

14. HyperText Markup Language

Z Wikipedie, otevřené encyklopedie

Skočit na: [Navigace](#), [Hledání](#)

HyperText Markup Language, označovaný zkratkou **HTML**, je [značkový jazyk](#) pro [hypertext](#). Je jedním z jazyků pro vytváření stránek v systému [World Wide Web](#), který umožňuje publikaci dokumentů na [Internetu](#).

Vývoj jazyka HTML

V roce [1989](#) spolupracovali [Tim Berners-Lee](#) a [Robert Caillau](#) na propojeném informačním systému pro [CERN](#), výzkumné centrum fyziky poblíž [Ženevy](#) ve [Švýcarsku](#). V té době se pro tvorbu dokumentů obvykle používaly jazyky [TeX](#), [PostScript](#) a také SGML. Berners-Lee si uvědomoval, že potřebují něco jednoduššího a v roce [1990](#) byl tedy navržen jazyk HTML a [protokol](#) pro jeho přenos v počítačové síti – [HTTP](#) (*HyperText Transfer Protocol – přenosový protokol hypertextu*). Zároveň také Tim Berners-Lee napsal první webový prohlížeč, který nazval [WorldWideWeb](#).

V roce [1991](#) CERN zprovoznil svůj web. Současně organizace [NCSA](#) (*National Center for Supercomputer Applications*) vybídla [Marca Andreessena](#) a [Erica Binu](#) k vytvoření prohlížeče [Mosaic](#); ten vznikl v roce [1993](#) ve verzích pro počítače [IBM PC](#) a [Macintosh](#) a měl obrovský úspěch. Byl to první prohlížeč s [grafickým uživatelským rozhraním](#).

Následoval rychlý rozvoj webu, takže bylo nutné pro HTML definovat [standardy](#).

Verze jazyka

Verze 0.9

Byla vydána zhruba v roce 1991. Nepodporuje grafický režim (verze, kterou vytvořil Tim Berners-Lee).

Verze 2.0

Zachycuje stav jazyka v polovině roku [1994](#). Přidává k původní specifikaci interaktivní formuláře a podporu grafiky ☺.

Verze 3.2

Byla vydána [14. ledna 1997](#) a zachycuje stav jazyka v roce [1996](#). Připravovaná verze HTML 3.0 nebyla nikdy přijata jako standard, protože byla příliš složitá a žádná firma nebyla schopna naprogramovat její podporu. Standard už vydalo [W3C](#), stejně jako následující verze. Přidává k jazyku tabulky, zarovnávání textu a stylové elementy pro ovlivňování vzhledu.

Verze 4.0

Byla vydána [18. prosince 1997](#). Do specifikace jazyka přibyly nové prvky pro tvorbu tabulek, formulářů a nově byly standardizovány rámy (*frames*). Tato verze se snaží dosáhnout původního účelu – prvky by měly vyznačovat význam (sémantiku) jednotlivých částí dokumentu, vzhled má být ovlivňován připojovanými [styly](#). Některé prezentační elementy byly zavrženy.

Verze 4.01

Byla vydána [24. prosince 1999](#). Tato verze opravuje některé chyby verze předchozí.^[1] Podle původního předpokladu se mělo jednat o poslední verzi – a my ji hojně používáme v AceHTML a PHP ☺, po které by se přešlo na [XHTML](#).

Verze 5

[7. března 2007](#) byla založena nová pracovní skupina HTML, jejíž cílem je vývoj nové verze HTML. V květnu 2007 bylo odhlasováno, že základem nové specifikace se stanou Web Applications 1.0 a Web Forms 2.0 ze specifikace [WHATWG](#). Jako název nové specifikace bylo odhlasováno HTML 5.0 a specifikace by měla být hotova v roce [2010](#) (odtud ji začnou vyvíjet webových aplikací používat, finalní verze zbavená všech chyb se však odhaduje až na rok [2022](#)^[2]).

Popis jazyka HTML

Koncepce

Jazyk HTML je od verze 2.0 aplikací [SGML](#). Je charakterizován množinou značek a jejich atributů definovaných pro danou verzi. Mezi značky se uzavírají části textu dokumentu a tím se určuje význam (*sémantika*) obsaženého textu. Názvy jednotlivých značek se uzavírají mezi úhlové závorky (< a >). Část dokumentu tvořená otevírací značkou, nějakým obsahem a odpovídající ukončovací značkou tvoří tzv. *element* (prvek) dokumentu. Například je otevírací značka pro zvýraznění textu a Červená Karkulka je element obsahující zvýrazněný text. Součástí obsahu elementu mohou být další vnořené elementy. Atributy jsou doplňující informace, které upřesňují vlastnosti elementu.

