CSE 480
Web Technologies

Rifat Ahmmad Rashid

Learning Goals

- Describe how Web servers and clients interact with each other.

- Request resources from web servers and understand the responses.

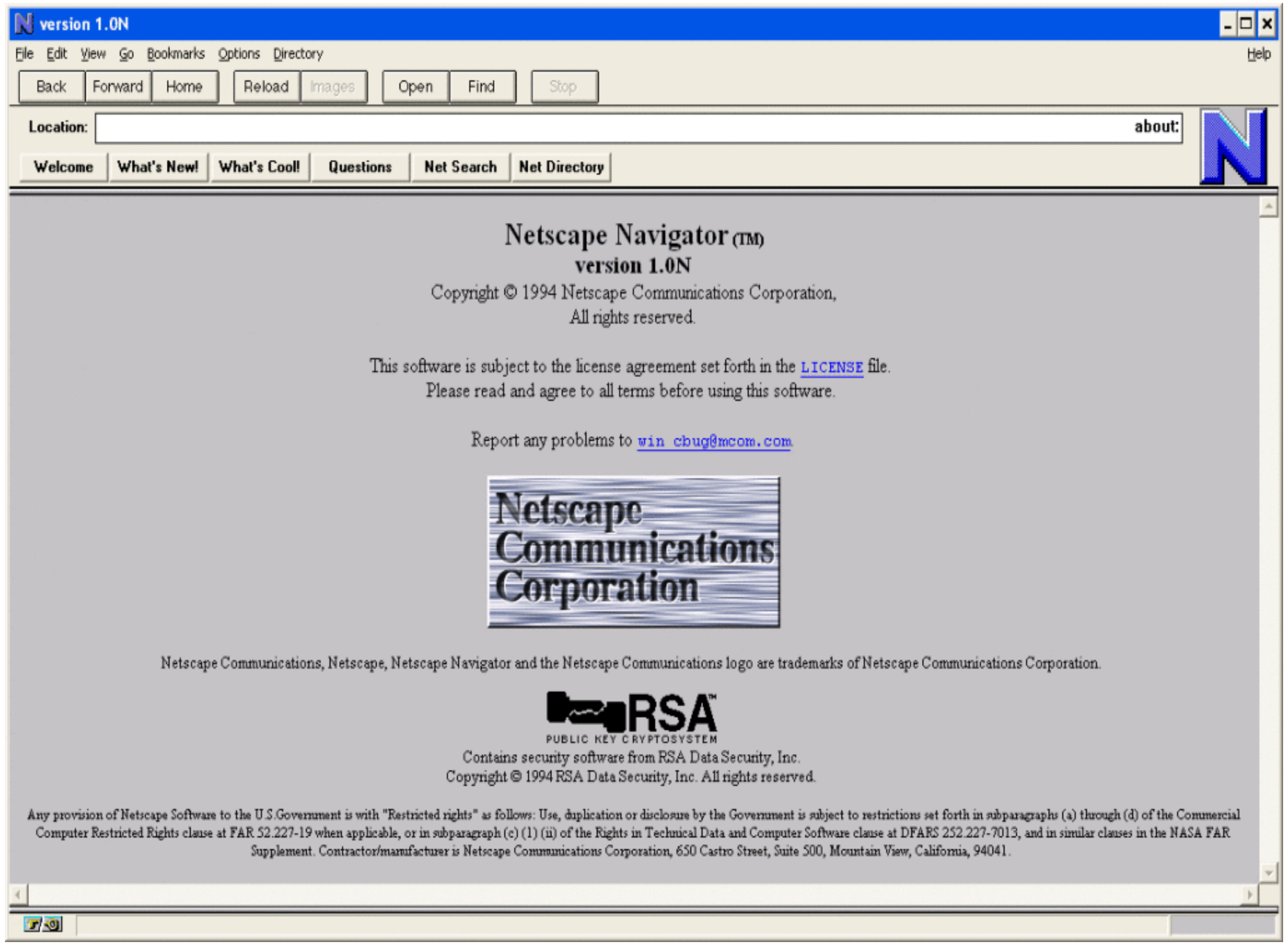
- Describe the different URL components

- Explain the difference between HTTP and HTTPS.

The Web: a brief history

World Wide Web: a global system of interconnected hypertext documents available via the Internet (envisioned already in 1945)

- 1960s: Precursor to the Internet (ARPANET) devised by the US department of Defense → Initial services: electronic mail, file transfer
- Late 1980s: Internet opened to commercial interests
- 1989: WWW created by Tim Berners-Lee (CERN)
- 1994: Netscape released its first Web browser
- 1995: Microsoft released Internet Explorer v1
- 1998: Google was founded
- 2002: Mozilla released Firefox v1


version 1.0N
File Edit View Go Bookmarks Options Directory Help
Back Forward Home Reload Images Open Find Stop
Location:
about:
Welcome What's New! What's Cool! Questions Net Search Net Directory

Netscape Navigator (TM)
version 1.0N
Copyright © 1994 Netscape Communications Corporation,
All rights reserved.

This software is subject to the license agreement set forth in the LICENSE file.
Please read and agree to all terms before using this software.

Report any problems to win_cbug@mcom.com

Netscape Communications Corporation

Netscape Communications, Netscape, Netscape Navigator and the Netscape Communications logo are trademarks of Netscape Communications Corporation.

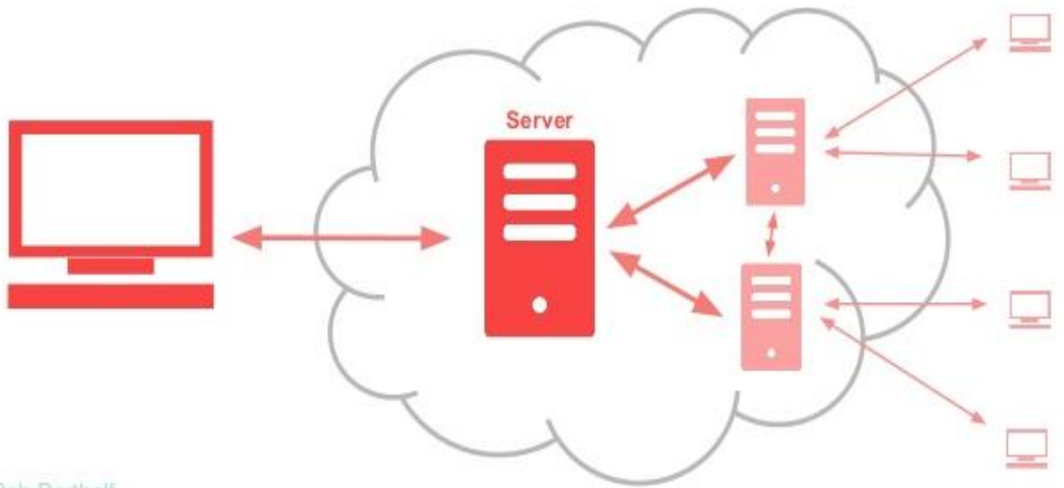
RSA
PUBLIC KEY CRYPTOSYSTEM
Contains security software from RSA Data Security, Inc.
Copyright © 1994 RSA Data Security, Inc. All rights reserved.

Any provision of Netscape Software to the U.S.Government is with "Restricted rights" as follows: Use, duplication or disclosure by the Government is subject to restrictions set forth in subparagraphs (a) through (d) of the Commercial Computer Restricted Rights clause at FAR 52.227-19 when applicable, or in subparagraph (c) (1) (ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.227-7013, and in similar clauses in the NASA FAR Supplement. Contractor/manufacturer is Netscape Communications Corporation, 650 Castro Street, Suite 500, Mountain View, California, 94041.

Key aspects of the Internet

Internet: interconnected computer networks that span the globe; communicating through a common standard (TCP/IP)

- Sub-networks function autonomously
- No centralised control
- Devices dynamically join/leave the network
- Devices interact through open standards
- Easy to use: server/client software widely available



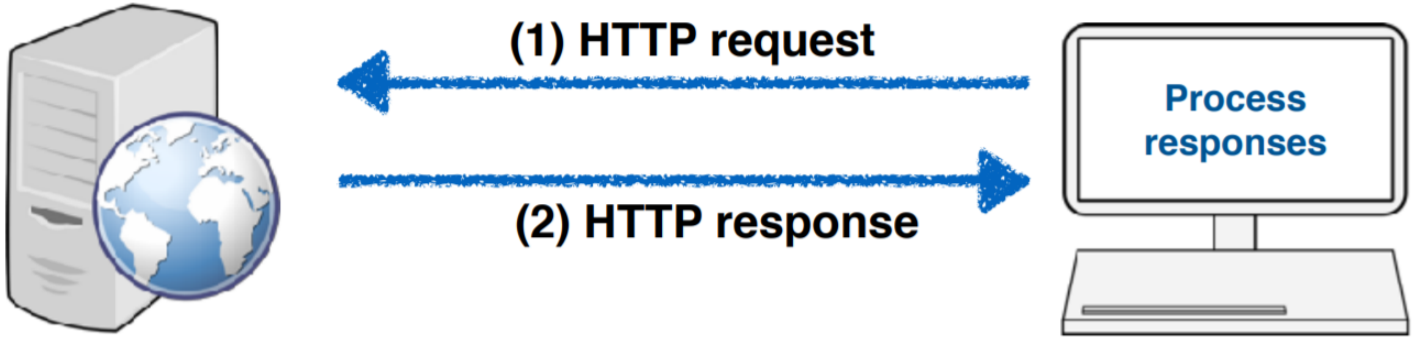
Rob Bertholf

Two important organisations

Internet Engineering Task Force (IETF) - "The mission of the IETF is to make the Internet work better by producing high quality, relevant technical documents that influence the way people design, use, and manage the Internet."

World Wide Web Consortium (W3C) - "The W3C mission is to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure the long-term growth of the Web."

