

Information Management Resource Kit

Module on Management of Electronic Documents

UNIT 6. NETWORKING DOCUMENTS AND DATABASES

LESSON 2. DYNAMIC WEBSITES: COMMON GATEWAY INTERFACE

NOTE

Please note that this PDF version does not have the interactive features offered through the IMARK courseware such as exercises with feedback, pop-ups, animations etc.

We recommend that you take the lesson using the interactive courseware environment, and use the PDF version for printing the lesson and to use as a reference after you have completed the course.



© FAO, 2003

Objectives

At the end of this lesson, you will be able to:

- understand **what CGI is**, and
- identify **capabilities and limitations** of CGI.



Introduction

If we want to allow document searches in our website, we have to connect it to a database...

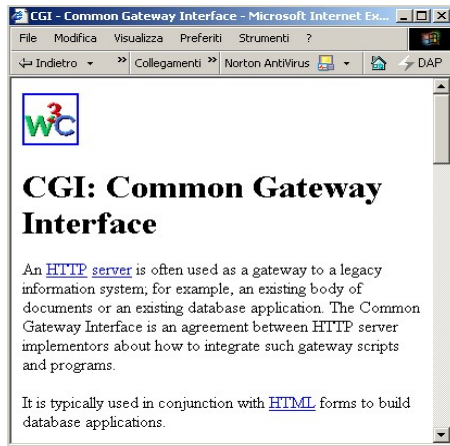


In order to overcome some of the limitations of static websites, the concept of web applications emerged in the 1980s. This is now the most common way to build a client-server application.

Instead of pointing to static HTML pages, the URIs in a web application point to fragments of executable code which **dynamically generate the HTML** that is sent back to the browser.

Very often these fragments of code retrieve data that is stored in **databases**.

What is CGI?



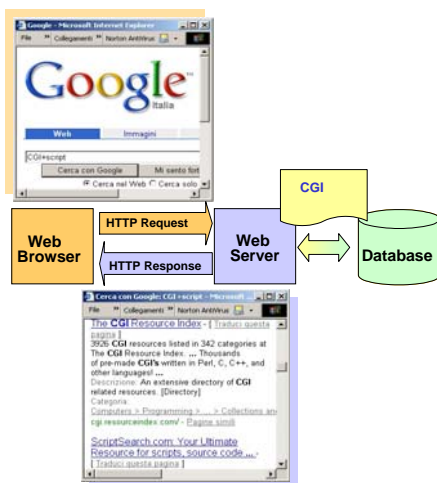
The original and very simple way to implement a web application was to use the CGI standard.

CGI (Common Gateway Interface) is a standard for **interfacing external applications with web servers**.

CGI can be used for a wide range of functions, but one of the most common usages is to **read or write data in backend databases** or other data sources over the World Wide Web.

This includes connecting the indexes used by search engines.

What is CGI?

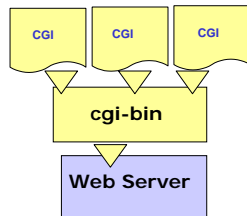


A CGI application (known as CGI or CGI script) acts as an interface between the client, the web server, the operating system, the hardware devices or other servers.

For example, if the user searches for a document on a database through a web interface, the server offering this service might use CGI in order to receive and translate the request by following the rules of the database language.

When CGI finds something, it allows the server to **generate a web page** "on the fly", which displays the results of the user request.

What is CGI?



CGIs can be used to perform a “**behind the scenes**” process such as processing a user request, writing a log, or sending an email. CGI scripts can also be used to generate **dynamic web pages**.

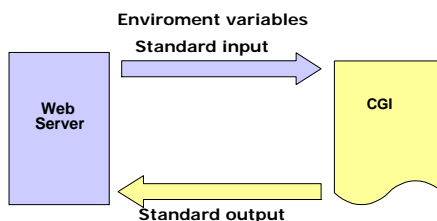
Where are CGIs located?

They are normally stored **on the web server** in a directory called **cgi-bin** (this is the most usual way to organize and set up the server, although it depends on the actual web server you are using).

How Does CGI Work?

How do CGI applications “talk” with the outside world?

As with any computer programme, a CGI script communicates with the outside world using **input and output parameters**.



Input to a CGI is done through two means:

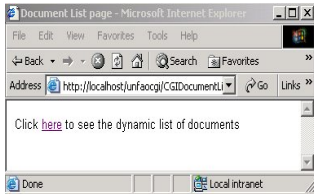
- by reading **environment variables** (e.g. the name of the server, the type of the content, etc.) that are set by the web server, and
- by reading the **standard input** stream.

Output is made by writing to the **standard output** stream.

The output is then channeled back through the web server to the web client that originally called the CGI.

How Does CGI Work?

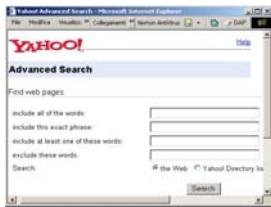
There are two ways that can be used to call a CGI through a web browser:



Using an HTML hyperlink

The user can execute a CGI by clicking on a link. The link points to a script instead of an HTML document.

In this case, the HTTP message uses the **GET** method.



Using an HTML form

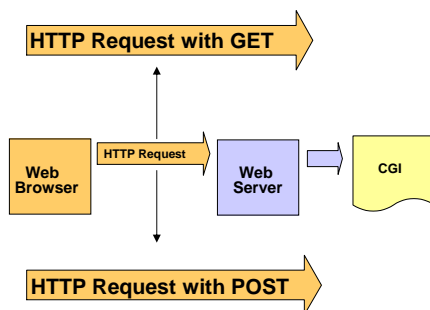
A form is a web page that contains input fields where the user can type their request then click one or more buttons to submit it.

