

# Web: Επισκόπηση

Αντώνης Σιδηρόπουλος

# Έννοιες – Ορολογία

- **Διαδίκτυο (Internet)**, είναι το φυσικό δίκτυο των δικτύων, στο οποίο είναι συνδεδεμένοι εκατομμύρια υπολογιστές και επιμέρους δίκτυα με διαφορετική τεχνολογία και λειτουργικά συστήματα, αλλά με τα ίδια πρωτόκολλα επικοινωνίας - TCP/IP.
- **Παγκόσμιος Ιστός (World Wide Web ή www ή Web)**, είναι μια συλλογή από διασυνδεδεμένα πολυμεσικά αρχεία-υπερκειμένου (hypertext) του διαδικτύου, προσβάσιμα μέσω του κοινού πρωτοκόλλου HTTP.
- **Υπερκείμενο (Hypertext)**, είναι ένα κείμενο στο οποίο κάποιες λέξεις (σύνδεσμοι) είναι λογικά συνδεδεμένες, με άλλα κείμενα του θέματος που παρουσιάζεται.
- **WWW ≠ Internet**  
(διαδίκτυο = hardware, web= μια από τις υπηρεσίες του Internet)

# Έννοιες – Ορολογία

- **Ιστοσελίδα (web page)**, ονομάζουμε μια σελίδα του Παγκόσμιου Ιστού στην οποία διαβάζουμε πληροφορίες. Πολλές φορές οι ιστοσελίδες είναι υπερκείμενα εμπλουτισμένα με εικόνες, ήχους, βίντεο, κινούμενα γραφικά, κινούμενο κείμενο.
- **Ιστότοπος (website)**, ονομάζουμε ένα σύνολο ιστοσελίδων που πραγματεύονται το ίδιο θέμα, είναι συνδεδεμένες μεταξύ τους με υπερσυνδέσεις και φιλοξενούνται στο ίδιο domain (περιοχή) του Παγκόσμιου Ιστού.
- **Φυλλομετρητής / πλοηγός (browser)**, είναι ένα πρόγραμμα με το οποίο μπορούμε να βλέπουμε ιστοσελίδες και να κάνουμε πλοήγηση στο διαδίκτυο.

# Ιστορικά

- **1989 – CERN:** **Tim Berners-Lee** wrote a proposal that referenced ENQUIRE, a database and software project he had built in 1980, and described a more elaborate information management system.
- **1990 – CERN:** **Tim Berners-Lee** και Robert Cailliau, he published a more formal proposal (on 12 November 1990) to build a "Hypertext project" called "WorldWideWeb" (one word, also "W3") as a "web" of "hypertext documents" to be viewed by "browsers" using a client-server architecture.  
defined:
  - universal document identifier (UDI), later known as uniform resource locator (URL) and uniform resource identifier (URI)
  - the publishing language HyperText Markup Language (HTML)
  - the Hypertext Transfer Protocol (HTTP)

# Ιστορικά

- **6 August 1991:** Berners-Lee posted a short summary of the World Wide Web project on the alt.hypertext newsgroup. This date also marked the debut of the Web as a publicly available service on the Internet, although new users only access it after **August 23**. For this reason this is considered the internaut's day.
- **December 1992:** The first server outside Europe was set up at the Stanford Linear Accelerator Center (SLAC) in Palo Alto, California
- .....