Web servers and clients



- Servers wait for data requests
- Answer thousands of clients simultaneously
- Host web resources (content with an identity)

- Clients are often browsers
- Application:
  Display, execute, music player, Acrobat Reader

Network communication

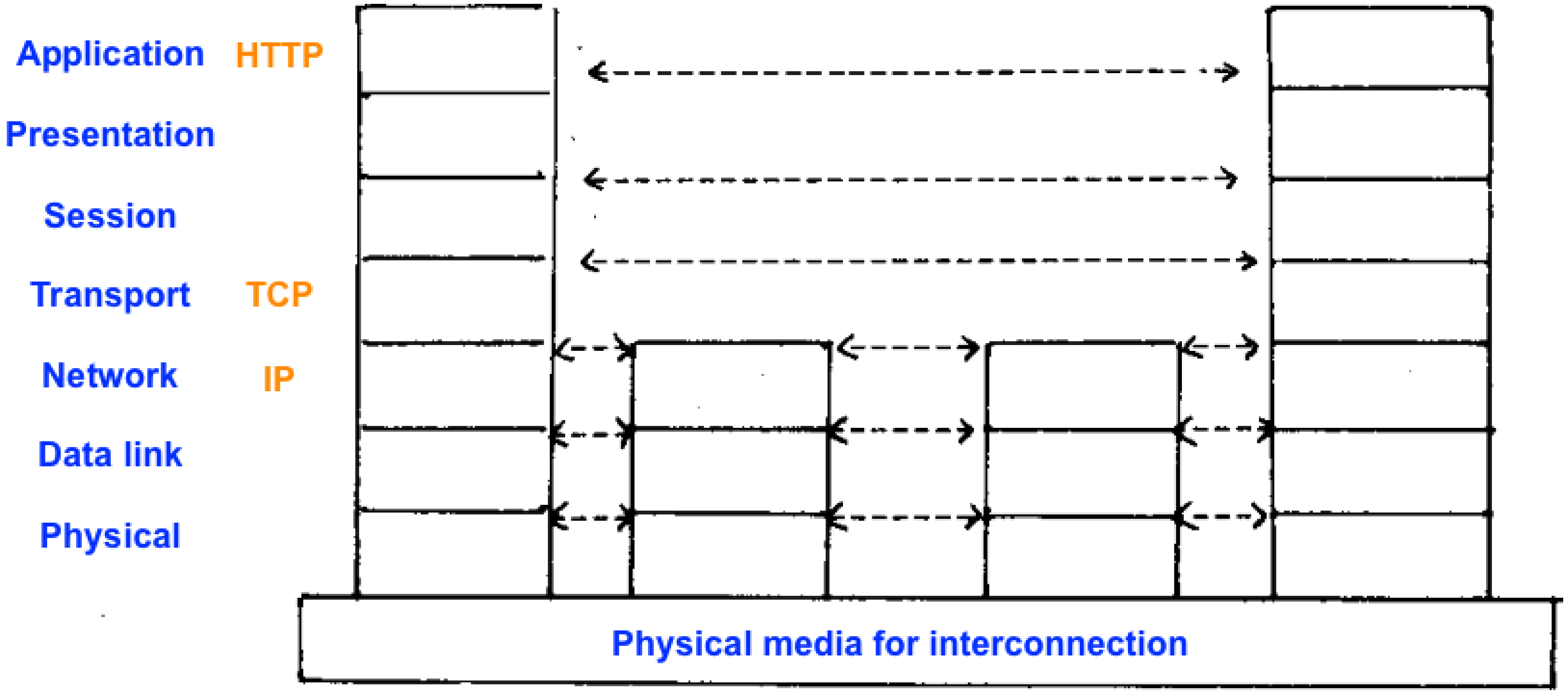
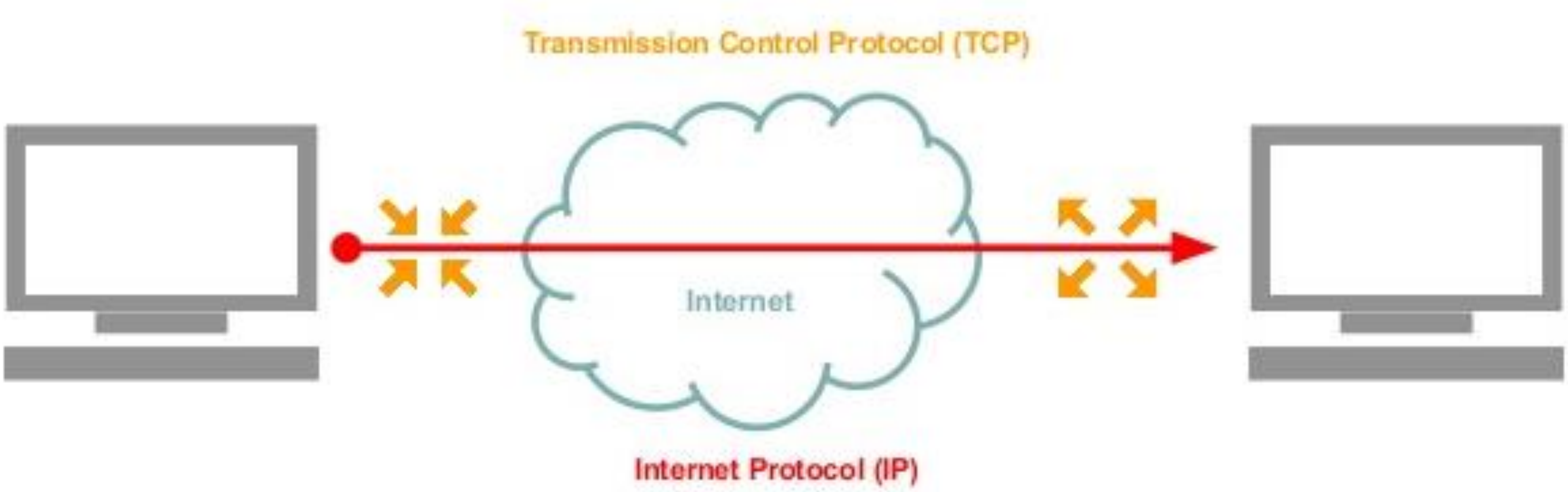


Image sourced from the OSI reference model paper

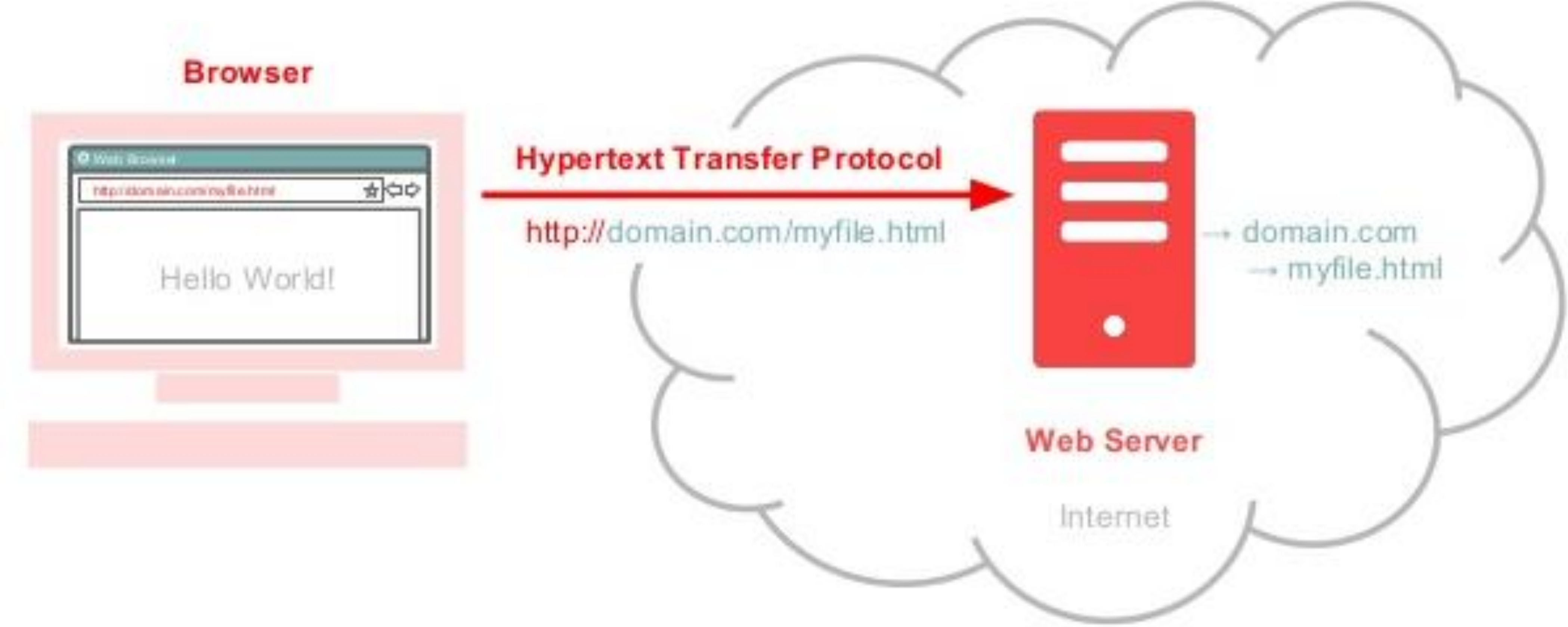
What is a Protocol ?

Set of rules that allow two electronic items to connect to and exchange information with one another.