In this case, you can choose the method (**GET or POST**) for your HTTP message.

Let's look at the difference between the POST and GET methods.

How Does CGI Work?

A **CGI** can be called through a web browser using an HTTP GET method or using an HTTP POST method.



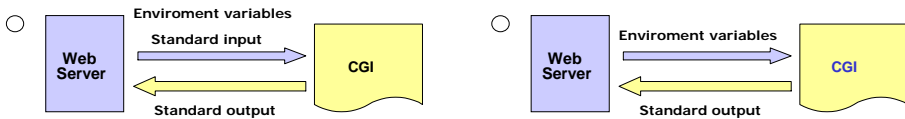
If the **GET method** is used, any input parameters passed to the CGI are appended to the URI and are set by the web server in the **environment variable** QUERY_STRING which can be read by the CGI.

If the **POST method** is used, the input parameters are passed in the entity body of the HTTP message and the web server passes them through on the **standard input** to be read by the CGI.

The **consequence** is that using the **POST method** allows for **no limitations** to the amount of information that can be passed.

How Does CGI Work?

In which of the following scenarios is the GET method used?

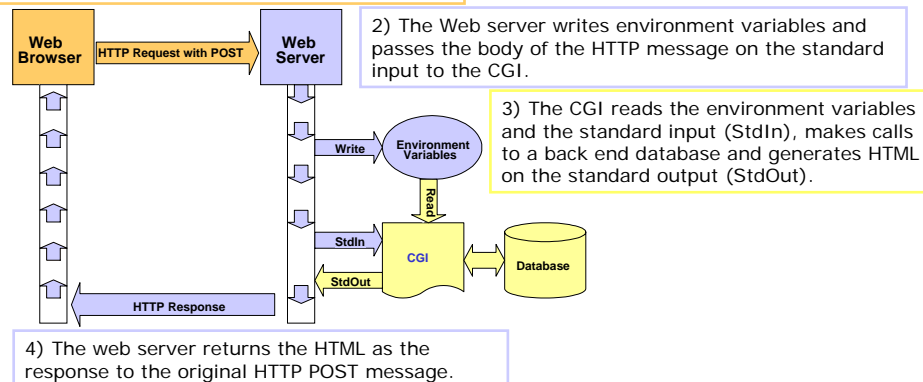


Please select the answer of your choice

How Does CGI Work?

The figure shows the sequence of steps in **calling a CGI using the HTTP POST** method:

1) The web browser sends the HTTP POST message to the web server.



Writing a CGI script

In order to write a CGI, you can use **any language** supported by the server.

Some of the most common choices of language for writing CGI scripts are:

- Perl
- ANSI C
- Shell scripts

You can find Perl CGI examples at:

<http://www.perl.com/>

<http://www.activestate.com/Solutions/Programmer/Perl.plex>

Advantages and Disadvantages of CGI

CGI can offer you some advantages. In fact:



It is a proven technology. CGI has been in widespread use since the mid 1990s, on websites across the world. A typical CGI may run faster than other solutions as it may be written in a compiled language (like C is); other platforms (e.g. Perl) have an optimized environment to minimize scripting parsing.

It can be written in a range of languages. Programmers don't have to use a language specific to one vendor, Web server or operating system.

Moreover:

It offers the full power of the host language. CGI's can be written using any of the features available in the language, not just the facilities offered by a proprietary scripting solution. For example, CGI's written in Perl have access to the full range of Perl functionality.

It is available on the Web. There are many CGI archives available on the Web. The most popular language for CGI's in these archives is Perl (e.g. <http://www.scriptarchive.com/>).

Advantages and Disadvantages of CGI

CGI has a number of disadvantages that have been addressed by subsequent technologies such as Servlets and Java Server Pages. For example:



Instances of CGI's run independently from each other.

This means that CGI's can be very "resource hungry", that is, they can cause a high usage of resources. However, different web server extensions have been created to optimize CGI runtime. For example "process pooling", a set of agents issued by the web server to reuse and optimize runtime/execution environment.

CGI's can be much harder to write.

While JSP, Java Servlets and ASP have a simplified environment, CGI's are normal programs that are executed within an http transaction. However, Perl CGI's can use many language specific functions that are quite similar to the simplified environment mentioned above.

Summary

- **CGI (Common Gateway Interface)** is a standard for interfacing external applications with web servers.
- CGI's can be used to perform a "**behind the scenes**" process such as processing a user request, writing a log, or sending an email; they can also be used to generate **dynamic web pages**.
- A CGI application can be called through a web browser using an **HTTP GET** method or using an **HTTP POST** method.
- In order to write a CGI application, you can use **any language** supported by the server.



Exercises

The following three exercises will allow you to test your understanding of the concepts described up to now.

Good luck!



Exercise 1

Can you order the stages of calling a CGI through a web browser?

- The web server passes the response message to the browser
- The web server writes environment variables and passes the body of the HTTP message to the CGI
- The CGI reads the environments variables and the standard input, calls the database and generates a standard output
- The web browser sends the HTTP message to the web server

Please select the answer of your choice

Exercise 2

Where are CGI applications located?

- On the browser.
- On the server.
- On the database.

Please select the answer of your choice

Exercise 3

Which of the following disadvantages is relevant to CGI?

- It can be harder to write than JSP, Java Servlets and ASP.
- It is a not-quite-proven technology.
- CGI archives are difficult to find on the Web.

Please select the answer of your choice

If you want to know more...

Information about CGI is available at <http://www.w3.org/CGI/>

The actual CGI specifications were originally developed by the National Center for Supercomputing Applications (NCSA) (<http://www.ncsa.uiuc.edu>)

You can find Perl CGI examples at: <http://www.perl.com/>
<http://www.activestate.com/Solutions/Programmer/Perl.plex>

