



# Základy informatiky

HTML, tvorba WWW stránek

Daniela Szturcová

Část převzata z přednášky P. Děrgela



# Obsah přednášky

- Historie, technologie
- Metaznaky, hlavička
- Kódování
- Publikování stránek
- Formuláře, formulářové prvky
- Dnešní a budoucí web
- CMS



# Prvky technologie WWW

## HTTP - Hypertext Transfer Protocol

- protokol pro přenos stránek mezi webovým serverem a prohlížečem
- jednoduchý aplikační protokol vystavený nad protokolem TCP
- několik verzí - HTTP 0.9, HTTP 1.1, HTTP 3.0 až HTTP 4.01
- dnes i https, což je stejný protokol jako http, pouze přenášená data jsou kryptována (zabezpečena)

## URL - Uniform Resource Locator

- URL je adresa, jednoznačně identifikuje zdroj v rámci Internetu
- URL se používá v různých službách: WWW, FTP, e-mail, telnet,

## HTML - Hypertext Markup Language

- značkovací jazyk založený na SGML pro popis struktury web stránek
- nezávislý na platformě



# Historie a vývoj HTML

## HTML 0.9 - 1991

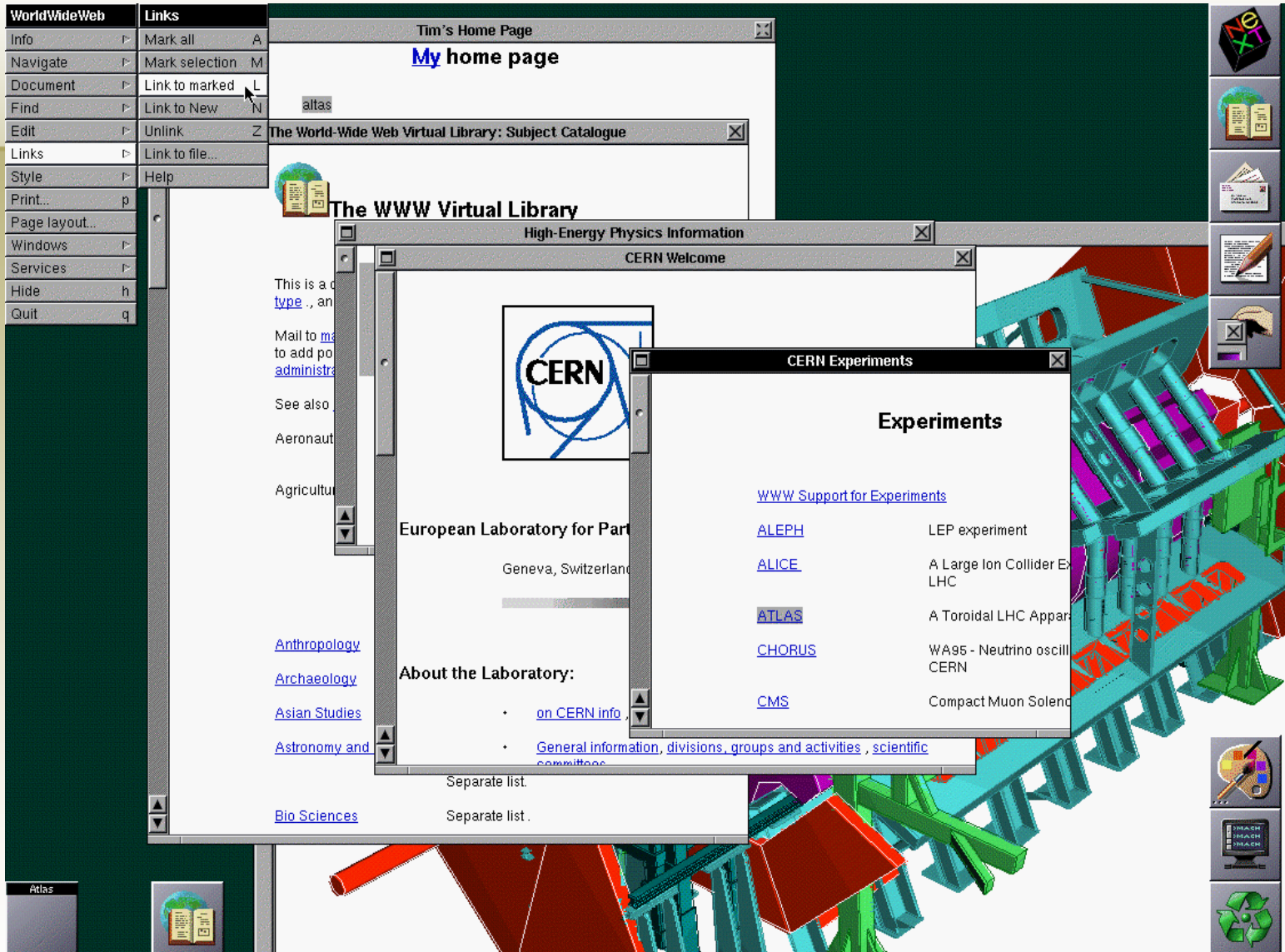
- Tim Berners-Lee - členění textu do několika logických úrovní,
- použití několika druhů zvýraznění textu, lze zařadit odkazy a obrázky.

## HTML 2.0

- první formální specifikace, návrh standardu (HTML 2.0), plně vyhovuje normě SGML (ISO 8879 z roku 1986),
- základní formátování a strukturování dokumentu, obrázky, formuláře.

## HTML 3.0 - 1995 návrh standardu

- Dave Raggett z laboratoří Hewlett-Packard formalizoval HTML+,
- vytvořil jeho deklaraci DTD (Document Type Declaration) v jazyku SGML (Standard Generalized Markup Language).





# Historie a vývoj HTML

## HTML 3.2 - 1996

- výrobci prohlížečů přidávají nestandardní rozšíření HTML,
- konsorcium W3C proto vybírá společnou podmnožinu rozšíření a schvaluje HTML 3.2,
- lepší formátování dokumentu, podpora tabulek, Java apletů.

## HTML 4.0 - 1997

- podpora kaskádových stylů, frames (rámy),
- skripty vložené do stránky, multimediální objekty,
- rozšířené možnosti tabulek, formulářů,
- vícejazyčná podpora,
- podporu HTML 4.0 mají prohlížeče (IE6, NN7, O7, M1.6).





# Historie a vývoj HTML

## HTML 4.01 - 1999

- opravy drobných chyb ve specifikaci HTML 4.0
- HTML 4.01 je definován v podobě tří DTD (Document Type Definition) v jazyce SGML:
  - Strict DTD obsahuje elementy, které nejsou ve standardu 4.0 označeny jako nedoporučované,
  - Transitional DTD obsahuje i nedoporučované elementy, včetně starších rozšířených elementů,
  - Frameset DTD je téměř totožná s Transitional, podporuje použití frames.



