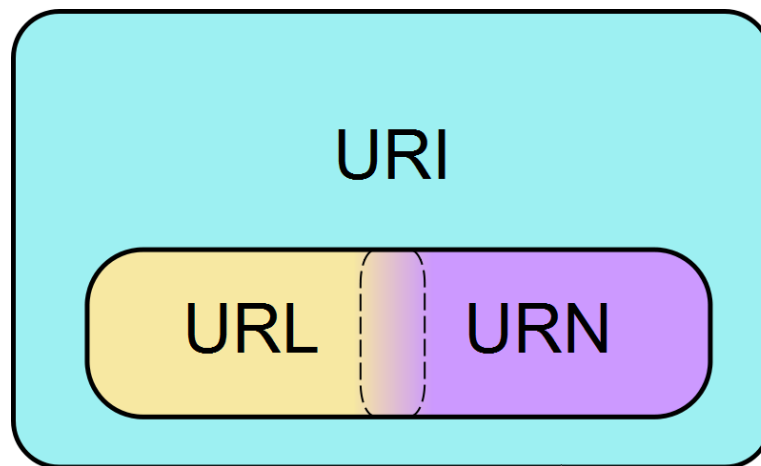


URL

- Uniform Resource Locator
- <http://www.zcu.cz/>



<http://en.wikipedia.org/wiki/URI>

- URI
 - Uniform Resource Identifier
 - Identifikuje soubor dat v Internetu
 - Bud' cestou k nim
 - Nebo jménem
 - Anebo obojím
- URN
 - Identifikace souboru dat
 - Např. ISBN knihy
- URL
 - Je URI, které definuje, jak získat URN
 - Např. kde, koupit knížku s daným ISBN
 - <http://www.zcu.cz/index.html>
 - Že k získání stránky index.html v doméně www.zcu.cz máme použít protokol http

- Syntaxe podle RFC
 - RFC je Request for Comments
 - Definiuje platné formáty
 - Zde jde konkrétně o RFC 3986
 - Zajímavost

schéma:hierarchická část[?parametry][#fragment]

- Části v hranatých závorkách nejsou povinné
- www.simpsons.com
 - hierarchická část – FQDN (DNS tvoří hierarchii)
 - pokud takto zadáme do prohlížeče, dosadí si výchozí protokol
 - může zde být přímo IP adresa
- <http://www.simpsons.com/>
- <mailto:homer@simpsons.com>
 - část před dvojtečkou zde určuje protokol
 - technicky však nemusí jít pouze o protokol, a proto se používá označení schéma
 - v KIV/ZPS se s tím ale nestkáme
 - `xml:lang="cs"`

- www.google.com/search?q=ZPS&ie=UTF-8
 - /search je cesta k souboru v doméně www.google.com
 - ? je oddělovač
 - q je parametr, který říká, co hledáme
 - ie je další parametr, který říká, v jaké znakové sadě byl dotaz zadán
 - parametry se oddělují ampersandem

- www.simpsons.com/cv/homer.html#beer
 - cv jako Curriculum Vitae
 - za prvním lomítkem následuje cesta k souboru s Homerovým životopisem v doméně www.simpsons.com
 - fragment oddělený mřížkou má v HTML funkci záložky
 - zde ukazuje na pozici ve stránce, kde se píše o Homerovo (kladném) vztahu k pivu

- URL však ještě může obsahovat i jméno a heslo, které jsou potřeba pro přístup k danému souboru dat
- Ale také může obsahovat i TCP port, pokud se používá jiný, než standardní

- Schéma URI definuje, jak interpretovat informace za dvojtečkou
- Pro URL používané HTTP protokolem platí tzv. generická syntaxe
 - Tj. schéma je http

`schéma:uživatel@heslo:doménaNeboIP:port
/cesta/soubor.přípona/?parametr#fragment`

- port
 - TCP port
 - Např. webový server poslouchá na portu 80, ale je možné ho nakonfigurovat, aby poslouchal i na jiném portu
- /cesta/soubor.přípona
 - Viz souborový systém
 - Kořenovým adresářem je doména, či IP adresa
 - Soubor lze vynechat
 - Existují jména souborů, které si server zkusí dosadit, pokud nejsou v URL obsažena
 - index.html
 - index.asp
 - default.htm
 - default.php
- `http://homer@koblížek:www.simpsons.com/private`
 - přístup do privátní oblasti webu chráněné uživatelským jménem a heslem
- `ftp://64.112.65.35:23`
 - přístup k FTP serveru daném IPv4 adresou
 - std. Port FTP je 21, zde je použit nestandardní port

