

# HTML po raz czwarty

Międzynarodowa organizacja World Wide Web Consortium tworzy kolejną wersję języka HTML 4.0. Jej zastosowanie da projektantom większe możliwości formatowania stron, a użytkownikom ułatwi nawigację po zasobach Internetu.

**W** przypadku HTML wszystko funkcjonuje trochę inaczej” powiedział Tim Berners-Lee w komentarzu do projektu wersji 4.0 Hypertext Markup Language (HTML), na temat którego toczą się obecnie liczne dyskusje. Berners-Lee jest twórcą sieci World Wide Web i dyrektorem World Wide Web Consortium (W3C), które dba o to, by Pajęczyna pozostała uniwersalna i niezależna od producentów oprogramowania. Wśród najważniejszych cech HTML 4.0 można wymienić:

- ▶ zestaw znaków rozszerzony o narodowe znaki diakrytyczne;
- ▶ ulepszone możliwości prezentacji i wyboru poszczególnych elementów na stronie WWW;

- ▶ skrypty wbudowane w stronę WWW, reagujące bezpośrednio na polecenia użytkownika;
- ▶ style („style sheets”);
- ▶ możliwość wstawienia na stronę dowolnego typu pliku jako obiektu.

Te i inne zmiany zostały wprowadzone nie tylko z myślą o użytkownikach, ale także o programistach HTML. W3C reaguje w ten sposób przede wszystkim na coraz większe wymagania ze strony właścicieli witryn internetowych, którzy pragną uatrakcyjnić swoje serwisy WWW. Wśród nowych elementów, jakimi dysponuje wersja 4.0, wymienić można między innymi dynamiczne formularze, tabele i ramki.

HTML 4.0 nie jest jeszcze standardem (patrz ramka „Zarząd World Wide Web: W3C”). W3C zaprezentowało go publicznie po raz pierwszy 8 lipca jako projekt nowego języka HTML, a nad jego ostatecznym kształtem ciągle jeszcze toczą się żarte spory. Wśród twórców nowej wersji są także Netscape i Microsoft – producenci najpopularniejszych przeglądarek internetowych. Trudno byłoby znaleźć inną dziedzinę oprogramowania, w której dwaj konkurenci spierają się tak zaciekle. Użytkownicy przyzwyczaili się już do tego, że każdy z producentów tworzy własne rozszerzenia języka HTML, które nie zostały oficjalnie zatwierdzone, i obsługiwane są tylko przez jego przeglądarkę.

Każda z konkurujących firm liczy na to, że to właśnie jej rozszerzenia zostaną przyjęte w nowej wersji HTML 4.0. W tej chwili w wyścigu przewagę uzyskał Microsoft. Wiele nowych mechanizmów wersji 4.0 bazuje na rozwiązaniach zawartych w systemie operacyjnym Windows. Z kolei Netscape obiecuje dostosować swoją przeglądarkę do nowego języka, gdy tylko W3C uzna go za standard. Rzecznik firmy Netscape Simone Droll zapewniła, że nastąpi to jeszcze przed końcem roku, ponieważ Netscape wychodzi z założenia, że w przygotowanym projekcie nie należy już wprowadzać istotnych zmian.

## Ładniejsze strony dzięki stylom

Nowy projekt języka HTML można uznać za prawdziwy przełom. Na szczególną uwagę zasługują style, które pozwalają opisać formaty poszczególnych elementów strony WWW. Znakomicie ułatwiają one programistom pracę nad przygotowaniem serwisu. Do tej pory, aby uzyskać zamierzony efekt, projektanci musieli stosować przeróżne sztuczki: niestandardowe (i nie obsługiwane przez wszystkie przeglądarki) rozszerzenia HTML, teksty w postaci obiektów graficznych albo niewidoczne grafiki i tabele, które pozwalały ułożyć tekst we właściwy sposób.

Teraz problemy te można rozwiązać za pomocą stylów. Z czasem zastąpią one wiele popularnych dotychczas atrybutów, takich jak *align*, *tt*, *i*, *b*, *big*, *small*, *strike*, *s* i *u*. W3C nie definiuje jednoznacznie języka stosowanego w stylach. Przykłady przedstawione w projekcie HTML 4.0 bazują na

Cascading Style Sheets (CSS1). Na tej podstawie można przypuszczać, że style oparte będą właśnie na CSS1.

