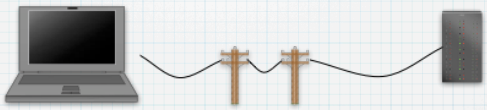
World Wide Web

WWW usage requires a combination of standards and protocols

DHCP	TCP/IP	DNS
HTTP	HTML	MIME

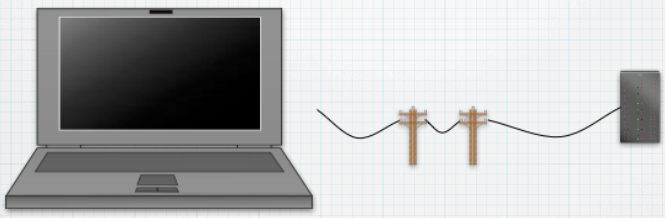
World Wide Web - how it works

User on a machine somewhere

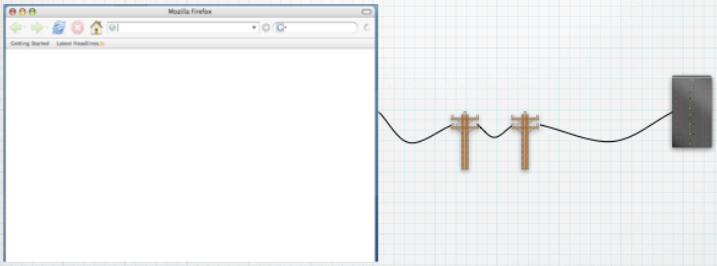


Server machine

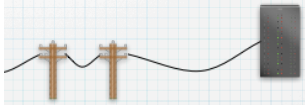
Being more specific...