- Schéma používané pro e-mailovou adresu
 - `mailto:jméno@doména?Subject=Předmět;Body=Tělo`
 - `mailto:homer@simpsons.com?Subject=Pivo`
- V HTML se u odkazů používá
 - Absolutní cesta – URL
 - Relativní cesta – URI
 - Viz souborový systém

Formátování textu

- Čistě textová informace jakéhokoliv textu je pouze posloupnost různých znaků, bez informace, jak je zobrazit
 - Tj. číslic, písmen a dalších symbolů
 - Tj. neobsahuje informace, kterou část textu zobrazit např. tučně, nebo kurzívou
- Dokumenty s formátovaným textem obsahují jak čistě textovou informaci, tak informaci, jak ji zobrazit (formátování)
 - Takové dokumenty mají buď binární strukturu
 - Naprosto mimo rozsah KIV/ZPS
 - Nebo mají textovou strukturu
 - Např. X/HTML
- X/HTML prokládá čistě textovou informaci tzv. formátovacími značkami
- Formátovací značka říká, jak se má text, který po ní následuje, zobrazit
 - Ukázkou `tučný` text.
 - `b` jako bold

HTML

- Hyper Text Markup Language
- Jazyk, ve kterém jsou vytvářeny webové prezentace

- Vznikl v letech 1989-1990
- Za vznikem stál Tim Berners-Lee v CERN
 - Evropská laboratoř pro fyziku částic
 - Potřeba sdílet data z experimentů
 - FTP nevyhovovalo
 - Ukaž a klikni

- 1994 – v NCSA Mar Andreessen stál za zrodem prohlížeče MOSAIC
 - National Center for Supercomputing Applications

- MOSAIC byl prvním grafickým prohlížečem
 - Volně přístupný
 - Později skončil jako Netscape
- 1995 – www komunikace se stává dominantní, založeno W3C konsorcium

- HTML je značkovací jazyk použitý při tvorbě dokumentů
 - Podmnožina SGML
 - Standard Generalized Markup Language

- HTTP je protokol, který umožňuje přenášet HTML dokumenty (a jimi odkazované soubory) z webového souboru na počítač klienta pomocí prohlížeče
 - Prohlížeč je anglicky browser a přednost českého překladu vyniká, když si pustíte audio nahrávku z webové stránky :-)

- Webový server je program spuštěný na počítači, který rozumí protokolu HTTP, hostuje HTML stránky a protokolem HTTP posílá soubory klientům, kteří si o ně řeknou tímto protokolem
 - Konkrétní soubor (stránka) je identifikovaná pomocí URL

- Dynamické HTML
 - Server-Side
 - Požadovaný HTML soubor je před odesláním klientovi upraven serverem
 - Např. doplnění aktuálního času serveru, výpis z databáze (STAG)
 - SSI (Server Side Include), PHP (Personal HomePage), ASP (Active Server Pages), JSP (Java Server Pages) a další

 - Client-Side
 - Požadovaný HTML soubor je upraven v prohlížeči, před jeho zobrazením uživateli
 - Např. zobrazení aktuálního času na počítači klienta, grafické efekty, menu
 - JS (JavaScript), VBS (Visual Basic Script)

 - Server-Side i Client-Side lze kombinovat

- HTML používá značky, tzv. tagy
 - Tagy se vnořují
 - Tagy se uvozují pomocí <název tagu>
 - Tagy jsou buď párové <p>paragraph</p>
 - Anebo nepárové

 - Tag může mít atributy
 - odkaz

- Existuje několik různých verzí HTML
 - Každá z verzí zná různou množinu tagů
 - S vyšší verzí se množina rozšiřuje počet podporovaných tagů
 - Některé byly v novějších označeny za neplatné