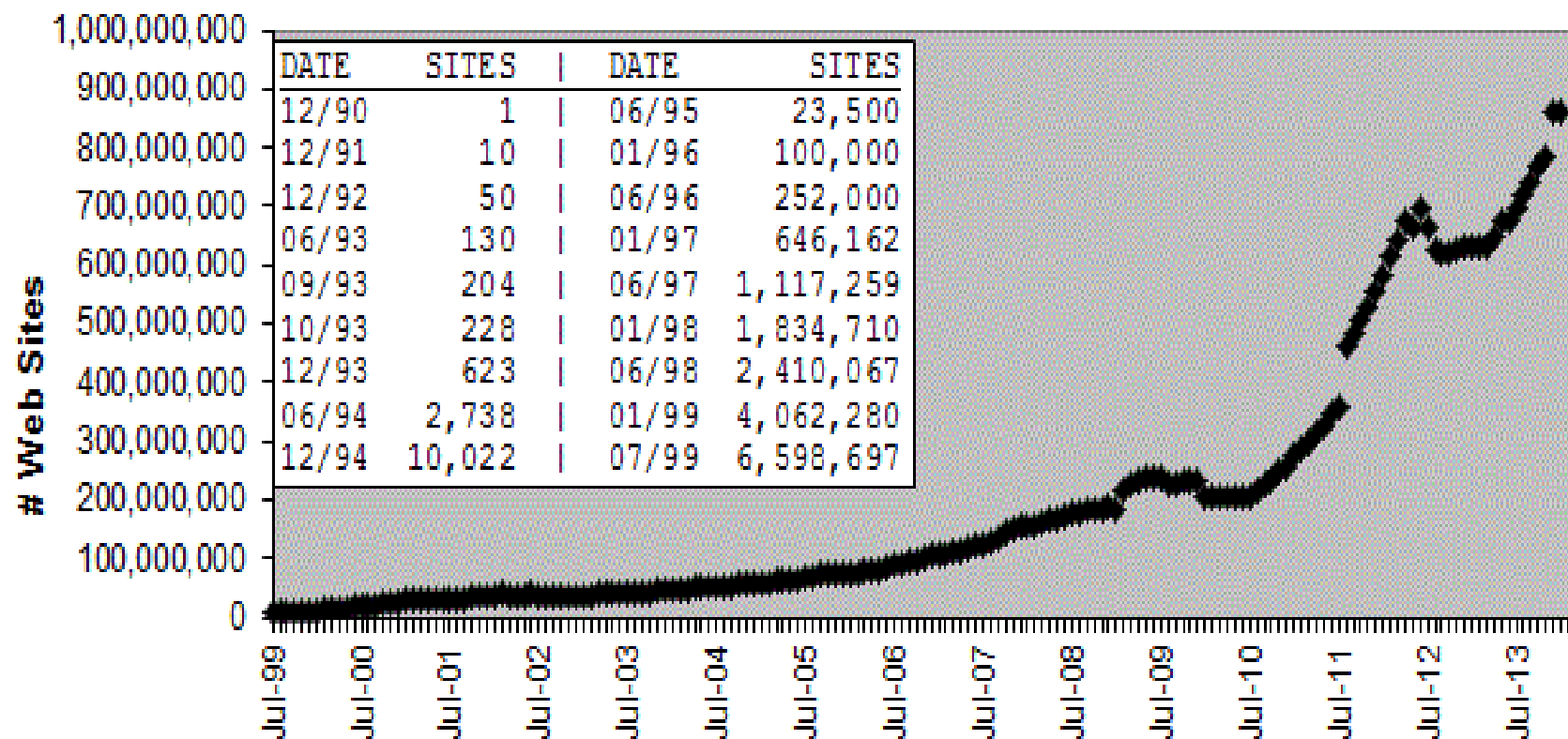


To logo tou Project

# Ιστορικά

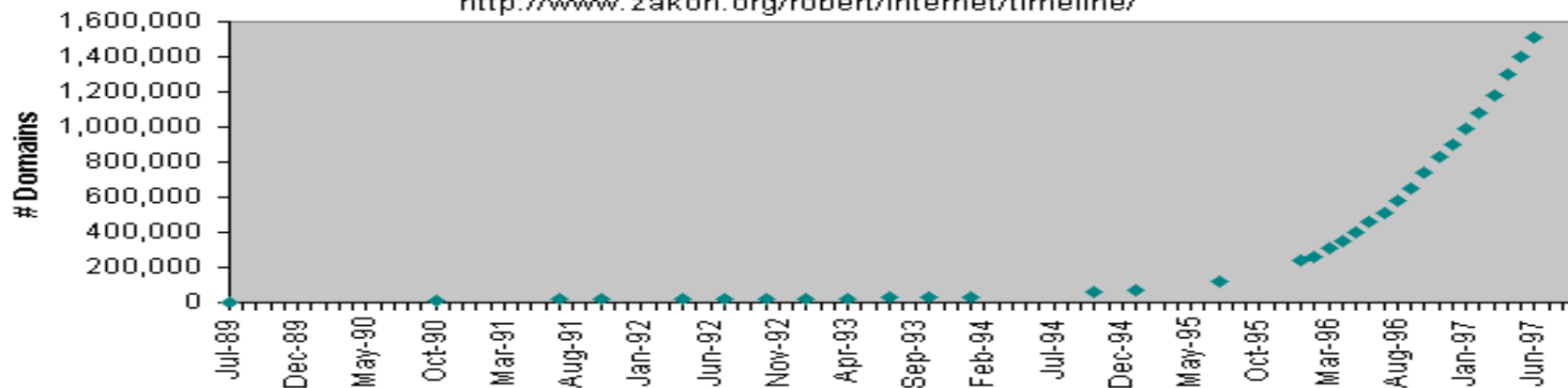
Hobbes' Internet Timeline Copyright ©2014 Robert H Zakon

<http://www.zakon.org/robert/internet/timeline/>

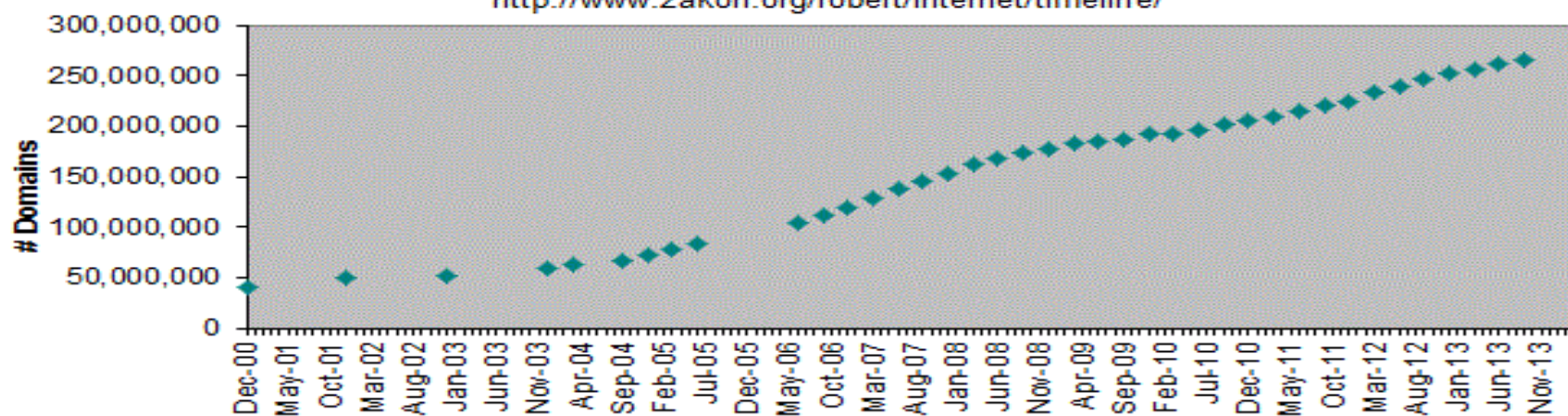


# Ιστορικά

Hobbes' Internet Timeline Copyright ©2003 Robert H Zakon  
<http://www.zakon.org/robert/internet/timeline/>

