

# Εισαγωγή στο HTTP

# HTTP

- HTTP : Hyper Text Transfer Protocol
- Πρωτόκολλο επικοινωνίας
- Ανήκει στο Επίπεδο Εφαρμογής (TCP/IP)
- Βάση για την λειτουργία του WWW
- Χρησιμοποιεί το TCP σαν επίπεδο μεταφοράς
- Εκδόσεις: 1.0 (πριν), 1.1 (τώρα), 2(σε εξέλιξη)
- Καθορίζει τύπο και σημασία μηνυμάτων πελάτη/εξυπηρετητή.

# Χαρακτηριστικά

- Αιτήσεις (HTTP Requests), Αποκρίσεις (HTTP Responses)
- Προεπιλεγμένη θύρα TCP : 80
- Πρωτόκολλο χωρίς διατήρηση κατάστασης (stateless). Χρήση cookies για διατήρηση κατάστασης.
- Αρχικά κάθε νέα αίτηση δημιουργούσε νέα σύνδεση TCP. Τώρα μια σύνδεση TCP μπορεί να υποστηρίζει πολλές αιτήσεις (ακόμη και πολλές αιτήσεις σε εξέλιξη).

# URI, URL Ονομασία Πόρων

- Uniform Resource Identifier (URI)
- Uniform Resource Locators (URL)
- URL υποσύνολο των URI. URL και URI δεν είναι μέρος του HTTP αλλά χρησιμοποιούνται για την λειτουργία του. Ο όρος URI έχει υπερισχύει.

`scheme:[//[user:password@]host[:port]][/]path[?query][#fragment]`

Παράδειγμα

`http://example.com/hello.html?name=mike`

# Http Request (Παράδειγμα GET)

GET /hello.html?par1=val1&par2=val2 HTTP/1.1

Host: www.example.com

User-Agent: Mozilla/5.0 (Windows NT 6.1; rv:49.0) Gecko/20100101 Firefox/49.0

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,\*/\*;q=0.8

Accept-Language: en-US,en;q=0.5

Accept-Encoding: gzip, deflate

Connection: keep-alive

# Http Response (Παράδειγμα POST)

POST /SomeProgram HTTP/1.1

Host: www.example.com

User-Agent: Mozilla/5.0 (Windows NT 6.1; rv:49.0) Gecko/20100101 Firefox/49.0

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,\*/\*;q=0.8

Accept-Language: en-US,en;q=0.5

Accept-Encoding: gzip, deflate

Connection: keep-alive

Content-Type: application/x-www-form-urlencoded

Content-Length: 33

firstName=George&lastName=Artinos

# Δομή Αιτήσεων

- Γραμμή αίτησης (Request Line)
  - Μέθοδο Αίτησης (GET, POST, HEAD, PUT, DELETE, TRACE, OPTIONS)
  - Μονοπάτι URI
  - Έκδοση HTTP
- Τμήμα Επικεφαλίδων (Header)
  - <Όνομα Επικεφαλίδας> : <Τιμή>
- Σώμα Μηνύματος (Message Body)
  - Ακολουθεί μετά από ένα κενό.

# Ενδεικτικές Επικεφαλίδες 1

- Host (Παράδειγμα: Host : `www.google.com`)
  - Με την επικεφαλίδα αυτή στέλνεται το όνομα τομέα για τον οποίο δημιουργήθηκε η σύνδεση. Η σημασία της βρίσκεται στο γεγονός ότι η σύνδεση που δημιουργήθηκε στον server έπρεπε να χρησιμοποιήσει διεύθυνση IP. Υπάρχουν περιπτώσεις που η ίδια IP μπορεί να χρησιμοποιηθεί για να εξυπηρετήσει διαφορετικά ονόματα τομέα. Έτσι για να ερμηνευτεί σε ποιο όνομα τομέα απευθύνεται ο πελάτης θα πρέπει η πληροφορία να περιέχεται στις επικεφαλίδες.
- User-Agent
- User-Agent: `Mozilla/5.0 (Windows NT 6.1; rv:49.0) Gecko/20100101 Firefox/49.0`
  - Αυτή η ετικέτα παρέχει στο server διάφορες πληροφορίες για την έκδοση του φυλλομετρητή που χρησιμοποιεί ο πελάτης και για το λειτουργικό σύστημα. Την πληροφορία αυτή μπορεί να την χρησιμοποιήσει ο server για να καθορίσει αν ο πελάτης είναι κινητή συσκευή έτσι ώστε να επιστρέψει μια απάντηση που ταιριάζει καλύτερα.