Web Browser software

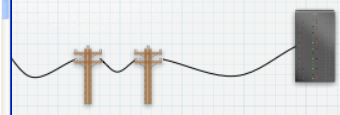


Web Browser software

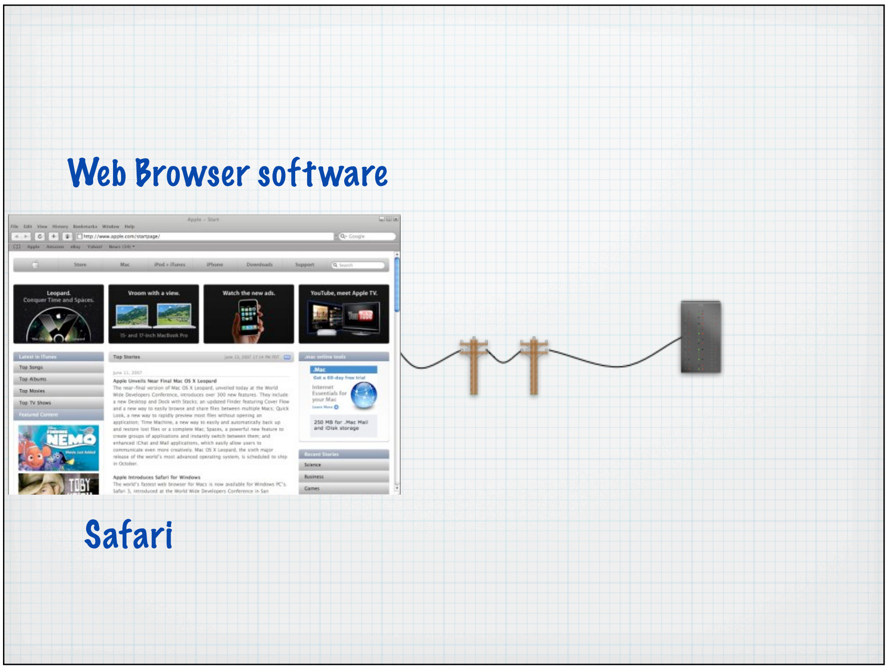


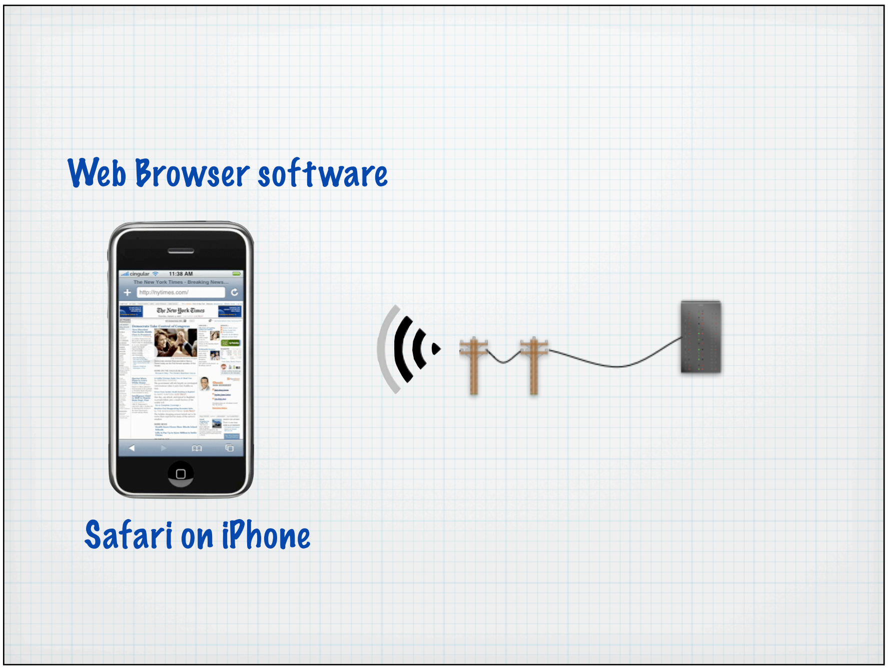
Firefox

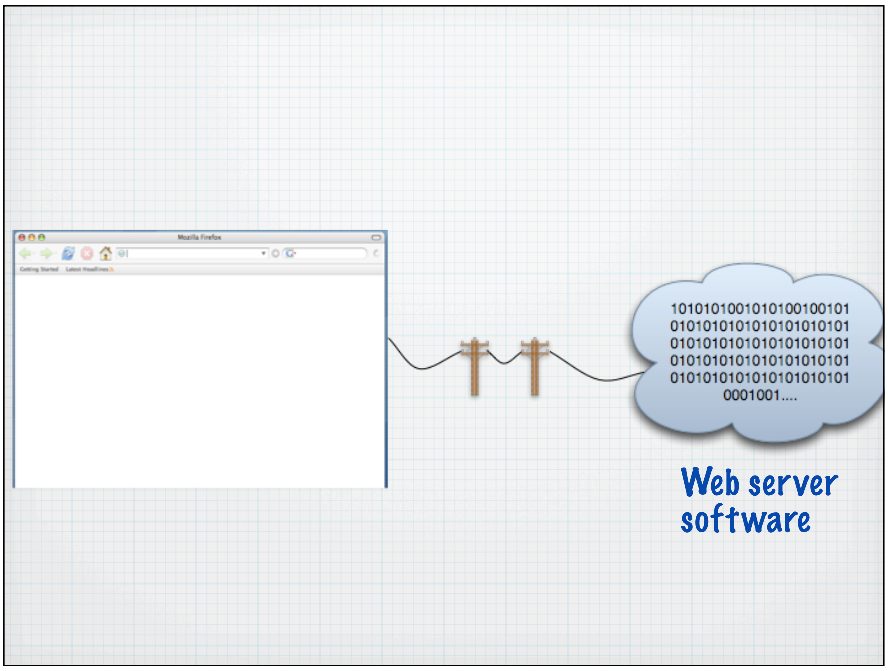
Web Browser software

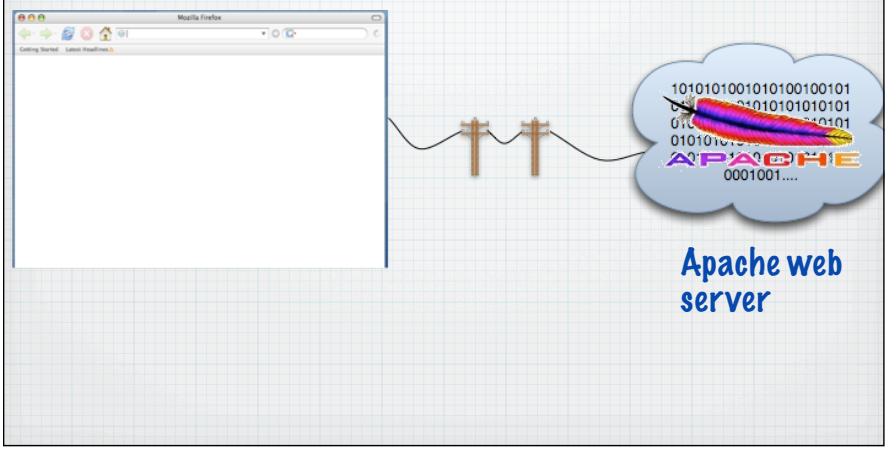


Internet Explorer

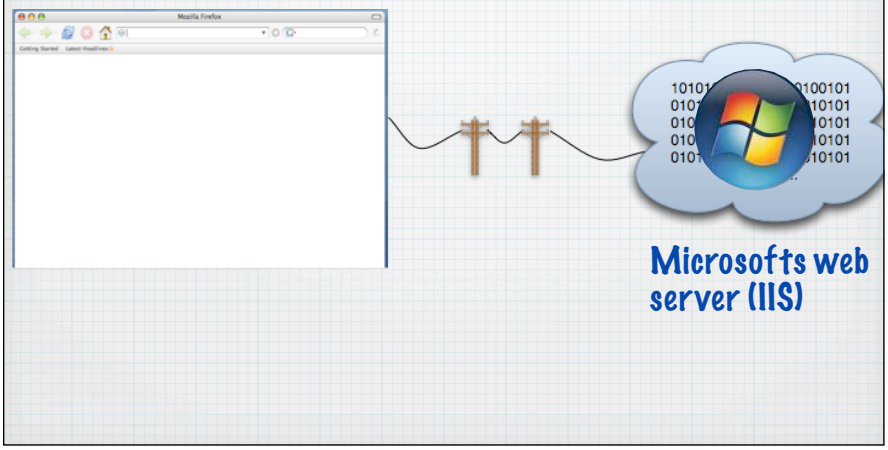

Web Browser software
Safari


Web Browser software
Safari on iPhone


1010101001010100100101
0101010101010101010101
0101010101010101010101
0101010101010101010101
0101010101010101010101
0001001....
Web server software

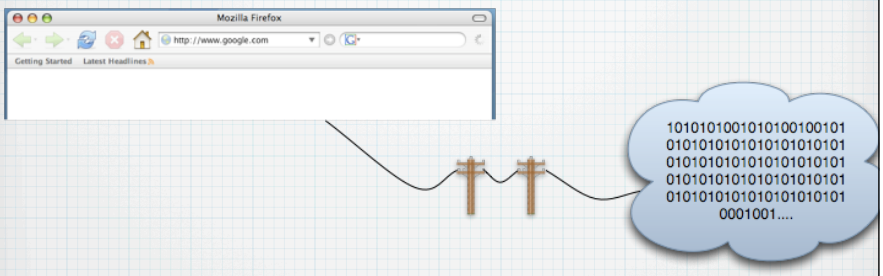


Apache web server

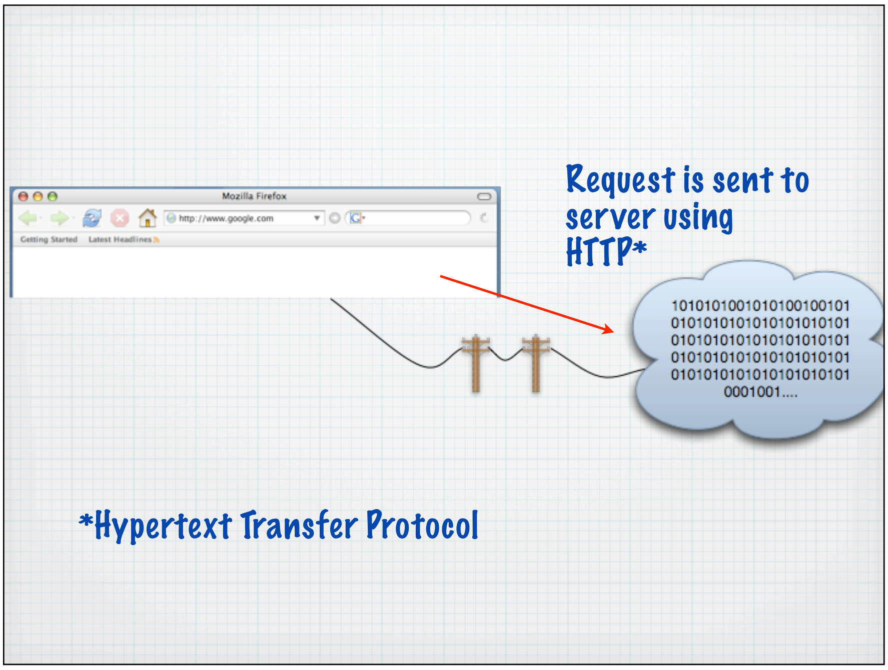


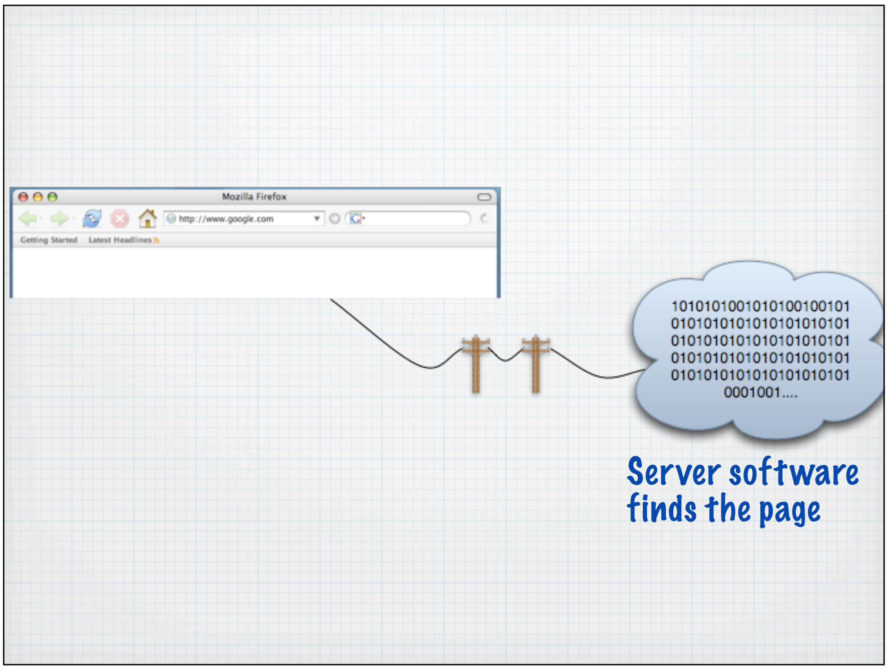
Microsofts web server (IIS)