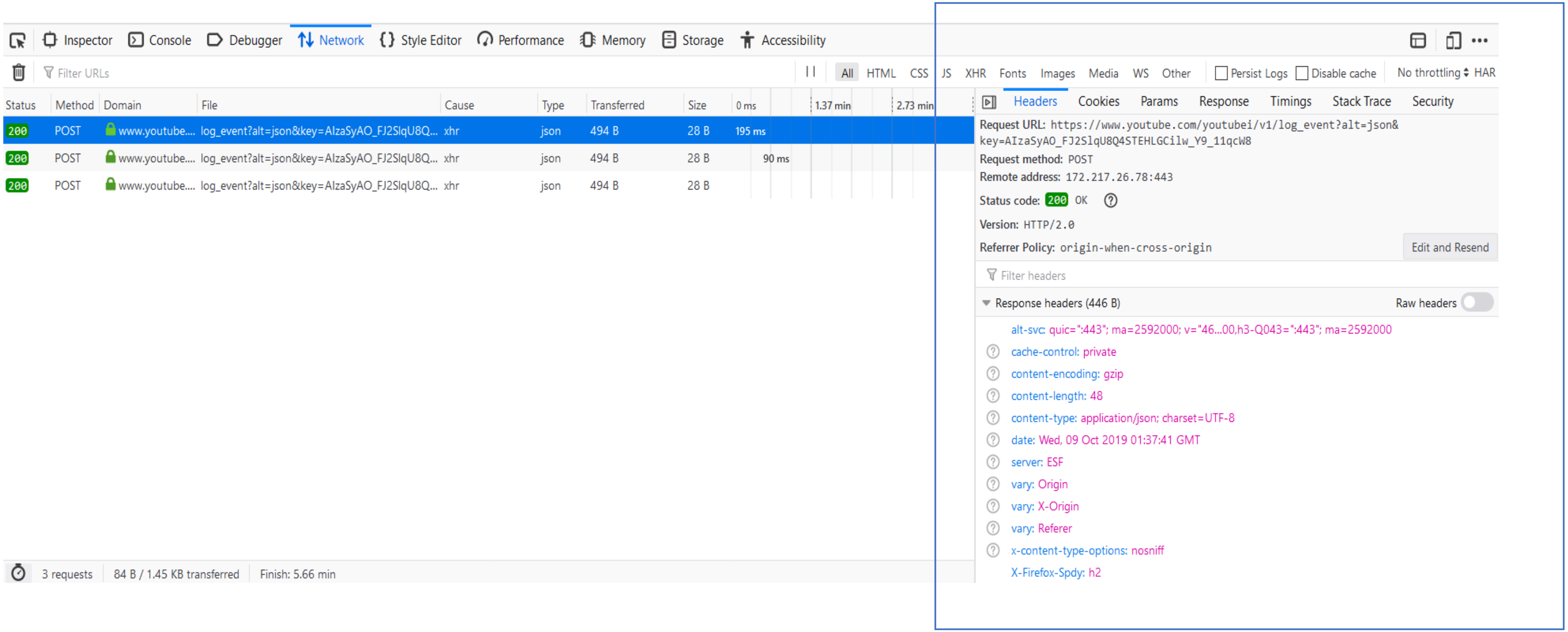
HTTP Protocol

Set of rules that allow two electronic items to connect to and exchange information with one another.


Browser
Web Browser
http://domain.com/myfile.html
Hello World!
Hypertext Transfer Protocol
http://domain.com/myfile.html
→ domain.com
→ myfile.html
Web Server
Internet

HTTP Protocol

Set of rules that allow two electronic items to connect to and exchange information with one another.



Status	Method	Domain	File	Cause	Type	Transferred	Size	0 ms 1.37 min 2.73 min
200	POST	www.youtube....	log_event?alt=json&key=AIzaSyAO_FJ2SlqU8Q...	xhr	json	494 B	28 B	195 ms
200	POST	www.youtube....	log_event?alt=json&key=AIzaSyAO_FJ2SlqU8Q...	xhr	json	494 B	28 B	90 ms
200	POST	www.youtube....	log_event?alt=json&key=AIzaSyAO_FJ2SlqU8Q...	xhr	json	494 B	28 B	

3 requests | 84 B / 1.45 KB transferred | Finish: 5.66 min

Request URL: https://www.youtube.com/youtubei/v1/log_event?alt=json&key=AIzaSyAO_FJ2SlqU8Q4STEHLGCilw_Y9_11qcW8

Request method: POST

Remote address: 172.217.26.78:443

Status code: 200 OK

Version: HTTP/2.0

Referrer Policy: origin-when-cross-origin

Response headers (446 B)

```
alt-svc: quic=":443"; ma=2592000; v="46...00,h3-Q043=":443"; ma=2592000
cache-control: private
content-encoding: gzip
content-length: 48
content-type: application/json; charset=UTF-8
date: Wed, 09 Oct 2019 01:37:41 GMT
server: ESF
vary: Origin
vary: X-Origin
vary: Referer
x-content-type-options: nosniff
X-Firefox-Spdy: h2
```

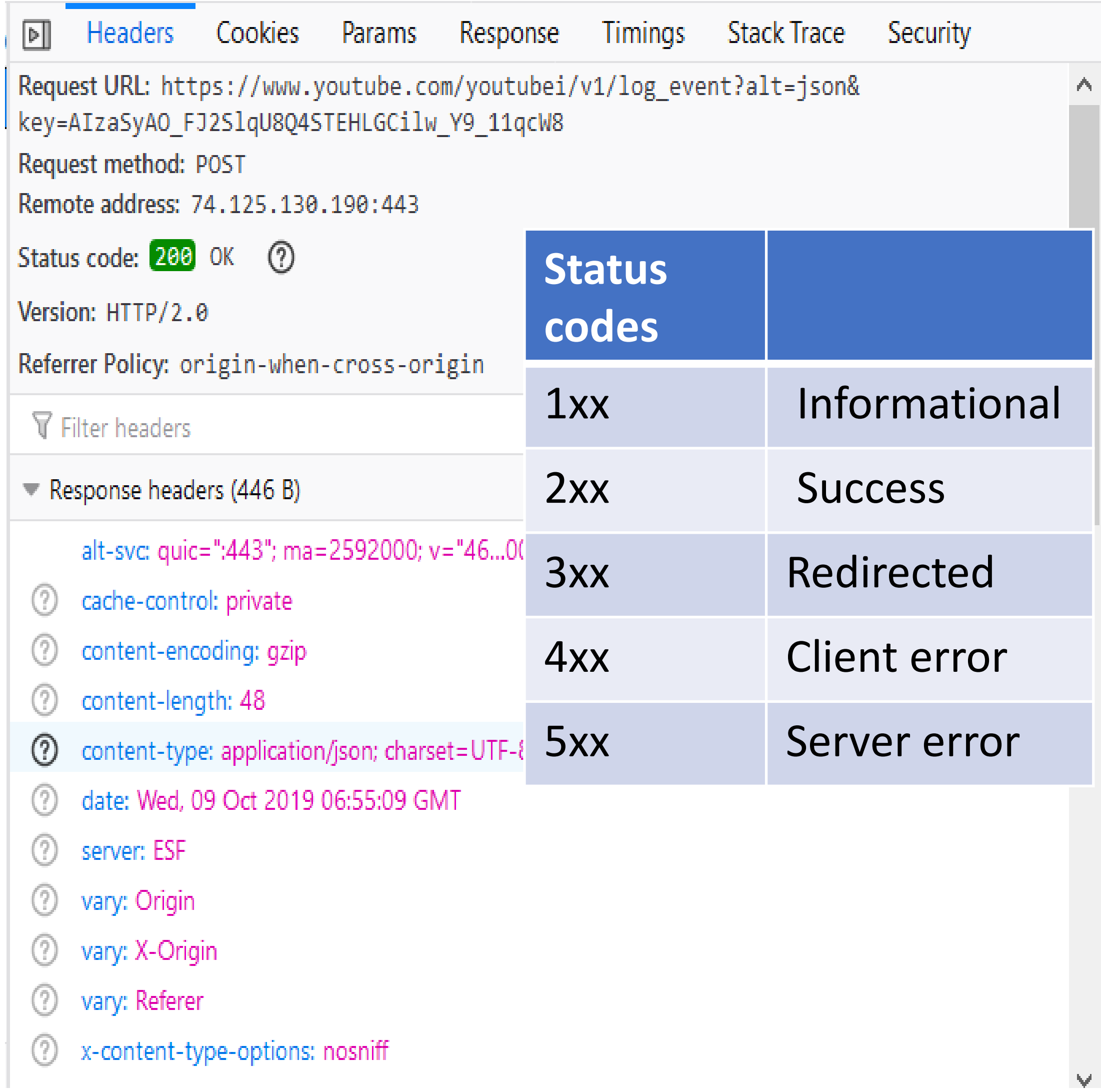
HTTP Protocol

HTTP request message

HTTP is a plain text protocol and line-oriented. The first line indicates what this message is about. In this case the keyword POST indicates that we are sending something.

HTTP response message

The status code indicates the status of the response. In this case, the server sends back the status 200 OK: everything is okay, the resource was found, you are allowed to receive it.



Request URL: https://www.youtube.com/youtubei/v1/log_event?alt=json&
key=AIzaSyAO_FJ2SlqU8Q4STEHLGCilw_Y9_11qcW8
Request method: POST
Remote address: 74.125.130.190:443
Status code: 200 OK
Version: HTTP/2.0
Referrer Policy: origin-when-cross-origin

Response headers (446 B)