# Ενδεικτικές Επικεφαλίδες 2

- Accept-Language
- Παράδειγμα
- Accept-Language: en-US,en;q=0.5
- Αυτή η επικεφαλίδα καθορίζει τις γλώσσες στις οποίες επιθυμεί ο πελάτης να δέχεται απαντήσεις. Ένας server που έχει την δυνατότητα να επιστέφει απαντήσεις σε διαφορετικές γλώσσες μπορεί να την χρησιμοποιήσει για να καθορίσει την γλώσσα. Μπορεί να παρέχονται διάφορες προτιμήσεις. Ορισμένες μπορεί να συνοδεύονται από μια έκφραση της μορφής  $q=0.5$ . Ο συντελεστής  $q$  (quality) μπορεί να συνοδεύεται από ένα αριθμό 0 έως 1 ο οποίος καθορίζει το βαθμό προτίμησης. Η προεπιλεγμένη τιμή είναι 1.
- Accept-Encoding
- Παράδειγμα
- Accept-Encoding: gzip, deflate
- Η επιλογή καθορίζει αν ο browser μπορεί να χειριστεί κάποια κωδικοποίηση συμπίεσης. Οι περισσότεροι φυλλομετρητές μπορούν να χειριστούν την κωδικοποίηση gzip. Το πλεονέκτημα είναι ότι η συμπιεσμένη απόκριση μπορεί να είναι κατά πολύ συντομότερη οπότε το μήνυμα μεταφέρεται πιο γρήγορα.

# Ενδεικτικές Επικεφαλίδες 3

- If-Modified-Since
- If-Modified-Since: Sat, 28 Oct 2016 16:32:15 GMT
  - Αν ένας browser έχει στην προσωρινή του μνήμη cache κάποιο έγγραφο που προσπαθεί να ανακτήσει πάλι μπορεί να συμπεριλάβει την επικεφαλίδα αυτή ώστε ο server να στείλει το έγγραφο αν έχει τροποποιηθεί από τότε. Σε περίπτωση που το έγγραφο δεν έχει τροποποιηθεί θα στείλει μια απόκριση της μορφής «304 Not Modified» και ο browser θα χρησιμοποιήσει το έγγραφο όπως είναι στην cache.
- Cookie
- Cookie: JSESSIONID=2340982308204823084024
  - Τα «μπισκότα» (cookies) είναι πληροφορίες της μορφής όνομα=τιμή οι οποίες στέλνονται σε αποκρίσεις των server στους πελάτες. Οι πληροφορίες αυτές μπορούν να επιστραφούν από τον πελάτη με επικεφαλίδες αυτές της μορφής. Με τον τρόπο αυτό ο server μπορεί να χρησιμοποιήσει πληροφορίες από προηγούμενες αιτήσεις κάτι που μπορεί να βελτιώσει την επεξεργασία της νέας αίτησης.
- Referer
- Referer : <http://ce.teiep.gr>
  - Όταν επισκεπτόμαστε μια ιστοσελίδα η οποία περιέχει άλλους υπερσυνδέσμους (hyperlinks) και επιλέξουμε κάποιον από αυτούς τότε στην νέα απαίτηση που θα σταλεί για να επιστραφεί το έγγραφο του υπερσυνδέσμου θα σταλεί με την επικεφαλίδα αυτή το URI της αρχική ιστοσελίδας στην οποία βρισκόμασταν. Κάτι τέτοιο είναι χρήσιμο στον server για να μαθαίνει, αν υπάρχει, από ποιον ιστότοπο κατευθύνονται απαιτήσεις στον server.

# HTTP Response (Παράδειγμα)

HTTP/1.1 200 OK

Date: Mon, 27 Jul 2009 12:28:53 GMT

Server: Apache/2.2.14 (Win32)

Last-Modified: Wed, 22 Jul 2009 19:15:56 GMT

Content-Length: 88

Content-Type: text/html

Connection: Closed

<html>

<body>

<h1>Hello, World!</h1>

</body>

</html>

# Δομή Απόκρισης

- Γραμμή Κατάστασης (Status Line)
  - Έκδοση HTTP
  - Κωδικός Απόκρισης
    - 1xx – Ενημέρωση – Η αίτηση παρελήφθη και συνεχίζεται η επεξεργασία
    - 2xx – Επιτυχία – Η αίτηση παρελήφθη, έγινε κατανοητή και αποδεκτή
    - 3xx – Ανακατεύθυνση – Επιπλέον ενέργειες πρέπει να συμβούν για να ολοκληρωθεί η αίτηση
    - 4xx – Λάθος πελάτη – Η αίτηση περιέχει σφάλματα ή δεν μπορεί να ικανοποιηθεί από το server
    - 5xx – Λάθος Εξυπηρετητή – Ο εξυπηρετητής απέτυχε να ικανοποιήσει μια απαίτηση που φαίνεται σωστή.
  - Επεξήγηση Κωδικού
- Επικεφαλίδες (Headers)
- Σώμα (Message Body)