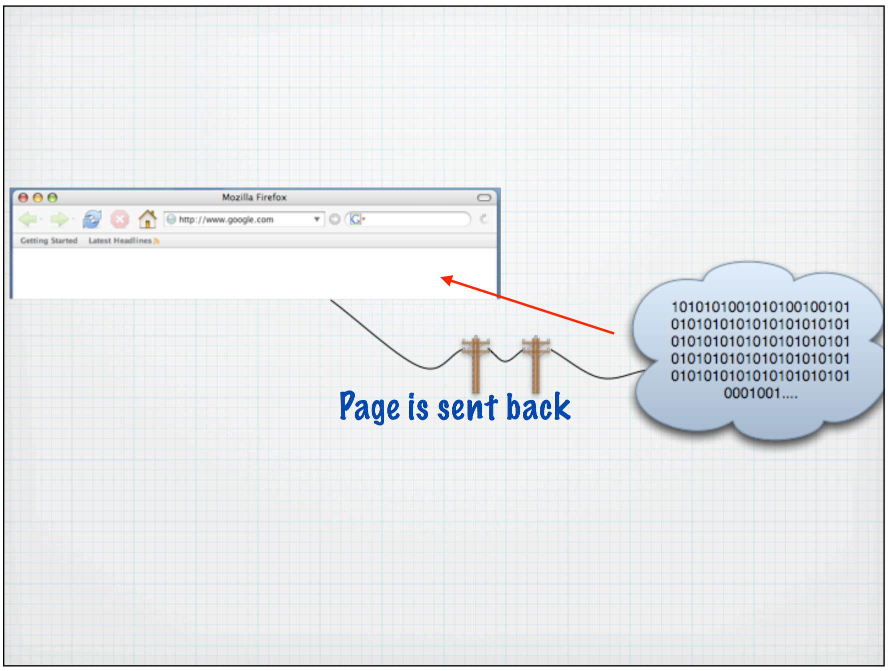
User types in a URL*



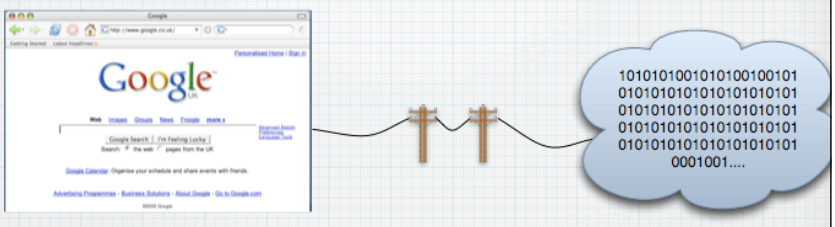
*Uniform Resource Locator


Mozilla Firefox
http://www.google.com
Getting Started
Latest Headlines
Request is sent to server using HTTP*
1010101001010100100101
0101010101010101010101
0101010101010101010101
0101010101010101010101
0101010101010101010101
0001001....
*Hypertext Transfer Protocol


Mozilla Firefox
http://www.google.com
Getting Started
Latest Headlines
1010101001010100100101
0101010101010101010101
0101010101010101010101
0101010101010101010101
0101010101010101010101
0001001....
Server software finds the page


Mozilla Firefox
http://www.google.com
Getting Started
Latest Headlines
1010101001010100100101
0101010101010101010101
0101010101010101010101
0101010101010101010101
0101010101010101010101
0001001....
Page is sent back

Browser displays the page

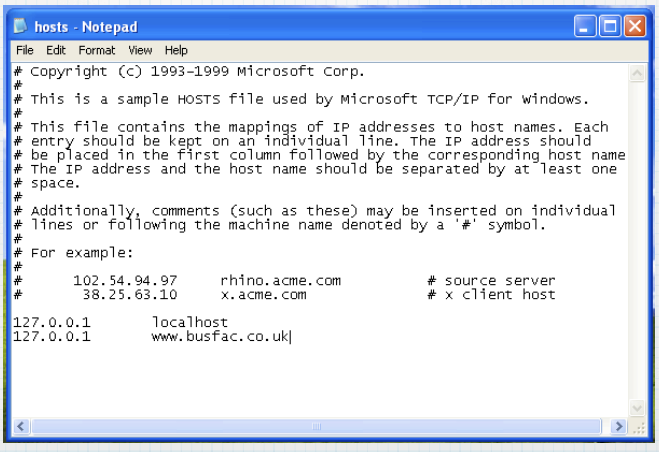


Demo

- Changing a hosts file to associate a name with an IP number
- Installing and running a web server
- Publishing some files in the public directory

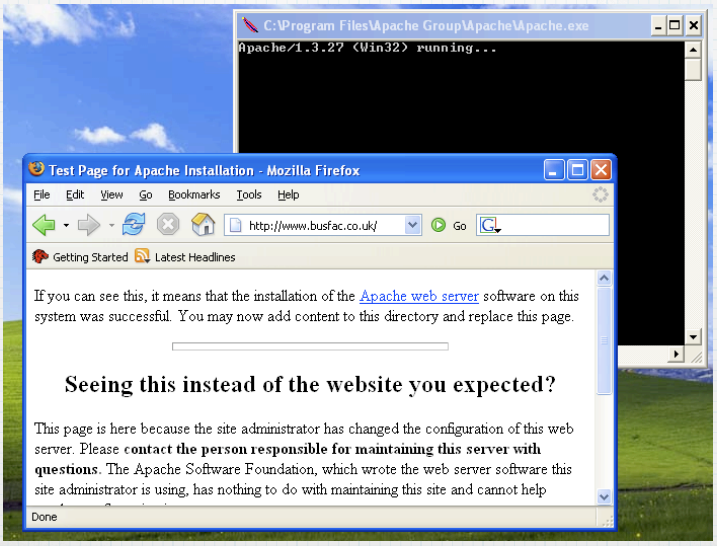
Demo

- Changing a hosts file to associate a name with an IP number



Demo

- Installing and running web server



Demo

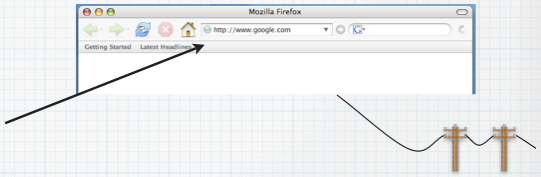
- Publishing some files in the public directory

In more detail then

1. User types in web page request using a special format
2. Browser sends HTTP request
3. Server sends HTTP response
4. Browser renders and displays page

