

# Tvorba webu

Úvod a základní principy



Martin Urza

# World Wide Web (WWW)



- World Wide Web (doslova „celosvětová pavučina“) je označení pro mnoho dokumentů rozmístěných na různých serverech po celém světě.
  - Tyto dokumenty jsou typicky hypertextové a vzájemně propojené (hypertextovými odkazy).
- Dokumenty jsou adresovány pomocí URL (unique resource locator), tedy jednoznačného identifikátoru serveru i odpovídajícího dokumentu.
- Pro přenos dokumentů se používá protokol HTTP.
- Fyzicky jsou dokumenty uloženy na mnoha různých serverech po celém světě, takže jsou typicky posílány po internetu (**world wide web není internet**).



# URL



- Zkratka unique/uniform/universal resource locator.
- Tvar URL adresy: protokol://server:port/dokument:
  - `http://vyuka.urza.cz:80/gis/prezentace/sem09.pdf`
    - Protokol je http, ale např. prohlížeče jej mají implicitní.
    - Server se jmenuje vyuka.urza.cz.
    - Číslo portu je 80, což je u HTTP protokolu implicitní.
    - Cesta k dokumentu je /gis/prezentace/sem09.pdf.
  - Před jménem serveru může být nepovinně uživatelské jméno a heslo ve tvaru `username:passwd@server`.
  - Za cestou k dokumentu mohou být ještě volitelné parametry, které jsou posílány serveru spolu s žádostí:
    - `dokument?param1=value1&param2=value2&...`

# DNS



- Zkratka slov domain name system/service.
- Servery mají své adresy, které si lidé těžko pamatují, takže byl zaveden DNS, aby mohly mít servery lépe zapamatovatelná jména.
  - Server může mít libovolný počet jmen (nebo žádné).
- Jména serverů se skládají ze jmen domén, které tvoří strom (kořenem je uzel „.“).
  - Např. vyuka.urza.cz. je server vyuka v doméně urza, jež leží v top level doméně cz (všechny TLD jsou pod .).
- Je-li v doméně webová služba, typicky se dotyčný server jmenuje www (to nemusí být podmínkou).
  - **Nicméně [www.foo.tld](#) je jiná adresa než [foo.tld](#)!!!!**





# Webové stránky

- Webovými stránkami se rozumí dokumenty, které jsou typicky zobrazovány webovými prohlížeči.
- Webové prohlížeče (neboli browsery) jsou počítačové programy (klasické aplikace) určené zejména pro zobrazování výše zmíněných dokumentů.
- V typickém případě je webová stránka umístěna kdesi v internetu a ten, kdo si tuto stránku prohlíží, ji stáhne do svého PC, kde ji zobrazí v prohlížeči.
  - To není pravidlem, úplně stejně lze pomocí prohlížeče zobrazovat dokumenty uložené v počítači.
  - Webová stránka je pro prohlížeč přibližně tím, čím je například video pro přehrávač.

<!DOCTYPE  
<HTML>  
<HEAD>  
<TITLE>RA  
<LINK REV  
<META NAM

# Jazyk HTML

- HTML (zkratka pro HyperText Markup Language) je formální jazyk, ve kterém lze (textově) popsat to, co má prohlížeč zobrazovat.
- Stránku v jazyce HTML lze „volně převyprávět“ v přirozeném jazyce například takto: „Vlevo nahoře na stránce má být velký nadpis, pod ním červený text, vedle kterého se má zobrazit obrázek, který má též sloužit jako odkaz na jinou HTML stránku.“
  - Prohlížeče přirozenému jazyku pochopitelně nerozumí, proto je nutné použít jazyk formální a říci jím totéž.
- Jazyk HTML byl vyvinut za tímto účelem, nicméně jeho vývoj byl poté zpětně ovlivněn prohlížeči.



# Divergence prohlížečů



- Navzdory tomu, že je HTML formální jazyk, různé prohlížeče mohou interpretovat tutéž HTML stránku jinak.
- Nejčastěji zobrazoval stránky špatně (tedy jinak, než by měl) Internet Explorer (zejména verze 6 a nižší), což se od verze 7 zlepšilo a verze 9 to snad vyřešila.
- Chyba zobrazení je leckdy primárně v prohlížeči a sekundárně na straně uživatele, který používá deset let starý prohlížeč, řešit to musí tvůrce stránek.
- **Vytvoříte-li nějaké webové stránky, vyzkoušejte jejich funkčnost v co největším počtu různých prohlížečů.**



# Validní HTML kód

- Problémům s různým zobrazováním stejného HTML dokumentu v rozdílných prohlížečích lze do značné míry (nikoliv zcela) předejít.
- Při tvorbě HTML stránek pište validní HTML kód, což jest takový kód, který odpovídá standardům W3C (zkratka pro World Wide Web Consortium).
- Psaní validního HTML kódu docílíte tak, že se budete naprosto přesně řídit vším, co se dozvíte v rámci tohoto kurzu.
- Již napsané stránky nechte zkontrolovat validátorem na adrese <http://validator.w3.org/>. Validátor Vám ukáže všechny chyby v dokumentu.