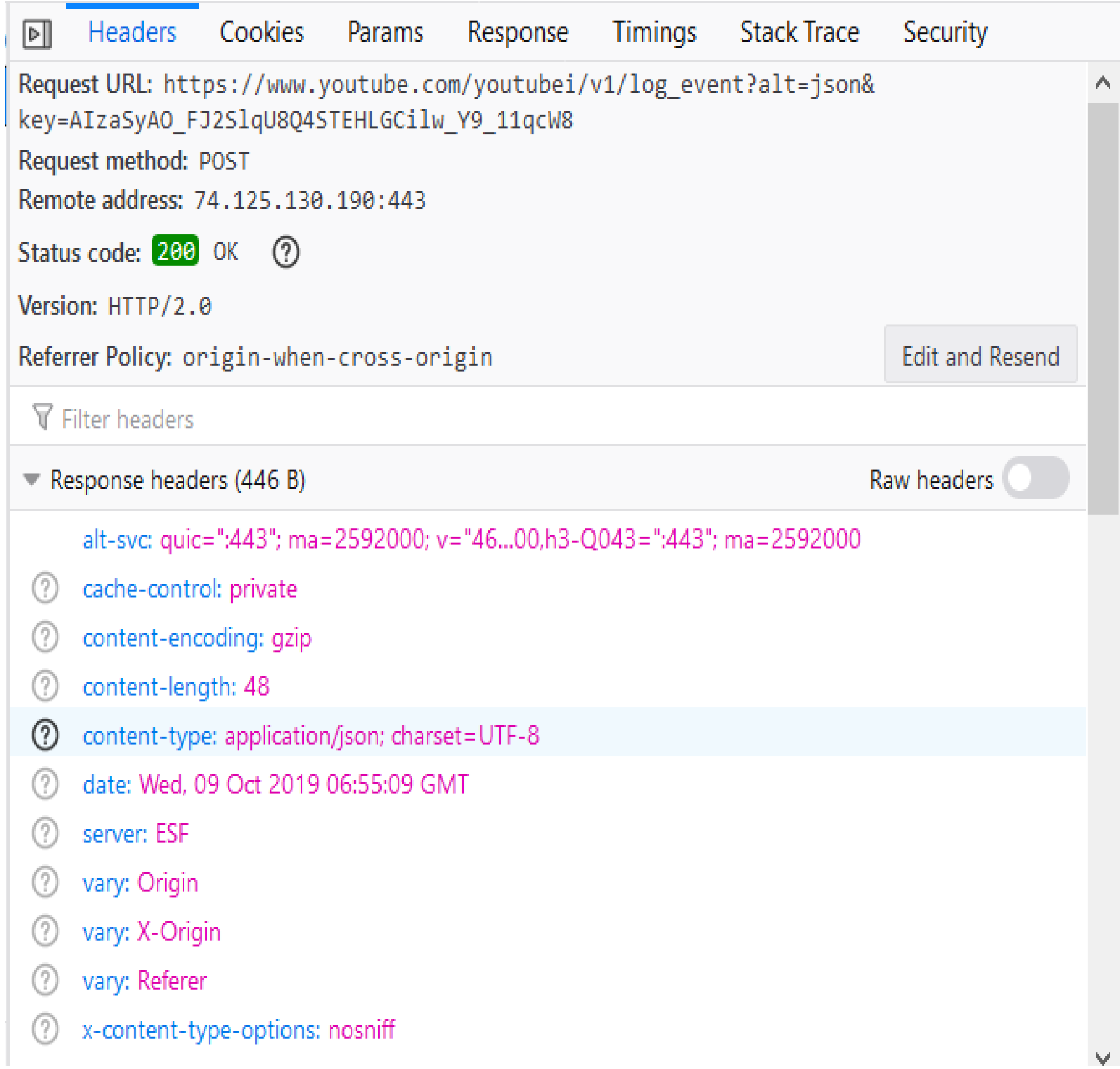
alt-svc: quic=":443"; ma=2592000; v="46...0(
cache-control: private
content-encoding: gzip
content-length: 48
content-type: application/json; charset=UTF-8
date: Wed, 09 Oct 2019 06:55:09 GMT
server: ESF
vary: Origin
vary: X-Origin
vary: Referer
x-content-type-options: nosniff

Status codes	
1xx	Informational
2xx	Success
3xx	Redirected
4xx	Client error
5xx	Server error

HTTP Protocol

Well-known header fields

Header field	Description
Content-Type	Entity type
Content-Length	Length/size of the message
Content-Language	Language of the entity sent
Content-Encoding	Data transformations applied to the entity
Content-MD5	Checksum of the content
Last-Modified	Date at which the entity will become stale
Connection & Upgrade	Protocol upgrade



Request URL: https://www.youtube.com/youtubei/v1/log_event?alt=json&key=AIzaSyAO_FJ2SlqU8Q4STEHLGCilw_Y9_11qcW8

Request method: POST

Remote address: 74.125.130.190:443

Status code: 200 OK

Version: HTTP/2.0

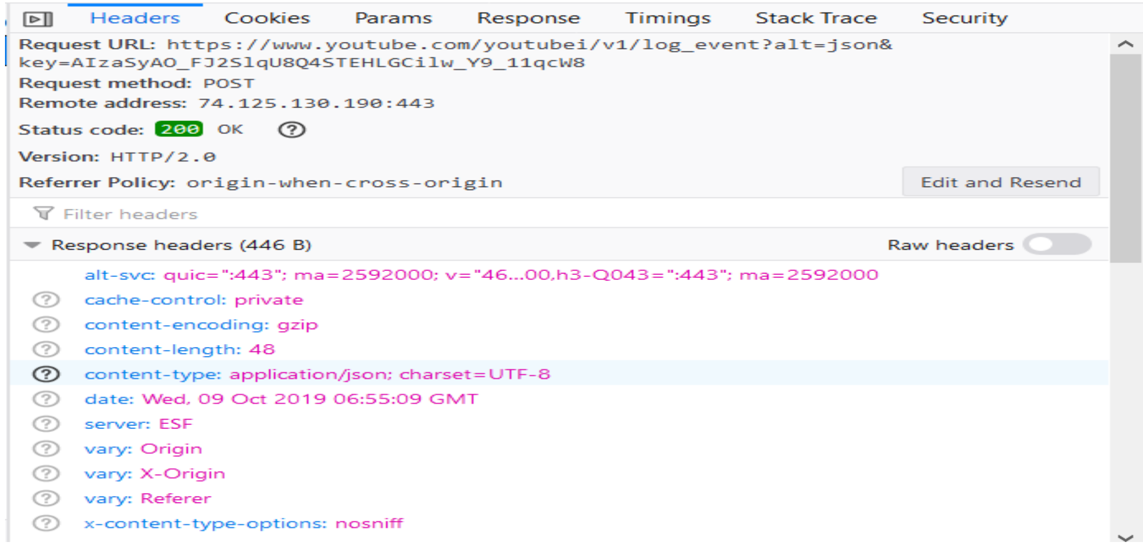
Referrer Policy: origin-when-cross-origin

Response headers (446 B)

```
alt-svc: quic=":443"; ma=2592000; v="46...00,h3-Q043=":443"; ma=2592000
cache-control: private
content-encoding: gzip
content-length: 48
content-type: application/json; charset=UTF-8
date: Wed, 09 Oct 2019 06:55:09 GMT
server: ESF
vary: Origin
vary: X-Origin
vary: Referer
x-content-type-options: nosniff
```

HTTP Protocol

Content-Type
MIME stands for *Multipurpose Internet Mail Extensions*
and was designed to solve problems when moving messages between electronic mail systems; it worked well and was adopted by HTTP to label its content.



```
Request URL: https://www.youtube.com/youtubei/v1/log_event?alt=json&
key=AIzaSyAO_FJ2SlqU8Q4STEHLGCilw_Y9_11qcW8
Request method: POST
Remote address: 74.125.130.190:443
Status code: 200 OK
Version: HTTP/2.0
Referrer Policy: origin-when-cross-origin
```

Response headers (446 B)

```
alt-svc: quic=":443"; ma=2592000; v="46...00,h3-Q043=":443"; ma=2592000
cache-control: private
content-encoding: gzip
content-length: 48
content-type: application/json; charset=UTF-8
date: Wed, 09 Oct 2019 06:55:09 GMT
server: ESF
vary: Origin
vary: X-Origin
vary: Referer
x-content-type-options: nosniff
```

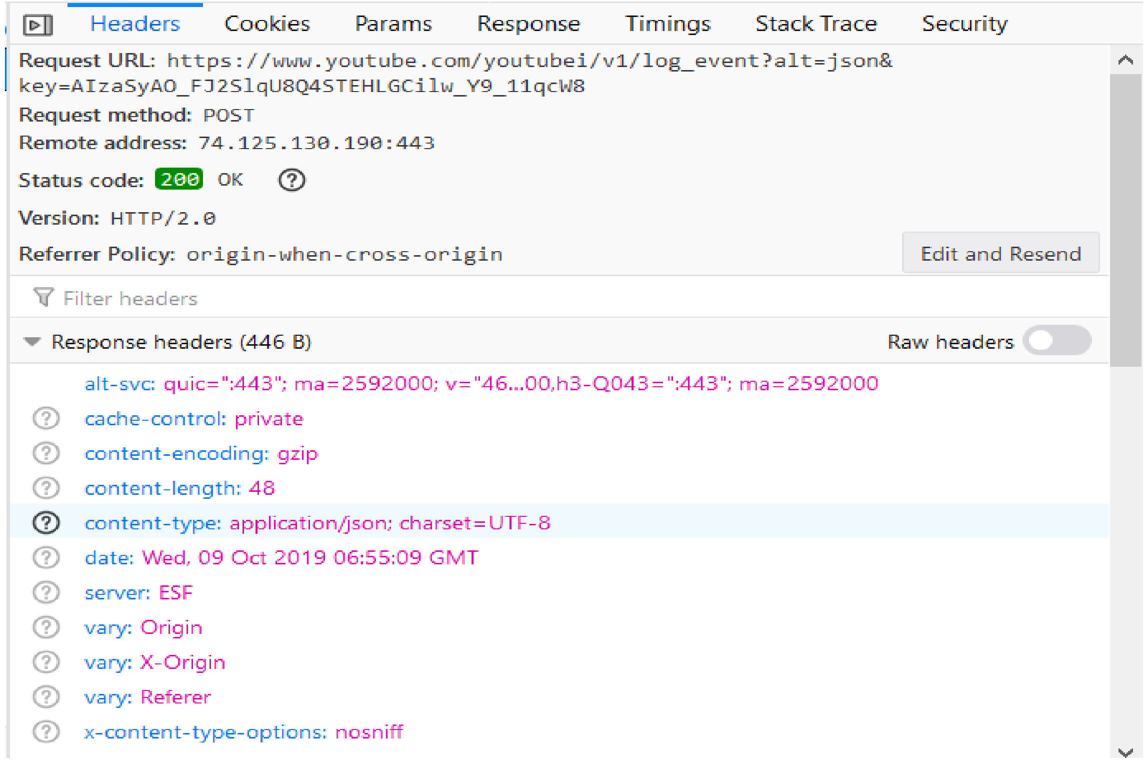
Most popular	Least popular
Text/html	application/pgp-keys
image/jpg	application/x-httpd-php4
text/xml	chemical/x-pdb
application/rss+xml	model/mesh
text/plain	application/x-perl
application/xml	audio/x_mpegurl
text/calendar	application/bib
application/pdf	application/postscript
application/atom+xml	application/x-msdos-program

HTTP Protocol

Content-Encoding

Content is often encoded, and in particular **compressed**. The four common encodings are:

(a) gzip (b) compress (c) deflate (d)identity (this encoding indicates that no encoding should be used)



```
Request URL: https://www.youtube.com/youtubei/v1/log_event?alt=json&
key=AIzaSyAO_FJ2SlqU8Q4STEHLGCilw_Y9_11qcW8
Request method: POST
Remote address: 74.125.130.190:443
Status code: 200 OK
Version: HTTP/2.0
Referrer Policy: origin-when-cross-origin

Response headers (446 B)
alt-svc: quic=":443"; ma=2592000; v="46...00,h3-Q043=":443"; ma=2592000
cache-control: private
content-encoding: gzip
content-length: 48
content-type: application/json; charset=UTF-8
date: Wed, 09 Oct 2019 06:55:09 GMT
server: ESF
vary: Origin
vary: X-Origin
vary: Referer
x-content-type-options: nosniff
```

How do client and server negotiate acceptable encodings?