In more detail then

1. User types in web page request using this format (structure specified in HTTP)



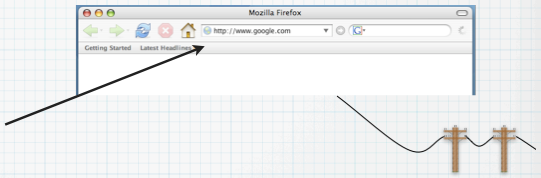
protocol: // server : port / directory / file

http: // www.oracle.com / downloads / list.html

The file list.html in the web server directory downloads on the HTTP server located on named machine www.oracle.com

In more detail then

1. User types in web page request using this format (structure specified in HTTP)



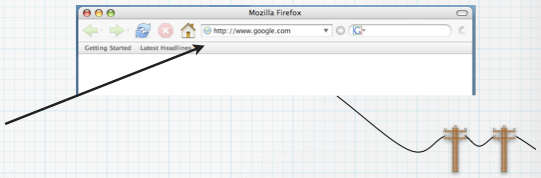
protocol: // server : port / directory / file

http: // www.kingston.ac.uk / index.html

The file index.html in the root web server directory on the HTTP server located on named machine www.kingston.ac.uk

In more detail then

1. User types in web page request using this format (structure specified in HTTP)



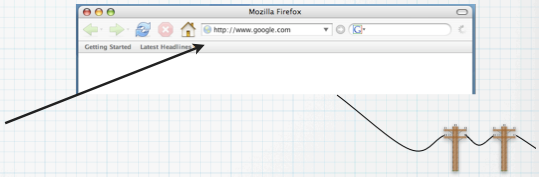
protocol: // server : port / directory / file

http: // www.kingston.ac.uk

The default file (generally called index.html) in the root web server directory - these filenames are specified in the Web server configuration file

In more detail then

1. User types in web page request using this format (structure specified in HTTP)



protocol: // server : port / directory / file

http: // www.kingston.ac.uk : 80 / index.html

The file index.html in the root directory of the HTTP server listening on port 80 on machine www.kingston.ac.uk

http: // www.kingston.ac.uk : 80 / index.html

Several software services may run on the same machine - the port number is used to distinguish them

80 is the default port number for HTTP servers (again specified in the web server configuration files)

Other examples:

http: // www.kingston.ac.uk / ~ku12492

The file index.html in the web server directory for user ku12492

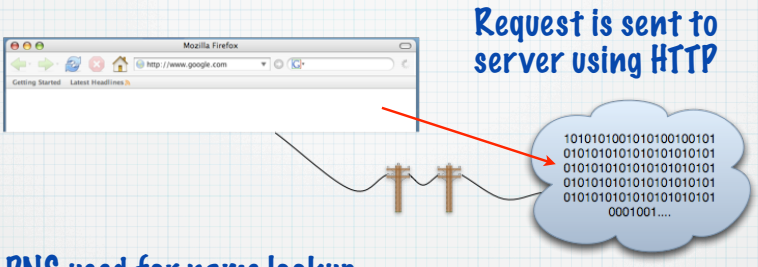
Typically ~ is used to provide a shortcut link to a sub directory on the web server rather than specifying the full path

Using port 80 by default (even though not specified)

Domain name registration

- Many different companies
- Quite cheap
- Can be difficult to get a good name
- Domain name squatters

2.Browser sends HTTP request



- DNS used for name lookup
- TCP session attempts to connect to web server
- If web server exists and can be connected to, GET request is issued

HTTP GET request

- Uses this syntax (specified in HTTP 1.1)

GET file_request protocol_version

followed by

browser and client information

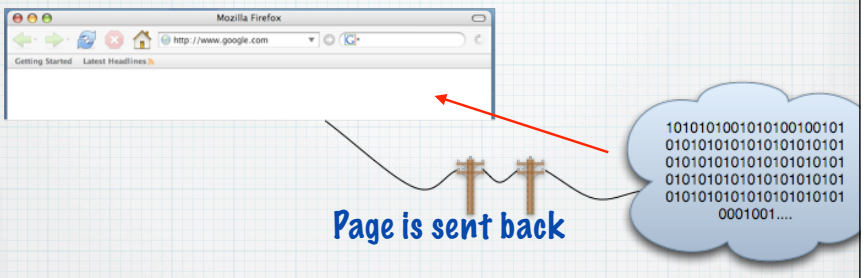
HTTP GET request

- Example

GET file_request protocol_version

```
GET /~ku12492/example/index.html HTTP/1.1
Host: www.kingston.ac.uk
User-Agent: Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.3a)
Accept: text/xml,application/xml,application/xhtml+xml,text/html;
Accept-Language: en-us,en;q=0.5
Accept-Encoding: gzip,deflate,compress;q=0.9
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive: 300
Connection: keep-alive
```

3.Server sends HTTP response



- HTTP servers listen and respond to HTTP GET requests
- Look for requested named page in file system
- Return the page, or an error number if the page does not exist

HTTP GET response

- Uses this syntax (specified in HTTP 1.1)

protocol status_code description

followed by

header block

document stream

HTTP GET response

- Example

```
HTTP/1.1 200 OK
Date: Sun, 26 Nov 2006 09:18:33 GMT
Server: Apache/1.3.27 (Unix)
Last-Modified: Tue, 14 Jan 2006 11:46:32 GMT
Accept-Ranges: bytes
Content-Length: 3181
Keep-Alive: timeout=15, max=99
Connection: Keep-Alive
Content-Type: text/html

<html>
<head>
<title>Example Page</title>
```

MIME types

- Multimedia Internet Mail Extensions
- Originally created to indicate what kind of file was attached to an email message...
- ...and how it could be transmitted and processed

```
text/css
text/javascript
image/gif
```

http://www.iana.org/assignments/media-types/

Web pages consist of html text, images and more

- HTTP can be used to transfer text, binary or other 'types' of files
- Requires viewers at the client end (accessible by the browser)

Web Server architecture

- HTTP processor
  - To receive and send HTTP encoded streams
- Script processor
  - Server side programming
  - May be module based so that any language can be switched in
- Security and Management system
  - User groups and passwords

Web Server architecture

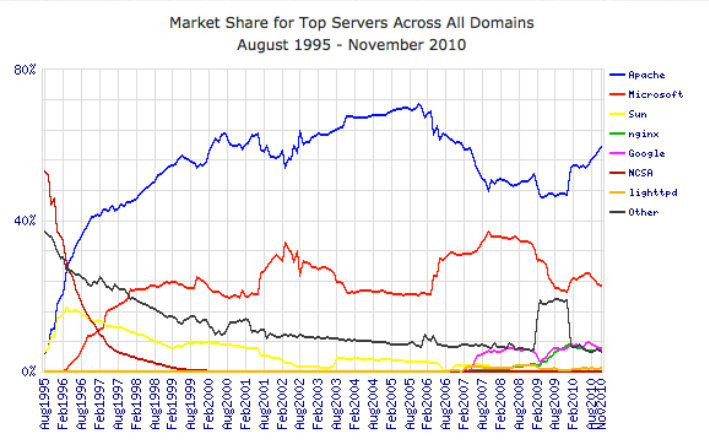
- Normal PC specification
  - Needs efficient networking, average graphics capability, large storage