Style mogą być stosowane w połączeniu z różnego rodzaju etykietami: akapit `<P>`, koniec wiersza `<BR>`, etykiety nagłówków `<H1>` do `<H6>`, etykieta `<BODY>` oraz elementy tabeli: `<TABLE>`, `<TR>`, `<TD>` itd. W zasadzie każdy sposób formatowania tekstu można zdefiniować jako styl. Jeśli na stronie pojawi się etykieta, wówczas występujący po niej tekst zostanie sformatowany zgodnie ze zdefiniowanymi uprzednio atrybutami.

Zarówno style, które zostały zdefiniowane globalnie w nagłówku strony `<HEAD>`, jak i te, które odnoszą się tylko do jednej etykiety, mogą być w razie potrzeby pominięte. Jeśli w nagłówku etykiety `<P>` został zdefiniowany pewien styl, ale w jednym akapicie chcemy zastosować inny format, wówczas jego etykieta powinna wyglądać następująco:

```
<P style="font-size: 18pt; color=red">
...tekst...
```

Tak zdefiniowany atrybut stylu pozwala uzyskać czerwony tekst o wielkości 18 punktów. Bardzo dużym udogodnieniem przy pracy z szablonami jest możliwość zapisu stylu w zewnętrznym pliku. Miejsce przechowywania danych opisujących styl określa atrybut `src`:

```
<STYLE src="formaty.htm">
```

Dzięki temu styl może być tworzony centralnie, bez potrzeby definiowania go oddzielnie dla każdej strony, co stanowi ogromne ułatwienie przy projektowaniu dużych serwisów WWW. Ponadto takie rozwiązanie pozwala efektywnie podzielić zadania pomiędzy redaktorami przygotowującymi treść serwisu a programistami i projektantami, którzy nadają stronom odpowiedni wygląd. Na stronie XXX w ramce „Ramki inline to nowe możliwości tworzenia stron” zobaczymy, w jaki sposób stosuje się szablony za pośrednictwem etykiet `<STYLE>`.

### Nowe ramki na środku strony

Nowy rodzaj ramek o nazwie „inline frames” to jedna z najbardziej rzucających się w oczy cech tej wersji HTML. W gruncie

rzeczy dopiero wersja 4.0 Hypertext Markup Language zawiera oficjalną definicję ramek, chociaż są one bardzo popularnym elementem, wykorzystywanym na stronach WWW, a W3C jedynie oficjalnie zaakceptowało istniejący de facto standard.

Ramki inline umieszczone na środku strony WWW, zostały rozpowszechnione przez Microsoft. Do ich definicji służy nowa etykieta `<IFRAME>`:

```
<IFRAME src="przyklad.html"
width="400" height="500">
```

Atrybut `src` wskazuje na plik HTML, którego zawartość ma być widoczna w ramce inline. `Width` i `height` określają rozmiary ramki w punktach, natomiast `frameborder` grubość linii.

### Małe, ale istotne zmiany w tabelach

Również table zostały w nowej wersji poprawione. Jednym z udogodnień jest możliwość stosowania w tabelach różnorodnych krawędzi. W3C zaproponowało w tym celu wprowadzenie dla etykiety `<TABLE>` atrybutu `frame`, który może przyjmować następujące wartości:

- ▶ `void`: brak obramowania
  - ▶ `above`: linia powyżej
  - ▶ `below`: linia poniżej
  - ▶ `hsides`: linie powyżej i poniżej
  - ▶ `vsides`: linie po obu stronach
  - ▶ `lhs`: linia po lewej stronie
  - ▶ `rhs`: linia po prawej stronie
  - ▶ `box`: linie ze wszystkich stron
  - ▶ `border`: również linie ze wszystkich stron
- Innym ciekawym rozwiązaniem jest możliwość grupowania kolumn tabeli. W tym celu na początku tabeli definiuje się za pomocą etykiety `<COLGROUP>` liczbę kolumn należących do jednej grupy:

```
<TABLE>
  <COLGROUP span=1>
  <COLGROUP span=2 bgcolor=red>
  <COLGROUP span=1>
...tabela...
</TABLE>
```

W powyższym przykładzie zdefiniowaliśmy czterokolumnową tabelę, w której dwie środkowe kolumny należą do jednej grupy. Grupy tworzone są kolejno od lewej do prawej strony za pomocą etykiety

## podstawy

### HTML 4.0 w skrócie

**Formularze:** Przyciski opatrzone tekstem lub grafiką, pola przeznaczone tylko do odczytu, możliwość grupowania pól, przypisywanie etykiet i skrótów klawiaturowych poszczególnym obiektom oraz tytułów każdemu elementowi aktywnemu w formularzu.