If the server would send content in an encoding for which the client requires specific software to decode but does not have, the client receives a blob of data but is unable to interpret it. To avoid this situation, the client sends in the HTTP request a list of encodings it can deal with. This happens in the

But why bother with encodings at all?

If an image or video is compressed by the server before it is sent to the client, **network bandwidth is saved**. There is a **tradeoff**, however: compressed content needs to be decompressed by the client, which **increases the processing costs**.

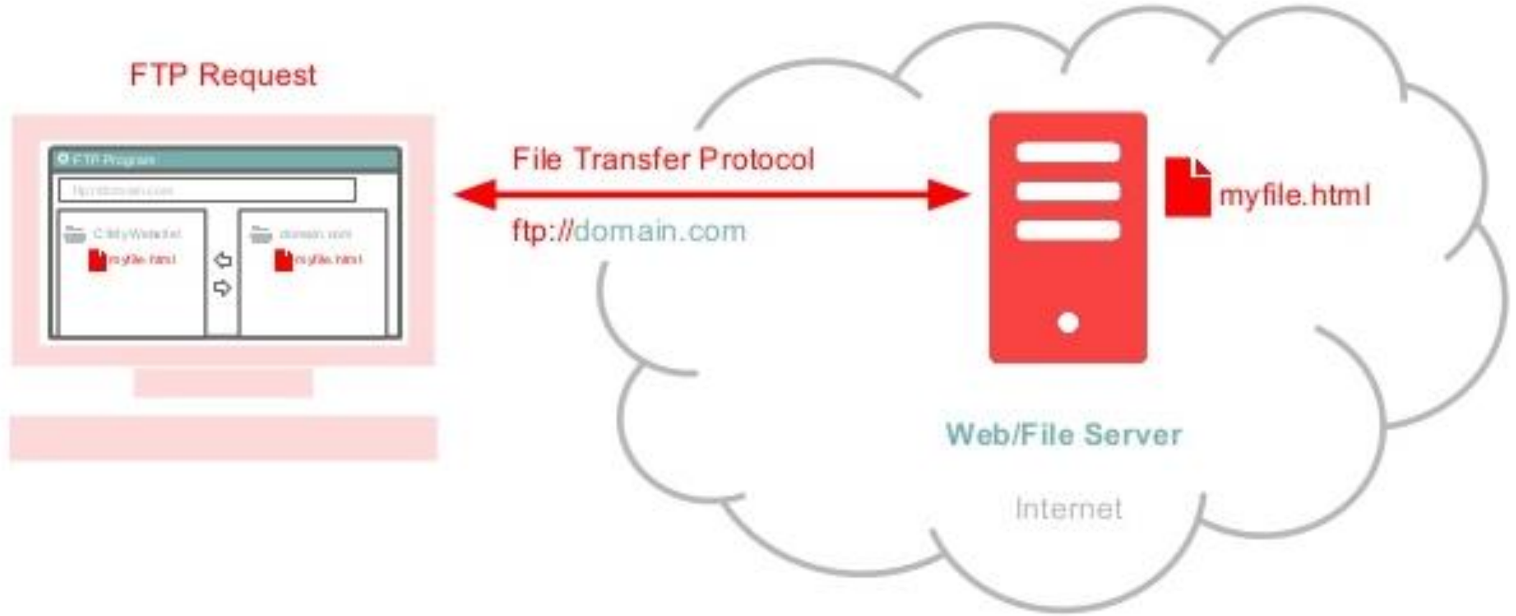
HTTP Protocol

Common HTTP methods

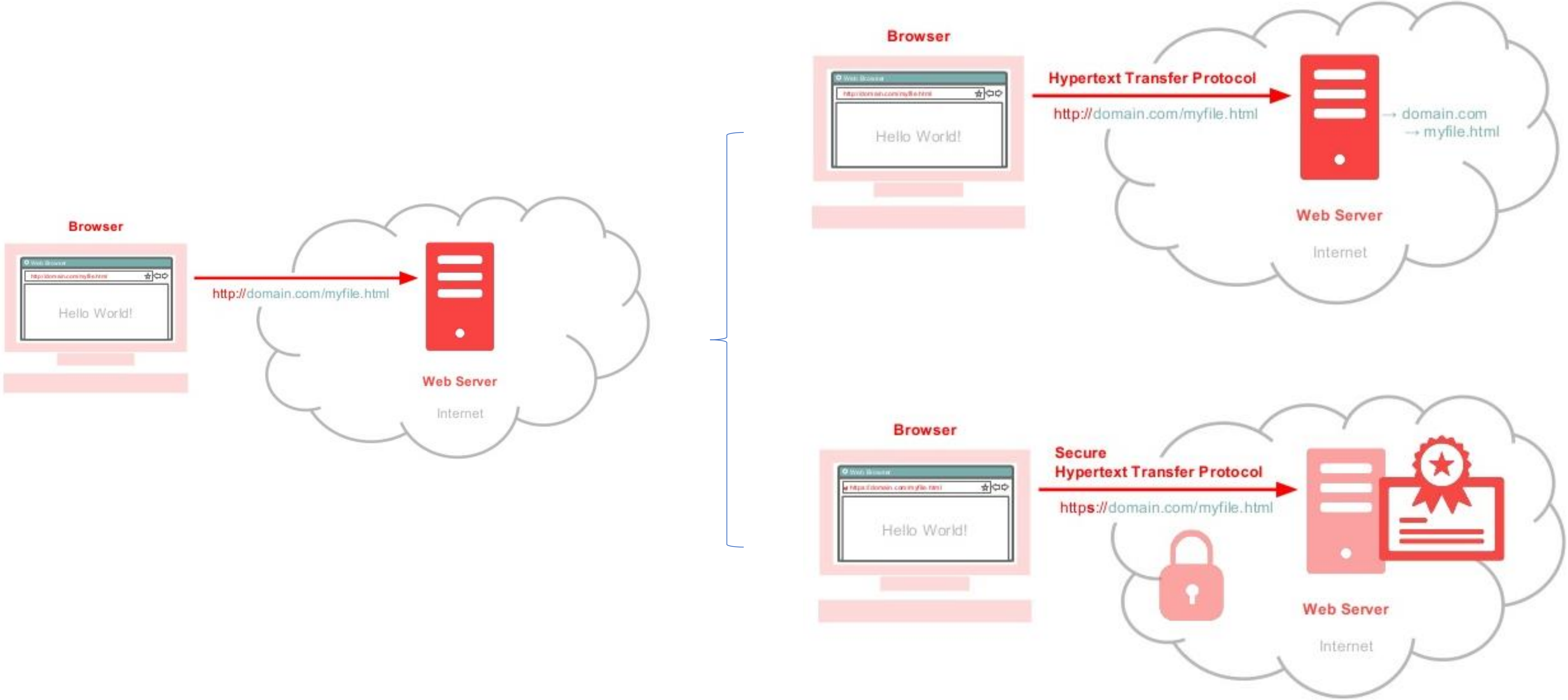
Methods	
GET	Request to get access to some web resource
HEAD	Returns the header of a HTTP response only (not the content)
POST	Sends data from the client to the server for processing
PUT	Saves the body of the request on the server; if you have ever used ftp you are already familiar with put
TRACE	Can be use to trace where a message passes through before arriving at the server
OPTIONS	Is helpful to determine what kind of methods a server supports
DELETE	Can be used to remove documents from a web server

File Transfer Protocol (FTP)

Set of rules that allow two electronic items to connect to and exchange information with one another.



Accessing the World Wide Web


Browser
Hello World!
http://domain.com/myfile.html
Web Server
Internet
Browser
Hello World!
Hypertext Transfer Protocol
http://domain.com/myfile.html
→ domain.com
→ myfile.html
Web Server
Internet
Browser
Hello World!
Secure
Hypertext Transfer Protocol
https://domain.com/myfile.html
Web Server
Internet

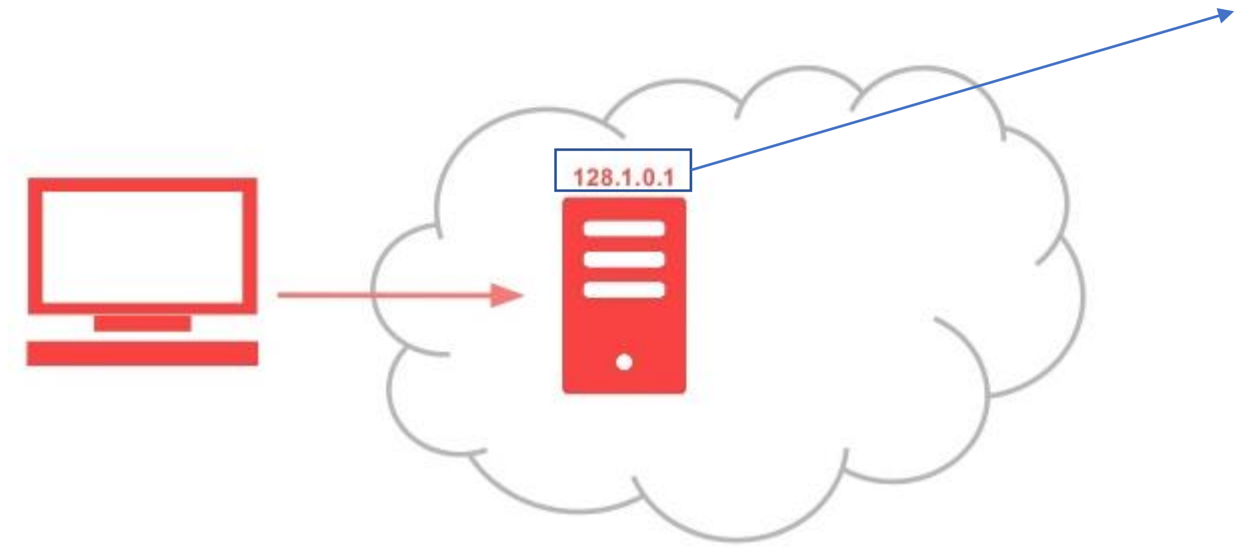
Protocol Recap

- We use Hypertext Transfer Protocol (http://) to access websites.