First web server machine

Now in Microcosm, CERN museum in Geneva, Switzerland

Web Server software

- Sites like Netcraft monitor the statistics
  - http://news.netcraft.com/
- Two competing products have 85% of the market share



Developer	October 2010	Percent	November 2010	Percent	Change
Apache	135,209,162	58.07%	148,085,963	59.36%	1.29
Microsoft	53,525,841	22.99%	56,637,980	22.70%	-0.28
nginx	14,130,907	6.07%	15,058,114	6.04%	-0.03
Google	14,971,028	6.43%	14,827,157	5.94%	-0.49
lighttpd	1,380,160	0.59%	2,070,300	0.83%	0.24

Internet Information Server

- Microsoft's Web Server (20-30% of Market)
- Native support for Active Server Pages
- Supports Microsoft's 'extensions'
- Integrates into Microsoft's Active Directory management system
- Windows OS specific (and integrated)

(continued)

- Easy to install (possibly a disadvantage)
- Possibly the most hacked software - ever
- Appalling security record
- Key part of .Net

Apache

- Open Source Web Server
- Free and most widely used (50-60% of market)
- Not bound to slow development and release cycle
- Cross platform
- Supports all server side languages (including ASP)

Web browser architecture

- HTTP processor
  - To create and receive HTTP encoded streams
- Script processor
  - Client side programming
- Mark up language processor (Parser)
  - HTML, XHTML or XML family

and ...

- Layout and Rendering engine
  - To show processed mark up on screen
- Object Model
  - Enables the document to be processed as a tree data structure
  - Now standardised by W3C (DOM)

Internet Explorer - History

- Originally licensed from NCSA Mosaic
- National Centre for Supercomputing Applications at the University of Illinois
- IE 4.0 beat Netscape 4.0 when bundled with Windows 98
- Originally a better browser on Windows (faster)
- Major application required for Microsoft's .Net initiative

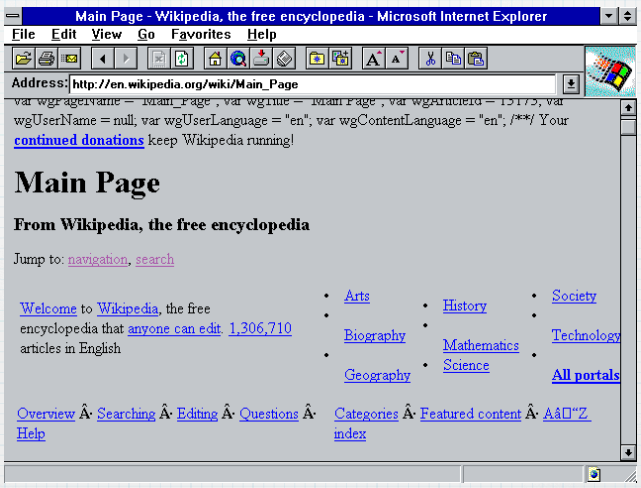
Internet Explorer

- Based on the Trident rendering engine

No name	IE4 Initial rendering engine
No name	IE5 improved CSS1.0 support and had sweeping changes in CSS2 rendering
No name	IE5.5 More corrections to CSS handling
No name	IE6 Corrected the box model and added quirks mode with DTD switching
Trident 3	IE7 fixed many CSS rendering issues and added partial PNG alpha support
Trident 4	IE8 first version to pass the Acid 2 test
Trident 5	IE9 Added support for SVG, XHTML, HTML5 and CSS3

Internet Explorer - Advantages

- Look and feel of Windows (whatever version)



Internet Explorer - Advantages

- Look and feel of Windows (whatever version)

Internet Explorer - Advantages

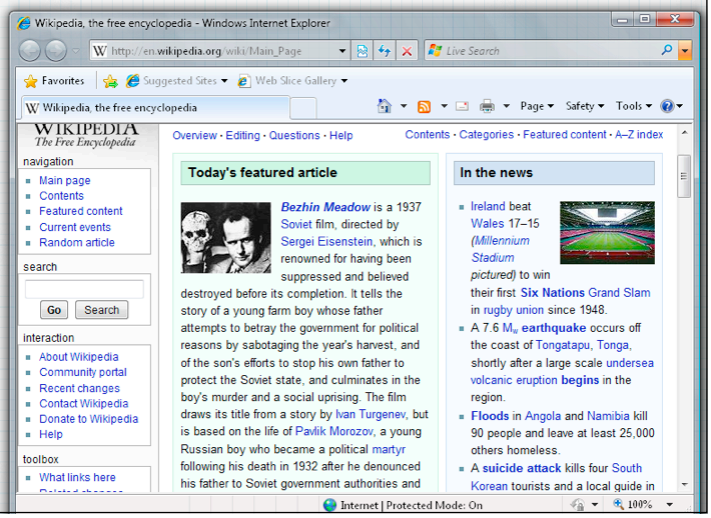
- Look and feel of Windows (whatever version)

Internet Explorer - Advantages

- Look and feel of Windows (whatever version)

Internet Explorer - Advantages

- Look and feel of Windows (whatever version)



Internet Explorer - Advantages

- Look and feel of Windows (whatever version)

Internet Explorer - Advantages

- Free
- Standard part of Windows so optimised
- Integrated into Operating System so HTTP engine available in other Windows applications
- Supports Microsoft's IIS extensions
- Supports client side programming language Basic and Active X

Internet Explorer - Disadvantages

- Not free (paid for in Windows cost)
- Can't be removed from Windows (limited platforms)
- Bloated software (large - slow support for new standards)
- Integrated into Operating System so HTTP engine available in other Windows applications
- Supports Microsoft's IIS extensions (not standard)

Bizarrely ...

- Microsoft has 4 different rendering engines

Trident

Microsoft Word HTMLengine

Microsoft Expression engine

Windows Mobile HTML engine

Trident

Developed for IE and now used widely in Windows

Internet Explorer for Windows from 4.0 onwards
Windows file manager/shell, Windows Explorer
Add/Remove Programs tool used to render the list of installed programs
Microsoft InfoPath
Microsoft Encarta and related products
Microsoft Compiled HTML Help
Microsoft Outlook (prior to 2007)
Microsoft Outlook Express
MSN Messenger
Windows Media Player, which uses Trident to render the "Media Information" pages

Microsoft Word HTMLengine

Developed for use Word to create/view HTML

Microsoft Office
Outlook 2007 onwards

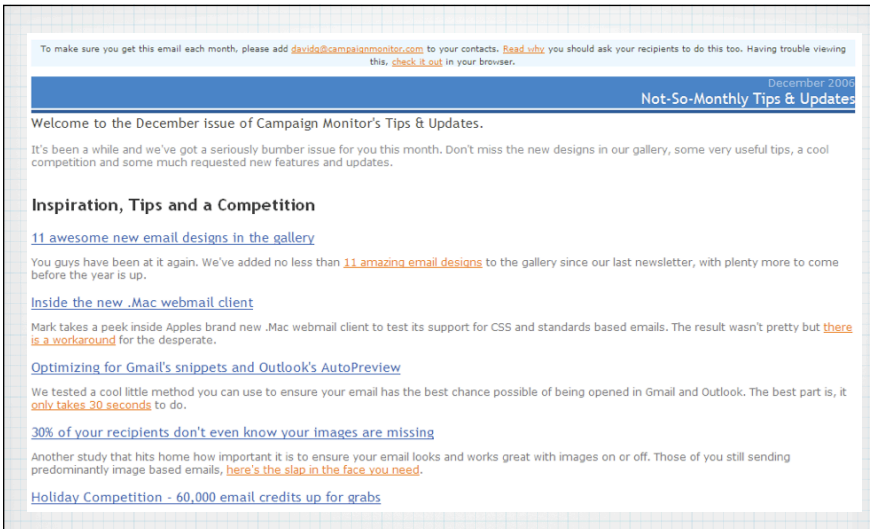
Many critics of the Outlook change

http://www.sitepoint.com/newsletter/viewissue.php?id=3&issue=156#5

http://blog.wired.com/monkeybites/2007/01/outlook_2007_ad.html

http://www.campaignmonitor.com/blog/archives/2007/01/microsoft_takes_email_design_b.html