Značky (zvané tagy) jsou obvykle párové (v [XHTML](#) jsou párové všechny), přičemž koncové značka je shodná se značkou počáteční, jen má před názvem znak [lomítko](#). Příklad pro označení odstavce:

```
<p>Text odstavce</p>
```

Některé značky jsou nepárové – nemají žádný obsah a nepoužívají koncovou značku. Příklad pro vykreslení vodorovné čáry:

```
<hr>
```

Tagy mohou obsahovat atributy, které popisují jejich vlastnosti nebo nesou jinou informaci. Příkladem může být odkaz (tag `a`), jehož atribut `href` říká, kam se uživatel po kliknutí na něj dostane (v tomto příkladu na stránku `http://example.com`):

```
<a href="http://example.com">text odkazu</a>
```

Jelikož je HTML aplikací SGML, existují i jiné, méně známé konstrukce pro tvoření elementů. Jedná se o tzv. *zkrácené HTML zápisy*.^{[3][4]}

- `<element/Obsah/` je totéž jako `<element>Obsah</element>`
- `<element>Obsah</>` je totéž jako `<element>Obsah</element>`
- `<el<el2>Obsah` je totéž jako `<el><el2>Obsah`
- `Adam<>Božena` je totéž jako `AdamBožena`

Všechny tyto zápisy jsou sice podle normy validní a zcela ekvivalentní, ale žádný ze známých prohlížečů zkrácené verze nepodporuje, takže se nedoporučuje je používat.

Pro každou verzi existuje definice pravidel [DTD](#) (*Document Type Definition*). Od verze 4.01 musí být odkaz na deklaraci DTD v dokumentu uveden pomocí klíčového slova `DOCTYPE`. DTD definuje pro určitou verzi elementy a atributy, které lze používat.

Dokument může mimo značkování obsahovat další prvky:

- Direktivy – začínají znaky `<!`, jsou určeny pro zpracovatele dokumentu (prohlížeč).
- Komentáře – pomocné texty pro programátora, nejsou součástí obsahu dokumentu a nezobrazují se (prohlížeč je ignoruje). Příklad komentáře je uveden níže.
- Kód skriptovacích jazyků.
- Definice událostí a kód pro jejich obsluhu.

Struktura dokumentu

Dokument v jazyku HTML má předepsanou strukturu:

- Deklarace DTD – je povinná až ve verzi 4.01, je uvedena direktivou `<!DOCTYPE`.
- Kořenový element – element `html` (značky `<html>` a `</html>`) reprezentuje celý dokument. Kořenový element je povinný, ale otevírací a ukončovací značka samotná povinná není (pokud tyto značky nebudou v těle dokumentu uvedeny, prohlížeč si je sám doplní podle kontextu).
- Hlavička elementu – obsahuje [metadata](#), která se vztahují k celému dokumentu. Definují např. název dokumentu, jazyk, [kódování](#), [klíčová slova](#), popis, použitý styl zobrazení. Hlavička je uzavřena mezi značky `<head>` a `</head>`. Element `head` je opět povinný, ale jeho otevírací a koncová značka povinná není, prohlížeč ji sám doplní podle kontextu.
- Tělo dokumentu – obsahuje vlastní text dokumentu. Vymezuje se značkami `<body>` a `</body>`. Element `body` je povinný, ale jeho otevírací a koncová značka povinná není, prohlížeč ji sám doplní podle kontextu.

Příklad zdrojového kódu

Příklad HTML dokumentu ve verzi 4.01:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<html>
  <!-- toto je komentář -->
  <head>
```

```
<title>Titulek stránky</title>
</head>
<!-- tělo dokumentu -->
<body>
  <h1>Nadpis stránky</h1>
  <p>Toto je tělo dokumentu</p>
</body>
</html>
```

Druhy značek

Značky lze z hlediska významu rozdělit na tři základní skupiny:

Strukturální značky

rozvrhují strukturu dokumentu, příkladem jsou odstavce (<p>), nadpisy (<h1>, <h2>). Dodávají dokumentu formu.

Popisné (sémantické) značky

popisují povahu obsahu elementu, příkladem je nadpis (<title>) nebo adresa (<address>). Současný trend je orientován právě na sémantické značky, které usnadňují automatizované zpracovávání dokumentů a vyhledávání informací v záplavě dokumentů na webu. Vyvrcholením této snahy je v současné době jazyk [XML](#).

Stylistické značky

určují vzhled elementu při zobrazení, typickým příkladem je značka pro [tučné písmo](#) (). Od tohoto druhu značek se postupně upouští, trendem je používání [kaskádových stylů](#), které vzhled popisují odděleně od obsahu dokumentu. Mezi důvody pro neužívání těchto značek patří především to, že tyto značky jsou orientovány na prohlížení na [obrazovce počítače](#), příliš se však nepočítá s používáním dokumentu jiným způsobem – alternativní prohlížeče pro postihžené (čtečky pro slepce), v mobilních zařízeních a podobně. Kaskádové styly umožňují definovat rozdílné zobrazení pro různá zařízení.

Parsování v prohlížečích

[Webové prohlížeče](#) jsou [programy](#), jejichž účelem je prezentovat dokument na zobrazovacím zařízení – nejčastěji monitoru počítače. Dokument je prohlížečem načítán a prováděna jeho rozklad ([syntaktická analýza](#)) na jednotlivé elementy. Prohlížeč obsahuje tabulku značek, které podporuje.