

[Henrik Frystyk](#), July 1994



# The World-Wide Web

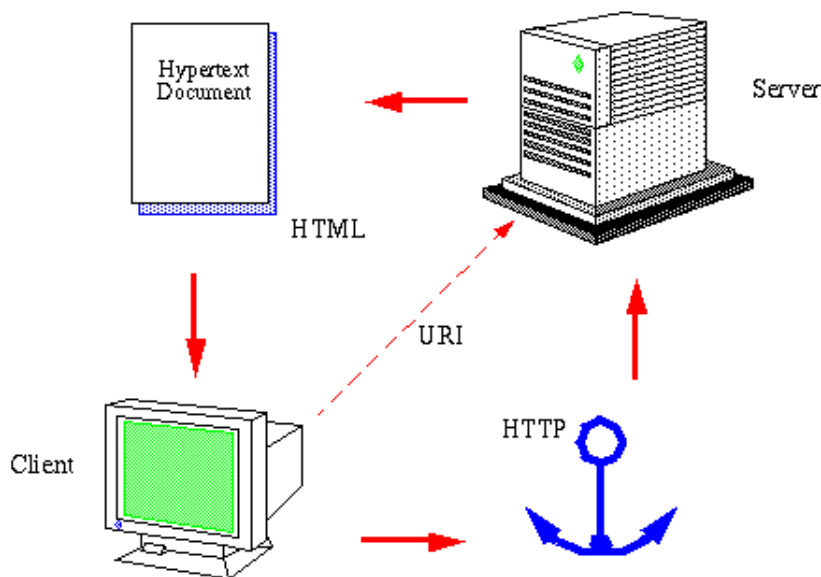


The official description of the [World-Wide Web \(WWW, W3\)](#) is a "wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents". It is a way of viewing all the on-line information available on the Internet as a seamless, browsable continuum. This section introduces the general concept of the World-Wide Web and the basic elements in the model. The content of the section is as follows:

1. [Basic World-Wide Web Model](#)
2. [Universal Resource Identifies](#)
3. [Hypertext Transfer Protocol](#)
4. [Hypertext Markup Language](#)
5. [Interactive World-Wide Web Model](#)

## Basic World-Wide Web Model

The basic idea behind the World-Wide Web is based on a client server application and hypertext documents as illustrated in the figure below. The model is simplified in that it only contains elements that exists within the World-Wide Web concept. Later the model will be expanded to a generic resource accessing model.



### [The Client](#)

The client is the user's interface to the Internet. Whatever type of service requested this interface stays the same, so users do not need to understand the differences between the many different access schemes in common use on the Internet. This principle is the same as is seen from other popular applications such as Microsoft Windows, Machintosh etc. where the user is always presented to the same GUI interface.

### [Uniform Resource Identifier URI](#)

The user initiates a request by specifying a Uniform Resource Identifier or a "hyperlink". This link can specify any accessible information or resource on the Internet as long as it can be uniquely identified as an object. The word "Web" refers to the combination of accessible objects and the links pointing to them throughout the Internet.

### The Server

The server is responsible for handling the request sent from the client. This can either be a local accessible resource or the server can request the resource from another server in which case the first server temporarily turns into a client.

### Hypertext Transport Protocol HTTP

The client sends of the user request to a WWW server using the Hypertext Transfer Protocol (HTTP). This is a typical client-server application based on a stateless connection between the client requesting the URI and the server handling the request.

### Hypertext Markup Language HTML

On a successful request, a data object is returned from the server to the client. The object is written in the Hypertext Markup Language (HTML) which is a hypertext language with the possibility of containing hyperlinks that the user can follow.

The model basically reflects the first version of the World-Wide Web as it is described in the HTTP Protocol version 0.9 and HTML version 1.0. However, the WWW specifications have been rapidly changing during the last 3-4 years, even though the current model is still based on a client-server approach. From being a HTML and HTTP based model, the World-Wide Web is now capable of handling virtually any existing data format on the Internet using a large set of access methods apart from HTTP such as FTP, [Gopher](#), [WAIS](#), Telnet etc. In other words, the World-Wide Web represents a generic information exchange tool capable of accessing information throughout the Internet. Though, before the more advanced model is presented, it is necessary to get an overview of the basic elements in the WWW model mentioned above.