**Ramki:** Oprócz znanych dotychczas ramek można w dowolnym miejscu na stronie WWW umieszczać ramki *inline*.

**Tabele:** Tabele można łączyć w grupy, a obramowania tabeli mogą przybierać różnorodne formy. Daje to większe możliwości projektantom stron WWW.

**Obsługa obiektów i skryptów:** Etykieta `<OBJECT>` pozwala programiście umieszczać na stronie dowolnego typu dane. Skrypty pozwalają użytkownikowi na prowadzenie dialogu, który sterowany jest przez klienta, a nie przez serwer.

**Szablony stylów (Style Sheets):** Pozwalają tworzyć wzorce formatów dla całych akapitów lub pojedynczych wyrazów oraz umożliwiają rozdzielanie prac redakcyjnych i edytorskich.

**Skróty klawiaturowe (Shortcuts):** Wszystkim elementom dialogowym, takim jak pola tekstowe, odsyłacze, przyciski i pola wyboru można przypisać skróty klawiaturowe. W ten sposób możliwe jest poruszanie się po Internecie za pośrednictwem klawiatury. Technika ta ułatwi korzystanie z Internetu przede wszystkim ludziom niepełnosprawnym i zgodnie z wolą W3C „Przybliży WWW milionom nowych użytkowników”.

**Internacjonalizacja:** Autorzy stron internetowych będą teraz mogli przygotowywać serwisy we wszystkich językach świata. Aby to umożliwić, HTML 4.0 wykorzystuje międzynarodowy zestaw znaków zgodny z ISO 10646. Ponadto w dokumencie HTML będzie można pisać teksty od strony prawej do lewej, tak

## podstawy

## Tabele ze stylami i grupami kolumn

HTML 4.0 zapewnia programistom znacznie większe niż dotychczas możliwości formatowania tabel. W stworzonej tu przykładowej tabeli ostatnie dwie z pięciu kolumn zostały za pośrednic-

twem czwartej etykiety `<COLGROUP>` przypisane do jednej grupy (różowy fragment listingu). Aby te dwie kolumny nie były rozdzielone linią, w etykiecie `<TABLE>` został podany atrybut `rules="groups"`.

Wygląd nagłówek kolumn określony został w czerwonym fragmencie listingu. Atrybut `style` łączy wszystkie nagłówki ramką oraz określa szary kolor tekstu i zielony kolor tła.

```
<TABLE border=5 rules="groups">
```

```
<COLGROUP span=1 width="60" bgcolor=orange>
```

```
<COLGROUP span=1 width="250">
```

```
<COLGROUP span=1 width="170">
```

```
<COLGROUP span=2 width="80">
```

```
<TR align=center style="border: solid; color=silver" bgcolor=gray>
```

```
<TH>Rang</TH><TH>Ger&auml;t</TH><TH>Processor</TH>
```

```
<TH>RAM</TH><TH>Festplatte</TH>
```

```
</TR>
```

```
<TR align=center bgcolor=red style="color: yellow">
```

```
<TD>1</TD>
```

```
<TD align=left>&nbsp;&nbsp;&nbsp;Tuncer TPX166 MMX</TD>
```

```
<TD>Pentium 166 MMX</TD><TD>32 MB</TD>
```

```
<TD>2,1 GB</TD>
```

```
</TR>
```

```
<TR align=center>
```

```
<TD>2</TD>
```

```
<TD align=left>&nbsp;&nbsp;&nbsp;Gateway P5-166 Elite</TD>
```

```
<TD>Pentium 166 MMX</TD><TD>32 MB</TD>
```

```
<TD>2,4 GB</TD>
```

```
</TR>
```

```
... reszta tabeli ...
```

Rang	Ger&auml;t	Processor	RAM	Festplatte
1	Tuncer TPX166 MMX	Pentium 166 MMX	32 MB	2,1 GB
2	Gateway P5-166 Elite	Pentium 166 MMX	32 MB	2,4 GB
3	Compaq Presario 1000	Pentium 166 MMX	32 MB	2,0 GB
4	Wahlbergl Phoenix	Pentium 166 MMX	32 MB	2,0 GB
5	Compaq Presario 1000	Pentium 166 MMX	32 MB	2,0 GB

