放在某一特定的目录中或使其带有特定的扩展名。  
Apache网络服务器配置在/var/www/cgi-bin里（如下图所示笔者电脑的目录位置）。C++编译的可执行文件可以转换成扩展名为.cgi的文件。  
更改初始配置的的方法：

```
<Directory"/var/www/cgi-bin">
    AllowOverride None
    Options ExecCGI
    Order allow,deny
    Allow from all
</Directory>
```

```
<Directory"/var/www/cgi-bin">
    Options All
</Directory>
```



6.CGI接口标准包括标准输入、环境变量、标准输出三部分。

	介绍
1.标准输入	CGI程序像其他可执行程序一样,可通过标准输入(stdin)从Web服务器得到输入信息,如Form中的数据,这就是所谓的向CGI程序传递数据的 <b>POST方法</b> 。这意味着在操作系统命令行状态可执行CGI程序,对CGI程序进行调试。POST方法是常用的方法。

入	
2.环境变量	操作系统提供了许多环境变量,它们定义了程序的执行环境,应用程序可以存取它们。Web服务器和CGI接口又另外设置了自己的一些环境变量,用来向CGI程序传递一些重要的参数。CGI的GET方法还通过环境变量QUERY-STRING向CGI程序传递Form中的数据。
3.标准输出	CGI程序通过标准输出(stdout)将输出信息传送给Web服务器。传送给Web服务器的信息可以用各种格式,通常是以纯文本或者HTML文本的形式,这样我们就可以在命令行状态调试CGI程序,并且得到它们的输出。

7.环境变量

环境变量是文本串(名字/值对),可以被OSShell或其他程序设置 ,也可以被其他程序访问。它们是Web服务器传递数据给CGI程序的简单手段,之所以称为环境变量是因为它们是全局变量,任何程序都可以存取它们。

下面是CGI程序设计中常常要用到的一些环境变量。

环境变量	意义
SERVER_NAME	CGI脚本运行时的主机名和IP地址.
SERVER_SOFTWARE	你的服务器的类型如： CERN/3.0 或 NCSA/1.3.
GATEWAY_INTERFACE	运行的CGI版本. 对于UNIX服务器, 这是CGI/1.1.
SERVER_PROTOCOL	服务器运行的HTTP协议. 这里当是HTTP/1.0.
SERVER_PORT	服务器运行的TCP口，通常Web服务器是80.
REQUEST_METHOD	POST 或 GET, 取决于你的表单是怎样递交的.
HTTP_ACCEPT	浏览器能直接接收的Content-types, 可以有HTTP Accept header定义.
HTTP_USER_AGENT	递交表单的浏览器的名称、版本 和其他平台性的附加信息。
HTTP_REFERER	递交表单的文本的 URL，不是所有的浏览器都发出这个信息，不要依赖它
PATH_INFO	附加的路径信息, 由浏览器通过GET方法发出.
PATH_TRANSLATED	在PATH_INFO中系统规定的路径信息.
SCRIPT_NAME	指向这个CGI脚本的路径, 是在URL中显示的(如, /cgi-bin/thescrypt).
QUERY_STRING	脚本参数或者表单输入项(如果是用GET递交). QUERY_STRING包含URL中间号后面的参数.
REMOTE_HOST	递交脚本的主机名，这个值不能被设置.
REMOTE_ADDR	递交脚本的主机IP地址.
REMOTE_USER	递交脚本的用户名. 如果服务器的authentication被激活，这个值可以设置。
REMOTE_IDENT	如果Web服务器是在ident (一种确认用户连接你的协议)运行, 递交表单的系统也在运行ident, 这个变量就含有ident返回值.
CONTENT_TYPE	如果表单是用POST递交, 这个值将是 application/x-www-form-urlencoded. 在上载文件的表单中, content-type 是个 multipart/form-data.
CONTENT_LENGTH	对于用POST递交的表单,标准输入口的字节数.
REQUEST-METHOD	指的是当Web服务器传递数据给CGI程序时所采用的方法,分为GET和POST两种方法。
【GET和POST方法的区别】:GET方法仅通过环境变量(如QUERY-STRING)传递数据给CGI程序,而POST方法通过环境变量和标准输入传递数据给CGI程	