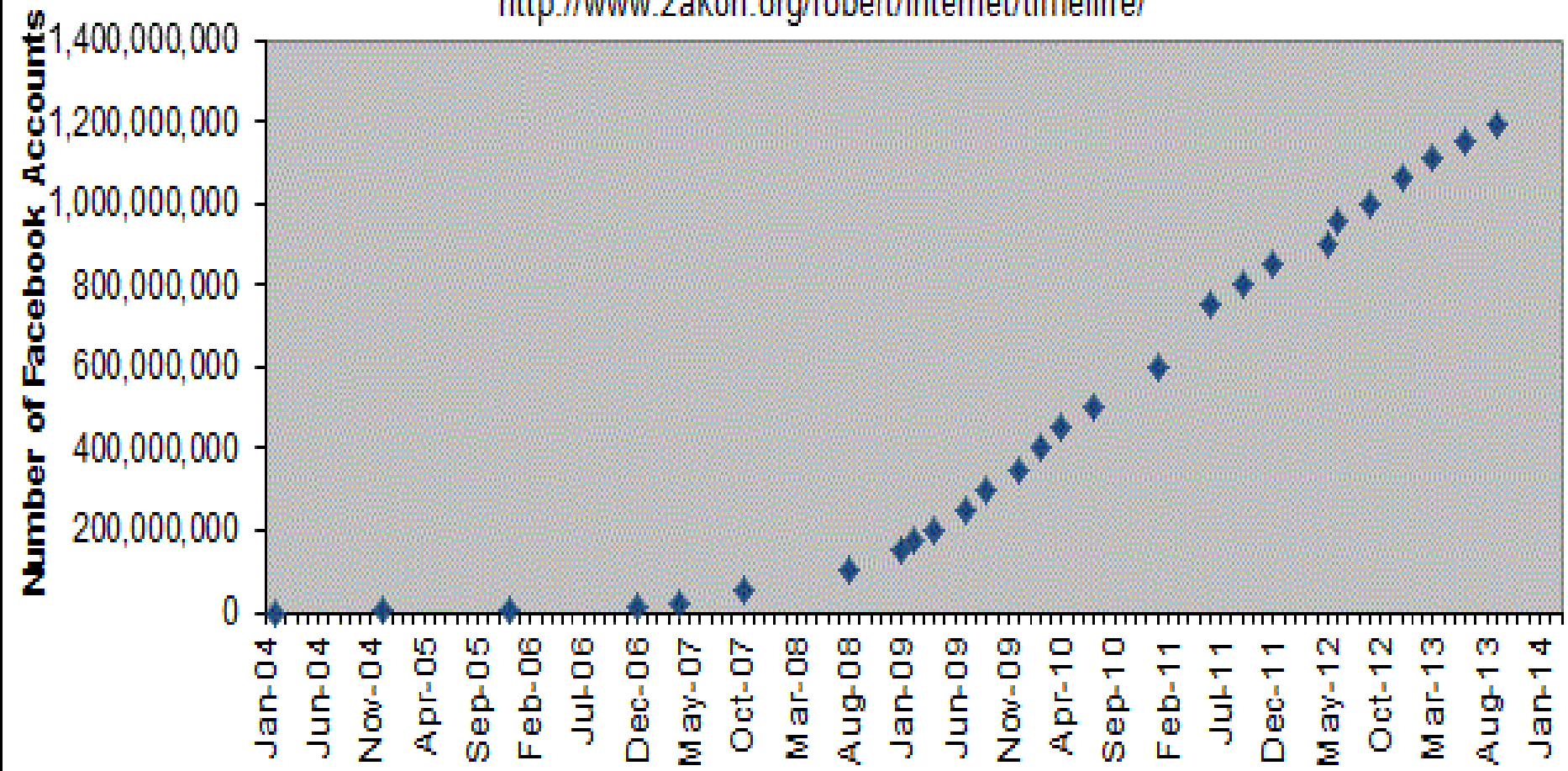


Hobbes' Internet Timeline Copyright ©2014 Robert H Zakon  
<http://www.zakon.org/robert/internet/timeline/>



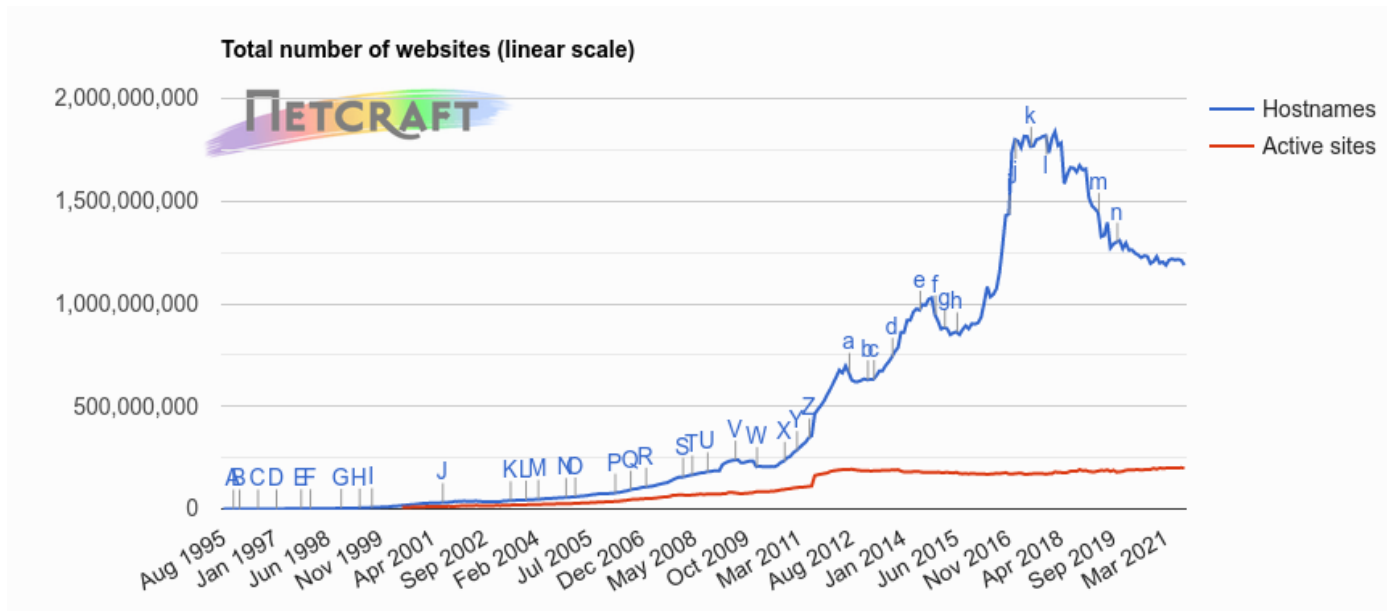
# Ιστορικά

Hobbes' Internet Timeline Copyright ©2014 Robert H Zakon  
<http://www.zakon.org/robert/internet/timeline/>



