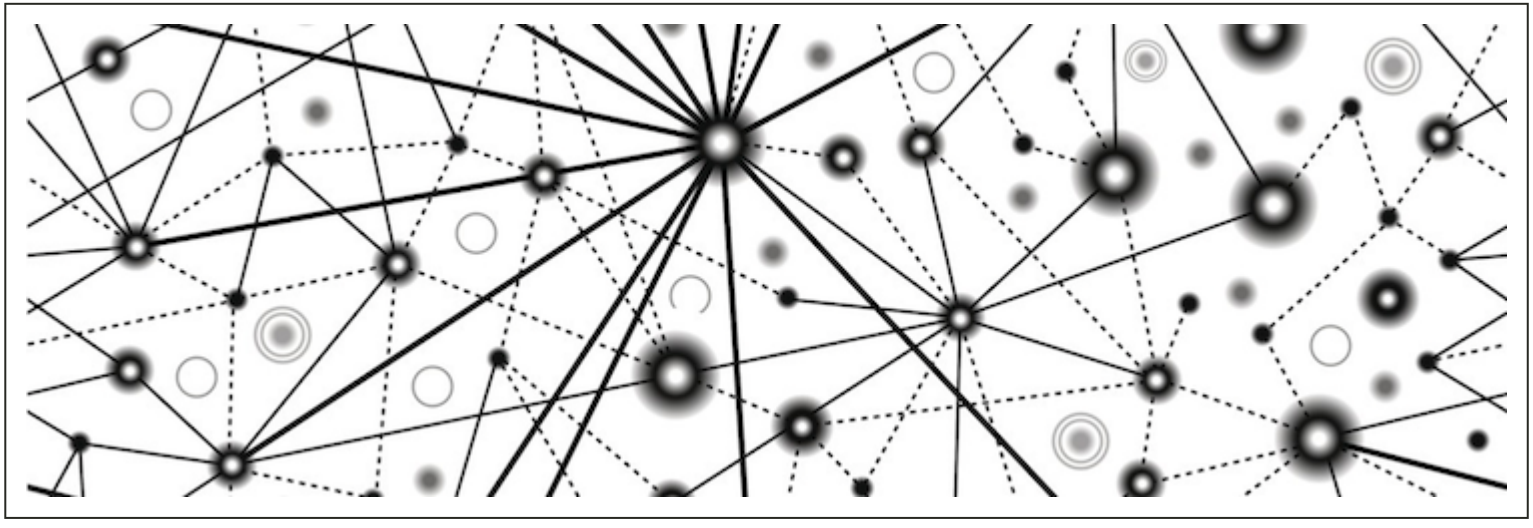
The Web

André Restivo

Web vs Internet

The Internet

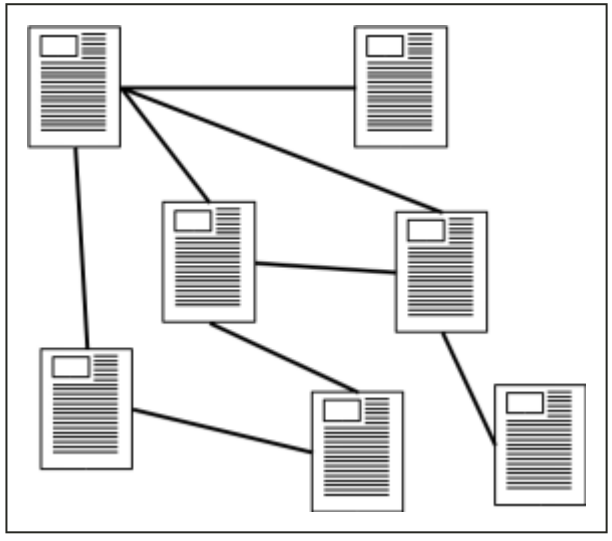
A global system of interconnected **computer networks** that use the standard *Internet protocol suite* to link several **billion** devices worldwide.



The Web

A system of interlinked **hypertext documents** that are accessed via the **Internet**.

Also known as the World Wide Web or WWW.



Web Origins

The Origins of the WWW

WWW was invented by **Tim Berners-Lee** at CERN (1989).

Three constituents: **HTML** + **URL** + **HTTP**:

- **URL** is an notation for locating resources on servers.
- **HTTP** is a high-level protocol for file transfers.
- **HTML** is an SGML language for hypertext.

