

24. **Хипертекстов протокол в Интернет**

Основното е хипертекста

Най-мощното приложение на Internet е **WWW** (world wide web).

Като технология и идея е създадено от физика **Тим Бърнър Лий** през **1990** година с цел да улесни комуникацията на група учени.

web-технологията се базира на **хипертекста**.

Това е текст, който съдържа в себе си **информация как да бъде изобразен на екрана**.

Изобразяването става чрез специална програма, наречена **хипертекстов browser**.

Хипертекст, URL, хиперлинк

Хипертекстът се оформя като **съвкупност от страници**.

Всяка страница си има уникално **URL** – уникален адрес, който еднозначно указва местоположението на страницата в Internet.

Хиперлинк – под част от текста стои URL на друга страница.

Т.е хипертекстовите страници съдържат **препратки** към други хипертекстови страници.

Не се изисква тези страници да се намират на същия сървър.

HTTP - за комуникация между сървър и клиент

Web на английски означава **паяжина** – страниците, които са пръснати и са свързани като паяжина чрез хиперлинковите връзки.

Browser е клиентът, който изтегля и изобразява страниците.

Web server е сървърът, който съхранява страниците.

За комуникация между browser-и и сървъри е създаден протокола **HTTP** (hyper text transfer protocol).

URL

URL съдържа протокол, име на домейн и пътя на страницата върху диска на сървъра.

При осъществяване на връзка между browser-а и сървъра по URL-то първо се търси по името на сървъра, а после по пътя върху диска на сървъра.

Страницата физически се изтегля от сървъра, предава се на browser-а и той я изобразява.

Една страница се прехвърля в рамките на една HTTP-сесия.

HTTP – TCP/80

Протоколът HTTP се базира на TCP.

HTTP клиентът отваря сесия на произволен порт с номер по-голям от 1024.

HTTP сървърът слуша за заявки на порт 80.

Фази на HTTP протокола

При HTTP протокола имаме
подготвителна фаза – прави се заявка
за HTTP сесия към сървъра.

След това се прави HELLO към сървъра,
потвърждава се от сървъра и се
изпращат методи на HTTP.

Методите са GET, HEAD, POST, PUT,
TRACE, CONNECT.

Основен метод **GET**

Основният метод е **GET**.

Като аргумент му се подава адреса на страницата върху диска на сървъра.

Страниците могат да се кешират върху проху-сървъри, затова в метода GET има *if-условие*.

Хипертекстова страница

Всяка хипертекстова страница има **заглавие**, което съдържа описание на възрастта на страницата, къде се намира, датата на последната модификация и др.

Методът **HEAD** взима само заглавието на страницата.

Той позволява на browser-а бързо да провери **дали я има физически страницата** и кога е модифицирана последно (спестява се тегленето на тялото на страницата).

Методи PUT и POST

Методът PUT служи за прехвърляне на страница върху сървъра.

Той е свързан с обмен на два етапа – първо се дава адреса на страницата, след това се прехвърля самата страница.

Методът **POST** е аналогичен на метода PUT с тази разлика, че той добавя новите данни към съществуващ адрес.

Методът TRACE

Методът TRACE връща от сървъра получените данни по заявката.

Той се използва за тестване – да видим дали сървърът е получил това, което сме изпратили.

Версия 1.0 на HTTP

При първите версии на HTTP протокола за изпълнението на всеки метод се прави отделна HTTP сесия.

Тя отваря TCP съединение, праща нещо, след това затваря последователно съединението и сесията.

С други думи имаме 1:1 – една сесия, едно съединение.

Няколко сесии върху едно съединение

Ако за всяко изпълнение на GET се затваря съединението, а предстои четене на серия от страници това е много неефективно.

Правенето на съединение означава resolving на URL-адреса, за да се вземе IP адрес, с който да се направи TCP съединение. На самото TCP ниво се договаря трикратно съединението.

След това се договаря HTTP сесията.

Так се генерира се много служебен

Колко време да е отворена сесията

От друга страна не може всяко обръщение към една страница да отваря сесия към някой сървър и да я държи за неопределено време.

Изглежда нормално сесията да приключи с изтеглянето на страницата.

Един сървър отговаря за много страници. Ако е претоварен, докато изпълни заявката, ще мине много време. През цялото това време сесията стои отворена и browserът е блокиран.

Версия 1.1 на HTTP

При версия 1.1 на HTTP на едно TCP съединение отговарят няколко сесии (няколко команди).

Тези команди касаят различни хипертекстови страници, но достъпът до тях се прави с едно TCP съединение.

За целта се създава съобщителен канал. По него в пълен дуплекс в едната посока текат заявки, а в обратна посока – отговори.

Така не се работи по метода спри и чакай за всяка заявка.

Виртуални сървъри

Друга особеност е, че за един IP адрес може да има няколко имена на сървъри (така наречените **виртуални сървъри**).

За клиента те са различни сървъри, но реално зад тях стои един и същ IP адрес.

По-късно те могат да мигрират към други компютри, но вътре в тях URL-то ще се запази.

Основен недостатък е, че не знае кога са обновени хиперлинковете. Затова е важно използването на виртуални сървъри.

HTTP сървъри...

В момента се използват два HTTP сървъра:

`apache` - в UNIX/Linux и

`IIS` на Microsoft.

...и браузъри

По-известни HTTP браузъри по ред на поява:

Mosaic – 1993 г.;

Netscape Navigator и Netscape Communicator,
1994

Internet Explorer 1, 1995

Opera, 1996

Mozilla Navigator, 2002

Safari (за Apple Macintosh), 2003

Mozilla Firefox, 2004

Google Chrome, 2008