

ΜΑΘΗΜΑ 2^ο - Το πρωτόκολλο HTTP

Το πρωτόκολλο HTTP (Hyper Text Transfer Protocol) (12')

<https://youtu.be/xCZKXKXZkPU>

Το HTTP είναι ουσιαστικά ο βασικός μηχανισμός επικοινωνίας των client με τους server στο παγκόσμιο ιστό, στο διαδίκτυο. Ένας browser ένας client, ο υπολογιστής δηλαδή ενός απλού χρήστη επικοινωνεί με κάποιο server στέλνοντας μηνύματα HTTP και λαμβάνοντας πίσω πάλι με το πρωτόκολλο HTTP το περιεχόμενο που έχει ζητήσει, παραδείγματος χάρη μια ιστοσελίδα.

Τα κύρια χαρακτηριστικά του πρωτοκόλλου HTTP.

Είναι ένα πρωτόκολλο πελάτη εξυπηρετητή. Χαρακτηρίζεται από δύο διαφορετικούς τύπους μηνυμάτων. Ο πρώτος τύπος είναι μηνύματα - αιτήματα, όταν ζητάμε κάτι, ο browser ζητάει κάτι από το server. Ο δεύτερος τύπος είναι μηνύματα απόκρισης όταν ο server αποκρίνεται και στέλνει πίσω κάτι. Θα πρέπει εδώ να σημειώσουμε ότι το πρωτόκολλο αυτό ορίστηκε από τον Tim Berners-Lee τον εφευρέτη όπως θα λέγαμε του **Παγκόσμιου Ιστού (World Wide Web)** ήδη από το 1989. Σήμερα είναι ευρέως χρησιμοποιούμενη η έκδοση 1.1 που έχει κάποια πλεονεκτήματα όσον αφορά την αποστολή πολλαπλών πόρων μέσα από την ίδια σύνδεση, και υπάρχει ακόμα και έχει οριστεί η έκδοση 2 με σημαντικές διαφορές οι οποίες όμως δεν είναι τόσο ευρέως διαδεδομένη.

Παραδείγματα μηνυμάτων που στέλνονται και λαμβάνονται σε ένα browser με χρήση του πρωτοκόλλου αυτού.

Μήνυμα αίτημα.

Ένα κλασικό μήνυμα αίτημα είναι το μήνυμα GET, το μήνυμα αυτό έχει την εξής δομή: ξεκινάει με τη λέξη GET που είναι ο τύπος του μηνύματος ακολουθεί η διεύθυνση του πόρου τον οποίο ζητάμε και στη συνέχεια με κάποιο κενό ανάμεσα ακολουθεί ο τύπος του πρωτοκόλλου, ας πούμε HTTP/1.1 Ας δούμε ένα παράδειγμα.

```
GET /index.php?name=nikos HTTP/1.1 [CRLF]
```

```
Host: www.upatras.gr:80 [CRLF]
```

```
Accept: image/gif, image/jpeg [CRLF]
```

```
User-Agent: Mozilla/4.0 [CRLF]
```

```
Connection: Keep-Alive [CRLF]
```

```
[CRLF]
```

Το μήνυμα που βλέπουμε εδώ είναι τύπου GET, στέλνεται σε αυτήν εδώ τη διεύθυνση και αυτό είναι το πρωτόκολλο που χρησιμοποιούμε. Η εντολή GET τερματίζει με την ακολουθία δύο non printable characters δύο χαρακτήρων δηλαδή που δεν τυπώνονται, που δεν μπορώ να τους δω. Ο ένας είναι ο χαρακτήρας carriage return κι ο άλλος είναι ο χαρακτήρας line feed, είναι δυο χαρακτήρες του πίνακα κωδικοποίησης χαρακτήρων ascii οι οποίοι χρησιμοποιούνται εδώ ως τερματικοί χαρακτήρες της εντολής. Στη συνέχεια ακολουθεί η κεφαλίδα του μηνύματος. Κεφαλίδα ορίζεται σε μία ακολουθία από λέξεις κλειδιά όπως εδώ είναι host, ακολουθούμενη μία τιμή που θα πάρει αυτή η λέξη κλειδί, εδώ είναι για παράδειγμα η τιμή www.upatras.gr. Η κεφαλίδα Accept προσδιορίζει τι τύπου δεδομένα θα δεχτούμε, User-agent ποιο είναι το πρόγραμμα που στέλνει το μήνυμα. Εδώ είναι ο browser Mozilla έκδοση 4. Τα μηνύματα τύπου αιτήματος μπορεί να είναι διαφορετικών τύπων, αν και ο πιο συνηθισμένος τύπος είναι ο τύπος GET. Μπορώ να έχω μηνύματα τύπου Post που χρησιμοποιείται συχνά για να στείλω το περιεχόμενο μιας φόρμας, put που είναι ένα μήνυμα για να μεταφέρω ένα αρχείο ή άλλα δεδομένα και head που ζητάω μόνο την κεφαλίδα χωρίς να ζητάω κάποιο περιεχόμενο ή σε κάποιο πόρο.