World Wide Web Consortium (W3C)

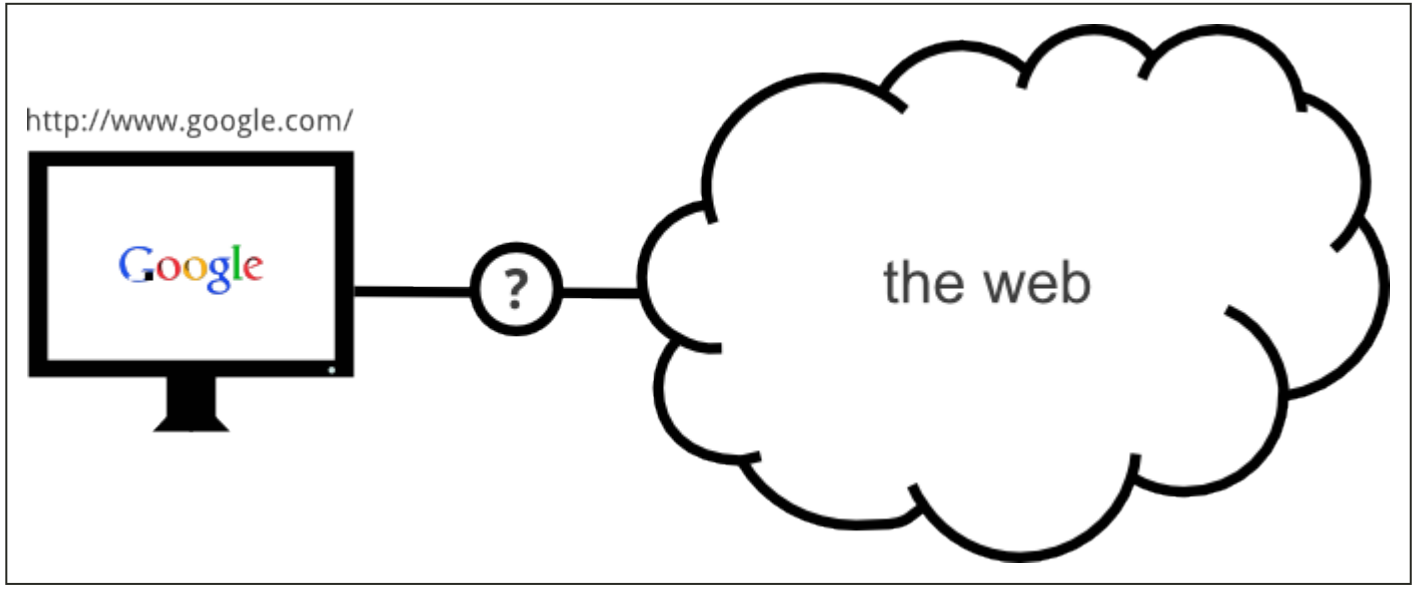
- Develops HTML, CSS, and most Web technologies.
- Founded in 1994.
- Has 380 companies and organizations as members.
- Is directed by Tim Berners-Lee.
- Located at MIT (US), Inria (France), Keiko (Japan).
- http://www.w3.org/

