

Εισαγωγή στο πρωτόκολλο HTTP

Το πρωτόκολλο HTTP (HyperText Transfer Protocol) είναι ένα πρωτόκολλο επικοινωνίας που ανήκει στο επίπεδο εφαρμογής και είναι το βασικό πρωτόκολλο στο οποίο στηρίζεται η λειτουργία του Παγκόσμιου Ιστού (World Wide Web).

Χρησιμοποιεί κυρίως το πρωτόκολλο μεταφοράς TCP για την δημιουργία συνδέσεων επιπέδου μεταφοράς.

Η τρέχουσα κύρια έκδοσή του είναι η 1.1 που αντικατέστησε την προηγούμενη έκδοση 1.0. Επίσης είναι σε εξέλιξη η προδιαγραφή και χρήση της επόμενης έκδοσης HTTP/2 η οποία ήδη υποστηρίζεται από πολλούς Εξυπηρετητές Ιστού (Web Servers) και Φυλλομετρητές (Web Browsers).

Η λειτουργία του πρωτοκόλλου βασίζεται σε ανταλλαγή μηνυμάτων μεταξύ μιας εφαρμογής πελάτη και μια εφαρμογής εξυπηρετητή και καθορίζει την δομή και την σημασιολογία αυτών των μηνυμάτων. Τα μηνύματα που στέλνονται από την εφαρμογή πελάτη λέγονται Απαιτήσεις HTTP (HTTP Requests) και τα μηνύματα που επιστρέφονται από τον Εξυπηρετητή λέγονται Αποκρίσεις HTTP (HTTP Responses)

Ένας εξυπηρετητής που υποστηρίζει το πρωτόκολλο HTTP μπορεί να χρησιμοποιεί την προεπιλεγμένη θύρα 80 όπου θα περιμένει για μηνύματα από κάποιο εφαρμογή πελάτη ή κάποια άλλη θύρα ή οποία θα πρέπει να αναφέρεται ρητά από την εφαρμογή πελάτη.

Το πρωτόκολλο HTTP είναι εξορισμού ένα πρωτόκολλο χωρίς διατήρηση κατάστασης (stateless protocol). Αυτό σημαίνει ότι κάθε νέα αίτηση που στέλνεται προς τον εξυπηρετητή είναι τελείως ανεξάρτητη από τυχόν προηγούμενες αιτήσεις και ο χειρισμός της δεν απαιτεί να διατηρούνται πληροφορίες από προηγούμενη ανταλλαγή μηνυμάτων είτε στον εξυπηρετητή είτε στον πελάτη.

Μιας μορφής διατήρησης κατάστασης μπορεί να επιτευχθεί με επιπρόσθετους μηχανισμούς όπως τα Cookies όπου στην ουσία οι εφαρμογές που τρέχουν στα δύο άκρα έχουν την δυνατότητα να στέλνουν επιπρόσθετες πληροφορίες που να τα βοηθούν να «θυμούνται» πληροφορίες από προηγούμενα μηνύματα.

Στις αρχικές υλοποιήσεις του HTTP (π.χ. 1.0) για να σταλεί μια νέα απαίτηση από τον πελάτη έπρεπε να δημιουργηθεί μια νέα σύνδεση TCP. Λόγω του ότι η δημιουργία μιας νέας TCP σύνδεσης είναι αρκετά χρονοβόρα (χρήση χειραψίας τριών βημάτων) οι τρέχουσες εκδόσεις του HTTP δίνουν την δυνατότητα να χρησιμοποιείται η ίδια σύνδεση TCP για την αποστολή και λήψη περισσότερων μηνυμάτων. Επίσης δίνεται και η δυνατότητα συνεχούς μεταφοράς (pipelining) αιτήσεων χωρίς την ανάγκη αναμονής για την ολοκλήρωση της προηγούμενης αίτησης (πολλοί browsers όπως ο firefox παρέχουν αυτή την δυνατότητα αλλά προεπιλεγμένα είναι απενεργοποιημένη).