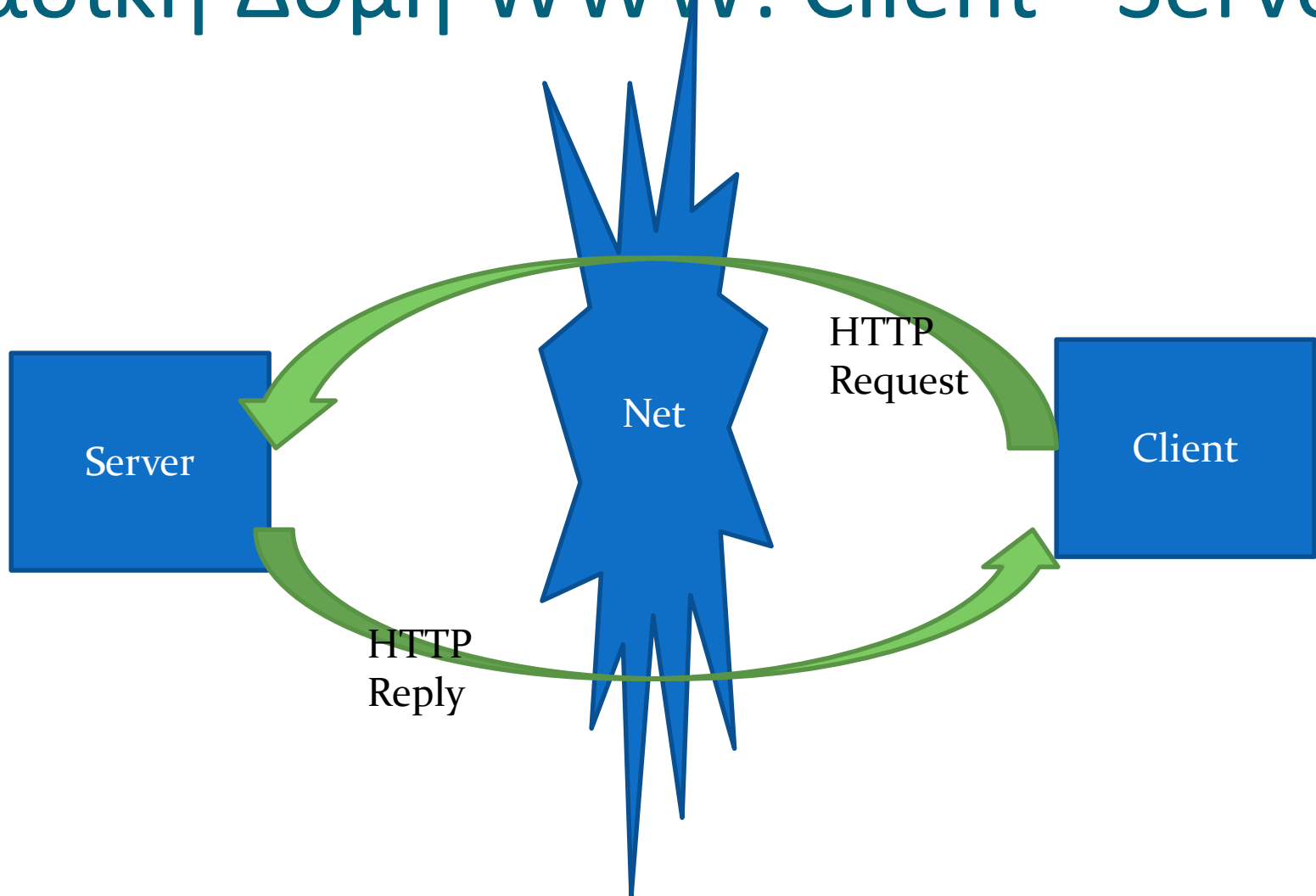


# Περισσότερα Στατιστικά



<http://news.netcraft.com/archives/2021/>

# Βασική Δομή WWW: Client - Server



- **HTTP Request:** Συνήθως είναι η αίτηση του πελάτη να λάβει (GET) ένα αντικείμενο (αρχείο) το οποίο προσδιορίζεται με το URL.
  - Άλλου τύπου αιτήσεις: HEAD (πληροφορίες), POST (αποστολή)
- **HTTP Reply:** Στέλνει ο server το αρχείο (ή πληροφορίες γι' αυτό) που ζήτησε ο πελάτης.
  - Το αρχείο μπορεί να είναι οτιδήποτε: HTML, IMAGE, BINARY, DOC, XLS, EXE, ZIP, ....

# Browsers

- Πρόγραμμα εμφάνισης και αλληλεπίδρασης του χρήστη με αρχεία υπερκειμένου (και όχι μόνο).
- Πρόγραμμα πελάτης (client).

# Browsers

- WorldWideWeb – Nexus (1991)
- Mosaic (1993 - NCSA)
- Netscape Navigator (1995 - ~2008), βασίστηκε στον Mosaic.
- Internet Explorer (1995 – σήμερα). Microsoft μαζί με τα Windows 95.
- Opera (1994, 1996 release – σήμερα). ερευνητικό έργο του εθνικού φορέα τηλεφωνίας της Νορβηγίας.
- Mozilla – Firefox (2002 – σήμερα). Το 1998 η Netscape Communications αποφάσισε να ξαναϋλοποιήσει το πρόγραμμα φυλλομέτρησης από την αρχή. Έτσι δημιουργήθηκε το πρόγραμμα Gecko, το οποίο όμως δεν κατάφερε να υλοποιήσει το στόχο του. Η επόμενη προσπάθεια της AOL δημιούργησε το πρόγραμμα Mozilla.
- Safari (2003, Apple). Το 2007 για Windows.
- Google chrome (2008, Google).
- + Πλειάδα από browsers ειδικά για κινητές συσκευές + tablets.