Outlook (pre 2007)



Outlook (2007 +)

Microsoft Expression Engine

For Microsofts new range of competitor products to Adobe

Windows Mobile HTML engine

Developed for browsers on mobiles

Pocket PC
Windows Mobile

Internet Explorer - Disadvantages

- Major problem - poor support for standards

v1	July 1995
v2	November 1995
v3	August 1996
v4	September 1997
v5	March 1999
v6	August 2001
v7	October 2006
v8	March 2009
v9	Public Beta Sept 2010

Internet Explorer - Disadvantages

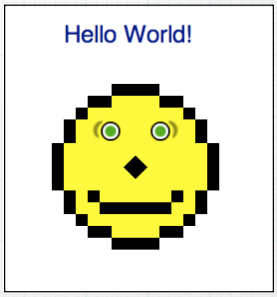
Many browser CSS comparison sites:

http://www.webdevout.net/browser-support-css

ACID 2

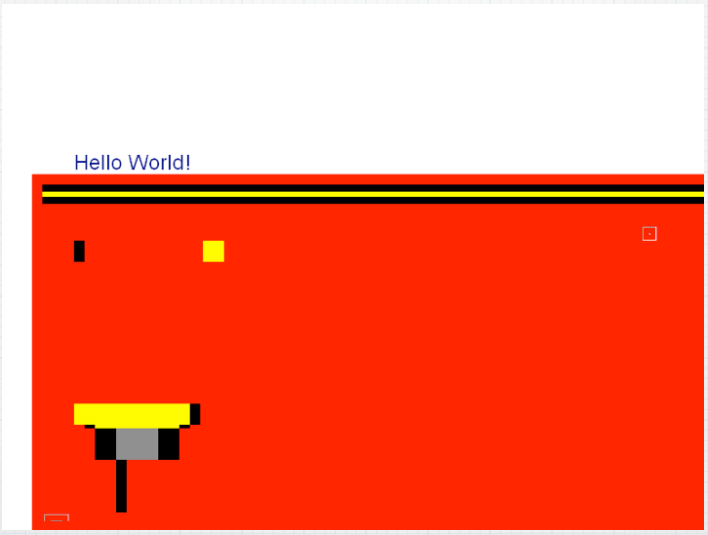
- ACID 2 test indicates nature of CSS standards compliance

Generates a picture created with XHTML and CSS

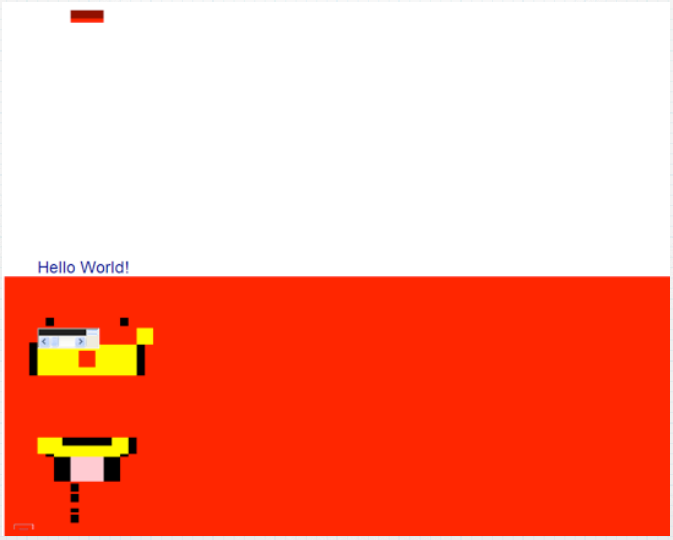


http://www.webstandards.org/action/acid2/

Internet Explorer Version 6 in ACID 2



Internet Explorer Version 7 in ACID 2



Internet Explorer Version 8 March 2009

IE Version 8 passes ACID 2 (finally)

Firefox - History

- Born out of Netscape / Mozilla
- Open Source
- Gecko rendering engine - good support for standards
- Written in the Mozilla development environment

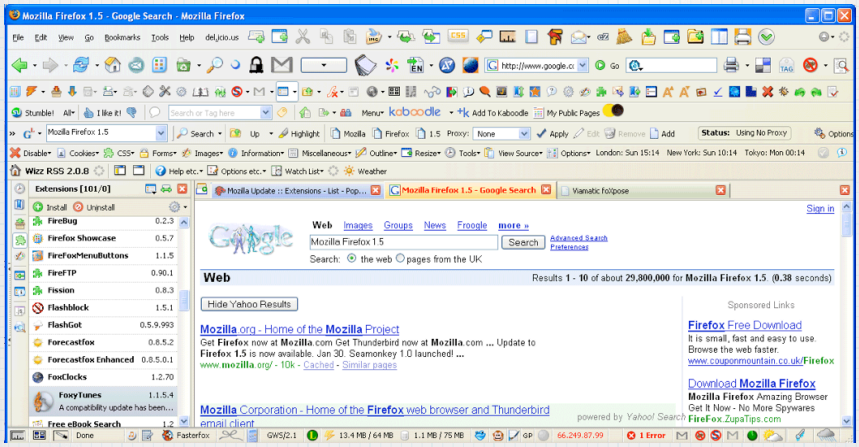
Firefox - History

- Very fast development cycle

v1	November 2004
v1.5	November 2005
v2	October 2006
v3	June 2008
v3.5	June 2009
v3.6	January 2010

- Design aim - keep the browser simple, small and fast
- Allow users to add the feature they want with extensions

- Perhaps too many here:



- Plug in example - Adblock [before]

- Plug in example - Adblock [after]

Firefox - Advantages

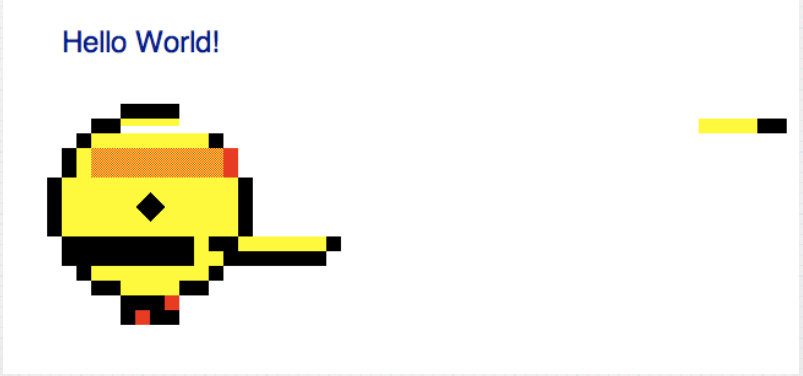
- Free apart from download costs
- Not tied to Operating System
- Nice design elements