Χρήση των URI (Universal Resource Identifier) ή URL (Universal Resource Locator)

Οι διαδικτυακές εφαρμογές βασίζονται στην αντίληψη ότι υπάρχουν στο διαδίκτυο διάφοροι πόροι (resources) που μπορούν να προσπελαστούν από απομακρυσμένα συστήματα. Για να προσπελαστούν αυτοί οι πόροι χρειάζεται ένας μηχανισμός προσδιορισμού τους.

Οι πόροι που είναι διαθέσιμοι στο διαδίκτυο μπορούν να προσδιοριστούν από ένα URI (Universal Resource Identifier) το οποίο είναι ένα σύνθετο όνομα που έχει την ακόλουθη δομή:

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

Τα πεδία που βρίσκονται ανάμεσα σε [] είναι προαιρετικά. Επίσης ένας χρήστης μιας εφαρμογής πελάτη μπορεί να αποφύγει το καθορισμό ενός τμήματος (όπως) το σχήμα (scheme) και να τον προσθέσει η εφαρμογή.

Η ακόλουθη είναι η περιγραφή των πεδίων:

scheme : Καθορίζει την προδιαγραφή (specification) που θα ερμηνεύσει το όνομα που ακολουθεί. Μπορεί να έχει τιμές: http, file, mailto, file, data κλπ

Οι δύο κάθετες απαιτούνται από ορισμένα σχήματα.

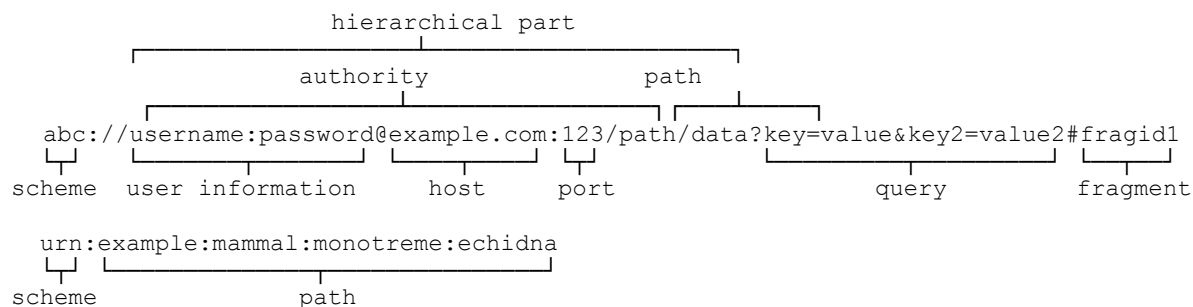
Τα επόμενα πεδία καθορίζουν την αρχή (authority) του πόρου. Αποτελείται (προαιρετικά) από πληροφορίες πιστοποίησης username και password, ένα όνομα τομέα (domain name) ή μια IP διεύθυνση και έναν αριθμό θύρας.

Το μονοπάτι (path) είναι το όνομα του πόρου στην αρχή. Μπορεί να έχει ιεραρχική δομή και να αποτελείται από τμήματα χωρισμένα με /.

Το ερώτημα (query) που ακολουθεί μετά από ένα ερωτηματικό αποτελείται από μη ιεραρχικές πληροφορίες που χρησιμοποιούνται στον προσδιορισμό του πόρου. Συνήθως υπάρχουν ζεύγη ονόματος τιμής (key=value) χωρισμένα με & ή ;.

Τέλος μπορεί να υπάρχει ο χαρακτήρας # ακολουθούμενος από τον προσδιορισμό ενός τμήματος (fragment) μέσα στον πόρο (π.χ. το συστατικό σε ένα HTML έγγραφο με κάποιο δεδομένο id).

Παραδείγματα URI



Σημείωση: Η προδιαγραφή των URI δεν είναι τμήμα του πρωτοκόλλου. Απλά το HTTP χρησιμοποιεί τμήματα ενός URI. Επίσης για ονόματα της προηγούμενης μορφής χρησιμοποιείται και όρος URL. Υπάρχουν κάποιες λεπτές διαφορές μεταξύ URI και URL. Τα URL θεωρούνται υποσύνολα των URI και είναι όσα URI καθορίζουν και την θέση (location) όπου μπορεί να βρεθεί κάποιος πόρος. Υπάρχουν ονόματα URI που δεν είναι URL.