How does it work

From the browser to the server and back

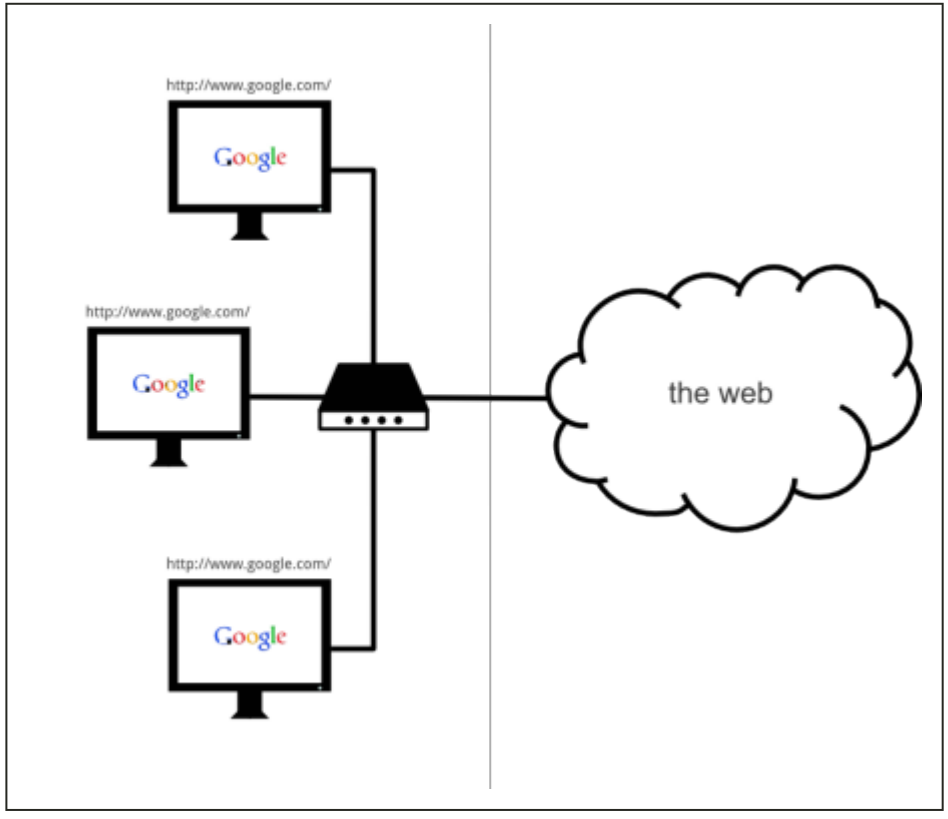
How does the web work?

What happens when you type http://www.google.com/ in the address bar of your browser?



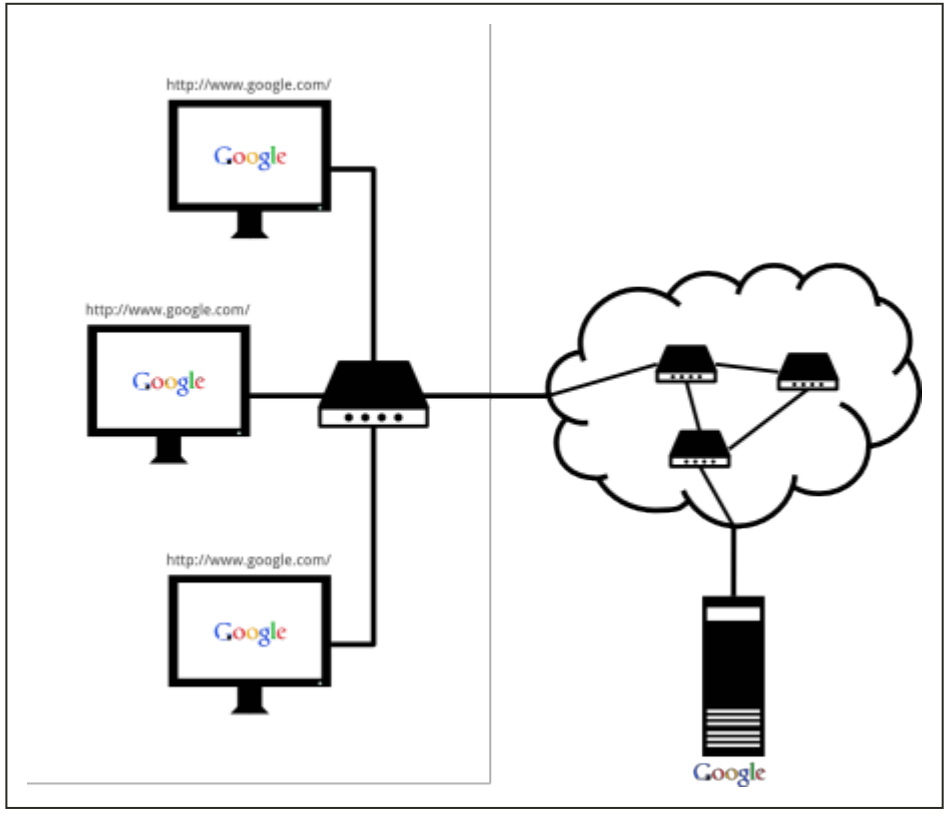
Routers

Computers are usually connected using other devices (such as routers)