Μήνυμα απόκρισης.

Ένα μήνυμα απόκρισης έχει συνήθως αυτήν εδώ τη δομή, αρχίζει το μήνυμα που είναι ξανά ένα μήνυμα χαρακτήρων κειμένου δηλαδή, αρχίζει με τον τύπο του πρωτοκόλλου που χρησιμοποιούμε, κενό και κωδικό κατάστασης. Αυτός κωδικός κατάστασης είναι πάρα πολύ σημαντικός γιατί με ενημερώνει αμέσως μόλις τον λάβω, αν το μήνυμά μου ικανοποιήθηκε ή όχι. Στη συνέχεια έχω μία φράση, μία περιγραφή της κωδικής κατάστασης, την κεφαλίδα του μηνύματος που πάλι έχει λέξη-κλειδί, τιμή ακολουθούμενη από carriage return-line feed, αυτούς τους δύο non printable characters μπορεί να είναι διάφορα στοιχεία της κεφαλίδας και ολοκληρώνεται η κεφαλίδα με το carriage return line feed, μία γραμμή δηλαδή περιέχει μόνο αυτούς τους δύο χαρακτήρες. Στη συνέχεια - και αυτό είναι πολύ σημαντικό σε ένα μήνυμα απόκρισης - έχω το κυρίως κείμενο του μηνύματος. Στο κυρίως κείμενο του μηνύματος μπορεί να είναι μία ιστοσελίδα που περιέχει html ή και άλλα στοιχεία, πρόσθετοι πόροι που χρειάζονται για μία ιστοσελίδα όπως είναι ένα αρχείο css μορφοποίησης του περιεχομένου, ένα αρχείο JavaScript που είναι το πρόγραμμα που θα τρέξει κατά τη λειτουργία της ιστοσελίδας.

Παράδειγμα

Το μήνυμα είναι HTTP κωδικός 1.x, κωδικός κατάστασης 200 που όπως θα δούμε σημαίνει ότι όλα πήγαν καλά, περιγράφεται με τη φράση Ok και στη συνέχεια έχω την κυρίως κεφαλίδα και εδώ κάτω είναι το κείμενο μηνύματος. Το συγκεκριμένο μήνυμα είναι τύπου text/plain, που σημαίνει ότι είναι απλό κείμενο δεν είναι html που είναι η πιο συνηθισμένη περίπτωση, απλά στάθηκαν κάποιοι χαρακτήρες, μία ακολουθία χαρακτήρων μία συμβολοσειρά.

HTTP/1.x 200 OK

Date: Sat, 08 Sep 2017 17:12:39 GMT

Server: Apache

Last-Modified: Sat, 08 Sep 2007 17:12:16 GMT

Accept-Ranges: bytes

Cache-Control: max-age=0

Expires: Sat, 08 Sep 2017 17:12:39 GMT

Vary: Accept-Encoding

Content-Encoding: gzip

Content-Length: 32

Content-Type: text/plain; charset=Windows-1253

Connection: keep-alive

Hello world!

Οι κωδικοί κατάστασης.

Οι δυνατοί κωδικοί κατάστασης αρχίζουν από χαρακτήρα 1 2 3 4 ή 5. Το 1 δείχνει ότι το μήνυμα είναι απλή πληροφορία, το 2 μας δείχνει ότι επιτυχώς ολοκληρώθηκε η επεξεργασία και στάλθηκε απόκριση. Όπως είδαμε λίγο πριν στο παράδειγμα 200 σημαίνει ότι όλα πήγαν καλά. Το 300, η ακολουθία 300 διάφορες περιπτώσεις σημαίνουν redirection δηλαδή το μήνυμα το πήρε ένας server αλλά το προώθησε σε έναν άλλο server που μπορεί να το εξυπηρετήσει. Τα μηνύματα τύπου 400 είναι σφάλματα που προέρχονται από το ίδιο το αίτημα. Ήταν ένα αίτημα που δεν μπορεί να ικανοποιηθεί. Κλασική περίπτωση τα μηνύματα 404 που λένε the requested url was not found on the server. Τέλος τα σφάλματα τύπου 500 είναι σφάλματα που οφείλονται στον ίδιο το server. Εδώ ο server παραδέχεται ότι δεν είναι σε θέση να ικανοποιήσει το συγκεκριμένο αίτημα.

Το πρωτόκολλο HTTP σε λειτουργία.