Απαιτήσεις HTTP (HTTP Requests)

Παραδείγματα Http Αιτήσεων

```
GET /hello.html 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
```

```
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
```

Μια απαίτηση HTTP αποτελείται από τρία βασικά τμήματα.

Την γραμμή αίτησης (Request Line), το Τμήμα Επικεφαλίδων (Header) και το Σώμα Μηνύματος (Message Body).

Η γραμμή αίτησης αποτελείται από τρία τμήματα χωρισμένα με ένα κενό χαρακτήρα: (α) την μέθοδο της αίτησης, (β) το **όνομα** του πόρου αίτησης (URI) και (γ) την έκδοση του HTTP πρωτοκόλλου που ακολουθεί ή αίτηση. Στο τέλος της γραμμής αίτησης υπάρχουν οι χαρακτήρες CarriageReturn (\r) και LineFeed (\n).

Η μέθοδος μπορεί να είναι μια από τις : GET, POST, HEAD, PUT, DELETE, TRACE, OPTIONS

Στην τάξη θα μας απασχολήσουν οι μέθοδοι GET και POST.

Το όνομα του πόρου συνήθως είναι το απόλυτο μονοπάτι (path) που υπάρχει σε ένα URI.

Η έκδοση του HTTP μπορεί να είναι: HTTP/1.0, HTTP/1.1 κλπ

Η επικεφαλίδα αποτελείται από γραμμές με την δομή

ΌνομαΕπικεφαλίδας: Τιμή CRLN

Οι επικεφαλίδες χρησιμοποιούνται για να στέλνονται περισσότερες πληροφορίες στον εξυπηρετητή έτσι ώστε να διευκολυνθεί στην εξυπηρέτηση της αίτησης και να προετοιμάσει κάποια απόκριση που είναι πιο κατάλληλη για τον πελάτη που στέλνει την αίτηση.

Θα αναφέρουμε ενδεικτικά ορισμένες από αυτές:

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

Παράδειγμα

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.

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 συμπίεσμένη απόκριση μπορεί να είναι κατά πολύ συντομότερη όποτε το μήνυμα μεταφέρεται πιο γρήγορα.

If-Modified-Since

Παράδειγμα

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

Παράδειγμα

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

Παράδειγμα

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.

Αποκρίσεις HTTP (HTTP Responses)

Οι αποκρίσεις HTTP είναι τα μηνύματα που στέλνουν οι εξυπηρετητές σαν απάντηση σε HTTP Αιτήσεις.

Παράδειγμα μιας Απόκρισης

```
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>
```

Τα μηνύματα αυτά αποτελούνται από τρία βασικά τμήματα: (α) την γραμμή κατάστασης (Status Line), (β) μια σειρά γραμμών που αποτελούν το τμήμα των επικεφαλίδων απόκρισης (Response Headers) και (γ) το σώμα του μηνύματος που τις περισσότερες φορές είναι το HTML κείμενο που επιστρέφεται στον πελάτη.

Η γραμμή κατάστασης αποτελείται από την έκδοση HTTP στην οποία συντάσσεται η απόκριση, έναν κωδικό κατάστασης αποκρισης (status code) και μια σύντομη φράση που επεξηγεί τον κωδικό κατάστασης αποκρισης (reason phrase).

Οι κωδικοί κατάστασης είναι τριψήφιοι ακέραιοι και το πρώτο ψηφίο υποδηλώνει την κλάση στην οποία ανήκουν. Υποστηρίζονται οι ακόλουθες κλάσεις.

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.

Ενδεικτικά αναφέρουμε ορισμένους κωδικούς κατάστασης:

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).

Επικεφαλίδες HTTP αποκρίσεων

Οι ακόλουθες είναι ορισμένες επικεφαλίδες HTTP αποκρίσεων

Cache-Control

Παραδείγματα

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

Content-Type

Καθορίζει τον τύπο MIME (Multipurpose Internet Mail Extensions) του εγγράφου ώστε ο φυλλομετρητής να το αναπαραστήσει σωστά

Παράδειγμα

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.

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 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.