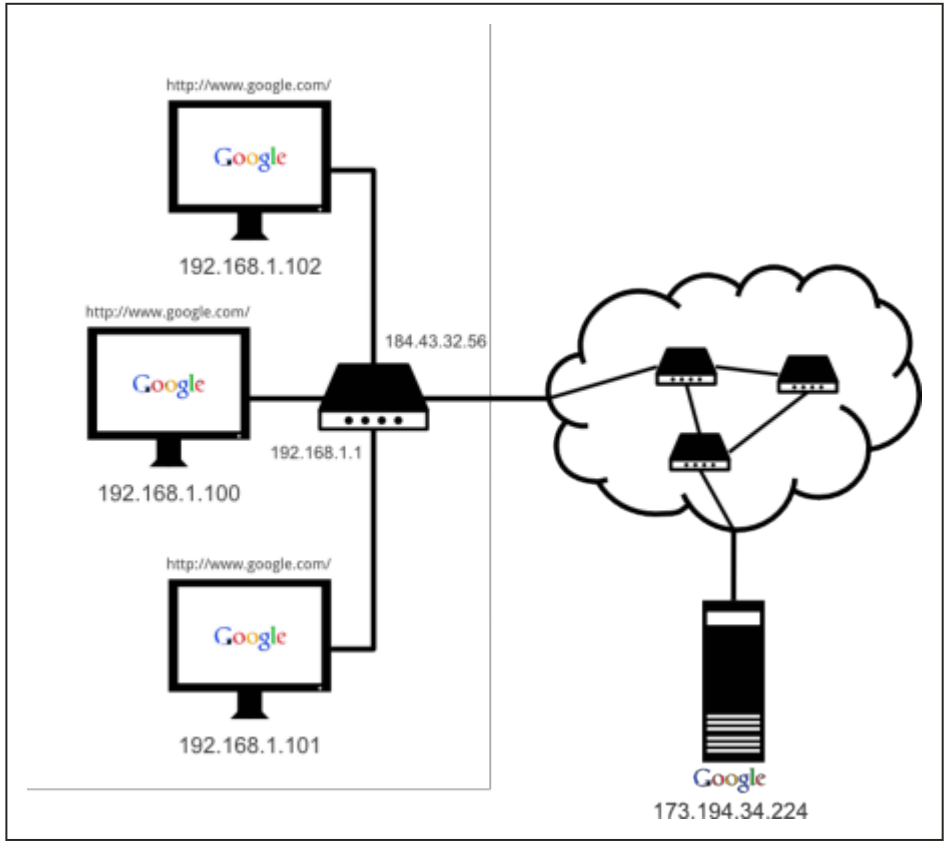
Internet Infrastructure

The Internet is a redundant network of networks that connects millions of hardware devices from laptops to servers.



IP Addresses

- Each connected device has at least one IP (Internet Protocol) address.
- Given an address, routers are able to calculate where they should send information to reach the desired device.



IPv4 and IPv6

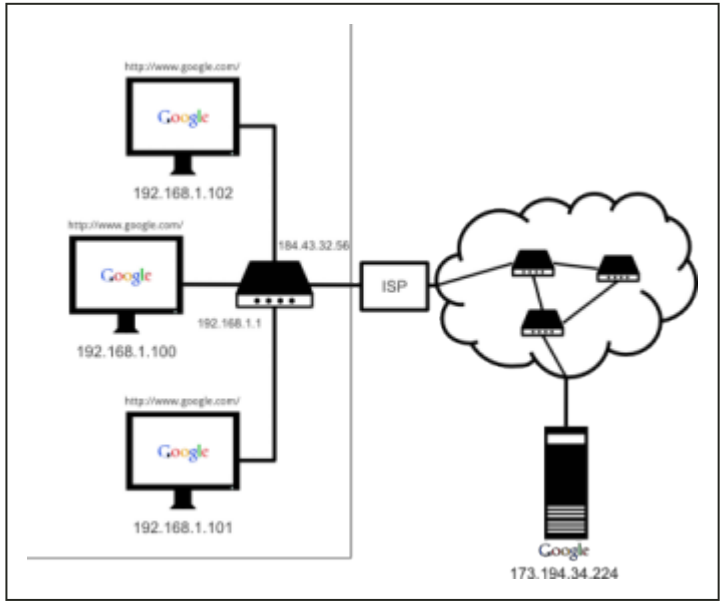
- IP addresses (IPv4) consist of four 8 bit numbers (0-255).
- There are 4,294,967,296 different possible IP addresses.
- Some IP addresses are exclusive for internal use: 10.x.x.x, 172.16-31.x.x and 192.168.x.x.
- IPv4 addresses are nearly exhaustion and are slowly being replaced by the new IPv6 standard (e.g. 2001:4860:4860::8888). There are 3.4×10^{38} possible IPv6 addresses.
- Some devices have static IP addresses while others have dynamic ones. Dynamic addresses are distributed using DHCP (Dynamic Host Configuration Protocol).