Ένας εύκολος τρόπος να το κάνουμε είναι να ξεκινήσουμε από command ή shell ανάλογα στον υπολογιστή που έχετε και να δώσετε την εντολή telnet. Να συνδεθείτε δε σε κάποιο υπολογιστή ας πούμε στο www.upatras.gr. Συνδεόμαστε στη θύρα 80 που είναι η θύρα για σύνδεση στον Παγκόσμιο ιστό στο web. Βλέπω ότι συνδέθηκα. Τώρα έχω λοιπόν τη δυνατότητα, telnet είναι να character based πρωτόκολλο δηλαδή μπορώ να στείλω εντολές χαρακτήρων. Το HTTP είναι ένα πρωτόκολλο χαρακτήρων άρα μπορώ να στείλω μία εντολή GET, ας πούμε ότι υπάρχει στο slash με ποιο πρωτόκολλο HTTP/1.1 και να δώσω τη δεύτερη όπως είπαμε εντολή κεφαλίδα HOST www.upatras.gr enter και enter ξανά.

Τι βλέπω εδώ; Πήρα μία απάντηση. Η απάντηση είναι αυτή: HTTP/1.1 κωδικός 301 moved permanently.

Γιατί; Γιατί αυτό που ζήτησα δεν είναι πια σε εκείνη τη διεύθυνση αλλά είναι σε αυτήν εδώ.

Ας προσπαθήσουμε να τρέξουμε ξανά το GET βάζοντας πια αυτό το /el στην ουσία είναι ένας φάκελος χρησιμοποιώντας το ίδιο πρωτόκολλο όπως 1.1 και δίνοντας πια ξανά την κεφαλίδα HOST www.upatras.gr enter και enter.

Τι βλέπω τώρα; Αυτό όλο είναι ένα μεγάλο κείμενο που στην ουσία είναι η ιστοσελίδα μου. Αυτό που βλέπετε εδώ είναι κώδικας html.

Το ίδιο ας προσπαθήσουμε να το κάνουμε μέσα από έναν browser. Πάμε σε ένα browser. Πατάω δεξί κλικ inspect και ξεκινάμε στην ουσία τα εργαλεία μελέτης της ιστοσελίδας, έχω τώρα δύο παράθυρα το δεξί είναι αυτό που θα δώ στο browser, αυτό δηλαδή που περιμένω και εδώ έχω αυτά τα λεγόμενα development Tools.

Να δούμε τι συμβαίνει και πώς θα φορτωθεί η ιστοσελίδα. Πατάω ξανά www.upatras.gr. Πατάω λοιπόν enter και τι βλέπω; Τώρα βλέπω ότι σιγά-σιγά κατεβαίνουν διάφορα κομμάτια της ιστοσελίδας, εδώ δεξιά βλέπω την ιστοσελίδα μου. Εδώ δε βλέπω σε χιλιοστά του δευτερολέπτου, όπως φορτώθηκε σιγά-σιγά η ιστοσελίδα.

Με περισσότερη λεπτομέρεια: Βλέπω ότι η πρώτη σελίδα είχε Status 301 σημαίνει ότι δεν τη βρήκε, όμως δεν μπήκε στον κόπο να μου τη ζητήσει να μου πει ξαναπροσπάθησε. Μόνος του προσπάθησε βρήκε την καινούργια διεύθυνση και το status είναι 200 Ok. Και από κει και κάτω φορτώθηκαν διάφορα άλλα αρχεία τα οποία χρειαζόντουσαν για αυτή τη σελίδα. Μία σύγχρονη ιστοσελίδα δεν είναι μόνο ένα αρχείο αλλά είναι πολλά διαδοχικά αρχεία.

Το πρωτόκολλο HTTPS.

Το HTTPS είναι το ίδιο το πρωτόκολλο HTTP όταν όμως μεταδίδεται σε συνεργασία με ένα άλλο πρωτόκολλο όπως είναι το Secure Sockets Layer (SSL) ή το Transport Layer Security (TLS) τα οποία μεριμνούν ώστε τα μηνύματα που στέλνω, τα μηνύματα HTTP να σταλούν κρυπτογραφημένα. Έτσι προστατεύουν τον αποστολέα του μηνύματος από πιθανή διαρροή του περιεχομένου τους σε ενδιάμεσους υπολογιστές. Μην ξεχνάμε ότι στο διαδίκτυο τα πακέτα, τα μηνύματα στέλνονται από υπολογιστή σε υπολογιστή. Άρα περνάνε από υπολογιστές στους οποίους δεν μπορούμε να εμπιστευτούμε. Το HTTPS δεν είναι διαφορετικό πρωτόκολλο. Είναι το ίδιο το HTTP απλώς τα μηνύματα είναι κωδικοποιημένα με τον τρόπο που περιγράφηκε, όμως αποστέλλονται κρυπτογραφημένα