## Universal Resource Identifies

In order to address a data object or more general, a resource, in the model above it is necessary to define a name space that not only contains information about hosts but also about resources available on each host. The World-Wide Web model defines [Uniform Resource Identifiers](#) or URIs that specifies a syntax for encoding the names and addresses of data objects on the Internet and how they can be accessed. The set of URIs covers



### Universal Resource Identifier (URI)

A generic set of all addresses in the address space of all resources on the Internet. They describe a hierarchical naming scheme that together with the [HTTP protocol](#) makes a significant difference between the World-Wide Web model and other Internet access schemes such as FTP that has a flat address space.

### Uniform Resource Locator (URL)

The term "URI" has been introduced by the [IETF](#) and is a a general description of all URL that are not persistent. In practice the URLs consist of the current set of Internet protocols supported by the WWW, i.e., HTTP, FTP, Gopher, WAIS, etc., followed by a directory path, a file name, and possibly a search directive.

### Uniform Resource Name (URN)

However, the ultimate goal for URIs is to be a persistent naming scheme independent of the mean of access, i.e., the protocol used and of the physical structure of resources on the specific host. The only way

to obtain this is to have a naming scheme like the [Internet Domain Name Service](#). URNs are currently under consideration in [IETF](#) but little is known about the status of the research.

#### Uniform Resource Citation (URC)

This is meta information about a URI. They consist of pairs of attribute/value which can contain information on the author, publisher etc. The URC are currently not used.

## Hypertext Transfer Protocol

The [Hypertext Transport Protocol \(HTTP\)](#) is a generic stateless presentation layer protocol with elements from other [Internet presentation layer protocols](#). The HTTP protocol is built on a client-server model where the client initiates a request and the server replies with a response.

The basic format of the HTTP protocol is based on the [MIME Protocol](#) with a set of [HTTP Headers](#) possibly followed by a message body containing a data object in any 7-bit or 8-bit accepted by the client. The client specifies what data format it can handle by having a list of [accept headers](#) in the request.

The [basic WWW-model](#) indicates that the client initiates a request and the server responds by sending a data object to the client. However, often the client wants to post a data object to the server, e.g. to post a mail message to an email address, to a news group, or to create a new file on the remote server. The HTTP protocol provides two methods for the client to transfer a data object to the server. Though, the client is not guaranteed that the request can be fulfilled - even on a successful return code. The action can at all times be cancelled by the responsible person of the remote server.

One of the characteristics of the HTTP protocol is that it is a superset of the other [Presentation Layer](#) supported by the WWW-model. This means that messages generated by other protocols can be handled by the HTTP protocol by wrapping a set of HTTP/MIME headers around the message. This is an essential feature for the [concept of Proxy servers](#).

The current version 1.0 of the HTTP protocol is built on top of the [TCP Protocol](#) that is a connection oriented protocol with a [3 way handshake](#) connection establishment. This causes an substantial overhead in a client-server oriented environment like the HTTP protocol. It would therefore be an excessive optimization if the HTTP protocol was moved to a lighter Transport Layer protocol such as the [Transactional TCP Protocol](#) which still provides a reliable stream transport service.