# Struktura HTML dokumentu



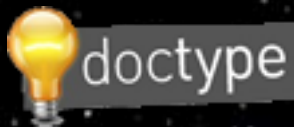
- HTML dokumenty se skládají z elementů. Element se zapisuje takto: `<jmeno atributy>obsah</jmeno>`, případně jen `<jmeno atributy />`, nemá-li obsah. Atributy jsou nepovinné (a obsah také). Elementy mohou obsahovat jiné elementy, ale nesmí se nijak křížit (`<a><b></b></a>` ano, `<a><b></a></b>` ne).
- Na počátku každého dokumentu by měl být uveden jeho typ (doctype).
  - Tímto typem oznamujeme programu, který dokument čte (typicky prohlížeč), že se jedná o HTML dokument.
- Celý HTML dokument se nachází v elementu HTML.
  - Element HTML obsahuje elementy HEAD a BODY.

# Párové a nepárové tagy



- HTML elementům se říká tagy.
  - Všude, kde je (bylo) použito slovo element, může být použito i slovo tag.
- Párové tagy jsou takové, které v sobě mohou mít nějaký obsah. Skládají se z otevíracího tagu `<jméno>`, obsahu a zavíracího tagu `</jméno>`.
  - Příklad: `<p>Laissez <b>faire</b>!</p>`
  - Příklad s atributy: `<div id="menu">Vyberte.</div>`
- Nepárové tagy nemohou mít žádný obsah. Skládají se pouze z jediného tagu `<jméno/>`.
  - Příklad: `<br/><hr/>`
  - Příklad s atributy: `<meta name='xxx' content='yyy' />`





# Doctype

- Doctype říká, o který typ dokumentu jde (v našem případě HTML dokument). Příklady zápisu:
  - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" 'http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd'>`
  - `<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">`
  - `<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN">`
  - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">`
  - `<!doctype html public "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">`
  - `<!DOCTYPE html>`
- Doctype obsahuje verzi HTML (či XHTML) a případně ještě DTD dokument, ve kterém je popsána validní struktura XML dokumentu.
  - HTML jazyk je speciálním případem XML.

# Elementy HTML, HEAD a BODY

- Celý dokument (krom doctype) je obsahem elementu html (tedy mezi `<html>` a `</html>`).
- Tento element může mít (jde-li se o XHTML) atributy:  
`<html xmlns='http://www.w3.org/1999/xhtml' xml:lang='cs' lang='cs'>`
- Element html obsahuje právě dva elementy, head a body:

```
<html xmlns='http://www.w3.org/1999/xhtml' xml:lang='cs' lang='cs'>  
  <head>  
    obsah elementu head  
  </head>  
  <body>  
    obsah elementu body  
  </body>  
</html>
```





# Element HEAD



- Tag HEAD obsahuje čtyři typy tagů:
  - `<title>titulek</title>` Titulek se zobrazuje v záhlaví prohlížeče, případně jako nadpis ve vyhledávačích.
  - `<base href='url' target='rám' />` Přesnější vysvětlení bude uvedeno později; base se beztak příliš nepoužívá.
  - `<link .... />` Užívá se k propojení s jinými dokumenty, např. soubory s kaskádovými styly.
  - `<meta .... />` Zde jsou informace o dokumentu určené pro strojové zpracování. Jedná se například o použité kódování, popis stránky a podobně, například:

```
<meta http-equiv='Content-Type' content='text/html; charset=windows-1250' />
<meta name='description' content='Null page.' />
<meta http-equiv='content-language' content='cs' />
```



# Element BODY

- Tag BODY odpovídá celé zobrazované HTML stránce, jak ji vidíme v prohlížeči.
- Atributů elementu BODY je celá řada, jednotlivě a podrobně budou zmíněny později, nicméně platí, že atributy tohoto tagu se vztahují pro celou stránku, lze zde tedy nastavovat například barva pozadí, fonty textu i odkazů a podobně.
- Obsahem tagu BODY je to, co má být zobrazeno na stránce, tedy různé tagy a/nebo text.
- Celé „programování stránky“ probíhá v tagu BODY.
  - Programování je dost silné slovo, **psaní HTML kódu je vlastně jen formátování textu!**





## Nadpisy a odstavce

- Základním stavebním kamenem dokumentu jsou odstavce, do kterých píšeme text.
- Tag odstavce je párový a jmenuje se p (což je první písmeno slova paragraph).
- Nadpisy jsou též reprezentovány párovými tagy, jež se jmenují h1, h2, h3, h4, h5 a h6. Tato čísla určují váhu nadpisu; čím nižší číslo, tím „větší“ (důležitější) nadpis.
  - Jak konkrétně který nadpis vypadá záleží na různých věcech (prohlížeči, stylech, ...), které budeme probírat později, nicméně nebylo-li něco „schválně“ změněno, nadpisy s nižšími čísly jsou větší, případně tučnější.

# Rekapitulace



- World wide web je soustava dokumentů a není to totéž jako internet (přes ten se dokumenty přenáší).
- URL je unikátní adresa, jež jednoznačně identifikuje dokument.
- Server je v URL identifikován svým jménem, které lze přeložit na adresu s využitím systému DNS.
- Webové stránky se zobrazují v prohlížečích, což jsou programy k tomu účelu určené.
  - Prohlížeče nerozlišují, zda zobrazují dokument stažený z internetu, nebo uložený na lokálním disku.
- Aby prohlížeče rozuměly dokumentům, jsou tyto psány ve formálním jazyce HTML.