X/HTML

- **eXtensible Hyper Text Markup Language**
 - na rozdíl od HTML vycházejícího přímo z SGML, X/HTML vychází z XML (které vychází z SGML)
- => Uzavírají se i nepárové tagy
 - `
`
- Důvodem vzniku byla postupně větší a větší složitost syntaxe HTML
 - X/HTML umožňuje snazší zpracování HTML stránek např. na mobilních zařízeních, které nemají až tak velký výpočetní výkon jako např. klasické stolní PC
- Na rozdíl od HTML se klade mnohem větší důraz na oddělení specifikace formátování a vlastní informace
 - Řada atributů, které jsou povoleny v HTML, v X/HTML nejsou povoleny
 - Např. `target` pro `a` v X/HTML Strict
- Na začátku HTML souboru se pouze očekává značka s tzv. typem dokumentu, v X/HTML tam musí být povinně

- Tagy jsou case-sensitive
 - Rozlišují se malá a velká písmena
- Tagy se musejí uzavírat ve správném pořadí
 - Špatně: `<i>text</i>`
 - Dobře: `<i>text</i>`
 - Jako závorky v matematice
- X/HTML zavedlo namespaces
 - Mimo rozsah KIV/ZPS, pro zajímavost
 - V nějakém odkazovaném dokumentu se mohou používat identifikátory
 - Mimo dokument se identifikátoru musí předřadit prefix dokumentu
 - V X/HTML je možné dokument vřadit do namespace identifikátoru

```
<h:html xmlns:z="http://www.kiv.zcu.cz/zps"
        xmlns:h="http://www.w3.org/1999/xhtml">
  <h:head>
    <h:title>Seznam přednášek</h:title>
  </h:head>
  <h:body>
    <z:lecture>
      <z:seq>6</z:seq>
      <h:p><z:title>X/HTML</z:title></h:p>
    </z:lecture>
  </h:body>
</h:html>
```

- V X/HTML existuje tag title, který už má nějaký význam, ale díky namespaces ho můžeme použít i pro naše účely – např. je pak lehčí napsat program, který projde dokument a zorientuje se v něm, než kdyby se měl orientovat pouze v textu HTML

- *Alespoň tohle byla pravda. Mrakoplaš uměl žadonit o smilování v devatenácti jazycích a jen obyčejně ječet v dalších čtyřiačtyřiceti. A To je velmi důležité. Nezkušený cestovatel si může myslet, že „Aúúú“ je univerzální, ale v be-trobštině to znamená „nanejvýš příjemné“ a v jakazačistánštině, podle místního dialektu, „rád bych snědl vaši nohu“, „vaše žena je velký hroch“ a koukněte, pan Červený kocour“. Jeden kmen získal děsivou pověst kvůli své strašlivé krutosti jen proto, že když zajatci křičí, příslušníci onoho malého kmene slyší: „Rychle! Další dávku vroucího oleje!“*
 - Terry Pratchett – Zajímavé časy
 - Jazyk je namespace

CSS

- Cascade Style Sheet
- X/HTML dokument obsahuje vlastní informaci ke sdělení, ale její grafická podoba je zapsána právě v externím souboru ve formátu CSS
 - Ideálně – v praxi se to stejně míchá do X/HTML
 - Jak bylo zvykem v původním HTML
- Mám-li jeden web a všechny jeho stránky používají jeden jediný CSS soubor, pak mohu změnit vzhled celého webu změnou jednoho jediného CSS souboru
 - Např. web pro ženy může tématicky s ročním obdobím měnit barvy
 - Pokud by ale formátování textu bylo natvrdo napsané v každém HTML souboru, tak by kvůli změně bylo nutné přepsat všechny HTML soubory

- Jako namespace obsahuje použitelné tagy, CSS obsahuje popis, jak budou zobrazovány – jejich grafickou podobu
 - Pokud není CSS v dokumentu specifikováno, prohlížeč použije vlastní s výchozími hodnotami
 - Také lze určit, kde budou elementy stránky zobrazeny
- Buď se nadefinuje styl se jménem existujícího tagu, a je použit automaticky
 - CSS: `p { height: 10pt; }`
 - X/HTML: `<p>text raz dva tři</p>`
- Anebo se nadefinuje styl s novým jménem a je pak přiřazen
 - CSS: `.highlight { color: white; background: yellow; }`
 - X/HTML: `<p>zvýrazněnýtext</p>`
- CSS je textový soubor, kde jsou jednotlivé styly uzavírány pomocí složených závorek a jejich atributy oddělovány středníkem
- Jméno stylu je uvozeno prefixem
 - Nic – selector
 - . – class selector
 - # – id selector
 - http://www.w3schools.com/css/css_syntax.asp