# Verze XHTML

- X ve zkratce XHTML vzniklo ze slova eXtensible, což by mělo znamenat rozšíření. Skutečnost je opačná, počet použitelných prvků se snížil.
- Je produktem přetvoření HTML 4 a XML 1.
- Dnes se používá XHTML ve třech verzích:
  - **XHTML 1.0 přechodové (transitional)**
  - **XHTML 1.0 striktní (strict)**
  - **XHTML 1.1**
- XHTML 1.0 strict je oproti trans pracnějšší, XHTML 1.1 je okleštěno proti HTML o nestandardní prvky.





# Historie a vývoj HTML

## HTML 5

V roce 2004 založena pracovní skupina WHATWG (The Web Hypertext Application Technology Working Group).

Je XHTML 2 je slepá ulička?

2007 se propojilo W3C a WHATWG → HTML5

Již se používá, zatím se mění. Část je již stabilní.

Dokončení se neočekává před rokem 2020.



# Historie a vývoj HTML

## HTML 5

DTD lze poznat dle uvedeného DTD na prvním řádku zdrojového dokumentu.

- HTML 4.01 Transitional

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01  
Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
```

- XHTML 1.1

```
<!DOCTYP html PUBLIC "-//W3C//DTD XHTML 1.1//EN"  
"http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd"> // XHTML 1.1
```

- HTML 5

```
<!DOCTYPE html>
```



# Historie a vývoj HTML

## HTML 5

```
<a href="http://gis.vsb.cz" target="_blank" hreflang="cs">  
  <h2>VÍTEJTE NA INTERNETOVÉM PORTÁLU INSTITUTU  
GEOINFORMATIKY</h2>  
  <p>Institut působí v oblasti geoinformatiky ...</p>  
</a>
```

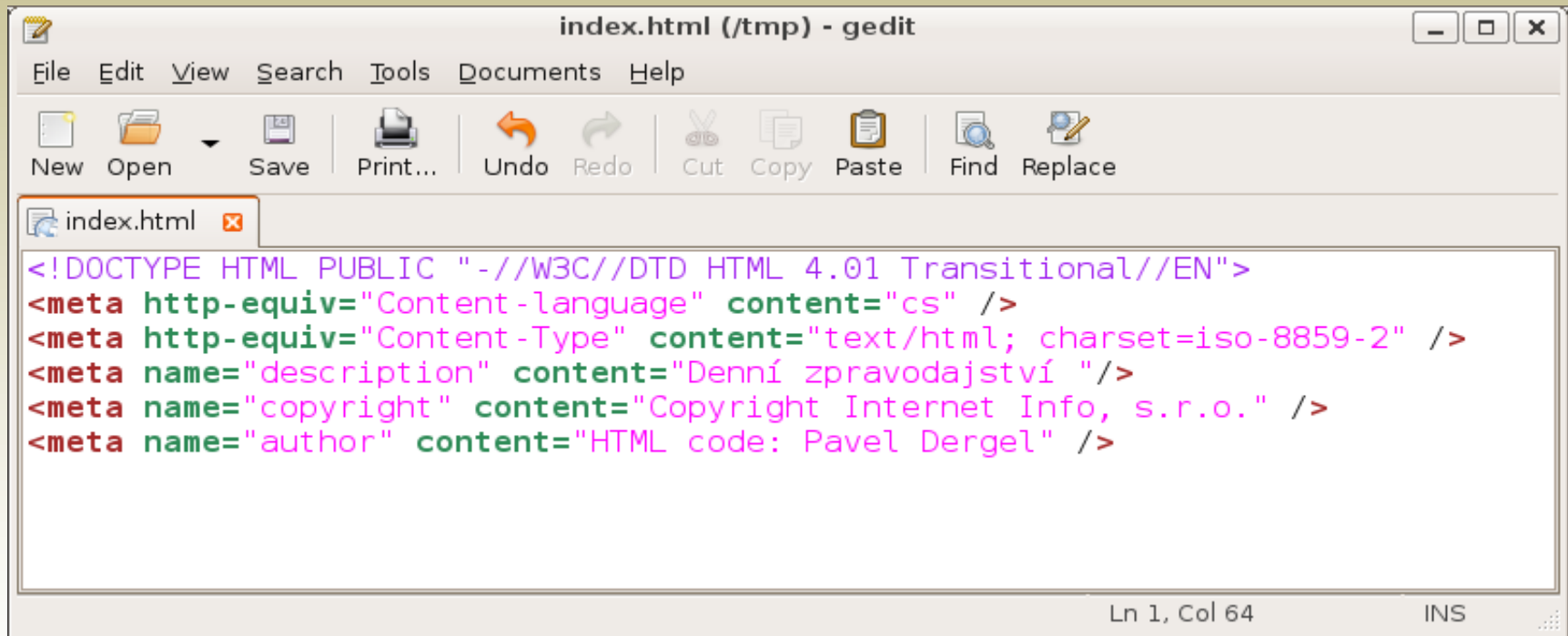


# Hlavička stránky

- `<head></head>`
- `<title>`Titulek`</title>` - Obsahuje text, který je chápán jako titulek stránky. Zobrazuje se v záhlaví okna prohlížeče. Je prioritní pro vyhledávače (zobrazí se jako odkaz).
- `<link>` - Využívá se k propojení s jiným souborem. Nepárový tag, vyskytuje se pouze v hlavičce. V současnosti se využívá především k propojení souboru se styly
  - **Příklad:** `<link rel="stylesheet" href="css/styl01.css">`
- `<meta>` - Umožňuje uložit informace o souboru. Nepárový tag, vyskytuje se pouze v hlavičce. Slouží třeba k uložení klíčových slov, autorovi, použitém jazyku a kódování apod.

# Metadata

- Běžná internetová stránka obsahuje ve své hlavičce kolem čtyř až šesti různých tagů `<meta>`.



The screenshot shows a gedit text editor window titled "index.html (/tmp) - gedit". The window contains the following HTML code:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<meta http-equiv="Content-language" content="cs" />
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-2" />
<meta name="description" content="Denní zpravodajství" />
<meta name="copyright" content="Copyright Internet Info, s.r.o." />
<meta name="author" content="HTML code: Pavel Dergel" />
```

The status bar at the bottom right of the window displays "Ln 1, Col 64" and "INS".



# Kódování stránky

- Každý znak (písmeno, číslice, . . .) je reprezentován jako číslo od 0 do 255 (jeden bajt). Američané a angličané (a programátoři) si vystačí se **128 znaky**, tzv. ASCII.
- Pozice 128 až do 255 jsou vyhrazeny pro diakritizované znaky jiných jazyků (např. kód Latin-1 - pro češtinu, použitelné jen pro á a í, jiné znaky se správně nezobrazí).
- Pro správnou interpretaci českých znaků je proto nutné v části `<head>` uvést patřičné kódování češtiny. *Kódováním* rozumíme přiřazení diakritizovaných znaků na pozicích 128 až 255.
- Problém je, že kódování češtiny nadefinovalo mnoho různých lidí, ale každý to udělal po svém.
- **Kódování češtiny má několik variant** a nesprávné uvedení či vynechání parametru `charset` může vést k chybné interpretaci českých znaků.
- Jaký je výběr? V jakých případech volíme danou variantu?