Routing Example

- A typical person has, in his house, several computers connected to a router.
- The router has two IP addresses: one internal (probably 192.168.1.1) and one external.
- When a computer connects to the router it is assigned an internal IP address (e.g. 192.168.1.100).
- When a computer tries to send a message to a computer outside the local network, the router knows it has to route the message through its external interface and into the Internet.

Internet Service Providers

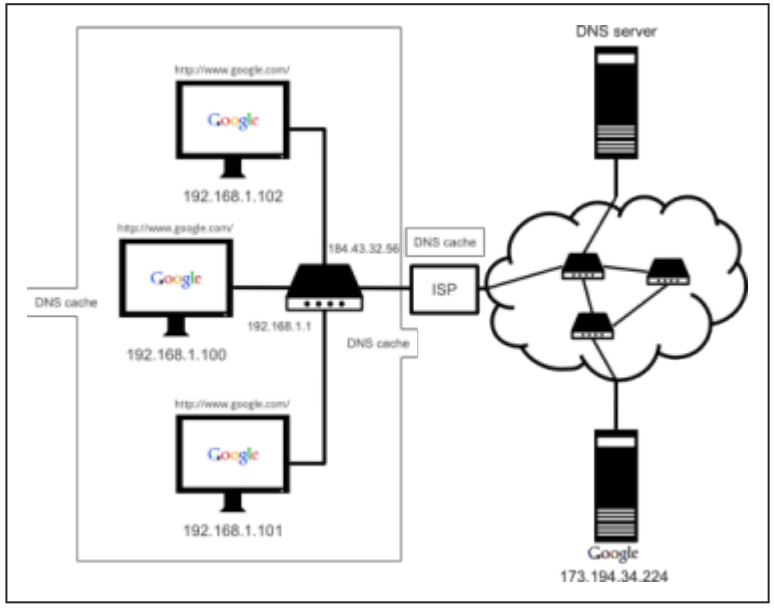
- ISPs are the organizations that connect users to the Internet.
- The external IP address on each router is, most of the times, assigned by the ISP using DHCP.



Name Resolution

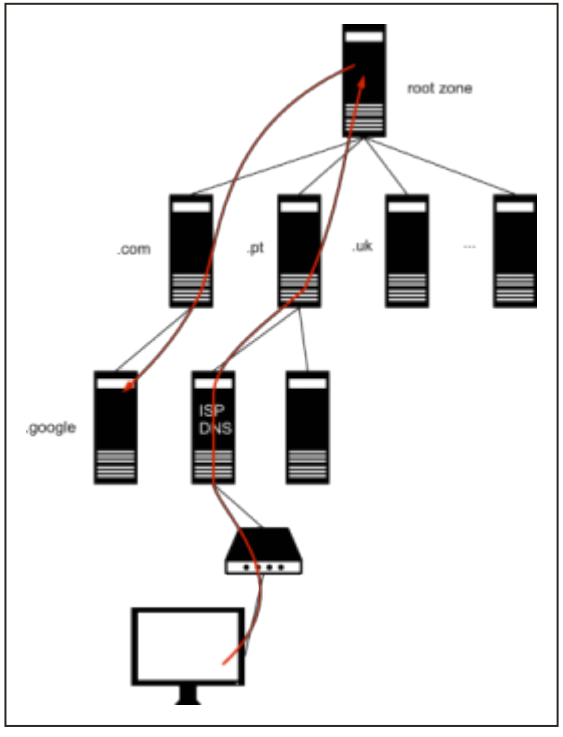
How do we go from www.google.com to 173.194.34.224?

The **Domain Name System (DNS)** is a hierarchical distributed naming system for computers connected to the Internet.



DNS Hierarchy

- DNS requests escalate the hierarchy until a DNS server has a record for the desired name.
- If the root zone DNS does not have the record, the request goes down until it reaches the responsible zone DNS.