Struktura X/HTML dokumentu

```
<?xml version="1.0" encoding="UTF-8" ?>  
<!DOCTYPE...>
```

```
<html>
```

```
  <head>
```

Vnořené tagy, které popisují stránku,
nebo se k ní nějak vztahují

```
  </head>
```

```
<body>
```

Vlastní informace, kterou stránka nese

```
</body>
```

```
</html>
```

- head – hlavička dokumentu
 - title – nadpis stránky
 - `<title>Nadpis stránky</title>`
 - popis stránky
 - `<meta name="description" content="Popis stránky" />`
 - klíčová slova
 - `<meta name="keywords" content="klíčová, slova" />`
 - Dnešní vyhledávače už tento údaj ignorují, protože býval často zneužíván

- Styl – CSS

- Bud' odkazem na externí CSS soubor

- ```
<link rel="stylesheet"
 type="text/css"
 href="/common/norm.css" />
```

- Nebo lze styl definovat přímo v hlavičce

- ```
<style type="text/css">
  .menuitem {
    font-family: sans-serif;
    font-weight: bold;
  }
</style>
```

- Kódování

- Je třeba nějak určit, ke kterým sekvencím bytů patří které znaky

- Není-li uvedeno, prohlížeč si nějakou dosadí

- Budeme-li psát cesky, tak je to jedno, ale kdo to ma potom cist?!
 - ```
<meta http-equiv="Content-Type"
 content="text/html;
 charset=utf-8" />
```

- Stránka je kódovaná v utf-8

- body

- tělo dokumentu, jednotlivé elementy stránky

- odstavce

- odrážky

- tabulky

- obrázky

- odkazy

- atd.

## Kódování znaků

- Existuje několik možností, jak kódovat znaky s diakritikou
  - Tzv. kódové stránky
- Střední Evropa
  - Windows-1250 – používaná ve Windows
  - iso-8859-2 – známá jako Latin-2, UNIX/Linux
  - x-mac-ce – Apple
  - ibm852 – DOS
  - KOI-8 – ČSN
  - Bratři Kameničtí – svépomoc

- Zobrazeno ve Windows s Windows-1250

```
75 6B E1 7A 6B 6F 76 FD|20 74 65 78 74 20 0D 0A | ukázkový text
E1 E8 EF E9 EC ED F3 F8|9A 9D FA F9 FD 9E | áčďěěíóřšťúůž
```

- Takto vypadá text ve Windows-1250

```
75 6B E1 7A 6B 6F 76 FD|20 74 65 78 74 20 0D 0A | ukázkový text
61 E8 EF E9 EC ED F3 F8|B9 BB FA F9 FD BE | ačďěěíóřř»úůžř
```

- Takto vypadá text v iso-8859-2

- Některé znaky mají shodné hodnoty svých bytů, některé však ne – v ukázce š, ť a ž

- Obě uvedené kódování používali na jeden znak jeden byte => maximálně 256 různých znaků
  - Ale co když budeme chtít v jednom dokumentu používat českou diakritiku řecké znaky, matematické symboly, atd.?

- Řešením je kódování znaků Unicode
- Jeden znak se kóduje na dvou bytech
  - Označuje se jako UTF-16, popř. UCS-2
  - 65 536 možných kombinací
    - Ale stejně vzniklo i UTF-32/UCS-4
- „Základním“ znakům Western Code Page zůstaly jejich „původní“ hodnoty
  - Pouze je před nimi byte s hodnotou nula (2 byty)
    - Big-endian
      - Byte vyššího řádu je uložen jako první
    - Některé znaky s diakritikou mají hodnoty větší než 256 => jejich první byte není nulový

```
00 75 00 6B 00 E1 00 7A|00 6B 00 6F 00 76 00 FD | u k á z k o u ý
00 20 00 74 00 65 00 78|00 74 00 20 00 0D 00 0A | t e x t
00 61 01 0D 01 0F 00 E9|01 1B 00 ED 00 F3 01 59 | a š é í ó ŷ
01 61 01 65 00 FA 01 6F|00 FD 01 7E | a e ú o ý ~
```

- Některé znaky jsou (vpravo) zobrazeny nesprávně, protože všechny byty byly interpretovány jako Windows-1250