序,因此POST方法可较方便地传递较多的数据给CGI程序。

		问题
GET方法	通过在URL中嵌入的形式传递参数。对CGI程序而言，在GET method中传递的参数要通过化境变量 “QUERY-STRING” 来接收。	1 ) 参数的内容作为URL信息，用户可以看到； 2 ) 有大小的限制。
POST方法	CGI程序从标准输入接收参数。与GET方法不同的是，参数的内容从URL信息中不能获得，对于大小也没有限制。	与GET方法问题1),2)完全相反。

**CONTENT-LENGTH**:传递给CGI程序的数据字符数(字节数)。

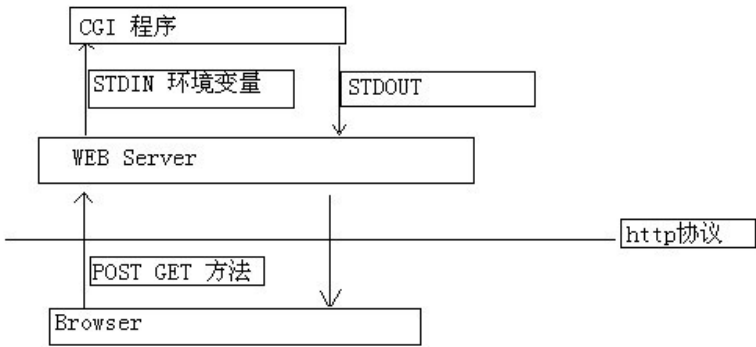
在C语言程序中,要访问环境变量,可使用getenv()库函数。例如:

```
if (getenv ("CONTENT-LENGTH"))n=atoi(getenv("CONTENT-LENGTH"));
```

请注意程序中最好调用两次getenv():第一次检查是否存在该环境变量,第二次再使用该环境变量。这是因为函数getenv()在给定的环境变量名不存在时,返回一个NULL(空)指针,如果你不首先检查而直接引用它,当该环境变量不存在时会引起CGI程序崩溃。

8. CGI的工作原理

**CGI**是一个WEB服务器提供信息服务的标准接口，通过这样一个接口，WEB服务器能够执行程序，并将程序输出的信息返回给浏览器。因为在WEB网上的数据都是静态的，通过CGI程序能够动态的处理浏览者的请求，如保存用户输入的信息，根据用户信息返回相关的资料等等。当客户端发送一个**CGI请求**给WEB服务器后，WEB服务器将根据CGI程序的类型决定数据向CGI程序的传送方式，一般来讲是通过**标准输入/输出流和环境变量**来与CGI程序间传递数据。



CGI结构示意图  
CGI输入输出原理

CGI的输入/输出方法：CGI程序通过标准输入（STDIN）和标准输出（STDOUT）来进行输入输出，STDIN和STDOUT是**两个预先定义好的文件指针**。你可以利用**文件读写函数**来对其进行操纵。

此外CGI程序还通过**环境变量**来得到输入，只不过环境变量中提供的是一些常用的信息，并且通常不包括用户在WEB页面中输入的信息（除使用下面讲的GET方法时，通过检查环境变量QUERY\_STRING来得到输入数据），而STDIN通常用来传递用户输入的信息。

在输入时所使用的POST/GET方法：在**WEB页面**向CGI发送数据时通常采用两种方法：GET/POST，**GET方法**将数据附加在URL后发送，如：/cgi/a\_cgi\_test.exe?your\_data，CGI程序通过检查环境变量QUERY\_STRING来得到输入数据。