# Παραδείγματα Κωδικών

- **200 OK**
  - Αυτός ο κωδικός στέλνεται όταν η απόκριση είναι επιτυχής. Το έγγραφο της απόκρισης βρίσκεται στο σώμα της απόκρισης.
- **301 Moved Permanently**
- **302 Found**
- **307 Temporary Redirect**
  - Οι κωδικοί 301,302,307 έχουν παραπλήσια σημασία. Ενημερώνουν τον αιτούντα ότι το έγγραφο που αναζητά είναι έγκυρο αλλά μπορεί να προσπελαστεί με κάποιο διαφορετικό URI. Το URI που πρέπει να χρησιμοποιήσει ο φυλλομετρητής επιστρέφεται με την επικεφαλίδα απόκρισης **Location**
- **404 Not Found**
  - Ο κωδικός αυτός επιστρέφεται όταν ο ζητούμενος πόρος (π.χ. έγγραφο ή ιστοσελίδα) δεν υπάρχει στον εξυπηρετητή.
- **500 Internal Server Error**
  - Αυτός ο κωδικός επιστρέφεται όταν η λειτουργία του server απέτυχε για κάποιο λόγο κατά την προσπάθεια χειρισμού της αίτησης (πχ. Λόγω έλλειψης, κάποιο πόρου, προγραμματιστικού σφάλματος, αδυναμία χειρισμού κάποιας εξαίρεσης κλπ).

# Ενδεικτικές Επικεφαλίδες 1

- **Cache-Control**
- **Cash-Control : max-age=3600, public**
  - Καθορίζει ότι η απόκριση μπορεί να κρατηθεί σε προσωρινή μνήμη (cache) από οποιοδήποτε σημείο στην αλυσίδα επικοινωνίας (πχ πύλη, διακομιστή μεσολάβησης κλπ). Η αποθηκευμένη απόκριση μπορεί να είναι διαθέσιμη για 3600 sec δηλαδή για μια ώρα.
- **Cash-Control : no-cache**
  - Απαγορεύεται η προσωρινή αποθήκευση. Μπορεί να χρησιμοποιείται από δυναμικές ιστοσελίδες που αναμένεται το περιεχόμενο της απάντησης να αλλάζει συχνά

# Ενδεικτικές Επικεφαλίδες 2

- Content-Type

- Καθορίζει τον τύπο MIME (Multipurpose Internet Mail Extensions) του εγγράφου ώστε ο φυλλομετρητής να το αναπαραστήσει σωστά
- Content-Type : text/html;charset=UTF-8

- Content-Length

- Ο Εξυπηρετητής μπορεί να προσδιορίσει το μέγεθος του περιεχομένου σε bytes. Είναι χρήσιμη πληροφορία για τον φυλλομετρητή γιατί μπορεί να προσδιορίζει την πρόοδο λήψης του εγγράφου και να την εμφανίζει στον χρήστη με μια μπάρα προόδου.

# Ενδεικτικές Επικεφαλίδες 3

- Last-Modified
  - Ο Εξυπηρετητής μπορεί να προσδιορίσει την Ημερομηνία τελευταίας ενημέρωσης του εγγράφου. Η πληροφορία μπορεί να χρησιμοποιηθεί από το φυλλομετρητή κατά την αποθήκευση του εγγράφου σε cache για υποβολή μελλοντικής αίτησης χρησιμοποιώντας την επικεφαλίδα If-Modified-Since.
- Location
  - Η επικεφαλίδα αυτή χρησιμοποιείται για ανακατεύθυνση σε συνδυασμό με τους κωδικούς απόκρισης 301 και 302. Όταν ο Εξυπηρετητής στείλει μια απόκριση ανακατεύθυνσης θα πρέπει στην επικεφαλίδα αυτή να αναφέρει την νέα θέση του εγγράφου. Παρατηρήστε ότι όταν προσπαθήσετε να προσπελάσετε την σελίδα [www.google.com](http://www.google.com) θα ανακατευθυνθείτε σε ένα νέο URL.
- Set-Cookie
  - Χρησιμοποιείται από το εξυπηρετητή για την αποστολή cookies στον φυλλομετρητή. Αυτές είναι πληροφορίες που μπορεί να διατηρήσει ο φυλλομετρητής και να τις επιστρέψει στον εξυπηρετητή σε μελλοντικές αιτήσεις. Ο εξυπηρετητής μπορεί να συσχετίσει με αυτόν τον τρόπο την νέα αίτηση με προηγούμενες.