- Posloupnost znaků tvoří tzv. řetězec
- Řetězec se v mnoha případech ukončuje znakem s hodnotou nula
- Obrovské množství programů předpokládá 8-bitové kódování znaků a nuly v 16-bitovém kódování by automaticky braly jako konec řetězce
  - Např. předchozí ukázkou v UTF-16 by vůbec nezobrazily, protože začíná nulou

- Z tohoto důvodu se používá kódování UTF-8
  - 16-bitový znak se kóduje na 1-3 bytech, podle své hodnoty
  - Např. anglický text v UTF-8 a 8-bitovém kódování je identický a tak s ním mohou pracovat právě i staré programy

```

75 6B C3 A1 7A 6B 6F 76 | C3 BD 20 74 65 78 74 20 | ukã~zkouã~ text
0D 0A 61 C4 8D C4 8F C3 | A9 C4 9B C3 AD C3 B3 C5 | aňřãžň@ň>ň-ňřč
99 C5 A1 C5 A5 C3 BA C5 | AF C3 BD C5 BE | "č~čřňščžňřčř

```

- Kódování UCS-2 do UTF-8
  - Pro zajímavost

znak Unicode	max. významových bitů	kódování UTF-8
0000–007F	7	0xxx xxxx
0080–07FF	11	110x xxxx 10xx xxxx
0800–FFFF	16	1110 xxxx 10xx xxxx 10xx xxxx

	Unicode				UTF-8			
B	0000	0000	0100	0010	4200	0100	0010	42
é	0000	0000	1110	1001	E900	1100	0011 1010 1001	A9C3
ď	0000	0001	0000	1111	0F01	1100	0100 1000 1111	8FC4
a	0000	0000	0110	0001	6100	0110	0001	61

- Tento zápis používá konvenci little-endian
  - Byty nižšího řádu jsou uloženy jako první
- Např. v MS Wordu můžete zkusit napsat následující čísla
  - Před každým číslem stisknout Alt a po napsání pustit
  - 66, 0233, 271, 97 (v tabulce jsou šestnáctkově)



- Na začátku souboru v Unicode bývá magické číslo, podle kterého lze poznat, zda soubor používá big, nebo little-endian a konkrétní varianty
  - Tzv. BOM – Byte Order Mask

<b>BOM</b>	<b>Endian</b>
00 00 FE FF	UTF-32, big-endian
FF FE 00 00	UTF-32, little-endian
FE FF	UTF-16, big-endian
FF FE	UTF-16, little-endian
EF BB BF	UTF-8

- Anebo to tam prostě není
  - A je to;-)
    - Hledej, šmudlo:-)
- Pokud váš editor zobrazuje „nějaké divné“ znaky, možná editujete soubor v jiné znakové sadě, než kterou daný program předpokládá/zvládá
  - Např. Windows Notepad a HTML soubor v iso-8859-2 vytvořený editorem vim v Linuxu

## Vybrané značky X/HTML

- Ty, které se používají v těle dokumentu
    - `<body> ... </body>`
  - `<p> ... </p>`
  - `<ol> ... </ol>` `<ul> ... </ul>` `<li> ... </li>`
  - `<table>` `<tr>` `<td> .. </td>` `</tr>` `</table>`
  - ``
  - `<a href="URI"> ... </a>` `<a name="" />`
  - `<span> ... </span>`
  - `<br />`
  - `<b> ... </b>` `<i> ... </i>` `<u> ... </u>`
  - `<sub> ... </sub>` `<sup> ... </sup>`
  - `&nbsp;`
  - `<div> ... </div>`
- 
- Detaily viz cvičení a samostatně vypracovaná semestrální práce

## Bezestavový protokol

- Pokaždé, když se serveru pošle příkaz, obsahuje úplný popis toho, co se má udělat
  - neřekne se, pokračuj, kde jsi skončil minule
- Line ([x, y] , [w, z])
  - Nakresli čáru z bodu [x, y] do bodu [w, z]
- MoveTo([x, y])
  - Přesuň pero do bodu [x, y]
- LineTo([x, y])
  - Nakresli čáru z aktuálního bodu do bodu [x, y]
- Jak nakreslit čtverec?
  - Bezestavový protokol (16 čísel k přenosu)
    - Line([0, 0], [0,1])
    - Line([0, 1], [1,1])
    - Line([1, 1], [1,0])
    - Line([1, 0], [0,0])
  - Stavový protokol (10 čísel k přenosu)
    - MoveTo([0, 0])
    - LineTo([0, 1])
    - LineTo([1, 1])
    - LineTo([1, 0])
    - LineTo([0, 0])
- Stavový protokol umožňuje přenášet méně parametrů (tj. menší objem se přenesou rychleji)
- Pokud se ale spojení neočekávaně přeruší, bezestavovému protokolu to nebude vadit
  - Stavovému se ztratí informace o předchozím stavu, kterou potřebuje