# Kódování češtiny

- **ASCII** čeština bez háčků a čárek, (hacky a carky odstraneny)
- **ISO-8859-2** (ISO Latin 2) mezinárodní kódovací norma pro jazyky střední Evropy, standard na Unixu a na Linuxu, u Microsoftu označován "Středoevropské jazyky (ISO)"
- **CP1250** (Windows-1250) základní kódování u firmy Microsoft, platforma Windows, označován "Středoevropské jazyky", vzniklo z ISO-8859-2, chybí plná kompatibilita
- **MAC** (Macintosh) standard užívaný na počítačích Apple Macintosh
- **CP852** (PC Latin 2) původní český a slovenský DOS
- **KEYBCS2** (Kamenici) kód bratří Kamenických, textový režim u MS-DOS
- **KOI8-CS** (KOI-8 CS) kdysi standard pro státy střední a východní Evropy
- Unicode, **UTF-8** užití pro všechny světové jazyky, mezinárodní kódování znaků založené na tabulce 256 používaných znaků





# Unicode - UTF-8

- **Unicode** je na rozdíl od dříve zmíněných znakových sad určeno pro všechny světové jazyky najednou.
- Znakům přiřazuje čísla od 0 až do 16 miliónů.
- UTF-8, UTF-16, apod. jsou různé způsoby zápisu Unicode znaků (liší se počtem bitů k vyjádření jednoho znaku).
- Jedná se o nejmodernější kódování (zatím bohužel ve většině aplikací nepodporované).
- V internetových prohlížečích je podporováno dobře.  
`<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">`



# Komentáře

- Do HTML kódu lze vložit popisný text, který *je prohlížečem ignorován*. Slouží k uložení poznámek, komentářů nebo jiných textů do HTML kódu.
- Komentář je vymezen znaky `<!-- -->`, cokoliv je uvnitř těchto značek je bráno jako komentář a ignorováno.
- Lze využít například k dočasnému zakomentování existujícího kódu.



# Správná úprava kódu

- Při psaní HTML kódu je třeba dbát na kvalitu a úhlednost kódu.
- Párové tagy je nutné uzavírat.
- Tagy se nesmí křížit! (`<b><i>překřížený tag</b></i>`).
- U obrázků používejte atribut `alt`, který zajistí korektní chování i v textových prohlížečích (``).
- Identifikátory elementů mohou být psány malými i velkými písmeny. Zápisy ve tvaru `<TITLE>`, `<TiTlE>`, `<tITLe>` i `<title>` prohlížeč chápe a interpretuje stejně.
- POZOR! Na začátku tagu nesmí být mezera!
- `< title>` je špatně!
- Pro speciální znaky používejte [znakové entity](#).



# Znakové entity

- Některé běžně používané znaky jsou chápány jako součást jazyka (např. `<>&“...`).
- Pokud chceme tyto znaky vložit na stránku, musíme použít speciální sekvence znaků – tzv. **znakové entity**.

Znak	Entita
Nedělitelná mezera	&nbsp;
<	&lt;
>	&gt;
&	&amp;
"	&quot;
±	&plusmn;
×	&times;
©	&copy;
€	&euro;



# Kontrola syntaxe

Kontrolu správné syntaxe HTML stránky lze provádět pomocí

- specializovaných programů a služeb,
- nebo je možné použít SGML parser.

Validační služby on-line:

- W3C HTML Validation Service na adrese "<http://validator.w3.org/>"
- Dr. Watson na adrese "<http://watson.addy.com/>"



# Editors HTML stránek

- **Strukturální** – upravují přímo výsledný kód. Vyžadují znalost jazyka HTML, většinou podporují alespoň zvýraznění syntaxe nebo doplňování tagů. Příklady takových editorů jsou třeba HomeSite, UltraEdit, EasyPad, PSPad atd. Za strukturální editory lze považovat i obyčejné textové editory typu Notepad, Gedit atd.
- **WYSIWYG** – What You See Is What You Get. Tyto editory zobrazují stránku již při psaní tak, jak bude vypadat v prohlížeči a HTML kód se generuje automaticky. K vytváření stránek není nutná žádná znalost jazyka HTML. Nevýhodou je, že tvůrce stránek nemá kontrolu nad výsledným zdrojovým kódem. Příkladem je třeba MS-FrontPage, Dreamweaver, OpenOffice, MS-Word, NVU a mnoho dalších.



# Zobrazení stránek

## Prohlížeče

- Stejná stránka se může v různých prohlížečích zobrazit různě.
- Dobrý zvyk – webové stránky vyzkoušet ve všech hlavních prohlížečích (MSIE, Mozilla, Firefox, Opera).

## Rozlišení

- Zobrazení umístěných prvků a rozdělení textu závisí na nastaveném rozlišení monitoru.
- Je vhodné vyzkoušet různá rozlišení monitoru (800x600, 1024x768, 1280x1024, 1600x1200).





# Soubory na webu

- Všechny HTML soubory by měly mít příponu [.html](#). Někdy se objevují i soubory s příponou [.htm](#) (nestandardní, bylo zavedeno kvůli MS Windows 3.1, kde soubory měly příponu max. 3 znaky).
- Názvy souborů by neměly obsahovat češtinu ani žádné jiné znaky s diakritikou. Problémy nastávají i při použití mezer.
- V názvu souborů se nesmí nacházet speciální znaky (`\ / * : ? # < >`)
- Doporučuji zadávat všechny názvy souborů i adresářů malými písmeny (URL adresy jsou case-senzitivní – záleží na velikosti písmen).
- Všechny soubory webu by měly být v jednom adresáři a podadresářích. Vyhybat se příliš zanořeným strukturám.
- Obrázky, skripty – speciální adresář (images atd.).
- Hlavní stránka s názvem [index.html](#) bude umístěna do kořenového adresáře webu.



# Publikování stránek na WWW server

## Zadarmo

- Možnost využití **freehostingů**. Takových serverů existují desítky ([www.sweb.cz](http://www.sweb.cz), [www.webpark.cz](http://www.webpark.cz), [www.webzdarma.cz](http://www.webzdarma.cz) atd.). Pozor! Na webu mohou vyskakovat různé reklamy.
- Možnost využít nějakého známého nebo firmu, který má server a nechá vás na něm publikovat.
- Využití školního serveru ([homel.vsb.cz](http://homel.vsb.cz), [gislinb.vsb.cz](http://gislinb.vsb.cz) atd.).