`<COLGROUPS>`. Atrybut `span` określa, ile kolumn wchodzi w skład danej grupy. Pojedyncze kolumny należy również traktować jako grupy, dlatego pierwsza kolumna została zdefiniowana jako pierwsza grupa, druga i trzecia kolumna jako druga grupa, a ostatnia kolumna jako trzecia grupa. Jeśli następnie dla wybranej grupy określimy kolor tła, będzie on obowiązywał dla wszystkich komórek z kolumn należących do tej grupy. Dokładny opis tych funkcji przedstawiony jest w ramce „Tabele ze stylami i grupami kolumn”.

Drugi atrybut, `rules`, definiuje obramowanie pojedynczych komórek. Dopuszczalne są tutaj wartości:

- ▶ *none*: brak linii pomiędzy komórkami
- ▶ *groups*: linie poprowadzone są pomiędzy grupami kolumn
- ▶ *rows*: linie pomiędzy wierszami tabeli
- ▶ *cols*: linie pomiędzy kolumnami tabeli
- ▶ *all*: linie pomiędzy wszystkimi komórkami. Odpowiada to wartości domyślnej, która przyjmowana jest wtedy, kiedy atrybut nie jest zdefiniowany.

Podobnie jak w dokumencie HTML również w tabeli można wyróżnić część nagłówek, część zasadniczą oraz stopkę przeznaczoną na komentarze i przypisy. Oznaczone są one etykietami `<THEAD>` (nagłówek tabeli), `<TBODY>` (część zasadnicza) i `<TFOOT>` (stopka). Wartość `<TFOOT>` musi być zdefiniowana wcześniej niż `<TBODY>`, tak aby przeglądarka wiedziała, w jaki sposób ma interpretować tabelę:

```
<TABLE>
  <THEAD>
    <TR>...nagłówek tabeli </TR>
  </THEAD>
  <TFOOT>
    <TF>...stopka tabeli </TF>
  </TFOOT>
  <TBODY>
    <TR>...pierwszy wiersz tabeli </TR>
    <TR>...drugi wiersz tabeli </TR>
  </TBODY>
  <TBODY>
    <TR>...pierwszy wiersz tabeli </TR>
```

```
<TR>...drugi wiersz tabeli </TR>
<TR>...trzeci wiersz tabeli </TR>
</TBODY>
</TABLE>
```

## Łatwe poruszanie się w wirtualnej przestrzeni

Bardzo ciekawą nowością wersji 4.0 jest możliwość dotarcia do elementów strony WWW za pomocą klawiatury. W3C wprowadziło w tym celu atrybut `accesskey`. Obowiązuje on między innymi dla etykiety zakotwiczenia `<A>`, która definiuje odsyłacz (link) do innej strony WWW, nadając dokumentom HTML hipertekstowy charakter:

```
<A accesskey="P" href="pomoc.htm">
Pomoc </A>
```

Napis *Pomoc* został zdefiniowany jako odsyłacz do strony `pomoc.htm`. Atrybut `accesskey` określa literę „P” jako skrót klawiaturowy do tego odsyłacza. Użytkownik nie musi klikać myszką napisu *Pomoc*, ▶ 133

zamiast tego może po prostu nacisnąć klawisz [Alt]+[p].

Jako skróty klawiaturowe wybiera się z reguły rzadko używane kombinacje przycisków. W przypadku systemu operacyjnego Windows HTML 4.0 wymaga kombinacji z klawiszem [Alt].

Autor strony WWW musi pamiętać, aby wyraźnie zaznaczyć, jaki klawisz odpowiada danemu skrótowi i wyróżnić odpowiednią literę. Wprawdzie Windows automatycznie podkreśla teksty, które zostały zdefiniowane jako odsyłacze, jednak w przypadku skrótów klawiaturowych programista musi sam się o to zatroszczyć. Aby w podanym wcześniej przykładzie litera „P” została, zgodnie z przyjętymi zasadami, podkreślona, program należy zmodyfikować w następujący sposób:

```
<A accesskey="P" HREF="pomoc.htm">
<u>P</u>omoc </A>
```

Teraz pod literą „P” w słowie „Pomoc” pojawi się podkreślenie. Pojawia się tu jednak problem z wyróżnieniem litery uaktywniającej skrypt. W przypadku odsyłacza tekstowego podkreślany jest cały wyraz, tak więc podkreślenie litery P nic nie da, gdyż i tak cały wyraz będzie w ten sposób wyróżniony. Pozostaje nam jedynie z zaciekawieniem oczekiwać, jaka konwencja przyjmie się tu z biegiem czasu.