# HTTP

- Hyper Text Transfer Protocol
- Bezstavový protokol
- Pro přenos dat využívá spojovaný protokol TCP
  - Tj. HTTP nad TCP nad IP
  - Přenosem dat se primárně rozumí HTML stránky, ale je možné přenášet i binární soubory
    - Např. obrázky, video, zvuk, archiv dokumentů...
- Verze 0.9
  - První verze
  - Umožňovala jenom stáhnout HTML stránku
  - Už se nepoužívá
    - Deprecated
- Verze 1.0
  - Umožňuje odesílat data na server
    - Metoda POST
  - Běžný uživatel nepozná rozdíl mezi verzemi 1.0 a 1.1
  - Pro každý požadavek se naváže samostatné TCP spojení
- Verze 1.1
  - Umožňuje odesílat několik požadavků v jednom TCP spojení (nižší režie)
  - Požadavky se nemusejí vyřizovat po jednom, ale je možné jich poslat několik „naráz“ (další urychlení)
  - Servery i klienti stále ještě nabízejí možnost použít verzi 1.0 pro případ „obskurního“ serveru, či klienta

- Verze 1.2
  - Hlavním tahákem mělo být Extension Mechanism for HTTP
    - Zkracováno na PEP jako Protocol Extension Protocol
  - Nakonec nebylo publikováno jako HTTP 1.2
- Základním požadavkem je načtení stránky
  - GET
  - Požadavek

```
GET /cz/people/homer.htm HTTP/1.1
Host: www.simpsons.com
User-agent: Mozilla (IE compatible)
```

- Odpověď, je-li vše v pořádku

```
HTTP/1.1 200 OK
Date: Fri, 10 Oct 2008 16:08:33 GMT
Server: Apache/2.0.55 (Unix) mod_ssl/2.0.55
OpenSSL/0.9.7k PHP/5.2.2 mod_jk/1.2.15 DAV/2SVN/1.2.3
X-Powered-By: PHP/5.2.2
Content-Length: 418
Content-Type: text/html
```

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE html PUBLIC
 "-//W3C//DTD XHTML 1.0 Strict//EN"
 "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
 xml:lang="cs" lang="cs">
```

```
<head>
<meta http-equiv="Content-Type"
 content="text/html; charset=utf-8" />
<title>Simpsonovi - Homer J.</title>
</head>
```

```
<body>
 <p>Homer vás zdraví!</p>
</body></html>
```

- Odpověď, pokud stránka neexistuje

```
HTTP/1.1 404 Not Found
Date: Fri, 10 Oct 2008 16:14:21 GMT
Server: Apache/2.0.55 (Unix) mod_ssl/2.0.55
OpenSSL/0.9.7k PHP/5.2.2 mod_jk/1.2.15 DAV/2
SVN/1.2.3
Content-Length: 368
Content-Type: text/html; charset=iso-8859-1
```

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>404 Not Found</title>
</head><body>
<h1>Not Found</h1>
<p>The requested URL /cz/people/homer.htm was not
found on this server.</p>
<hr>
<address>Apache/2.0.55 (Unix) mod_ssl/2.0.55
OpenSSL/0.9.7k PHP/5.2.2 mod_jk/1.2.15 DAV/2
SVN/1.2.3 Server at www.simpsons.com Port
80</address>
</body></html>
```

- Jsou možné ještě i jiné, odpovědi
  - Mimo rámeček KIV/ZPS
  - Např. stránka byla přesunuta
- Dalším důležitým požadavkem je odeslání dat formuláře webové stránky
  - Např. stisknutím tlačítka vyhledat
  - POST

```
POST /search.php HTTP/1.1
Host: www.simpsons.com
Content-Length: 13
Content-Type: application/x-www-form-urlencoded
```