URL

Uniform Resource Locators

- A **Uniform Resource Locator (URL)** is a character string that constitutes a reference to an Internet resource.
- It always starts with a **scheme name** followed by a colon and two slashes.
- In the case of the HTTP scheme it is followed by a **server name** (or an IP address) and, optionally, a **port number**, the **path** of the resource to be fetched, a **query** string, and an **fragment** identifier.
- Before the server name it is also possible to add an **username** and a **password**.
- Other common schemes: https, file, ftp, smtp, ...

Uniform Resource Locator Examples

```
http://www.google.com/
http://username:password@www.example.com/path/image.jpg
http://www.example.com:80/path?query_string#fragment_id
```

- The port is 80 by default.
- The query string allows one to pass parameters to the resource.
- The fragment id indicates a specific point on the resource.

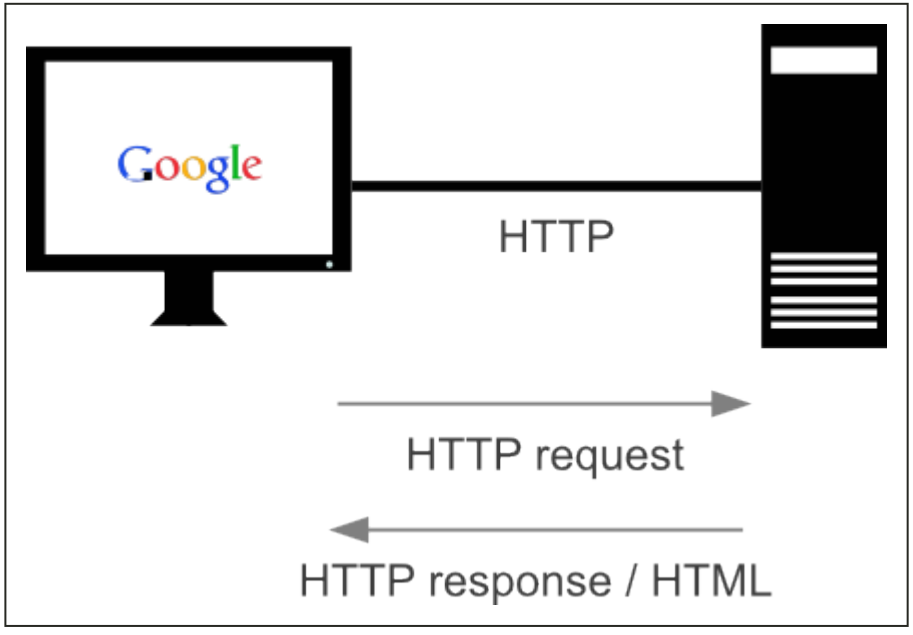
HTTP

Hypertext Transfer Protocol

- The **Hypertext Transfer Protocol (HTTP)** is a protocol that mediates the flow of information between a client computer (generally in the form of a browser) and a web server.
- When a certain URL is introduced into the browser location bar, the browser creates an HTTP connection to the desired server and requests the resource represented by the URL.
- It is the responsibility of the server to return that resource to the browser via the same connection (or produce an error).
- The browser then presents the resource to the user.

Hypertext Transfer Protocol

- Resources can be of various types.
- The most common are HTML pages but they can also be images, style sheets, PDF files, ...
- The browser is responsible for presenting them in the most convenient way to the user.