- We use File Transfer Protocol (ftp://) to store and retrieve files from the internet.

- SSL Certificates secure the HTTP protocol (https://) to guard data passed through.

How to find a web server?



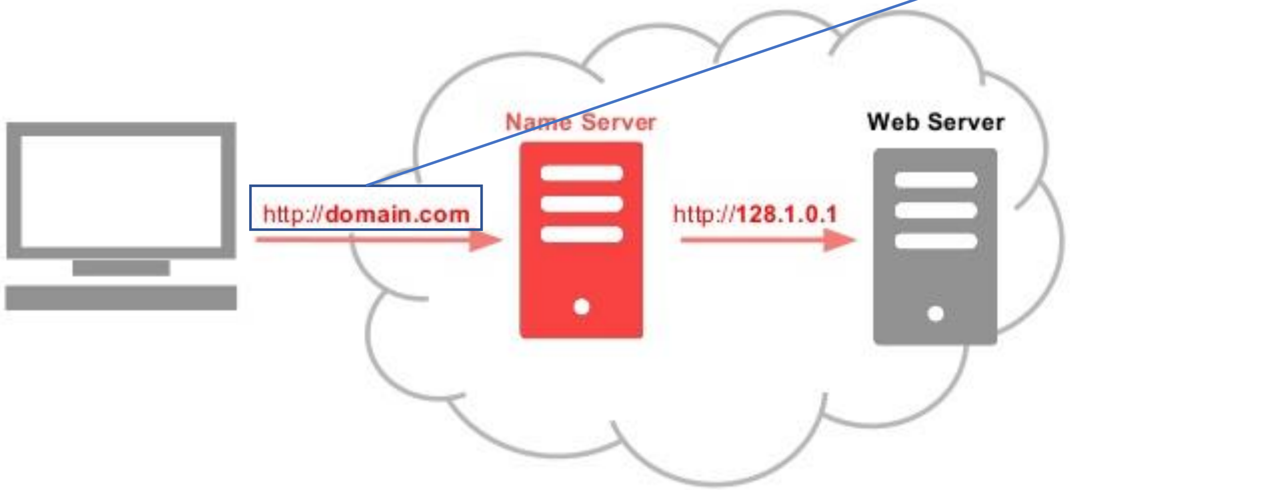
Internet Protocol (IP) address - a unique string of numbers separated by periods that identifies each computer using the Internet Protocol to communicate over a network.

→ Internet locations have IP addresses – 128.1.0.1

IP address – localhost – 127.0.0.1

- The IP address 127.0.0.1 is the loopback Internet protocol (IP) address also referred to as the “localhost”
- If you wanted to use your web browser to access your local web server this would be the IP address used to establish an IP connection to the same machine or computer being used by the end-user.

How to a Domain Name?



Domain Name – A domain name is an identification string that defines a realm of administrative autonomy, authority or control within the Internet.

Top level domain extensions
.com = Commercial
.org = Organization
.net =Network
.gov = Government
.edu = Education

Domain Name System (DNS) – a system for naming computers and network services that is organized into a hierarchy of domains. DNS naming is used in TCP/IP networks, such as the Internet, to locate computers and services through user-friendly names.

Uniform Resource Locators (URLs)

Let's now take a closer look at the format of *Uniform Resource Locators*, more commonly known by their abbreviation URLs.

URLs are the common way to access any resource on the Internet; the format of URLs is standardized. You should already be relatively familiar with the format of URLs accessing resources through HTTP and HTTPS. Resource access in other protocols (e.g. ftp) is similar, with only small variations.

In general, a URL consists of up to 9 parts:

```
<scheme>://<user>:<password>@<host>:<port>/<path>;<params>?<query>#<frag>
```

Uniform Resource Locators (URLs)

<scheme>://<user>:<password>@<host>:<port>/<path>;<params>?<query>#<frag>

From back to front

Methods	
<frag>	The name of a piece of a resource. Only used by the client - the fragment is not transmitted to the server.
<query>	Parameters passed to gateway resources, i.e. applications [identified by the path] such as search engines.
<params>	Additional input parameters applications may require to access a resource on the server correctly. Can be set per path segment.
<path>	the local path to the resource

Uniform Resource Locators (URLs)

<scheme>://<user>:<password>@<host>:<port>/<path>;<params>?<query>#<frag>

From back to front

Methods	
<port>	the port on which the server is expecting requests for the resource (ports enable multiplexing: multiple services are available on one location)
<host>	domain name (host name) or numeric IP address of the server
<user>:<password>	the username/password (may be necessary to access a resource)
<scheme>	determines the protocol to use when connecting to the server.

Uniform Resource Locators (URLs)

URL syntax: query

One of the most important URL types for us is the syntax for us is the syntax for a query.
What does that mean?

https://www.youtube.com/watch?**v=X-eea-Vf0Yw**

Schemes: more than just HTTP(S)

http and https differ in their encryption – http does not offer encryption, while https does. mailto is the email protocol, ftp is the file transfer protocol

Relative vs. absolute URLs:

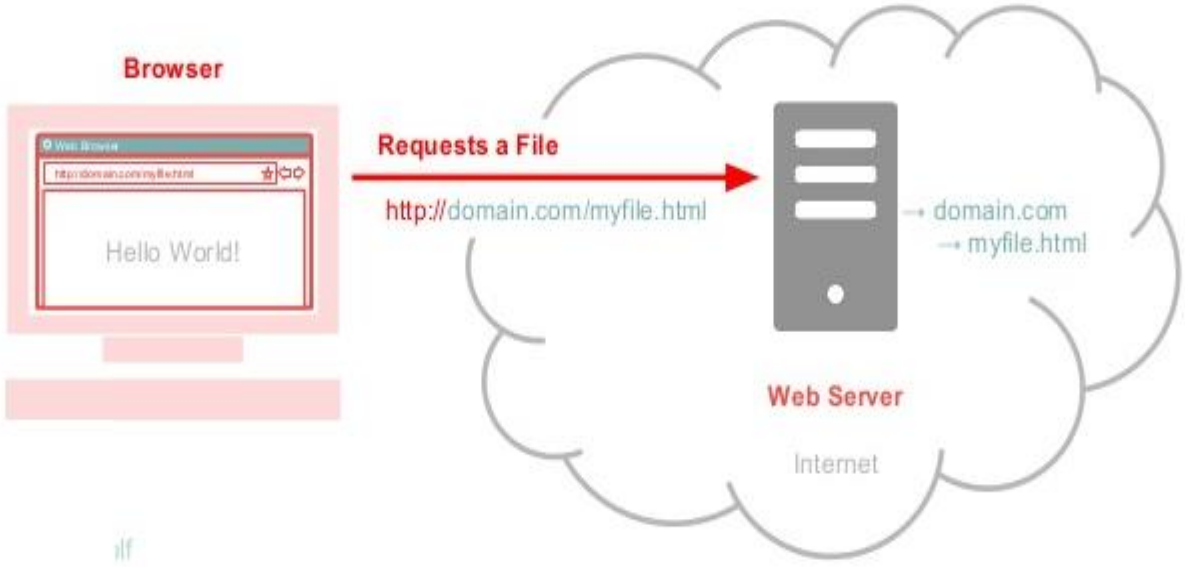
URLs can either be **absolute** or **relative**

```
https://ajax.googleapis.com/ajax/lib
s/jquery/3.1.0/jquery.min.js
```

```
<script src="static/js/form.js"></script>
```

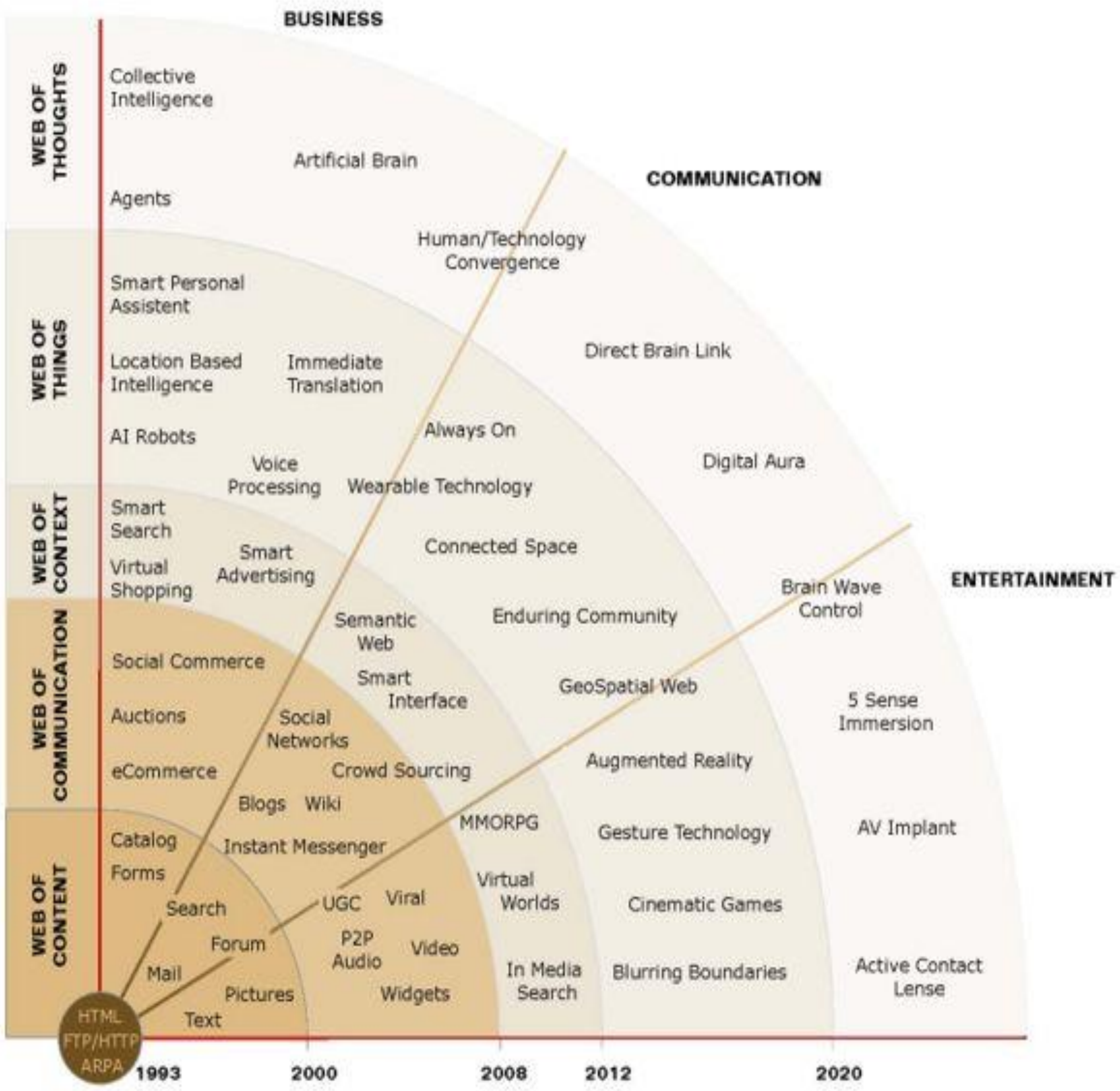
What is a Web Browser?