### Obiekty pozwalają wstawiać dowolne typy plików

Strony WWW mogą oprócz tekstu zawierać grafiki, programy Javy, pliki dźwiękowe i inne obiekty. Do umieszczania na stronie grafik od dawna stosuje się etykietę <IMG>. Również etykieta <APPLET> pozwalająca włączyć do dokumentu HTML programy Javy jest obecnie powszechnie używana.

Aby dla każdego typu obiektu nie trzeba było tworzyć nowej etykiety, HTML

4.0 wprowadza nową etykietę <OBJECT>. Za jej pomocą można umieścić na stronie dowolny element wraz z programem do jego obsługi.

Również plik graficzny GIF można wstawić do dokumentu HTML, posługując się etykietą <OBJECT> zamiast tradycyjną <IMG>. Poniższy przykład pokazuje, w jaki sposób wykorzystać etykietę <OBJECT> do wstawienia na stronę GIF-a. Tradycyjne rozwiązanie:

```
<IMG src="logo.gif" alt="Tutaj znajduje się logo.">
```

można teraz zastąpić następującym wyrażeniem:

```
<OBJECT data="logo.gif"
type="image/gif">
Tutaj znajduje się logo.
</OBJECT>
```

► 134

## podstawy

### Ramki inline to nowe możliwości tworzenia stron

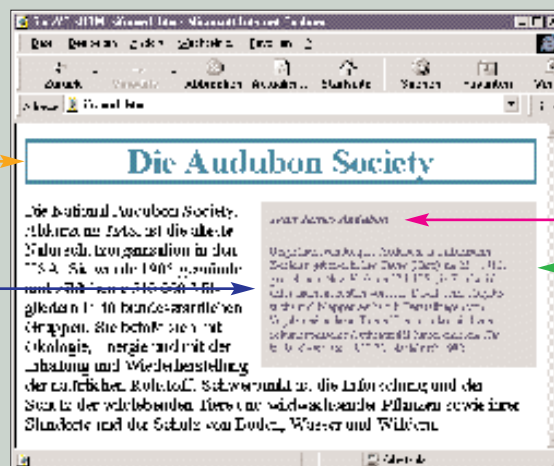
Nowe ramki inline dają programiście znacznie większą swobodę podczas edycji stron WWW. Za ich pomocą można na przykład stworzyć okienka

tekstowe, które otaczane są przez tekst główny. Pozwalają one bez trudu uzyskać efekty, które wcześniej wymagały posługiwania się skomplikowanymi

tabelami. Poniższy przykład pokazuje zastosowanie ramek inline oraz przedstawia, w jaki sposób zastosować szablon stylu do tekstu w ramce.

Listing programu głównego:

```
<HTML>
<HEAD>
<STYLE>
H1 {border: solid;
text-align: center;
color: teal;}
</STYLE>
</HEAD>
<BODY>
<H1>Die Audubon Society</H1>
<IFRAME src="iframe2.htm"
frameborder=0
marginwidth=10
marginheight=10
height=180 align=right>
</IFRAME>
Die National Audubon Society
(...) und der Schutz von
Boden, Wasser und
Wäuml;ldern.
</BODY>
</HTML>
```



Listing okienka tekstowego (iframe2.htm):

```
<HTML>
<HEAD>
<STYLE>
h1 {font-style:
italic;
font-size: 14}
body {font-size: 12;
color: navy}
</STYLE>
</HEAD>
<BODY bgcolor=silver>
<H1> John James Audubon
</H1>
Urspruuml;lich Jean
Jacques Audubon,
amerikanischer Zeichner
(...) sind von
dokumentarischer
Authentizitauml;t (unter
anderem <i>The birds of
America</i>, 1827-38;
Nachdruck 1982).
</BODY>
</HTML>
```

Nowa etykieta `<OBJECT>` pozwala połączyć ze stroną WWW również zewnętrzne programy obsługi danych. Oznacza to, że `<OBJECT>` może uruchomić program na serwerze WWW, a następnie przekazać mu dane, które mają być przedstawione. Program taki można porównać do zdalnego plug-ina, udostępnianego przeglądarce.