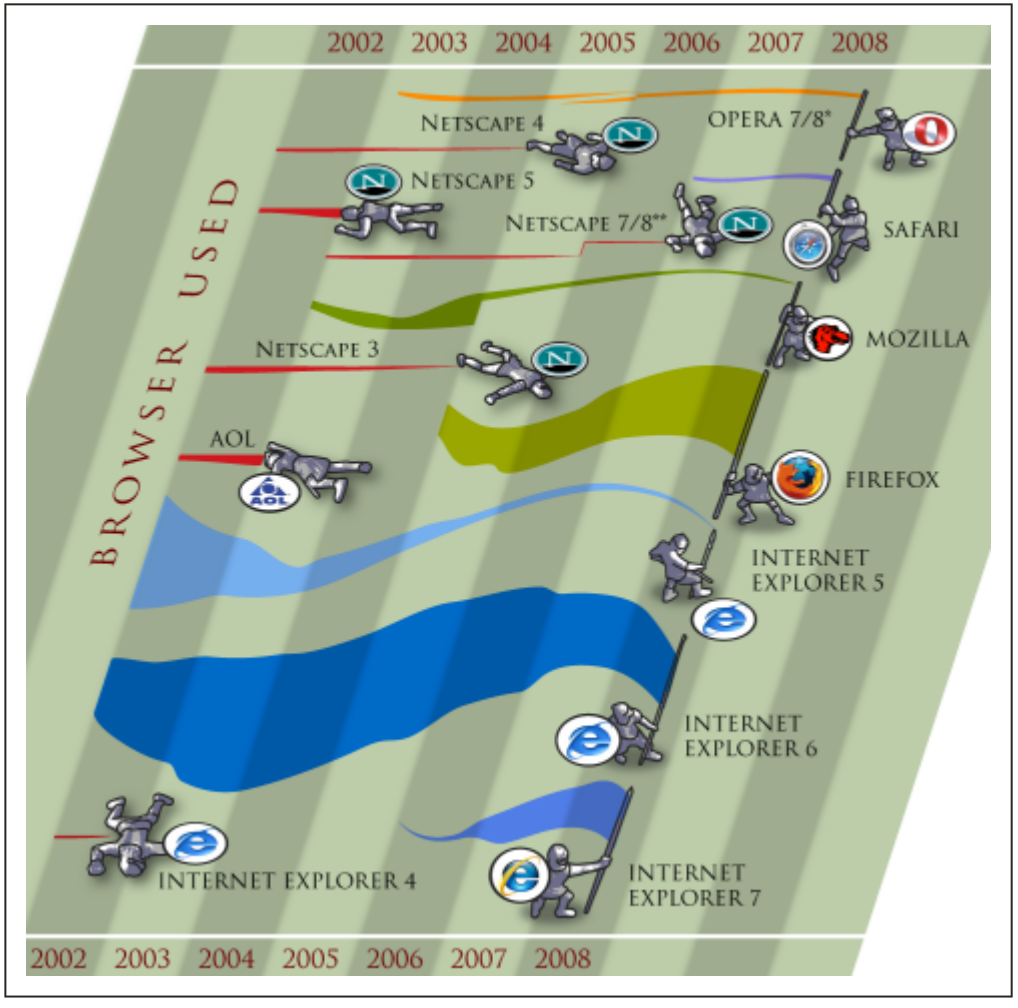
HTML

An introduction

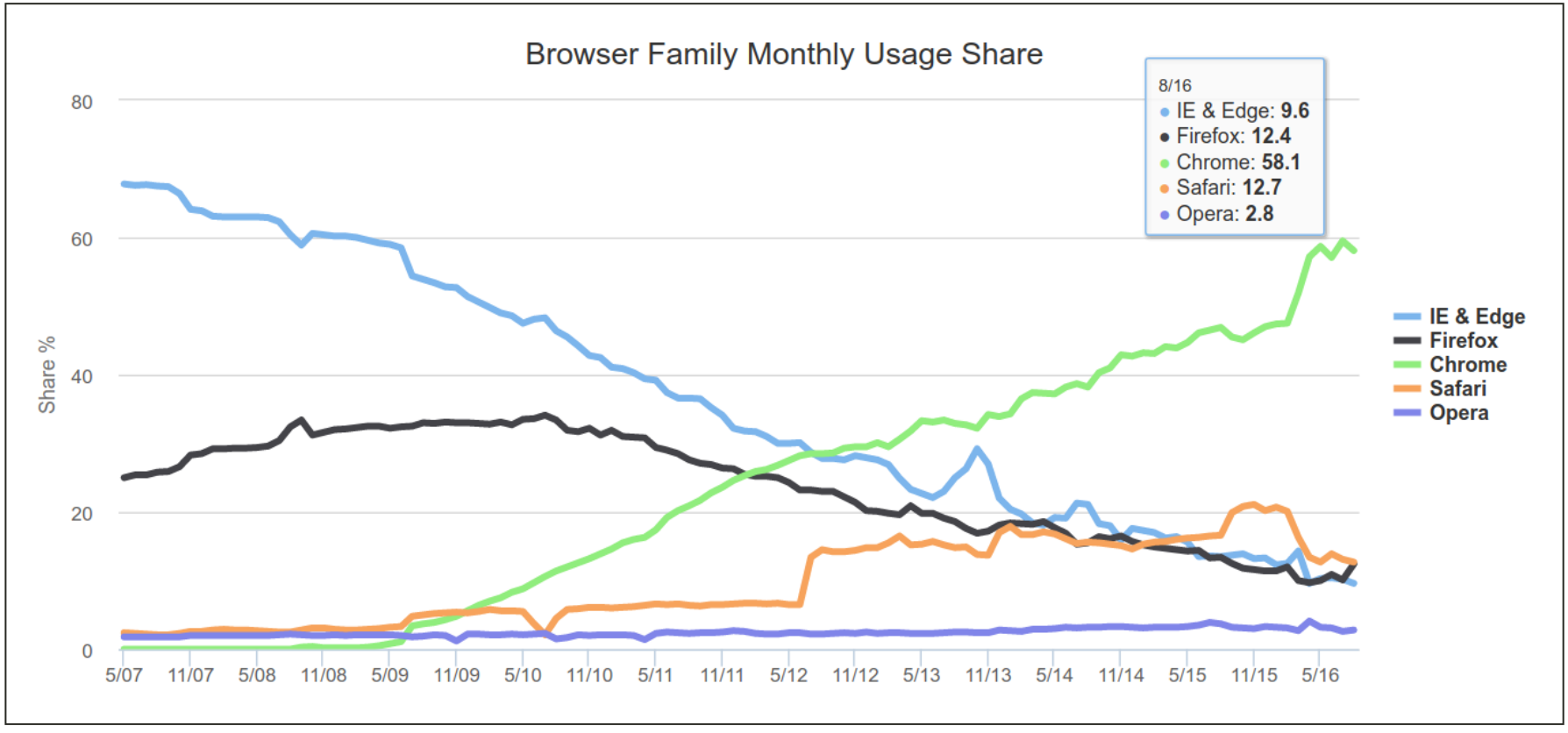
The History of HTML

- 1992: **HTML 1.0**, Tim Berners-Lee original proposal
- 1993: **HTML+**, Dave Raggett's competing standard
- 1994: **HTML 2.0**, tables, file upload, …
- 1995: Non-standard Netscape features
- 1996: Competing Netscape and Internet Explorer features
- 1996: **HTML 3.2**, W3C standard, the Browser Wars end
- 1997: **HTML 4.0**, stylesheets are introduced
- 1999: **HTML 4.01**, we have a winner!
- 2000: **XHTML 1.0**, an XML version of HTML 4.01
- 2001: **XHTML 1.1**, modularization
- 2008: **HTML 5**, reduces the need for proprietary plug-in based apps

Browser Wars



