

What is a URL?

A **URL** (*Uniform Resource Locator*) is a universal naming format used to indicate a resource on the Internet. It is a printable **ASCII** character string which breaks down into five parts:

- **The name of the protocol:** i.e. in a way the language used to communicate over the network. The most widely used protocol is the **HTTP protocol** (*HyperText Transfer Protocol*), the protocol which makes it possible to change Web pages into **HTML** format. However, many other protocols can be used (FTP, News, Mailto, Gopher, ...).
- **Login and password:** enables the access parameters for a secure server to be specified. This option is unadvisable because the password is visible in the URL
- **The name of the server:** This is a **domain name** for the computer hosting the requested resource. It is worth noting that it is possible to use the server's **IP address** which conversely makes the URL less readable.
- **The number of the port:** this is a number related to a service allowing the server to know what type of resource is requested. The default port related to the protocol is port 80. So, when the Web service of the server is associated to port number 80, the port number is optional
- **The access path to the resource:** This last part allows the server to know where the resource is located, i.e. generally the site (directory) and the name of the file requested.

A URL therefore has the following structure:

Protocol	Password (optional)	Server name	Port (optional if 80)	Path
http://	user:password@	www.commentcamarche.net.	:80	/glossair/glossair.php3

For example, the following protocols may be used through URL:

- [http](#), for looking at web pages
- [ftp](#), for looking at FTP sites
- [telnet](#), to connect to a remote terminal
- [mailto](#), for sending an email
- [wais](#)
- [gopher](#)

The name of the file in the URL can be followed by a question mark then data in the ASCII format, this is additional data sent as parameters for an application on the server (a CGI script for example). The URL will then resemble a character string like this:

<http://en.kioskea.net/forum/?cat=1&page=2>

Encoding a URL

Considering that the URL is a medium for sending information over the Internet (to send data with a CGI script for example), it must be able to send special characters, yet URLs cannot contain special characters. In addition, certain characters are reserved because they have a meaning (the slash enables sub-directories

to be specified, the characters & and ? are used to send data via forms..). Finally, URLs can be included in an HTML document, which makes it difficult to insert characters such as < or > in the URL.

That is why encoding is necessary! Encoding consists of replacing special characters with the character % (itself also becoming a special character) followed by the [ASCII code](#) of the character to be encoded in [hexadecimal notation](#).

Here is the list of characters which require particular encoding:

Character	URL encoding
Tabulation	%09
Space	%20
"	%22
#	%23
%	%25
&	%26
(%28
)	%29
+	%2B
,	%2C
.	%2E
/	%2F
:	%3A
;	%3B
<	%3C
=	%3D
>	%3E
?	%3F
@	%40
[%5B
\	%5C

]	%5D
^	%5E
'	%60
{	%7B
	%7C
}	%7D
~	%7E