Na pytanie, w jakim języku powinny być pisane programy tego rodzaju, W3C nie udzieliło do tej pory żadnej odpowiedzi. Prawdopodobnie producenci przeglądarek zostaną poproszeni o wyposażenie swoich produktów w odpowiednie interpretery, podobnie jak to się stało z interpreterami Javy.

W praktyce nie będzie to prawdopodobnie miało na razie większego znaczenia. Dopiero gdyby razem z danymi trzeba było przesyłać programy obsługi, oznaczałoby to kolejną próbę wydajności dla Internetu.

### Skrypty ożywiają strony WWW

W HTML 4.0 więcej uwagi poświęcono problemowi skryptów. Wprawdzie w specyfikacji HTML 3.2 wymieniana jest etykieta `<SCRIPT>`, jednak nie została ona ściśle zdefiniowana. World Wide Web Consortium zarezerwowało ją jedynie „dla przyszłego wykorzystania z językami skryptowymi”.

Przeglądarki firm Netscape i Microsoft od dawna wykorzystują tę etykietę i każdy z potentatów próbował ustanowić własny quasi-standard. W3C uznało za obowiązujący standard język *JavaScript* firmy Netscape. Napisany w tym języku skrypt rozpoczyna się następująco:

```
<SCRIPT language="JavaScript">
```

W HTML 4.0 przewidziano dla etykiety `<SCRIPT>` atrybuty *type*, *language* i *src*. *Type* i *language* określają język, w jakim napisany jest skrypt, przy czym *language* zachowany został jedynie jako model opracowany wcześniej przez Netscape'a. Definicja skryptu za pośrednictwem *type* wygląda następująco:

```
<SCRIPT type="text/JavaScript"
  src="http://www.mydomain.com/
  ascript.html">...
</SCRIPT>
```

## W3C

### Tak funkcjonuje W3C

World Wide Web Consortium, w skrócie W3C, jest konsorcjum przemysłowym, co oznacza, że w jego skład wchodzi przedstawiciele różnych firm produkcyjnych. Obecnie organizacja ta liczy ponad 180 członków.

Dyrektor organizacji nie jest jednak menedżerem dbającym o interesy firmy, ale naukowcem. Stanowisko to piastuje Tim Berners-Lee, który w 1990 roku stworzył World Wide Web. Razem ze swymi współpracownikami czuwa on nad tym, aby sieć WWW rozwijała się jako otwarty standard, niezależny od żadnego producenta oprogramowania.

Prace W3C koordynowane są w USA przez MIT Laboratory for Computer Sciences (LCS) (<http://www.w3.org>), we Francji przez Institut National de Recherche en Informatique (INRIA) (<http://www.inria.fr>) i w Japonii przez Uniwersytet w Keio (<http://www.keio.ac.jp>).

Nowe specyfikacje muszą zawsze być zatwierdzone jednogłośnie. Do tej pory opracowano pięć takich specyfikacji nazywanych „zaleceniami”. Zanim nowy standard (taki jak HTML 4.0)

zostanie przyjęty przez społeczność internetowa, a przede wszystkim W3C, dyskutuje się o nim na łamach grup dyskusyjnych. Udział w dyskusji może wziąć w zasadzie każdy.

Jako projekt otwarty („public working draft”) po upływie określonego czasu przedstawiane są w Internecie wyniki prac W3C do dalszej dyskusji. Projekt jest już wstępnie opracowany, a ewentualne uwagi i propozycje mogą być uwzględnione podczas dalszych prac nad standardem.

Zalecenie W3C („W3C recommendation”) jest kolejnym etapem w procesie standaryzacji, na którym podejmowane są ostateczne decyzje. Pomimo przyjęcia terminu „zalecenie” dokumenty tego typu uznawane są w środowisku World Wide Web za standard. Zgodnie z nimi rozwija się Web, a producenci powinni go ściśle przestrzegać jeśli chcą, aby ich produkty, takie jak przeglądarki WWW, były zgodne z normami obowiązującymi w Internecie.