## Za peníze

- Pokud chcete doménu druhé úrovně ([neco.cz](http://neco.cz)), musíte za to zaplatit. Poplatek činí většinou několik set korun ročně, liší se v závislosti na množství poskytovaných služeb, rychlosti, diskové kapacitě atd.



# Umístění stránek na gislinb.vsb.cz

Přenos stránek pomocí scp či jiného podobného sw.

1) Ve svém domovském adresáři na serveru gislinb vytvořte adresář [public\\_html](#).

2) Ověřte nastavení přístupových práv na tento adresář:

[vlastník](#): (čtení, zápis, spouštění),

[skupina](#): (spouštění),

[ostatní](#): (spouštění).

Výsledek musí vypadat alespoň takto: `rwX__X__X`

3) Umístěte své webové stránky do adresáře [public\\_html](#).

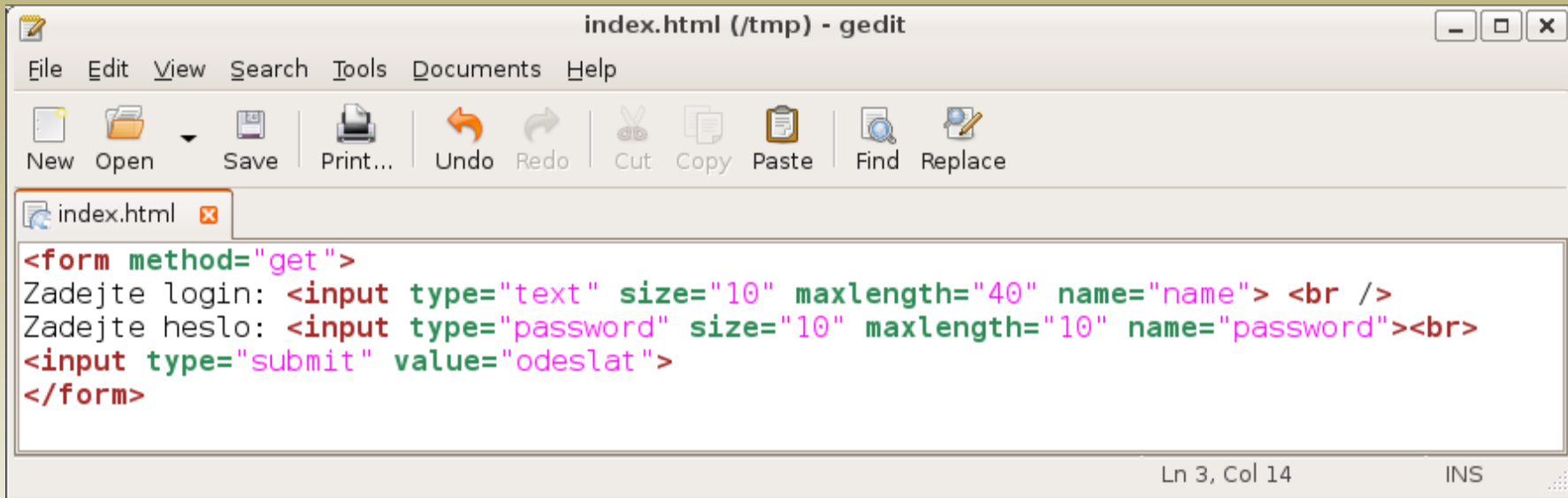
4) Váš web bude přístupný odkudkoliv z internetu přes URL adresu:  
<http://gisak.vsb.cz/~osobnicislo/>



# Formuláře

- Formuláře slouží pro vstup dat od uživatele (zadání Jména, adresy, výběr z několika možností apod.)
- HTML poskytuje k tomuto účelu vstupní textová pole, zaškrtačací tlačítka, přepínací tlačítka atd.
- Formulář se vytváří pomocí tagu `<form></form>`
- Má několik atributů, které ovlivňují způsob zpracování vstupních informací
  - **action** – skript, který bude data zpracovávat (obvykle se jedná o CGI nebo PHP skript, mapserver atd. ). Není-li uvedeno nic, jsou data odeslána téže stránce
  - **method** – výchozí metoda tohoto atributu je **get** (říká, že data se budou předávat jako součást URL adresy, jsou tedy viditelné v okně prohlížeče). Metoda **post** slouží k nezávislému zabalení přenášených dat, takže nejsou vidět (vhodné pro větší formuláře)
  - **enctype** – většinou není potřeba nastavovat. Využívá se v případě odesílání souborů.

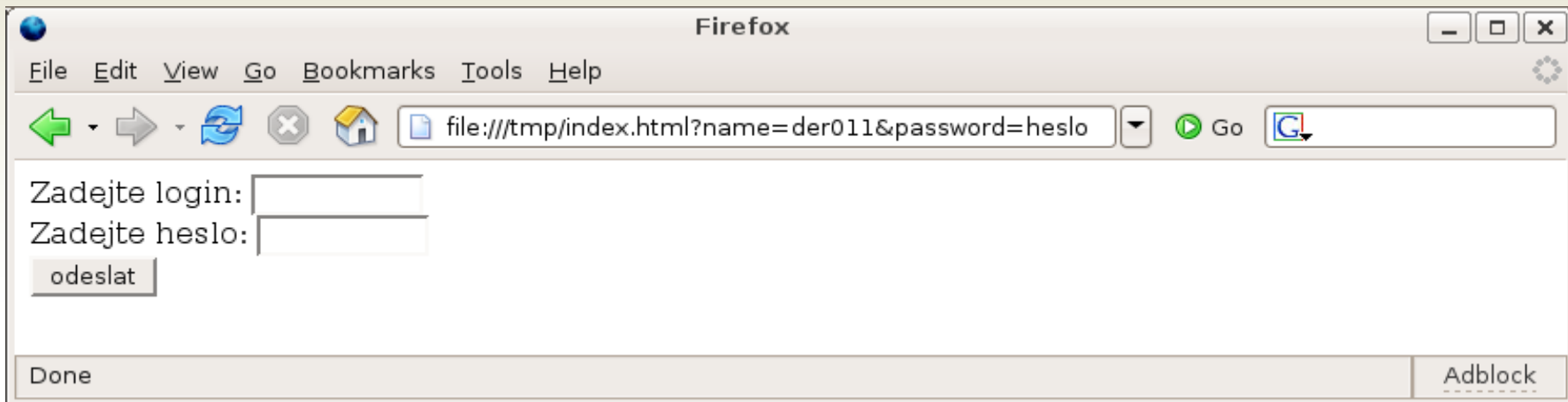
# Textové pole



The screenshot shows the gedit text editor window titled "index.html (/tmp) - gedit". The menu bar includes File, Edit, View, Search, Tools, Documents, and Help. The toolbar contains icons for New, Open, Save, Print..., Undo, Redo, Cut, Copy, Paste, Find, and Replace. A single tab labeled "index.html" is open. The editor contains the following HTML code:

```
<form method="get">  
Zadejte login: <input type="text" size="10" maxlength="40" name="name"> <br />  
Zadejte heslo: <input type="password" size="10" maxlength="10" name="password"><br>  
<input type="submit" value="odeslat">  
</form>
```

The status bar at the bottom right indicates "Ln 3, Col 14" and "INS".