**示例一、下图即是GET方法!**



上图的对应程序为：

### //2012-12-5 GET c程序示例..

```
[cpp]
01. void main(void)
02. { // 本程序将用户输入的数据打印出来
03.     fprintf(stdout, "content-type:text/plain\n\n");
04.     // 输出一个CGI标题，这行代码的意义后面会讲解
05.
06.     char *pszMethod;
07.     pszMethod = getenv("REQUEST_METHOD");
08.     if(strcmp(pszMethod, "GET") == 0)
09.     { //GET method
10.         //读取环境变量来获取数据
11.         printf("This is GETMETHOD!\n");
12.         printf("SERVER_NAME:%s\n", getenv("SERVER_NAME"));
13.         printf("REMOTE_ADDR:%s\n", getenv("REMOTE_ADDR"));
14.         fprintf(stdout, "input data is:%s\n", getenv("QUERY_STRING"));
15.     }
16.     else
17.     { // POST method
18.         //读取STDIN来获取数据
19.         int iLength=atoi(getenv("CONTENT_LENGTH"));
20.         printf("This is POSTMETHOD!\n");
21.
22.         fprintf(stdout, "input data is:\n");
23.         for(int i=0; i<iLength; i++)
24.         {
25.             char cGet=fgetc(stdin);
26.             fputc(cGet, stdout);
27.         }
28.     }
29. }
```

### 示例二、下图即是POST程序示例:

地址 http://10.166.178.225/post.html

---

Your input (80 chars max.):

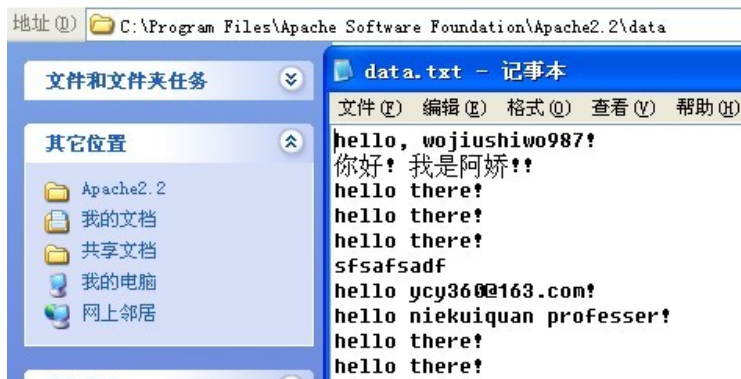
---

地址 http://10.166.178.225/cgi-bin/post.cgi

---

CONTENT\_LENGTH = 19

Thank you! Your contribution has been stored.



```
[cpp]   loa

01. void unencode(char *src, char *last, char *dest)
02. {
03.     // str = hello+there%21 此处跳过data=...
04.     // last = ; 已到末尾.
05.     // dest= ; 空串.
06.
07.     //解码原则
08.     //原则1: '+'变' ';
09.     //原则2: '%xx'变成对应的16进制ASCII码值;
10.     for(; src != last; src++, dest++)
11.     {
12.         if(*src == '+')
13.         {
14.             *dest = ' ';
15.         }
16.         else if(*src == '%')
17.         {
18.             int code;
19.             if(sscanf(src+1, "%2x", &code) != 1)
20.             {
21.                 code = '?';
22.             }
23.             *dest = code;
24.             src +=2;
25.         }
26.         else
27.         {
28.             *dest = *src;
29.         }
30.     }
31.
32.     *dest = '\n';
33.     *++dest = '\0';
34. }
35.
36. intmain(void)
37. {
38.     char *lenstr;
39.     char input[MAXINPUT], data[MAXINPUT];
40.     long len;
41.     printf("%s%c\n", "Content-Type:text/html;charset=iso-8859-1",13,10);
42.     printf("<TITLE>Response</TITLE>\n");
43.
44.     lenstr =getenv("CONTENT_LENGTH");
45.     printf("CONTENT_LENGTH =%s\n", lenstr);
46.
47.     if(lenstr == NULL ||sscanf(lenstr,"%ld",&len)!=1 || len > MAXLEN)
48.     {
49.         printf("<P>Error in invocation - wrong FORM probably.");
50.     }
51.     else
52.     {
53.         FILE *f;
```

```

54.         fgets(input, len+1, stdin);           //add by ycy从输入流中获取字符串.
55.         unencode(input+EXTRA, input+len,data);
56.
57.         f = fopen(DATAFILE, "a");
58.         if(f == NULL)
59.         {
60.             printf("<P>Sorry, cannot store your data.");
61.         }
62.         else
63.         {
64.             fputs(data, f); //add byycy 将数据存储在对对应的文件中.
65.         }
66.         fclose(f);
67.         printf("<P>Thank you!Your contribution has been stored.");
68.     }
69.
70.     return 0;
71. }