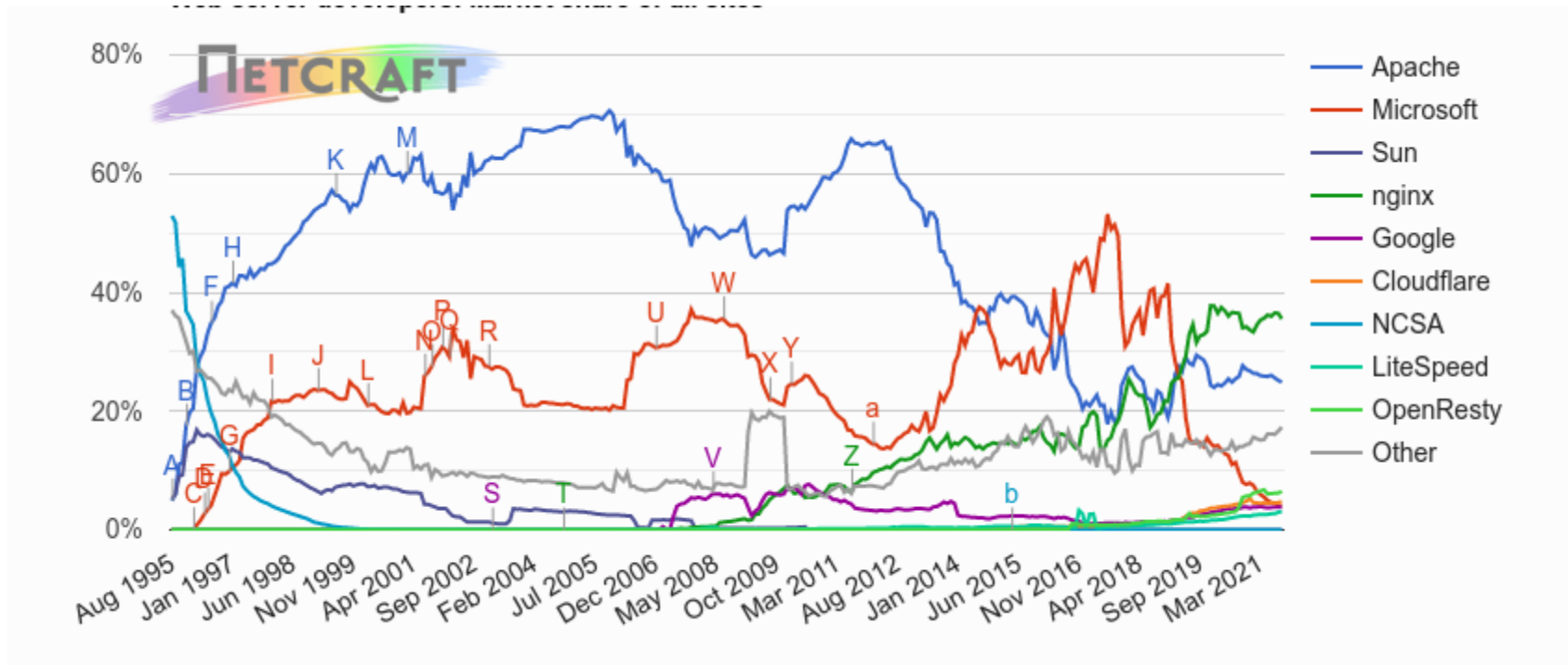
# Τι είναι Web Server?

- The term web server can refer to either the hardware (the computer) or **the software** (the computer application) that helps to deliver web content that can be accessed through the Internet.
- The primary function of a web server is to deliver web pages (ή οποιονδήποτε άλλου τύπου αρχείο) to clients.
- The communication between client and server takes place using the Hypertext Transfer Protocol (HTTP).
- Many generic web servers also support server-side scripting.
- The world's first web server, later known as CERN httpd, which ran on NeXTSTEP

# Web Servers

- Οι πλέον διαδεδομένοι Web Servers:
  - IIS (Internet Information Server) – Microsoft
  - Apache, Apache Tomcat - Apache
  - nginx - NGINX, Inc.
- Άλλοι Web Servers:
  - Cherokee HTTP Server, Google Web Server (GWS), Hiawatha HTTP Server, IBM HTTP Server, Jetty, lighttpd, LiteSpeed Web Server, Mongoose, NaviServer, OpenLiteSpeed, Oracle HTTP Server, Oracle WebLogic Server, thttpd, Tornado, TUX web server, Wakanda Server, WEBrick, Yaws, Zope,...

# Web Servers





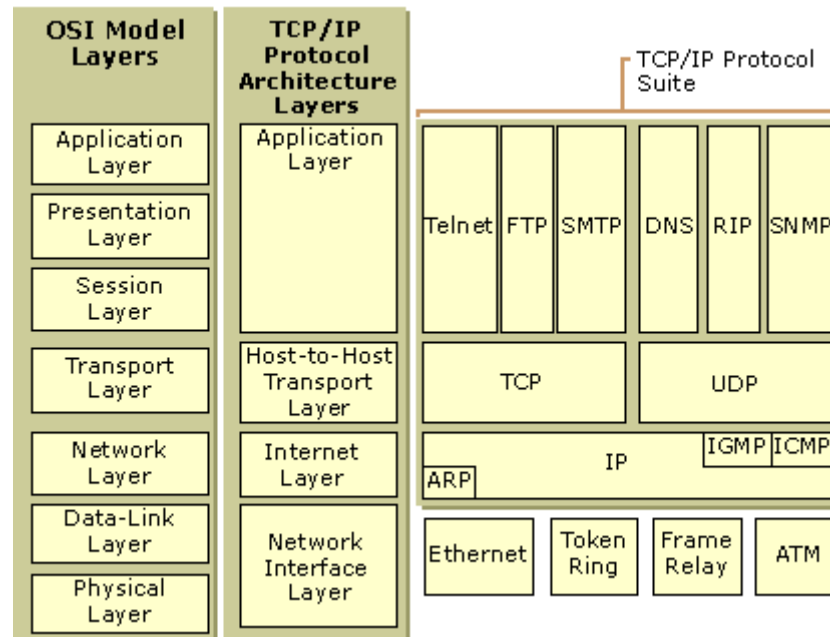
# Είδη Ιστοσελίδων (Στατικές vs. Δυναμικές)

- Στατικές, απλά HTML-έγγραφα, το περιεχόμενο των οποίων δεν αλλάζει και δεν ανανεώνεται παρά μόνο αν κάποιος επεξεργαστεί τα αντίστοιχα αρχεία (χρήση της HTML).
- Δυναμικές, δεν είναι απλά HTML-έγγραφα, αλλά προγράμματα ή μίξη HTML και προγραμμάτων
  - Τα προγράμματα δημιουργούν on-demand το HTML αρχείο που στέλνει ο server στον client.
  - Υπάρχουν διάφοροι μέθοδοι /τεχνολογίες δημιουργίας δυναμικών ιστοσελίδων (CGI, FastCGI, php module, mod\_perl, .... )
  - Υπάρχουν διάφορες γλώσσες προγραμματισμού που μπορούν να χρησιμοποιηθούν για την δημιουργία δυναμικών ιστοσελίδων – ουσιαστικά οποιαδήποτε γλώσσα προγραμματισμού (php, perl, java, python, C, C++, shell script...)