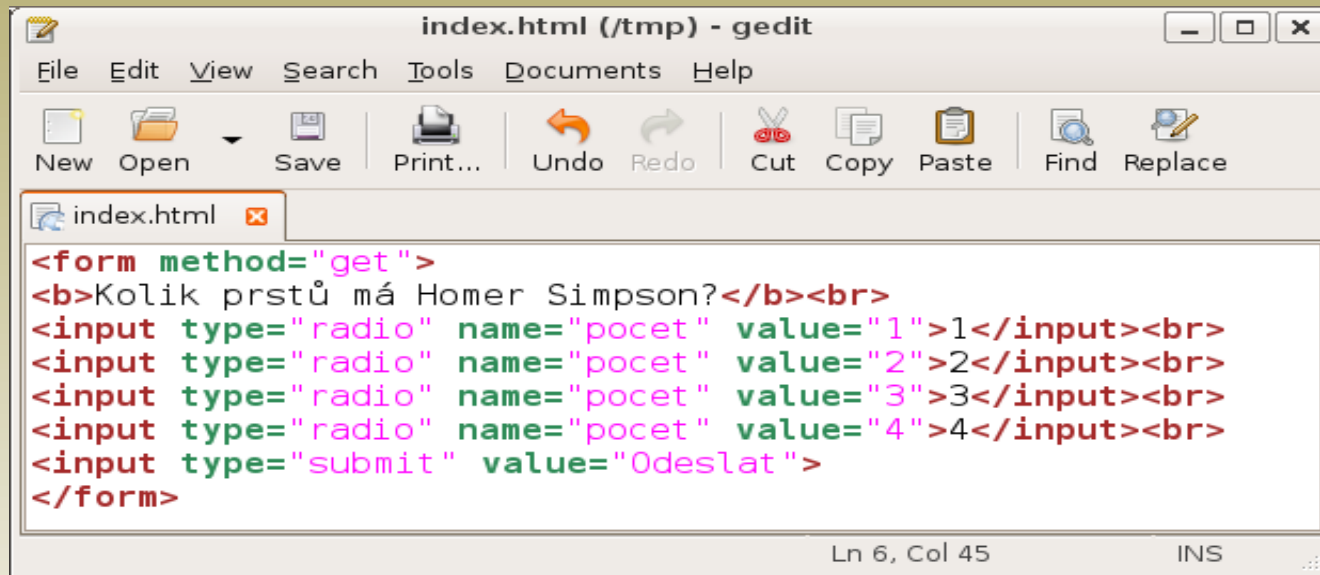
The screenshot shows the Firefox browser window titled "Firefox". The menu bar includes File, Edit, View, Go, Bookmarks, Tools, and Help. The address bar shows the file path: "file:///tmp/index.html?name=der011&password=heslo". The browser displays the rendered HTML form from the previous screenshot:

Zadejte login:

Zadejte heslo:

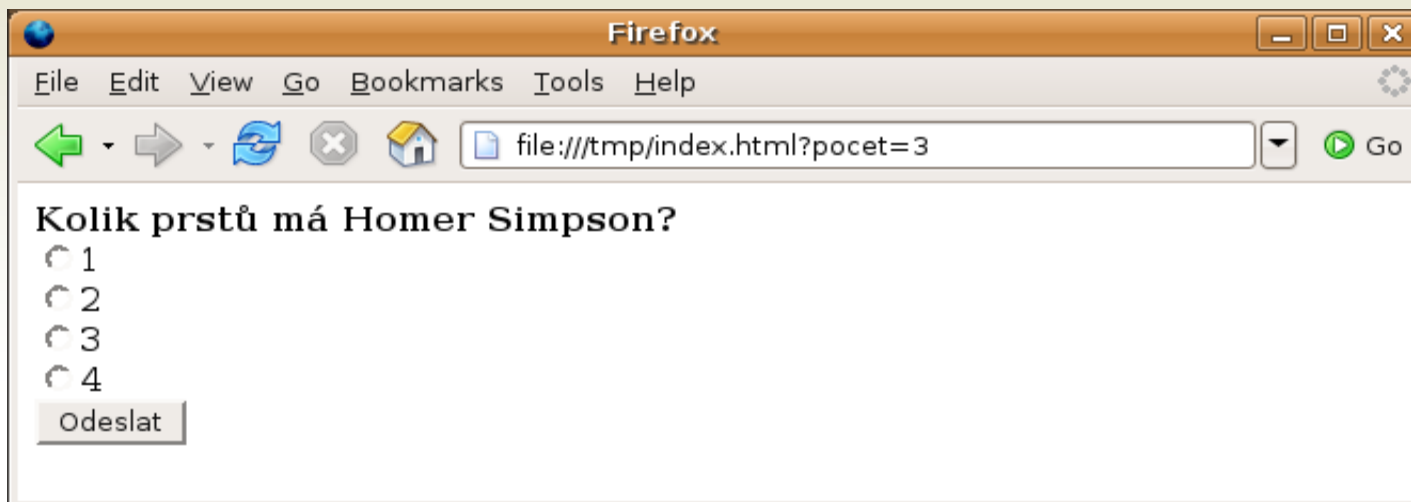
The status bar at the bottom shows "Done" and "Adblock".