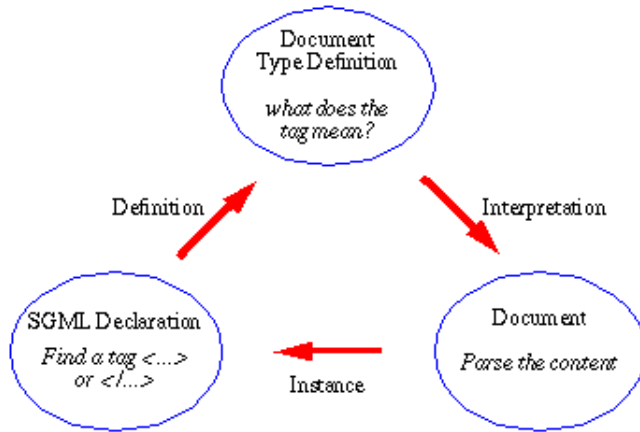
## Hypertext Markup Language

[The Hypertext Markup language \(HTML\)](#) is the users interface to create information on the World-Wide Web. The description of the World-Wide Web has until now focused on the technology that due to specifications and conventions provide the functionality necessary to request and serve information across the Internet. HTML is defined to be the hypertext language of communication which actually flows over the network. There is no requirement that files are stored in HTML. Servers may store files in any other formats and then generate a HTML on the fly upon a client request. This gives the possibility of having virtual documents instead of static documents on rapidly changing information like weather reports etc. HTML can be used to represent:

- Hypertext news, mail, online documentation, and collaborative hypermedia
- Menus and options
- Database query results
- Simple structured documents with inlined multi media elements like images, audio and movie
- URI-Links to other resources on the Internet.

HTML is built on top of the International Standard ISO 8879 Standard Generalized Markup Language (SGML). SGML is a system for defining structured document types and markup languages to represent instances of the document types. That is, HTML is a [Document Type Definition \(DTD\)](#) used on top of a SGML parser. Every

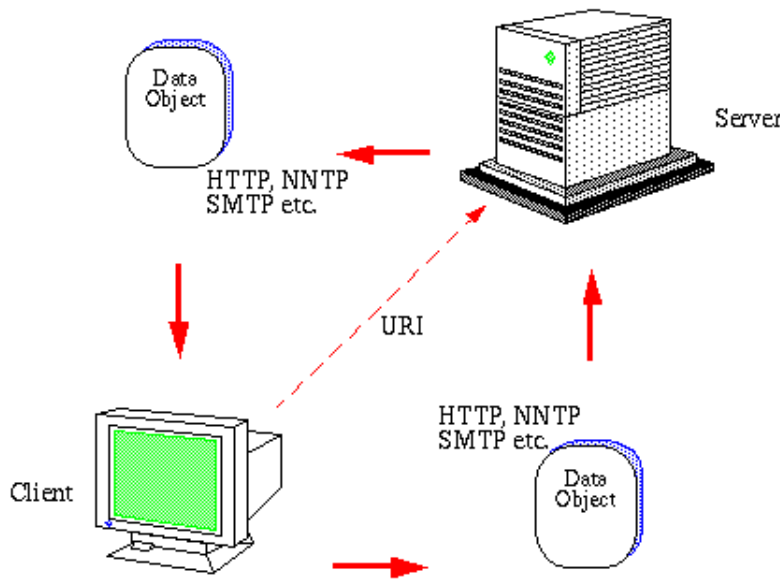
SGML based document contains three elements as illustrated in the figure:



HTML is now superseded by [HTML+](#) that is an enriched DTD with possibilities of handling tables, math, images etc. Currently many browsers support a subset of the HTML+ specifications in addition to the basic HTML features.

### Interactive World-Wide Web Model

The description of the [Universal Resource Identifiers](#), the [Hypertext Transfer Protocol](#), and the [Hypertext Markup Language](#) now calls for an update of the [Basic WWW-model](#) as illustrated in the figure.



This model is a generic resource exchange model based on a client-server concept. Instead of the limited model with data flowing only from the server to the client, the client is capable of posting data to the server if the server allows this kind of service. Furthermore, the format of the data transferred in the message body can have any format from 7-bit ASCII text to 8-bit binary data. The transfer carrier can be any protocol supported by the World-Wide Web but the main protocol is HTTP as it can be used to encapsulate the other protocols supported, even a FTP message that is a highly state dependent protocol.



*Henrik Frystyk, July 1994*