# Είδη Ιστοσελίδων (Στατικές vs. Δυναμικές)

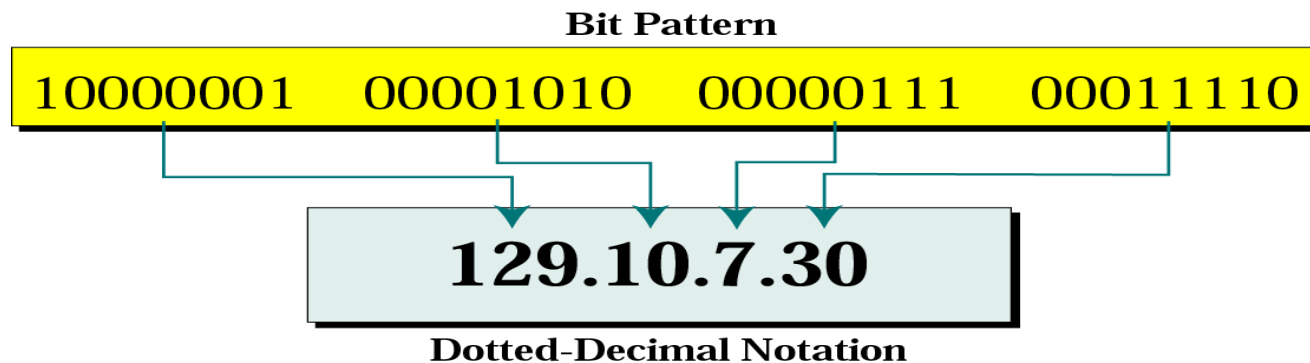
- Στις δυναμικές ιστοσελίδες, το περιεχόμενο της ιστοσελίδας,
  - συνήθως αποθηκεύεται και αντλείται δυναμικά από μια ή περισσότερες βάσεις δεδομένων (π.χ. MySQL) – ως βάση δεδομένων μπορούμε να θεωρήσουμε και ένα απλό αρχείο.
  - Δημιουργείται χρησιμοποιώντας άλλες υπηρεσίες (πχ <http://whois.net/> )
  - Δημιουργείται επικοινωνώντας με άλλους web –servers ( πχ: <http://www.tripadvisor.com.gr/>)

# Περί Πρωτοκόλλων



# IP Διευθύνσεις στο Διαδίκτυο

- Η IP διεύθυνση (IPv4), είναι ένας 32-bit αριθμός οργανωμένος σε τέσσερα δεκαδικά ψηφία χωρισμένα με τελεία, της μορφής π.χ.



- Μια διεύθυνση IP περιλαμβάνει δύο τμήματα:
  - την ταυτότητα του δικτύου (network identifier)
  - την ταυτότητα του κόμβου (host identifier) (π.χ. υπολογιστής) που βρίσκεται συνδεδεμένος στο συγκεκριμένο τμήμα του δικτύου.

# Σύστημα Ονομασίας Περιοχών (Domain Name System ή DNS )

- Οι διευθύνσεις είναι πιο ευκολομνημόνευτες σε μορφή λέξεων, π.χ. `www.in.gr`
- Η υπηρεσία της μετάφρασης από διευθύνσεις με λέξεις, σε διευθύνσεις με αριθμούς ονομάζεται Σύστημα Ονομασίας Περιοχών και υλοποιείται από διασκορπισμένους στο διαδίκτυο εξυπηρετητές/ διακομιστές (servers).
- Όμως μας τελειώνουν οι διαθέσιμες διευθύνσεις. Η λύση είναι η έκδοση IPv6, με διευθύνσεις 128 bits αντί των σημερινών 32 bits.

# IPv6

- Όμως μας τελειώνουν οι διαθέσιμες διευθύνσεις. Η λύση είναι η έκδοση IPv6, με διευθύνσεις 128 bits αντί των σημερινών 32 bits.
- Μια διεύθυνση IPv6 έχει την μορφή:  
**2111:0dc9:95a3:07d341319:8a2e:0370:7344**

# URLs

- Uniform Resource Locator
- Είναι το αναγνωριστικό μιας ιστοσελίδας.
- Περιέχει την πληροφορία
  - Πως επικοινωνούμε με τον server που διαθέτει το αρχείο (ή την υπηρεσία) την οποία θέλουμε να προσπελάσουμε
  - Ποιο είναι το αρχείο μέσα στον server.
- υπηρεσία://χρήστης:κωδικός@δικτυακή.διεύθυνση:πόρτα/διαδρομήΦακέλου/πόρος

# URLs

- Το σύστημα διευθυνσιοδότησης με URLs υποστηρίζει και άλλα πρωτόκολλα εκτός από το http.
  - Γενική σύνταξη:  
Πρωτόκολλο://χρήστης:κωδικός@δικτυακή.διεύθυνση:πόρτα/διαδρομή/Φακέλου/πόρος  
Πχ:
    - `http://asidirop:lala@www.teithe.gr:4000/public`
    - `ftp://asidirop:lala@www.teithe.gr/public`
    - `http://www.teithe.gr:80/public`
    - `http://www.teithe.gr/public`



# URLs

- Για να επικοινωνήσει ένα πρόγραμμα (πχ browser) με μια υπηρεσία του δικτύου χρειάζεται 2 πληροφορίες:
  - IP
  - Port number (όταν παραλείπεται σε ένα URL, εννοείται το default για την αντίστοιχη υπηρεσία. Για το HTTP=80, ftp=21, (gopher=70)  
...

# URLs

- Το τελευταίο τμήμα ενός URL μπορεί να περιέχει και ορίσματα προς την υπηρεσία:
  - `http://www.nytimes.com/article?id=1034&page=2&use  
rid=A4DE55F`
    - `?`: Διαχωριστής των ορισμάτων με το υπόλοιπο URL
    - `&`: Διαχωριστής των ορισμάτων μεταξύ τους

# URLs

- Ειδικοί χαρακτήρες όπως: /, ?, &, +, space απαγορεύονται μέσα σε διαδρομή, όνομα αρχείου ή ορίσματα και αντικαθίστανται με άλλους:
  - <https://www.google.gr/search?q=antonis+sidiropoulos>
- Το space “ ” αντικαθίσταται με το “+”
- Όλοι οι υπόλοιποι χαρακτήρες αντικαθίστανται με τον αντίστοιχο UTF-8 αριθμό σε δεκαεξαδική μορφή:

# URLs - αντικαταστάσεις

- Παραδείγματα χρησιμοποιώντας το search της google:

Query	URL
a sidiropoulos	<a href="https://www.google.com/#q=a+sidiropoulos">https://www.google.com/#q=a+sidiropoulos</a>
test?	<a href="https://www.google.com/#q=test%3F">https://www.google.com/#q=test%3F</a>
,.>?;:.'"}[{\%	<a href="https://www.google.com/#q=%2C.%3E%3F%3B: '%5D%7D%5B%7B%25">https://www.google.com/#q=%2C.%3E%3F%3B: '%5D%7D%5B%7B%25</a>

- Οι χαρακτήρες που δεν επιτρέπονται, μετατρέπονται στην μορφή %XX, όπου XX ο UTF-8 κωδικός του χαρακτήρα σε δεκαεξαδική μορφή.