```

请求过程即是：

(1) send发送按钮--->(2) 调用post.cgi--->(3) 将数据存储在对data\data.txt里面。

### 示例三、下图即是GET/POST程序示例。

综合实例：在上面两个程序上的扩展(应用Get方法及QUERY\_STRING) ,

地址 http://10.166.178.225/cgi-bin/submit\_test.cgi

**Your First Name:**

**Your Last Name:**

**Check if you are a student:**

☒

**What is your gender?**

☒ Male ☐ Female

**How many moons are there on your planet?**

A Fine Triplet

Four Celestial Bodies

Between Five and Eight

We Have Between Nine and Twelve

Too Many To Count!

**Comments:**

This is Comments!

**Hidden Data!**

**Submit this Form**

Send Data Now!

**Reset this Form**

Clear all my input now

地址 http://10.166.178.225/cgi-bin/6.exe?FirstName=zhang&LastName=zhiyi&isStudent=on&Gender=isMale&numMoons=3&Secret=Invisible

This is GET METHOD!  
SERVER\_NAME: 10.166.178.225  
REMOTE\_ADDR: 10.166.178.225  
input data is :FirstName=zhang&LastName=zhiyi&isStudent=on&Gender=isMale&numMoons=3&Secret=Invisible

如上两图所示，在表单（cgi接口的表单只是在html语言的基础上用C或C++实现的扩展操作而已）的基础上，提交按钮对应的另一个CGI接口（6.exe或6.cgi），这样通过Get方法及QUERY\_STRING参量就可以完成输出操作。

而POST方法则会将数据送入CGI程序的**STDIN输入流**。在表单（FORM）中的各个变量都会成为name=value的形式向WEB服务器发送，多个数据间用&分隔，如：name=value&name2=value2。其中名字（name，name2）是Form中定义的INPUT、SELECT或TEXTAREA等标置(Tag)名字,值是用户输入或选择的标置值。

- 如上面说讲，在CGI程序输出时必须先输出一个CGI标题，标题共有以下三类：
- Location: 标题，指明输出另一个文档的URL，例如 `fprintf(stdout,"Location:http://www.vchelp.net/\n\n");`
  - Content-Type: 标题，指明发送的数据的MIME类型，例如 `fprintf(stdout,"Content-Type:text/html\n\n");`



- Status: 标题, 指明HTTP状态码, 例如 `fprintf(stdout, "Status:200\n\n");`

注意每种标题后都必须跟一个换行和一个空行。

MIME类型以类型/子类型的形式来表示, 下面是一些常用的类型/子类型的组合:

- Text/plain 普通文本类型
- Text/html HTML格式的文本类型
- Audio/basic 八位声音文件格式, 后缀为.au
- Video/mpeg MPEG文件格式
- Video/quicktime QuickTime文件格式
- Image/gif GIF图形文件
- Image/jpeg JPEG图形文件
- Image/x-xbitmap X bitmap图形文件, 后缀为.xbm

有了上面的知识我们就可以写出一些CGI程序, 首先需要对输入数据进行分析, 方法为: 每当找到字符=, 标志着一个Form变量名字的结束; 每当找到字符&, 标志着一个Form变量值的结束。请注意输入数据的最后一个变量的值不以&结束。这样我们可以将输入数据分解为一组一组的值。

但随后会发现CGI的输入并不规则, 例如有时会类似下面格式的输入字符串: `filename=hello&cmd=world+I%27`, 这是因为浏览器对一些上传的特殊字符进行了编码, 所以在将数据分解开后需要进行解码,

#### 解码规则为:

- 1) +: 将+转换成空格符;
- 2) %xx: 用其十六进制ASCII码值表示的特殊字符 ( %作为转意符 )。根据值xx将其转换成相应的ASCII字符。对Form变量名和变量值都要进行这种转换。