A web browser (commonly referred to as a browser) is a software application for retrieving, presenting, and traversing information resources on the World Wide Web.

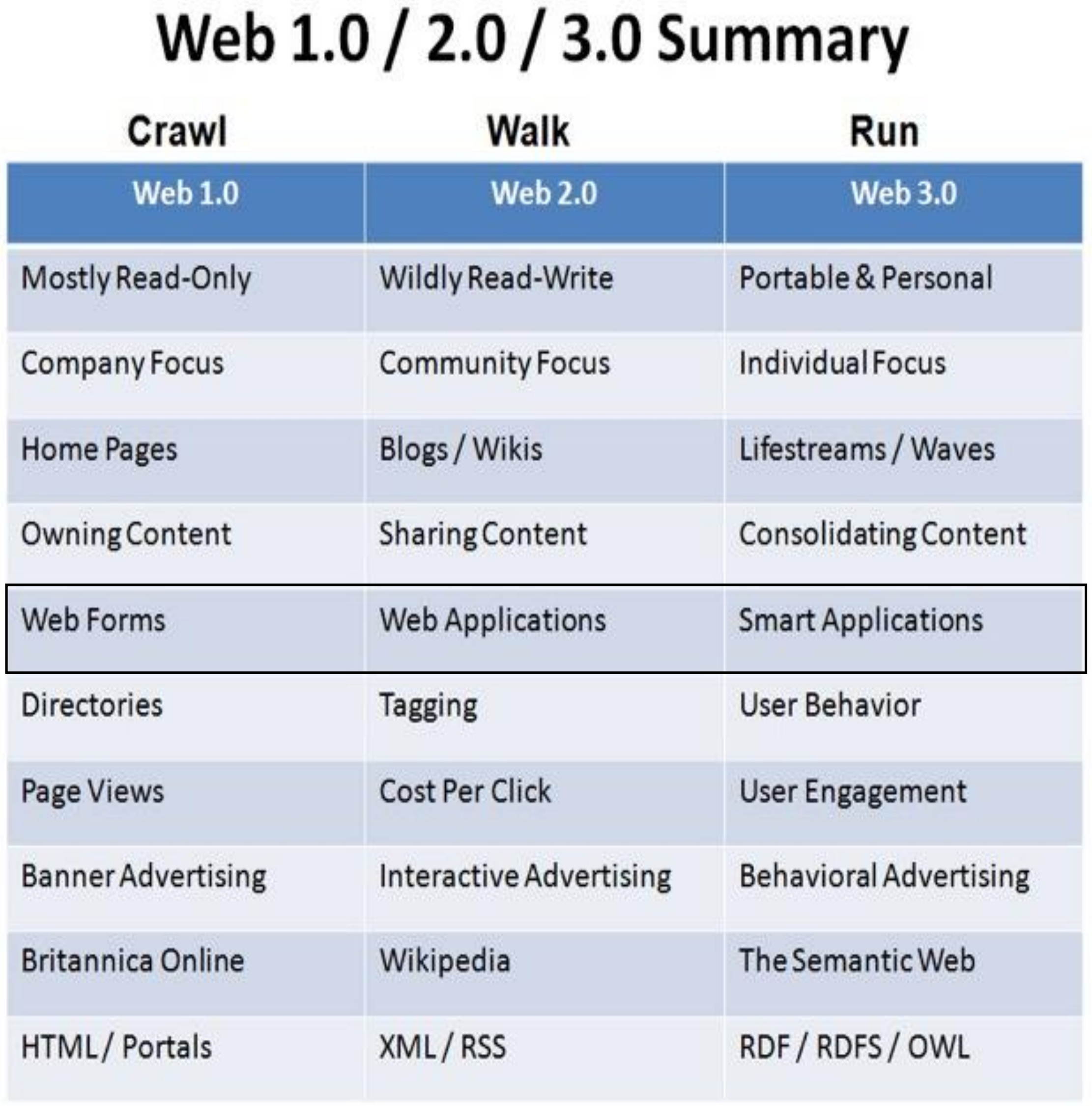


trendone
NILS MÜLLER COMMUNICATIONS

THE WEB EXPANSION
FROM WEB OF THINGS TO WEB OF THOUGHTS



©TrendONE 2008 by Nils Müller
www.TrendONE.de All rights reserved



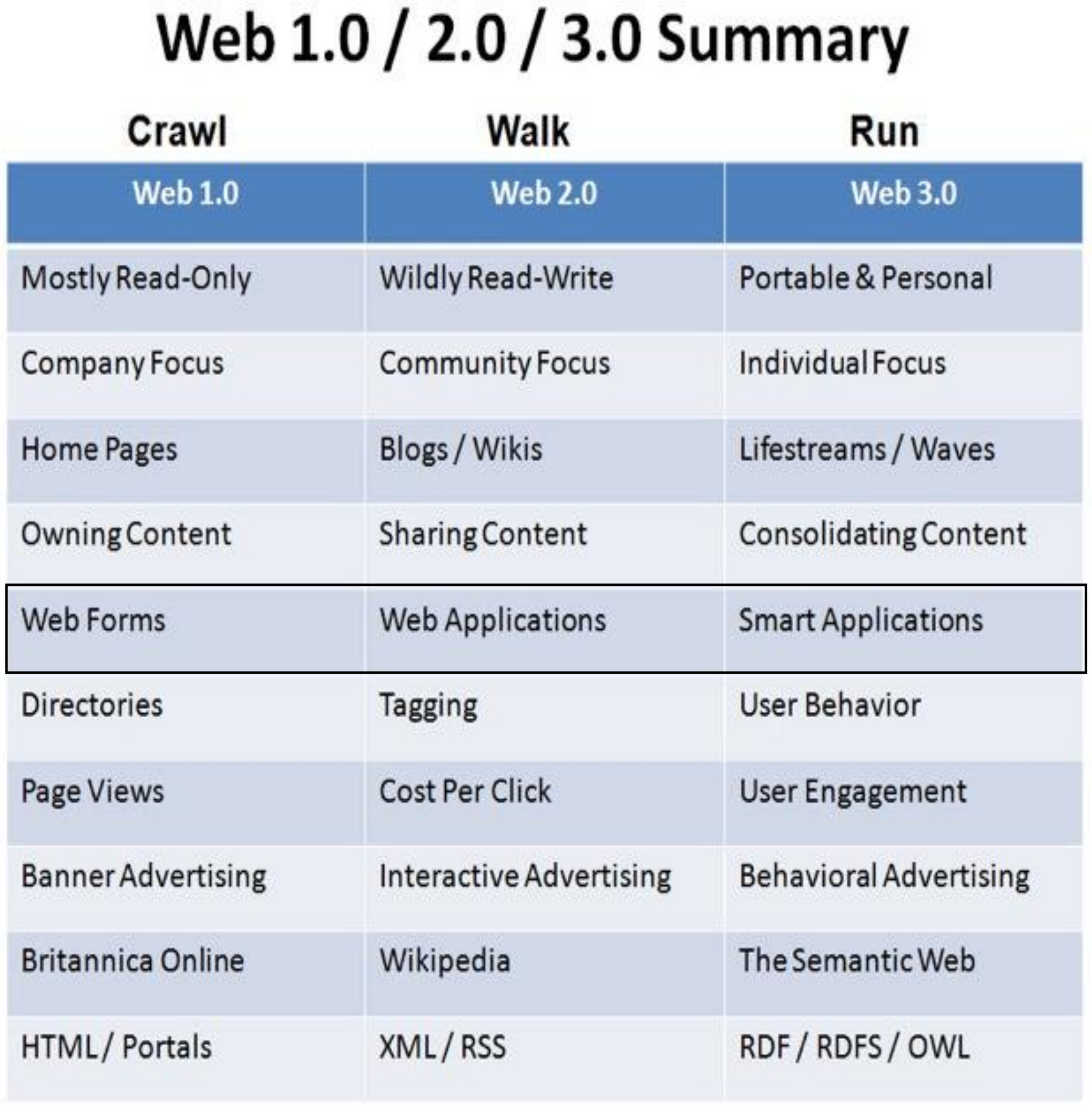
Web 1.0 / 2.0 / 3.0 Summary

Crawl	Walk	Run
Web 1.0	**Web 2.0**	**Web 3.0**
Mostly Read-Only	Wildly Read-Write	Portable & Personal
Company Focus	Community Focus	Individual Focus
Home Pages	Blogs / Wikis	Lifestreams / Waves
Owning Content	Sharing Content	Consolidating Content
Web Forms	Web Applications	Smart Applications
Directories	Tagging	User Behavior
Page Views	Cost Per Click	User Engagement
Banner Advertising	Interactive Advertising	Behavioral Advertising
Britannica Online	Wikipedia	The Semantic Web
HTML / Portals	XML / RSS	RDF / RDFS / OWL

Image source: https://flatworldbusiness.wordpress.com/flat-education/previously/web-1-0-vs-web-2-0-vs-web-3-0-a-bird-eye-on-the-definition/

Web 2.0 (Collaboration)
The second stage of development of the World Wide Web, characterized especially by the change from static Web pages to dynamic or user-generated content and the growth of social media.