# URLs - αντικαταστάσεις

- Παραδείγματα χρησιμοποιώντας το search της google:

Query	URL
a sidiropoulos	<a href="https://www.google.com/#q=a+sidiropoulos">https://www.google.com/#q=a+sidiropoulos</a>
test?	<a href="https://www.google.com/#q=test%3F">https://www.google.com/#q=test%3F</a>
,.>?;:.' '][{%	<a href="https://www.google.com/#q=%2C.%3E%3F%3B%27%5D%7D%5B%7B%25">https://www.google.com/#q=%2C.%3E%3F%3B%27%5D%7D%5B%7B%25</a>

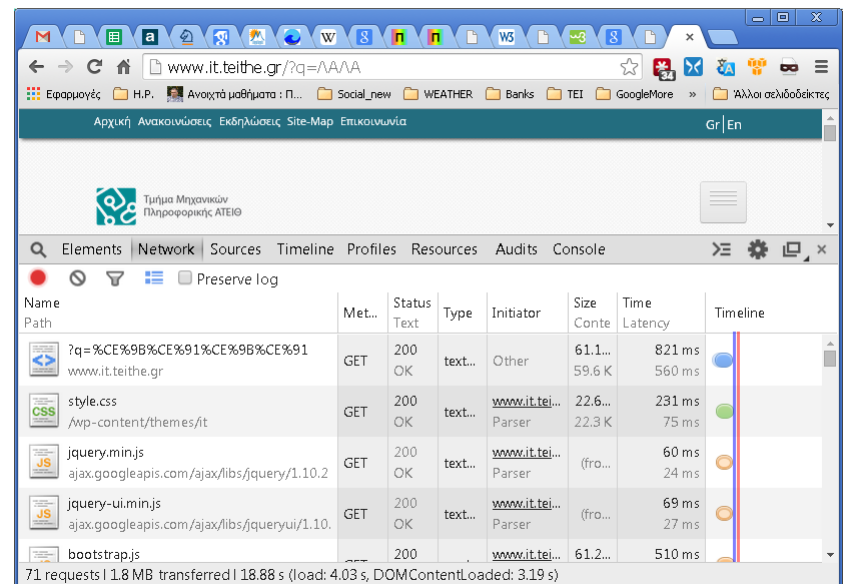
- Οι χαρακτήρες που δεν επιτρέπονται, μετατρέπονται στην μορφή %XX, όπου XX ο UTF-8 κωδικός του χαρακτήρα σε δεκαεξαδική μορφή.

Query	URL
<b>A</b> Σιδηρόπουλος	<a href="https://www.google.com/#q=%CE%91+%CE%A3%CE%B9%CE%B4%CE%B7%CF%81%CF%8C%CF%80%CE%BF%CF%85%CE%BB%CE%BF%CF%82">https://www.google.com/#q=%CE%91+%CE%A3%CE%B9%CE%B4%CE%B7%CF%81%CF%8C%CF%80%CE%BF%CF%85%CE%BB%CE%BF%CF%82</a>

- Στο UTF8 οι Ελληνικοί χαρακτήρες κωδικοποιούνται σε 2 bytes
- Εάν σε ένα URL εισάγονται Ελληνικοί χαρακτήρες τι συμβαίνει?

# URLs - αντικαταστάσεις

- Εάν σε ένα URL εισάγονται Ελληνικοί χαρακτήρες τι συμβαίνει?
- Από τους περισσότερους browsers μετατρέπονται «κρυφά» στους αντίστοιχους εκαεξαδικούς αριθμούς



UR

- Ε
- σ
- 

The screenshot shows a web browser window with the address bar containing `www.it.teithe.gr/?q=^AAA`. The browser's taskbar at the top shows various open applications. The browser's address bar and the Network tab's request list are circled in red.

The Network tab is active, showing a list of requests. The first request, `?q= %CE%9B%CE%91%CE%9B%CE%91 www.it.teithe.gr`, is highlighted with a red circle. This request is a GET request with a status of 200 OK, a type of text, and a size of 61.1... 59.6 K. It was initiated by 'Other' and took 821 ms to complete, with a latency of 560 ms.

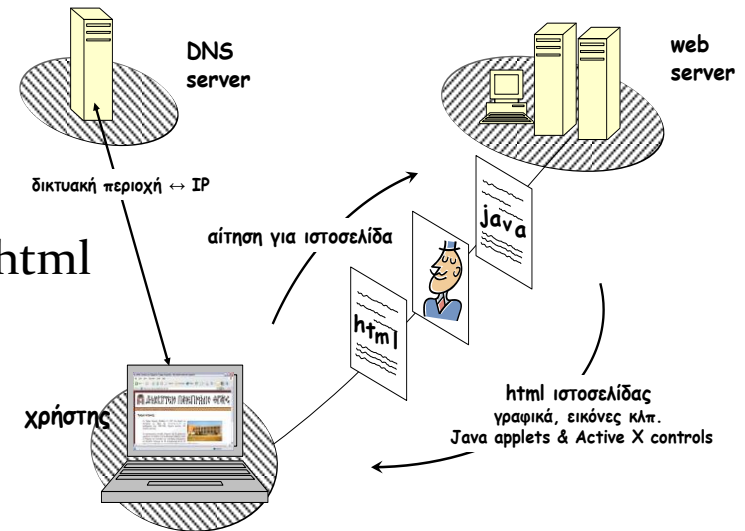
Name Path	Met...	Status Text	Type	Initiator	Size Conte	Time Latency	Timeline
<code>?q= %CE%9B%CE%91%CE%9B%CE%91 www.it.teithe.gr</code>	GET	200 OK	text...	Other	61.1... 59.6 K	821 ms 560 ms	
<code>style.css /wp-content/themes/it</code>	GET	200 OK	text...	<code>www.it.tei...</code> Parser	22.6... 22.3 K	231 ms 75 ms	
<code>jquery.min.js ajax.googleapis.com/ajax/libs/jquery/1.10.2</code>	GET	200 OK	text...	<code>www.it.tei...</code> Parser	(fro...	60 ms 24 ms	
<code>jquery-ui.min.js ajax.googleapis.com/ajax/libs/jqueryui/1.10.</code>	GET	200 OK	text...	<code>www.it.tei...</code> Parser	(fro...	69 ms 27 ms	
<code>bootstrap.js</code>	GET	200		<code>www.it.tei...</code>	61.2...	510 ms	

71 requests | 1.8 MB transferred | 18.88 s (load: 4.03 s, DOMContentLoaded: 3.19 s)