  Tabbed browsing, pop up window configuration, anti-phishing

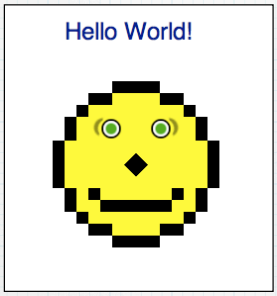
Firefox - Disadvantages

- Separate download
- No support for Microsoft non standard extensions
- Not integrated into the Windows platform (Outlook and Office)

Firefox 2.0 in ACID 2



Firefox in ACID 2 (April 2006)



(IE8 March 2009)

Safari

- Apple
- Open Source
- WebKit rendering engine derived from KHTML
- Mac OS X and PC
- iPhone and iPad

Opera

- Opera Software
- Propriety/Moving to Open Source
- Presto rendering engine
- Multiplatform

Safari - History

- Produced to counter Microsofts "we'll stop supporting IE on Mac" threat

v1	June 2003
v2	April 2005
v3	June 2007
v4	June 2008
v5	June 2010

- Forked from the open source KHTML rendering engine on Linux (not Mozilla)
- First browser to pass ACID 2 in 2005
- Same code base used on Mac OS X, iPhone and iPad

Opera - History

v2	June 1996
...	...
v9.5	Oct 2007

- v1 - research project at Telenor (telecommunications company in Norway)
- Fast, multiplatform, large market share on mobiles
- One of the first browsers to offer CSS support

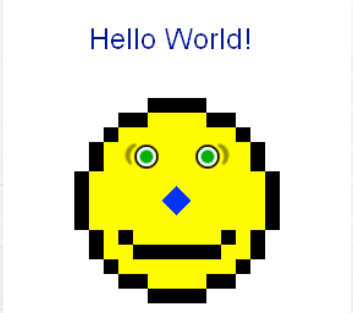
Chief technical officer of the Opera Software company
Håkon Wium Lie - creator of CSS web standard

Google Chrome

v0.2	September 08
...	...
v5.0.307	Jan 2010
...	...
v9.0.587	November 2010

- Very different architecture
- Process Independent tabs (reducing crashes)
- Open source and Web-Kit rendering engine
- Look for the Google Chrome comic online, which explains the architecture

Google Chrome V 0.2 - passes ACID 2 from beginning



Mobile web

- Many different mobile browsers

Mobile web

- Generally don't support all web standards
- Offer 'cut down' experience
- Better ones are derived from proper PC browser rendering engines
- Expensive data plan, charged by the MB

iPhone Safari

- Closest so far to standard web experience

Synchronised bookmarks
Javascript, CSS support
Touch screen zoom
Inclusive data plan
Derived from Safari and webkit

ACID 3

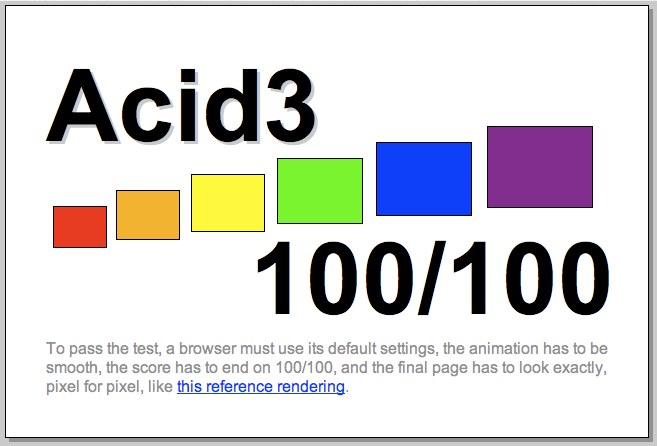
ACID 3 increases focus on DOM

- DOM2 Core
- DOM2 Events
- DOM2 HTML
- DOM2 Range
- DOM2 Style (getComputedStyle, ...)
- DOM2 Traversal (NodeIterator, TreeWalker)
- DOM2 Views (defaultView)
- ECMAScript
- HTML4 (<object>, <iframe>, ...)
- HTTP (Content-Type, 404, ...)
- Media Queries
- Selectors (:lang, :nth-child(), combinators, dynamic changes, ...)
- XHTML 1.0
- CSS2 (@font-face)
- CSS2.1 ('inline-block', 'pre-wrap', parsing...)
- CSS3 Color (rgba(), hsla(), ...)
- CSS3 UI ('cursor')
- data: URIs