# Přepínací tlačítka

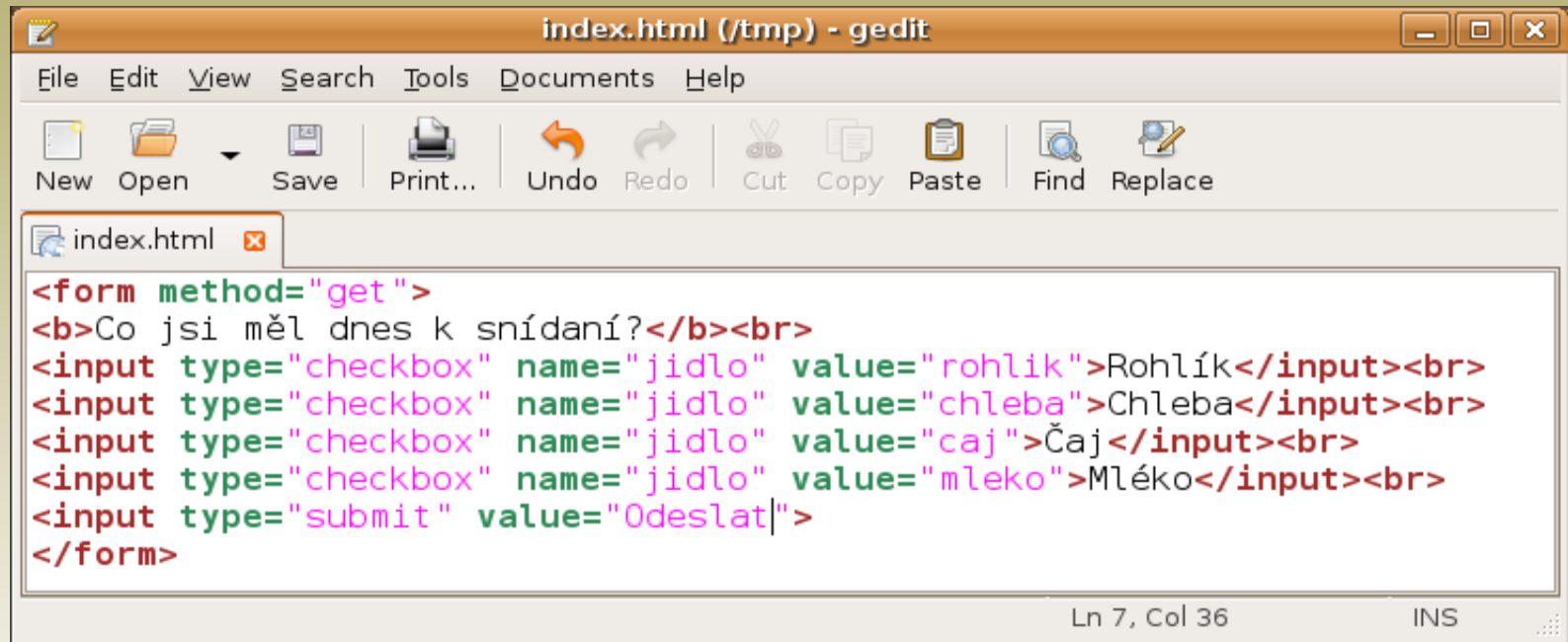


```
<form method="get">
<b>Kolik prstů má Homer Simpson?</b><br>
<input type="radio" name="pocet" value="1">1</input><br>
<input type="radio" name="pocet" value="2">2</input><br>
<input type="radio" name="pocet" value="3">3</input><br>
<input type="radio" name="pocet" value="4">4</input><br>
<input type="submit" value="Odeslat">
</form>
```

Ln 6, Col 45      INS

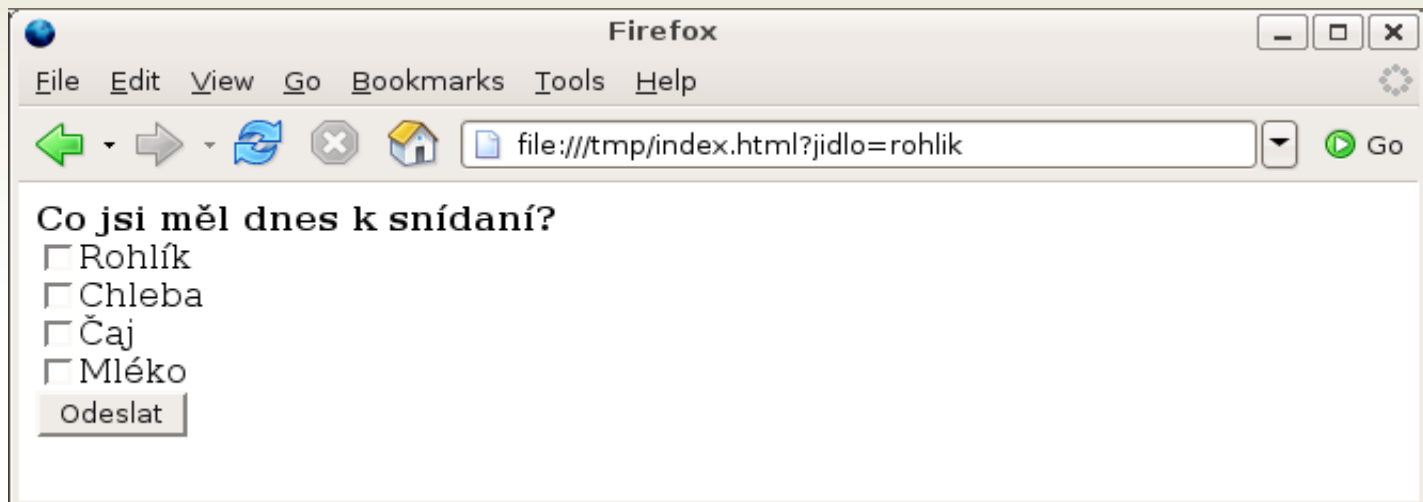


# Zaškrtávací políčka



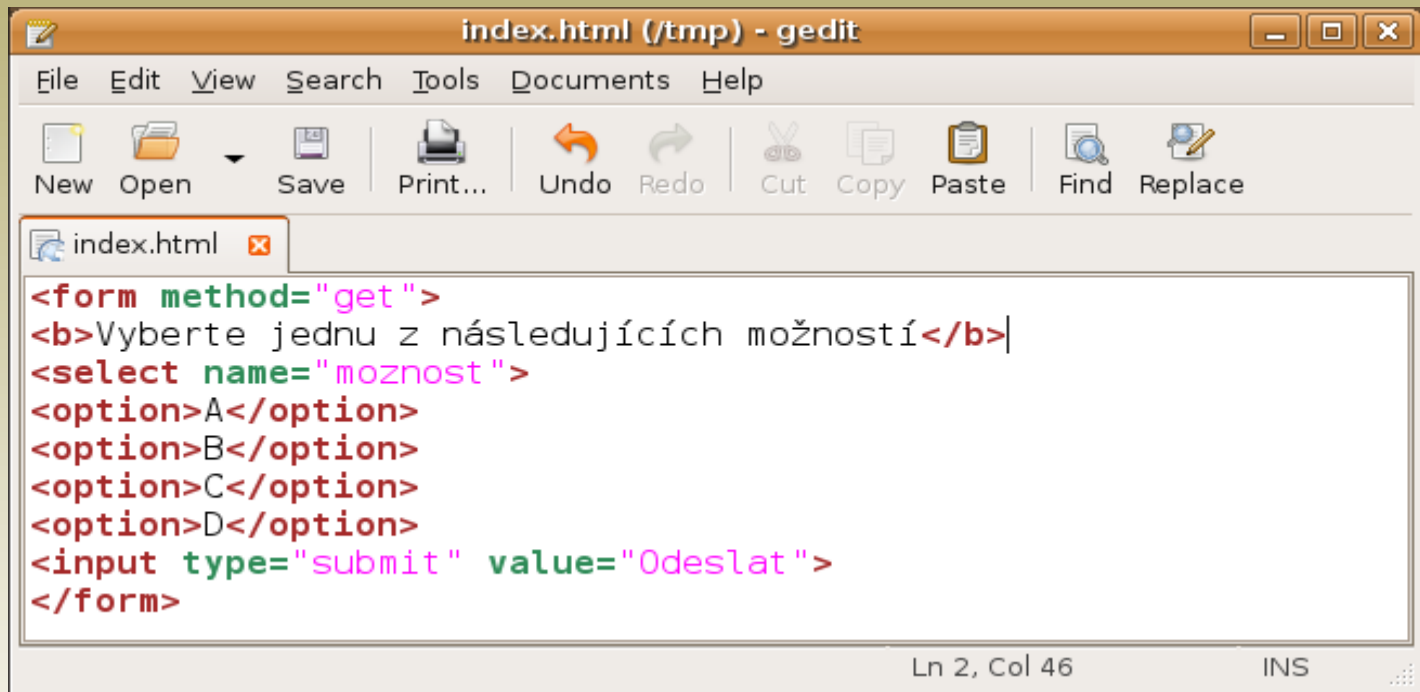
```
<form method="get">
<b>Co jsi měl dnes k snídání?</b><br>
<input type="checkbox" name="jidlo" value="rohlík">Rohlík</input><br>
<input type="checkbox" name="jidlo" value="chleba">Chleba</input><br>
<input type="checkbox" name="jidlo" value="čaj">Čaj</input><br>
<input type="checkbox" name="jidlo" value="mléko">Mléko</input><br>
<input type="submit" value="Odeslat">
</form>
```