Browser Share



Source: http://www.w3counter.com/trends

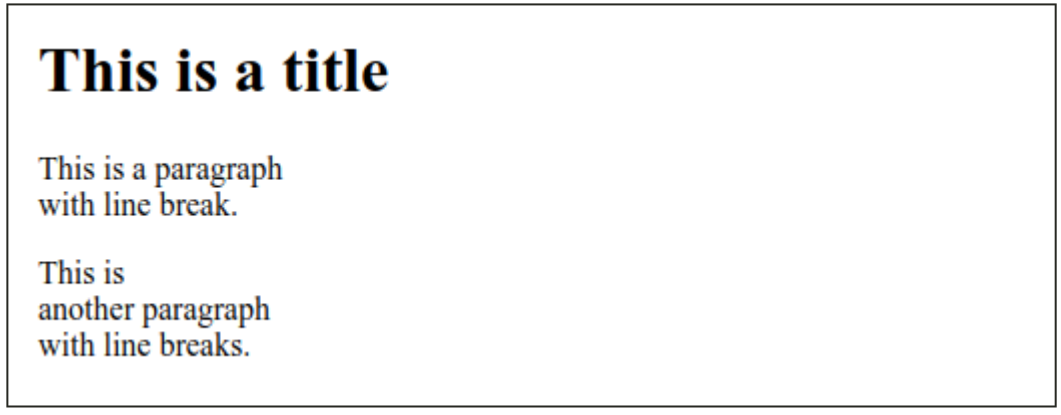
HTML – Hypertext Markup Language

- Simple, purist design principles.
- HTML describes the logical structure of a document.
- Browsers are free to interpret tags differently.
- HTML is a lightweight file format.

HTML Example

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <title>The Title of the Document</title>
</head>
<body>
    <h1>This is a title</h1>
    <p>This is a paragraph <br> with line break.</p>
    <p>This is <br>another paragraph <br> with line breaks.</p>
</body>
</html>
```

Result


This is a title
This is a paragraph
with line break.
This is
another paragraph
with line breaks.