Query=Homer+Simpson

- Další požadavky jsou
  - HEAD
    - Vrací jenom hlavičku
    - Šetří se tak přenosová kapacita, protože není třeba přenášet i vlastní dokument
  - PUT
    - Nahraje soubor na server
  - DELETE
    - Smaže data zadaná URI
  - TRACE
    - Server vrátí zpět požadavek, takže klient vidí, zda ho někdo po cestě nezměnil
  - OPTIONS
    - Server vrací, jaké nabízí možnosti k získání dané URI
  - CONNECT
    - Vytvoří tunel, který lze šifrovat
    - Prochází-li spojení přes nějaký proxy server
      - Např. vyrovnávací paměť
- HTTP server v odpovědi vždy vrací číselný kód
  - 1xx – požadavek byl přijat a zpracovává se
  - 2xx – požadavek byl úspěšně proveden
  - 3xx – přesměrováno, požadovaná data jsou jinde
  - 4xx – chyba na straně klienta
  - 5xx – chyba na straně serveru
  
  - xx jsou specifické kódy upřesňující odpověď

- Cookies
  - HTTP protokol je sám o sobě bezstavový
  - Cookie je textová informace, která se uchovává na počítači klienta a server si ji může vyžádat, či uložit
  - Prostředek, jak uložit stav návštěvy webu
    - Nepřečtené příspěvky na fóru
    - Uživatelsky nastavitelný vzhled webu
    - Stav nákupního košíku
    - Apod.
  
- Referer
  - V hlavičce HTTP požadavku se také udává předchozí stránka, kterou prohlížeč navštívil
  - Do standardu se skutečně dostalo Referer namísto anglicky-správně Referrer

```
GET /shop/checkout.html HTTP/1.1
Host: www.simpsons.com
```

```
Referer: www.simpsons.com/shop/cart.php
Cookie: item1=DuffKeg
Cookie: item2=Skateboard
Cookie: creditcard=0123-4567-89AB-CDEF
Accept: */*
```

- Bez referer a cookies mohou některé stránky přestat fungovat
  - Alias neumožní jinak na nich obvyklou činnost
  
- WebDAV a OpenCMS
  - Web-based Distributed Authoring and Versioning
  - Content Management System
  - Další rozšíření HTTP o příkazy, které umožňují několika uživatelům editovat jeden web současně



## Zobrazení HTML stránky

- HTTP/1.1
- Uživatel zadá URL stránky
- Prohlížeč z ní extrahuje doménové jméno a pomocí DNS ho převede na IP adresu
- Prohlížeč naváže TCP spojení se serverem na zjištěné IP adrese
  - Do hlavičky napíše doménové jméno, protože na jedné IP adrese může být hostováno několik webů
- Slušný prohlížeč si nejprve požadavkem GET z www serveru stáhne soubor robots.txt
  - Bývá v kořenovém adresáři domény
  - Říká, které stránky jsou přístupné a které ne
    - Např. aby Google a jiné vyhledávače zbytečně neindexovaly soubory, kde to nemá smysl
  - Např. software na tzv. mirroring webů ho pochopitelně nabízí ignorovat
    - Protože je v robots.txt obvykle napsáno, ať nestahuje

```
User-agent: *
Allow: *
Disallow: /search
```

```
User-agent: WebCopier
Disallow: *
```

- Pro zajímavost, mimo rozsah KIV/ZPS
- Prohlížeč si požadavkem GET řekne o požadovanou URL

- Jestliže server dostal žádost GET s URL bez konkrétního souboru, pak
  - Server se pokusí v adresáři daném URL najít soubor s některým předdefinovaným jménem
    - index.html
    - index.htm
    - default.asp
    - default.php
  - jména a priority souborů jsou konfigurovatelné
  - Pokud takový soubor našel, vrátí ho
  - Jestliže však takový soubor nenajde, pak
    - buď vytvoří HTML stránku, která bude výpisem adresáře
    - anebo odešle chybové hlášení jako HTML stránku
      - Např. 403 Forbidden (odmítnutí přístupu/výpisu adresáře)
- Prohlížeč projde přijatý HTML kód (zparsuje ho) a odešle požadavky na všechny vnořené URI, mimo tagu `<a href=““></a>`
  - CSS a img
  - Všechny relativní URI jdou stále tím samým spojením, protože patří do stejné domény jako zadaná URL
  - Všechny URL vedoucí mimo doménu jsou získána jako HTML kód původního požadovaného URL
- Prohlížeč zobrazí, co už má načteno
- Další části stránky zobrazí tak, jak je postupně obdrží