Ln 7, Col 36 INS





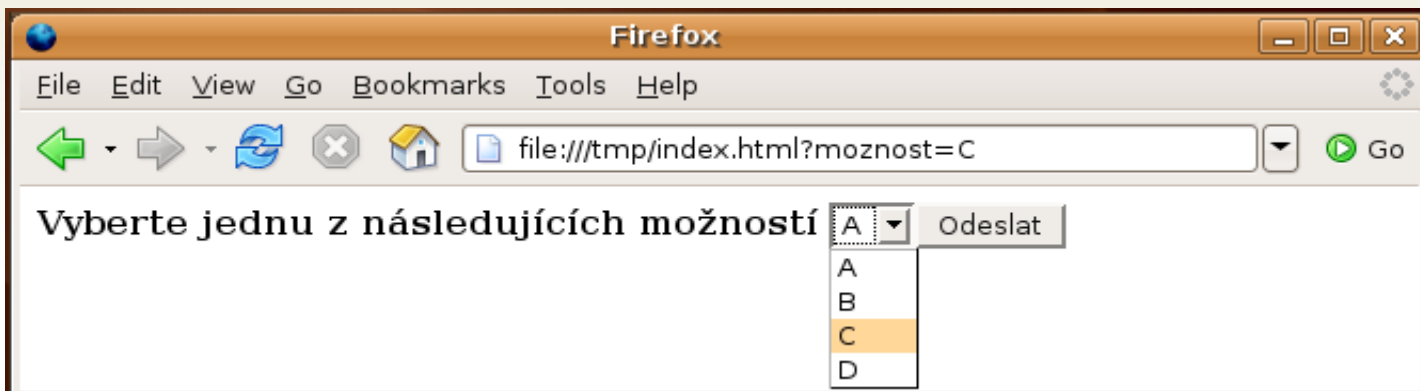
# Menu



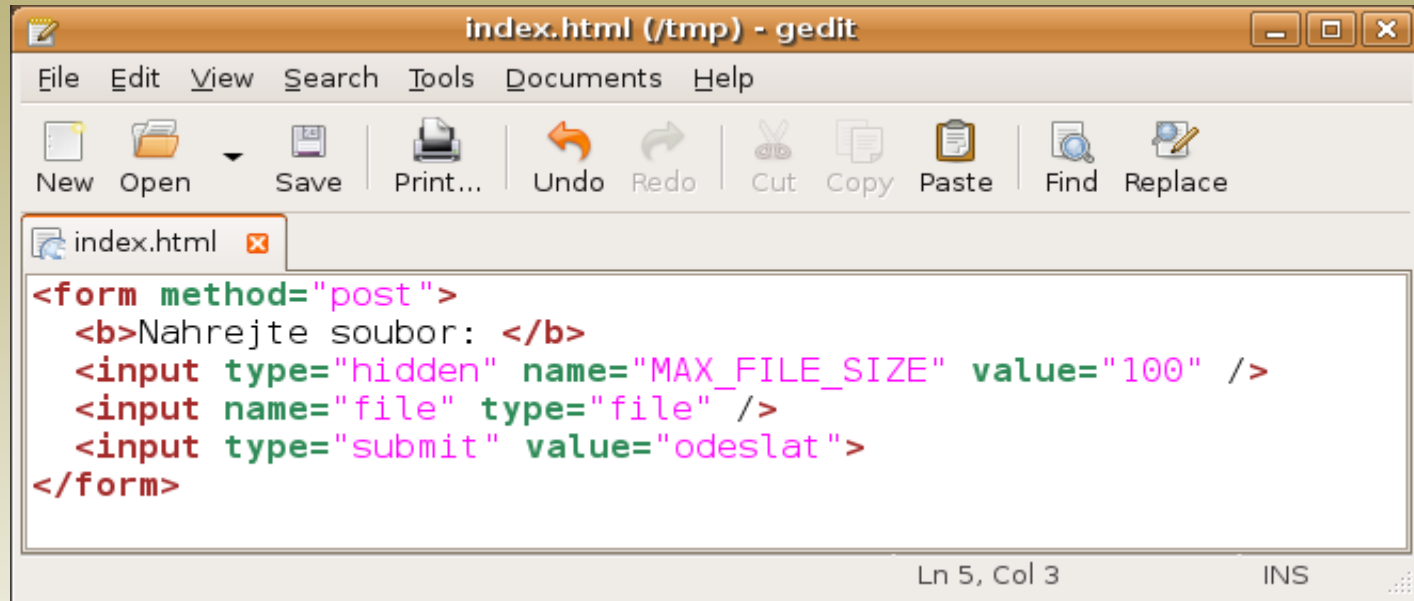
The screenshot shows the gedit text editor window titled "index.html (/tmp) - gedit". The menu bar includes File, Edit, View, Search, Tools, Documents, and Help. The toolbar contains icons for New, Open, Save, Print..., Undo, Redo, Cut, Copy, Paste, Find, and Replace. A single tab labeled "index.html" is open. The text area contains the following HTML code:

```
<form method="get">
<b>Vyberte jednu z následujících možností</b>|
<select name="moznost">
<option>A</option>
<option>B</option>
<option>C</option>
<option>D</option>
<input type="submit" value="Odeslat">
</form>
```

The status bar at the bottom right indicates "Ln 2, Col 46" and "INS".