Prócz prac standaryzacyjnych W3C udostępnia w sieci WWW obszerny zasoby informacji, z których korzystać mogą użytkownicy i programiści.

Atrybut *src* wskazuje na inny plik HTML, zawierający właściwy skrypt. Ponieważ skrypt nie musi być zawarty w dokumencie HTML, z którego jest wywoływany, można w łatwy sposób oddzielić program skryptu od kodu HTML, dzięki czemu nawet rozbudowane strony stają się bardziej przejrzyste.

Wartości, jakie przyjmuje *type*, powinny zgodnie z konwencją „Internet Media Types” odpowiadać multimedialnym typom stosowanym w Internecie. Oznacza to, że najpierw podawany jest typ MIME w jakim zapisany jest obiekt – na przykład *text*, a następnie po znaku „/” podgrupa. Podgrupa służy do określenia języka skryptowego. Taki sposób notacji pozwoli w przyszłości stosować inne niż czysto tekstowe typy MIME – na przykład skompilowane pliki binarne lub bajtowe kody Javy.

Projekt języka HTML 4.0 mówi o trzech rodzajach typów i proponuje dla

nich następujące oznaczenia: *TCL* (*text/tcl*), *JavaScript* (*text/javascript*), *Visual Basic Script* (*text/vbscript*).

HTML 4.0 przewiduje również możliwość zdefiniowania zdarzeń, których zajście będzie powodowało uruchomienie przez przeglądarkę określonego skryptu. Zdarzenia te będą dotyczyły operacji myśzą, rozpoznawania znaków z klawiatury, ładowania i zamykania stron WWW lub powiększania wybranych elementów strony. Wykorzystują one takie elementy, jak `<INPUT>`, `<SELECT>`, `<BUTTON>`, `<TEXTAREA>` i nową etykietę `<LABEL>`.

### Unicode czyni Pajęczynę międzynarodową

Pracując nad HTML 4.0, W3C potraktowało wreszcie poważnie problem dostosowania WWW do wymogów międzynarodowych. Spełnienie tego warunku jest ► 137

niezbędne, aby Internet można było traktować jako sieć globalną. Istnieje już techniczna możliwość zestawienia połączenia z Internetem w dowolnym miejscu na Ziemi, brakuje jednak technologii, która pozwoliłaby użytkownikowi obejrzeć w swojej ulubionej przeglądarce stronę napisaną w dowolnym języku.

Jak do tej pory największym ograniczeniem jest konieczność przesyłania znaków przez Internet w postaci kodów ASCII. Skrót ASCII pochodzi od American Standard Code of Information Interchange i jest standardem opisującym sposób reprezentacji znaków w systemach wymiany danych. Każdemu znakowi, jaki znajduje się na stronie WWW, odpowiada określona liczba, będąca jego kodem.

Za pośrednictwem kodów ASCII można jednak przedstawić tylko 128 znaków. Po uwzględnieniu wszystkich liter, cyfr i specjalnych znaków sterujących ASCII okazuje się, że o ile liczba ta jest wystarczająca do przedstawienia wszystkich liter alfabetu anglosaskiego, o tyle zapis polskich ogonków, niemieckich przegłosów, francuskich akcentów i skandynawskich liter, takich jak Å lub Ø, jest o wiele bardziej skomplikowany. Z tego względu powszechnie stosuje się obecnie zapis 8-bitowy, który pozwala zapisać 256 różnych znaków.

Pośród tych 256 znaków jedynie pierwszych 128 zgodnych jest ze standardem ASCII, kodowanie pozostałych nie zostało ujednoczone. Najgorsze jest to, że w obszarze tym leżą wszystkie znaki narodowe, a ponieważ nie obowiązuje tu żadna norma, temu samemu kodowi w różnych systemach operacyjnych odpowiadają często różne znaki.

W Internecie jak na razie obowiązuje jedynie standard ASCII i rozpoznawanych jest tylko 128 znaków. Jednak HTML proponuje w tym wypadku pewne rozwiązanie. Znaki specjalne przedstawiane są w programie HTML jako kombinacja podstawowych znaków ASCII. I tak ó może być zapisane jako „&okreska;”, a Ø jako „&Oslash;”. Wyraz „wzgórze” zapisany w takim kodzie HTML miałby postać „wzg&okreska;rze”.

