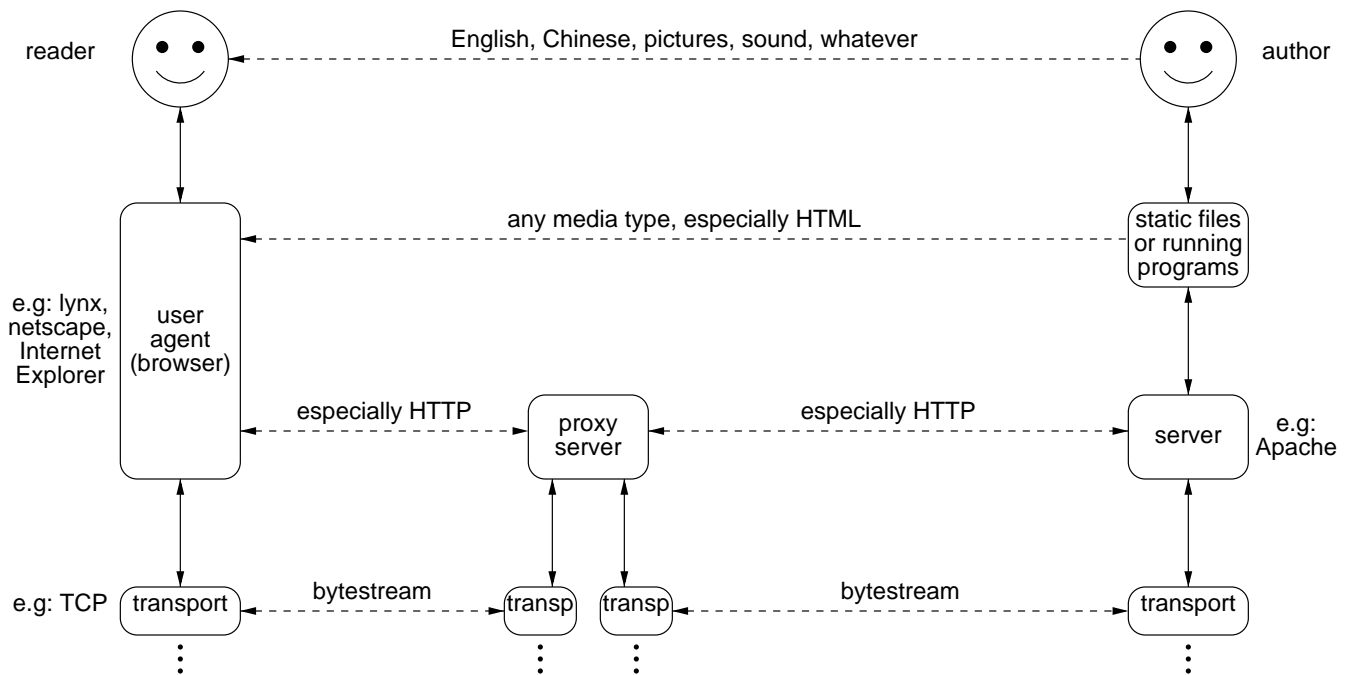


# World Wide Web

Adam M. Costello <amc@cs.berkeley.edu> 1999-Sep-02-Thu



Resources (like web pages) are referred to by Uniform Resource Indicators (URIs), which have the form:

scheme:scheme-specific-part

Many URIs are structured this way:

scheme://authority/p/a/t/h?query  
e.g: <http://www.m-w.com/cgi-bin/mweb?va=spam>

URIs can be Uniform Resource Names (URNs), e.g:

urn:isbn:0-256-17404-0

URNs are independent of a resource's location, but they have not been standardized and are not yet used. URIs can also be Uniform Resource Locators (URLs), which tell where to find a resource, for example:

<http://www.cs.berkeley.edu/~amc/monroe/>  
<ftp://ftp.isi.edu/in-notes/rfc2396.txt>

In these examples the scheme tells what protocol to use, the authority names a host running a server, and the path names a file on the server. See also RFC 2396.

Resources that are objects can be any MIME media type (see RFC 2045), like text/plain or image/jpeg, but the HyperText Markup Language (HTML, text/html) is especially for web pages. An HTML document consists of elements delimited by tags of the form:

```
<tag-name attribute-name=attribute-value ...>
contents of the element
</tag-name>
```

```
<html>
<head><title>Example HTML Document</title></head>
<body>
<h1>This is a Heading</h1>
<p>This is a paragraph containing a link to the <a href =
"http://www.w3.org/pub/WWW/TR/REC-html40">HTML
specification</a>. Next is an unordered list.</p>
<ul>
<li>first item</li>
<li>second item</li>
</ul>
</body>
</html>
```

Resources can be retrieved by a variety of protocols, but the HyperText Transfer Protocol (HTTP) is especially for the web. The browser sends a request to the server, which sends back a response. Both requests and responses contain an RFC-822-like header optionally followed by an arbitrary body (not necessarily text). For example:

```
request: GET /~amc/monroe/ HTTP/1.1      (request-line)
Host: www.cs.berkeley.edu
User-Agent: ACME Super Web Browser 2.1.3
Accept: text/html;q=1, text/plain;q=0.4, image/jpeg
Accept-Charset: iso-8859-1, iso-2022-jp, utf-8
Accept-Encoding: gzip, compress
Accept-Language: en, ja
```

header

(GET requests have no body)

```
response: HTTP/1.1 200 OK                  (status-line)
Server: Apache/1.3.4 (Unix)
Date: Wed, 01 Sep 1999 05:52:34 GMT
ETag: "31b0e8-6ec-3739d086;31b29e-87-36..."
Last-Modified: Wed, 12 May 1999 19:03:34 GMT
Vary: accept-encoding
Content-Type: text/html
Content-Length: 1772
Content-Encoding: gzip
Content-Language: en
```

header

```
<html>
<head>
<title>Monroe</title>
</head>
<body>
<h1>Monroe</h1>
<p>Monroe is a cat belonging to <a
href="http://www.cs.berkeley.edu/~amc/">Adam
Costello</a>...
```

body

See also RFC 2068.

In order to access the web from behind a firewall, browsers can be configured to send the request to an HTTP proxy server, which relays the request to the actual server, then relays the response back to the browser. Proxies are also useful for caching: if the proxy server can return a cached response, it can avoid a request to an overloaded or far away server.