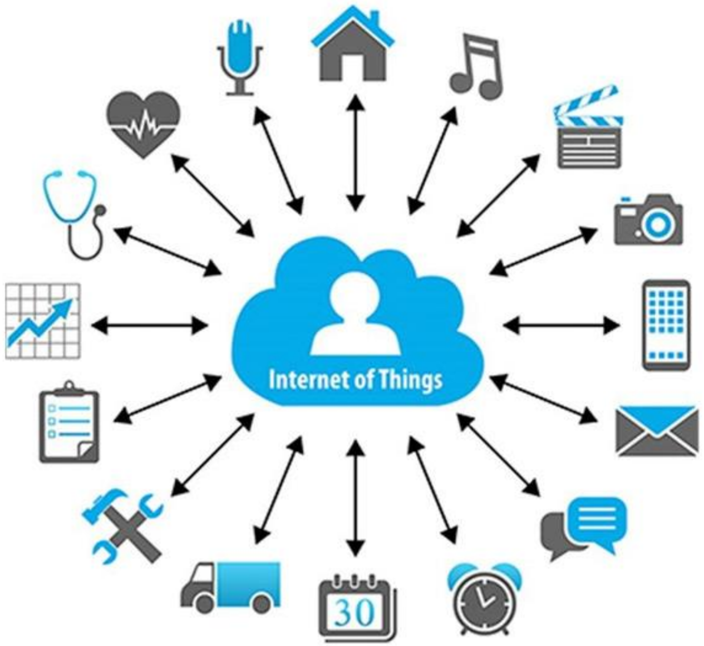
Web 3.0 (Context)
This third evolution of the web is identified as the Semantic web which loosely is the mapping of understanding of the relationship of all content and participation.



Web 1.0 / 2.0 / 3.0 Summary

Crawl	Walk	Run
Web 1.0	Web 2.0	Web 3.0
Mostly Read-Only	Wildly Read-Write	Portable & Personal
Company Focus	Community Focus	Individual Focus
Home Pages	Blogs / Wikis	Lifestreams / Waves
Owning Content	Sharing Content	Consolidating Content
Web Forms	Web Applications	Smart Applications
Directories	Tagging	User Behavior
Page Views	Cost Per Click	User Engagement
Banner Advertising	Interactive Advertising	Behavioral Advertising
Britannica Online	Wikipedia	The Semantic Web
HTML / Portals	XML / RSS	RDF / RDFS / OWL

Image source: https://flatworldbusiness.wordpress.com/flat-education/previously/web-1-0-vs-web-2-0-vs-web-3-0-a-bird-eye-on-the-definition/

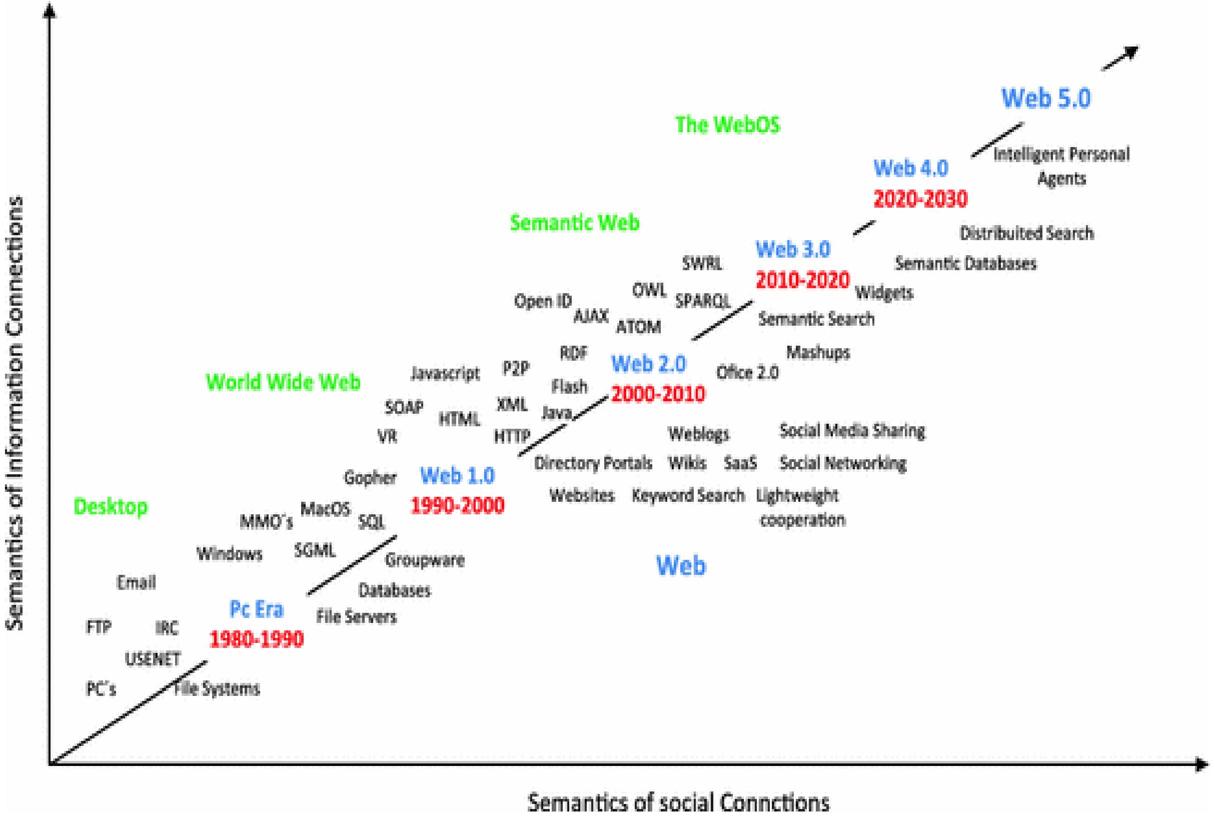


Web 4.0 - Internet of Things (IoT)

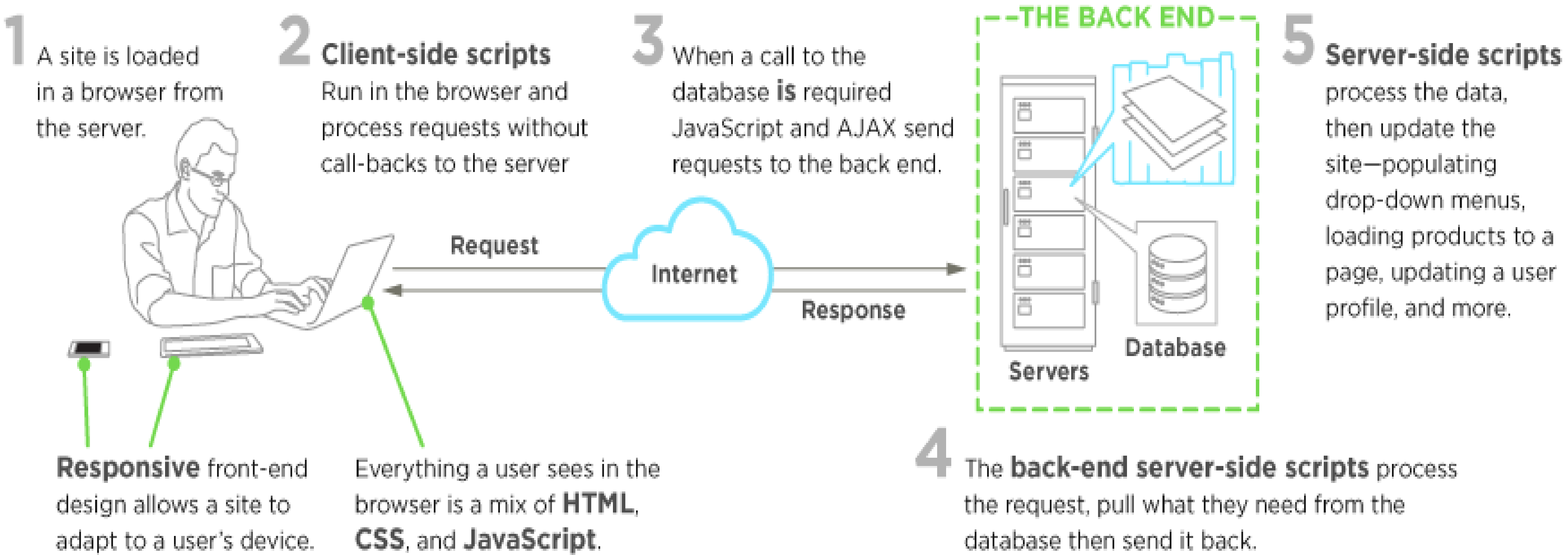
This term refers to a network of objects not historically connected.
The evolution of the internet will include everyday objects withy network connectivity able to send and receive data.

Web 5.0 - emotional web

The next big evolution will be about the (emotional) interaction between humans and computers. For the moment web is “emotionally” neutral, which means web does not perceive the users feel and emotions. This will change with web 5.0 – emotional web



Basic design and implementation of websites



1 A site is loaded in a browser from the server.

2 **Client-side scripts** Run in the browser and process requests without call-backs to the server

3 When a call to the database **is** required JavaScript and AJAX send requests to the back end.

5 **Server-side scripts** process the data, then update the site—populating drop-down menus, loading products to a page, updating a user profile, and more.

Responsive front-end design allows a site to adapt to a user's device.

Everything a user sees in the browser is a mix of **HTML**, **CSS**, and **JavaScript**.

4 The **back-end server-side scripts** process the request, pull what they need from the database then send it back.

Image source: https://www.upwork.com/hiring/development/front-end-developer/