# Επικοινωνία web client (browser)

## – web server

- ο χρήστης δίνει την διεύθυνση του δικτυακού πόρου (URL), π.χ. `http://www.it.teithe.gr/~asidirop/index.html`
- ο browser αναλύει την διεύθυνση σε 3 μέρη:
  - πρωτόκολλο υπηρεσίας (π.χ. http)
  - όνομα της δικτυακής περιοχής (π.χ. `www.it.teithe.gr`)
  - μονοπάτι και όνομα πόρου (π.χ. `/~asidirop/index.html`)
- ο web browser βρίσκει ποια IP διεύθυνση αντιστοιχεί στο όνομα του web server (π.χ. `www.it.teithe.gr = 195.251.123.232`) είτε ψάχνοντας τοπικά σε ειδική προσωρινή μνήμη, είτε ρωτώντας τον DNS server που του αντιστοιχεί
- ο web browser σχηματίζει σύνδεση με την συγκεκριμένη διεύθυνση IP
- ο web browser στέλνει στον web server μια αίτηση, στη μορφή του πρωτοκόλλου HTTP, για το ζητούμενο αρχείο (π.χ. για το `index.html` από τον φάκελο `/~asidirop`)





# Επικοινωνία web client (browser)

## – web server

- ο web browser αναλύει το html αρχείο που έλαβε και το εμφανίζει στην γνωστή μορφή της ιστοσελίδας
- αν κατά την ανάλυση του html αρχείου προκύψει ότι η ιστοσελίδα περιέχει και άλλα αντικείμενα εκτός κειμένου (π.χ. εικόνες, Java Applets, Active X controls, κλπ), ο web browser:
  - στέλνει ανάλογες αιτήσεις για τα επιπλέον αρχεία στον web server
  - λαμβάνει διαδοχικά τα επιπλέον αρχεία
  - τα ενσωματώνει στην τελική ιστοσελίδα
- όταν ο χρήστης ενεργοποιήσει έναν υπερσύνδεσμο από την σελίδα, ο web browser αναγνωρίζει το URL που αντιστοιχεί στον υπερσύνδεσμο και επιστρέφει στο βήμα 2.
- αν αλλάξει κάτι στα αρχεία του web server, ο χρήστης θα δει την αλλαγή μόνο μετά από δική του αίτηση για ανανέωση της ιστοσελίδας (refresh = 'ανανέωση').

# Επικοινωνία web client (browser)

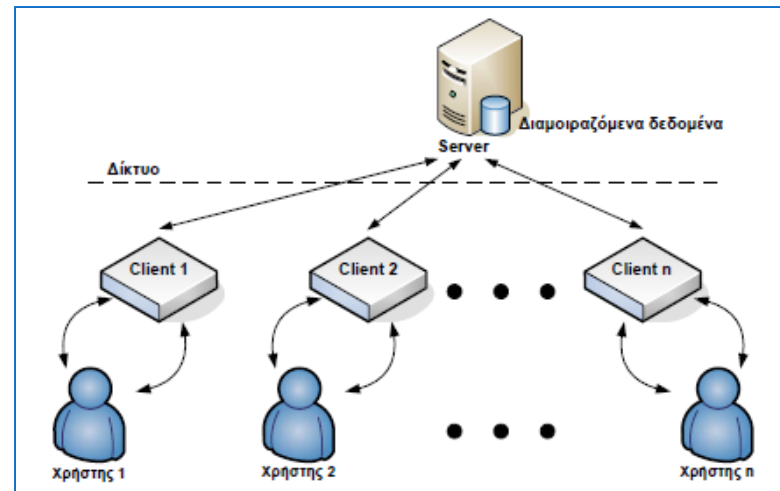
## – web server (2/3)

- Οι Clients (browsers) δεν μπορούν να επικοινωνήσουν μεταξύ τους.
- Το σύνολο των πελατών (clients) μπορούν να επικοινωνήσουν μεταξύ τους ΜΟΝΟ μέσω ενός (ή περισσότερων) εξυπηρετητών (servers). Οι πελάτες δεν ανταλλάσσουν απευθείας μηνύματα μεταξύ τους, αλλά στέλνουν τα μηνύματα αυτά στους εξυπηρετητές, οι οποίοι τα προωθούν στους υπόλοιπους πελάτες (και εξυπηρετητές). Όλοι οι χρήστες χρησιμοποιούν μια εφαρμογή που βρίσκεται σε έναν κεντρικό εξυπηρετητή.

# Επικοινωνία web client (browser)

## – web server (3/3)

- Πλεονεκτήματα:
  - Δεν χρειάζεται μεγάλη υπολογιστική ισχύ στον client (μόνο ο server).
  - Η χρήση των εξυπηρετητών κάνει εύκολη την παροχή νέων υπηρεσιών στους χρήστες.
- Μειονεκτήματα:
  - Το σύστημα εξαρτάται άμεσα από την ισχύ του εξυπηρετητή
  - Ένα κεντρικό σημείο αστοχίας του συστήματος
  - Bottleneck στον server



# HTTP (HyperText Transfer Protocol)

- Το **πρωτόκολλο μεταφοράς υπερκειμένου (HTTP)** είναι ένα σύνολο κανόνων που καθορίζουν τον τρόπο πρόσβασης και μεταφοράς αρχείων στον παγκόσμιο ιστό.
- Καθορίζει τον τρόπο επικοινωνίας των web clients με τους web servers, δηλ. την μορφοποίηση και τρόπο μετάδοσης των μηνυμάτων (πακέτων).
- Όταν εισάγουμε μια URL σε ένα φυλλομετρητή, στέλνεται μια εντολή μέσω του πρωτοκόλλου HTTP σε κάποιον web server προστάζοντας τον να βρει και να αποστείλει μια συγκεκριμένη ιστοσελίδα, δηλ. ένα HTML αρχείο.
- Το HTTP είναι ένα **stateless** –πρωτόκολλο, δηλ. χειρίζεται κάθε αίτημα σαν ξεχωριστή και ανεξάρτητη διαδικασία – δεν υπάρχει μνήμη. Έτσι για ολοκλήρωση της αποστολής μιας ιστοσελίδας με κείμενο και εικόνες απαιτούνται πολλαπλές συνδέσεις (αιτήσεις και αποκρίσεις). – έχουν αναπτυχθεί διάφορες τεχνικές για να «αποθηκεύεται» το state (όπως cookies). – (θα μιλήσουμε σε επόμενο μάθημα).