Still in development

http://www.webstandards.org/action/acid3

ACID 3



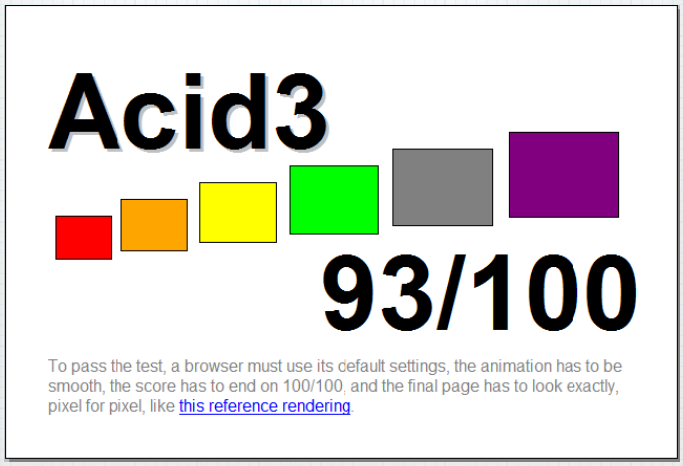
http://annevankesteren.nl/2008/01/acid3

ACID 3

http://annevankesteren.nl/2008/01/acid3

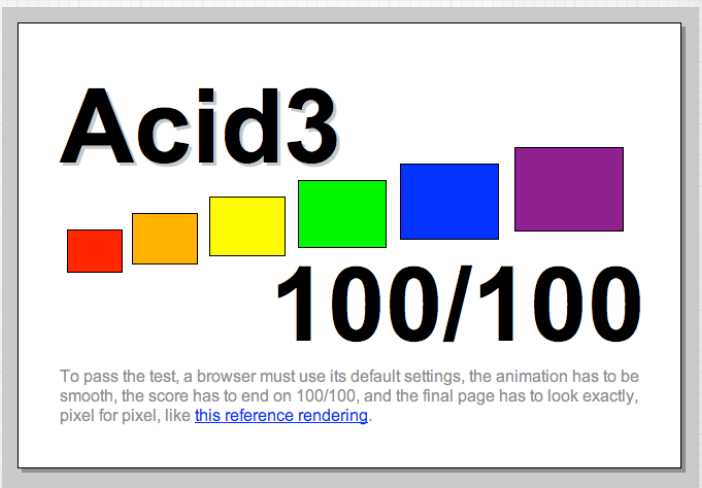
ACID 3

Firefox 3.6



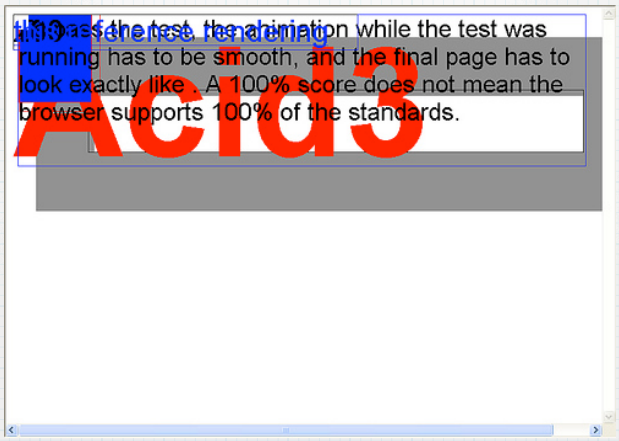
ACID 3

Safari 4



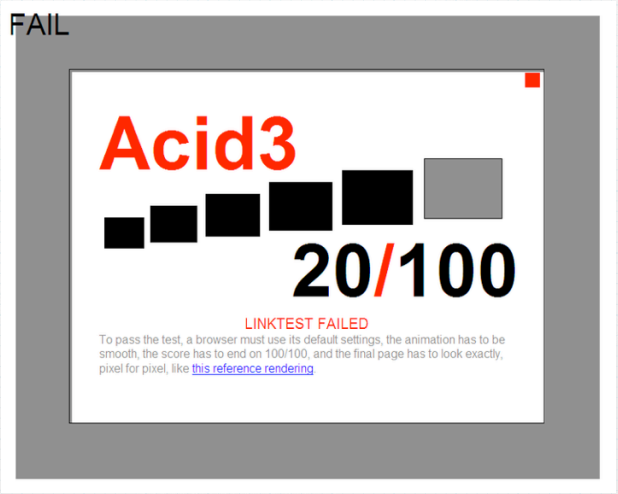
ACID 3

IE 7.0



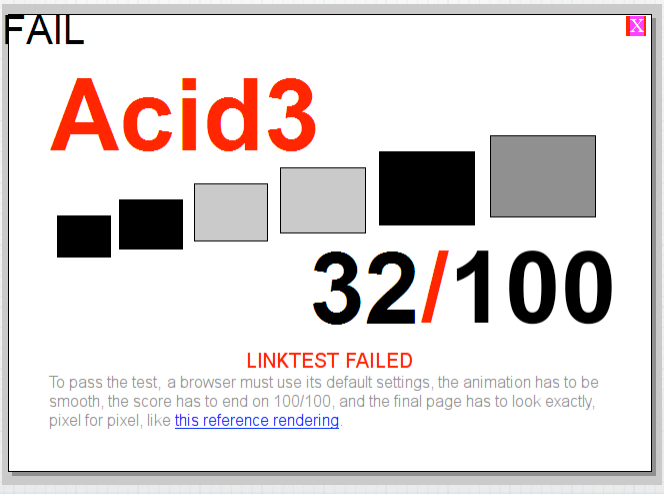
ACID 3

IE 8.0



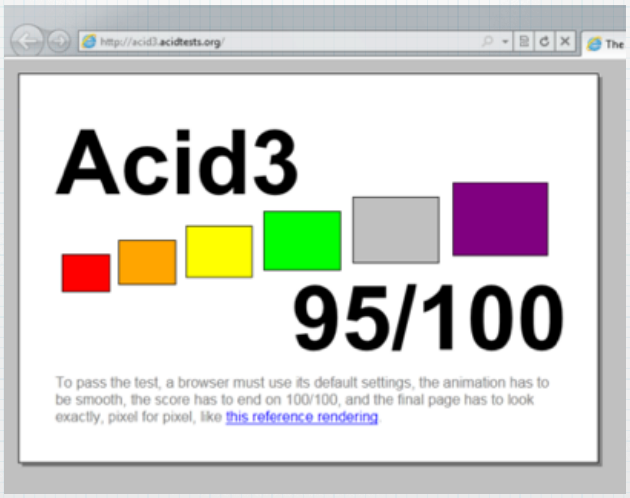
ACID 3

IE 9.0 Pre-beta



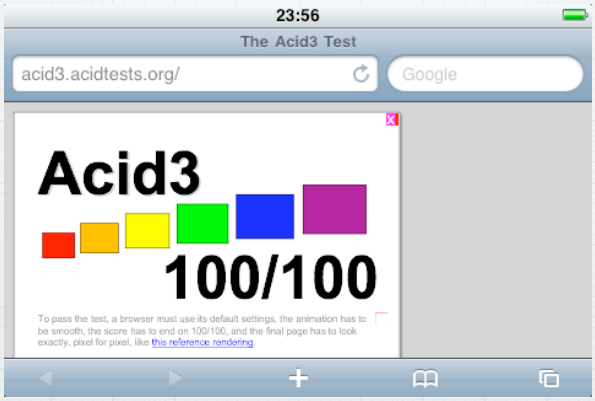
ACID 3

IE 9.0 Beta



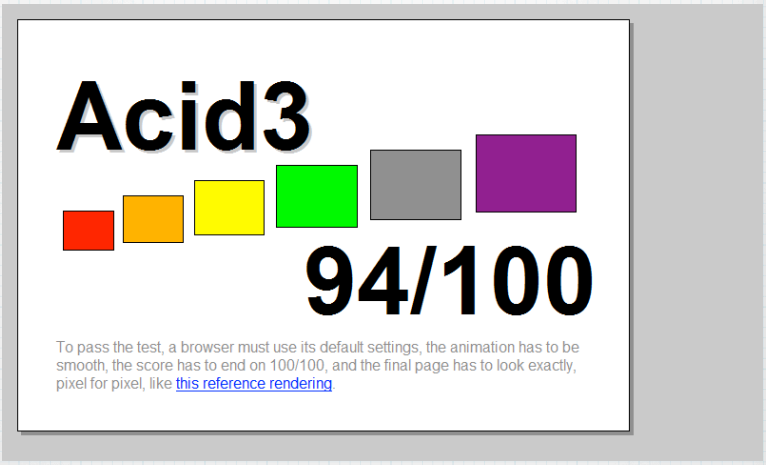
ACID 3

Mobile Safari



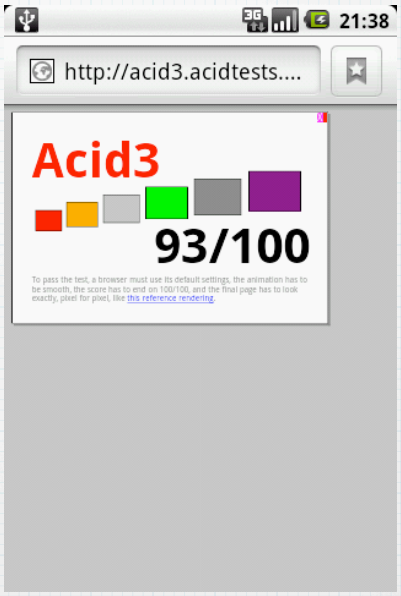
ACID 3

Mobile Firefox



ACID 3

Android



ACID 3

Internet Explorer
Windows Mobile 7