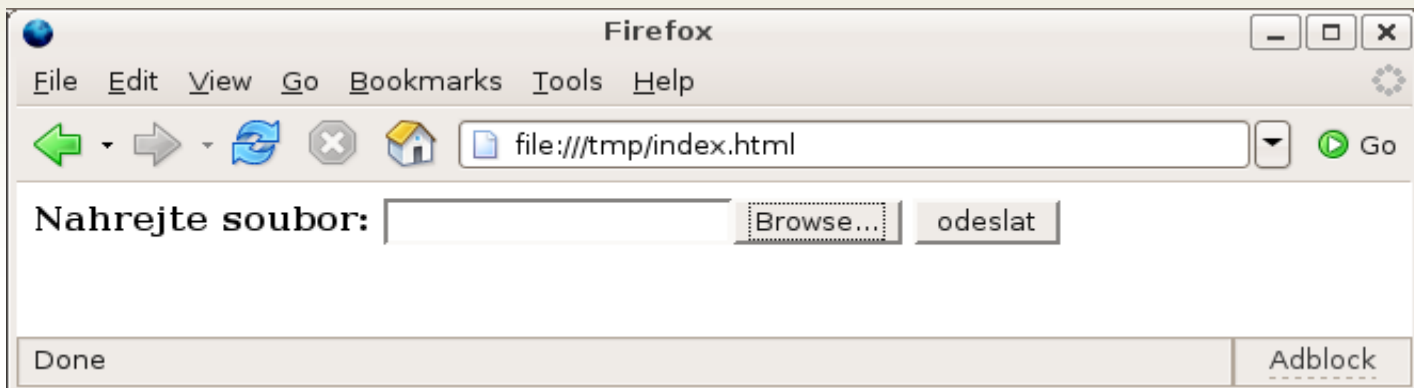
# Upload souboru



The screenshot shows a gedit editor window titled "index.html (/tmp) - gedit". The menu bar includes File, Edit, View, Search, Tools, Documents, and Help. The toolbar contains icons for New, Open, Save, Print..., Undo, Redo, Cut, Copy, Paste, Find, and Replace. The main text area contains the following HTML code:

```
<form method="post">
  <b>Nahrejte soubor: </b>
  <input type="hidden" name="MAX_FILE_SIZE" value="100" />
  <input name="file" type="file" />
  <input type="submit" value="odeslat">
</form>
```

The status bar at the bottom right indicates "Ln 5, Col 3" and "INS".



The screenshot shows a Firefox browser window titled "Firefox". The menu bar includes File, Edit, View, Go, Bookmarks, Tools, and Help. The address bar shows "file:///tmp/index.html". The main content area displays the rendered form:

Nahrejte soubor:

The status bar at the bottom left shows "Done" and the bottom right shows "Adblock".



# Problémy dnešního webu

- Informační zahlcení,
- nutnost lepšího vyhledávání,
- nestandardní rozšíření HTML jednotlivými výrobci prohlížečů.

Řešení problému? XML - eXtensible Markup Language

- pomocí DTD můžeme vytvářet nové jazyky založené na XML,
- vhodně zvolené elementy přidají dokumentu informační hodnotu a usnadní prohledávání Webu,
- XML má striktnější syntaxi než HTML (neukončené tagy, ohraničení atributů uvozovkami apod.) ,
- snazší implementace.



# CMS — Content Management System

Systemy pro správu obsahu umožňují uživatelům bez znalosti web technologie publikovat různé texty i obrazový materiál. Uživatel s nimi pracuje přes webové rozhraní, které je určeno pro správu obsahu, manipulaci zajišťují běžné webové prohlížeče.

Hlavní funkce poskytované redakčními systémy

- vytváření, editace a vlastní publikování text. dokumentů na webu,
- tvorba pomocí WYSIWYG editoru - generovaný HTML kód ovšem bývá větší než při ruční tvorbě,
- správa obrazového materiálu, vytváření interních i veřejných galerií,
- možnost statistiky pomocí funkcí k tomu určených,
- správa diskusních fór,
- správa souborů,
- podpora funkcí na uživatelské a administrátorské úrovni.

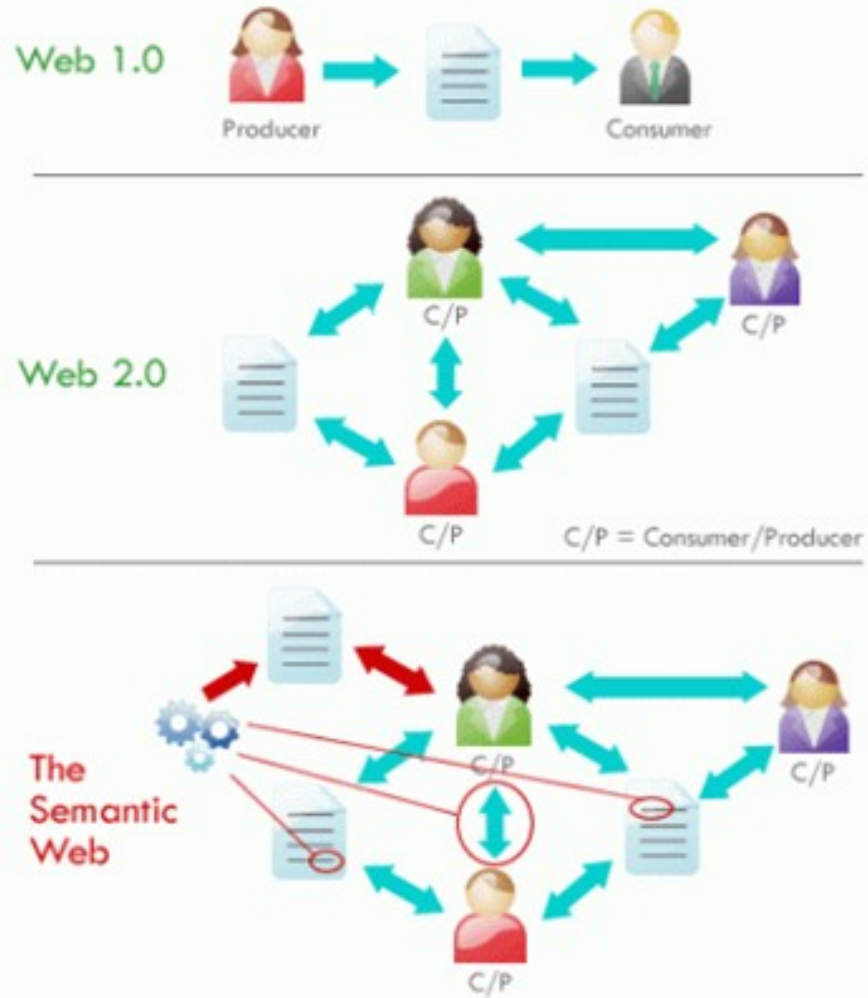


# CMS — Content Management System

## Hlavní představitelé

- Wiki
- PHP-Nuke (česká verze United-Nuke)
- MAMBO CMS
- Plone
- Joomla!
- Drupal

# Budoucnost webu?





# Zdroje

- [www.jakpsatweb.cz/](http://www.jakpsatweb.cz/)
- <http://www.onlio.com/>