Pomimo zastosowania kodowania HTML wciąż jeszcze wiele znaków specjalnych nie ma swojej reprezentacji, dotyczy to przede wszystkim znaków

narodowych kultur euroazjatyckich. Dopiero HTML 4.0 porządkuje tę sytuację. W jego specyfikacji zawarty jest opis kodowania znaków na podstawie normy ISO 10646.

Zgodnie z tą normą znaki reprezentowane są za pośrednictwem 31 bitów, co daje ponad dwa miliardy ( $2^{31}$ , a więc 2147483648) dostępnych znaków. Dochodzi do tego możliwość definiowania kierunku pisma, co pozwala na zapis tekstu w języku hebrajskim lub arabskim ze strony prawej do lewej.

Podstawowy język strony WWW definiuje się za pomocą atrybutu *lang*, który przyjmuje dwuliterowe wartości. Aby określić podstawowy język dokumentu jako francuski należy dodać atrybut *lang* do etykiety HTML:

```
<HTML lang="fr">
  <BODY>
    ..tekst po francusku...
  <P lang="es"> ..tekst po hiszpańsku...
  <P lang="fr"> ..tekst po francusku...
  </BODY>
</HTML>
```

Atrybut *lang* można stosować w odniesieniu do niemal wszystkich elementów tekstowych języka HTML. W połączeniu z etykietą *P* możliwa jest zmiana języka jedynie dla jednego akapitu na stronie WWW.

Dwuliterowe kody języka można uściślić, podając po myślniku nawet oznaczenie dialektu. Kody języków rejestrowane są przez IANA (Internet Assigned Numbers Authority) – organizację, która sprawuje kontrolę nad poprawnością adresów internetowych i nazw domen.

O tym, czy tekst pisany jest ze strony lewej do prawej, czy też z prawej do lewej decyduje wartość atrybutu *dir*. Atrybut *dir* może przyjmować wartości *rtl* („right to left”) lub *ltr* („left to right”) i można go stosować w połączeniu z niemal każdą etykietą. Jego wprowadzenie znacznie ułatwia pisanie w takich językach jak hebrajski lub arabski.

Jeśli chcemy, aby na polskojęzycznej stronie WWW tekst następnego akapitu pisany był od prawej do lewej, możemy to zrobić w następujący sposób:

```
<P dir="rtl">
```

Aby w następnym akapicie powrócić do normalnego stanu rzeczy, należy wpisać:

```
<P dir="ltr">
```

## HTML 4.0 zamienia WWW w gazetę online

Centralnym punktem projektu języka HTML 4.0 są szablony stylów. Ponadto do HTML wpisano oficjalnie ramki, uzupełniając je ramkami inline, które można umieszczać dowolnie na stronie WWW. Podsumowując, można powiedzieć, że twórcy stron internetowych mają znacznie większe możliwości edycji. Osadzone obiekty, języki skryptowe oraz rozszerzony zestaw znaków dopełniają całości projektu.

Brian Kelly z *UK Office for Library and Information Networking* w Anglii dostrzega wyraźne korzyści, wynikające z coraz większej współpracy pomiędzy producentami oprogramowania: „W ostatnich latach konkurencja pomiędzy producentami software’u hamowała rozwój WWW. Teraz, dzięki temu że Internet staje się coraz bardziej „unormowany”, urzeczywistnia się marzenie Tima Berners-Lee: uniwersalny dostęp do zasobów sieciowych w sposób niezależny od platformy sprzętowej i systemowej”.

Możemy się pocieszyć, że nowy Hypertext Markup Language pojawi się na licznych serwerach WWW natychmiast po tym, jak tylko zostanie on uznany za standard. Użytkownicy – przynajmniej ci, których przeglądarki obsługują wersję HTML 4.0 – będą z tego faktu z całą pewnością zadowoleni. Dzięki HTML 4.0 graficzne rozrywki internetowe będą w przyszłości bardziej atrakcyjne, a dane, jakie będziemy przy tej okazji ściągać, będą miały znacznie mniejszą objętość.

oprac. Marcin Pawlak (jp)

## info

Informacje o HTML 4.0 znaleźć można pod następującymi adresami:

### Download:

Konsorcjum WWW:  
<http://WWW.w3.org/>  
 Microsoft: <http://WWW.microsoft.com/>  
 Netscape: <http://WWW.netscape.com/>  
 ZDNet Anchordesk:  
<http://www.zdnet.com/anchordesk/mad/>