

HTML が記述する構造と表現

タグがもつ二つの意味

HTML はもともとは文書の構造を示すために作られた技術から生まれている。例えば<H1>..<</H1>というタグはそれに囲まれた部分が「章」の見出しであることを意味しており、<H2>..<</H2>というタグは、もう一つ小さな区分の「節」の見出しであることを示す。（「部」と「章」でも、ともかくそういう関係を示す。）WWW ブラウザはこれを視覚的にわかりやすく表現しようとする。例えば多くの場合は、見出しらしく大きく太めの文字で、左に寄せ、前後に一行程度の空間をあけて文字を表示する。また、H1とH2の上下関係を表現するために、<H1>を大きく、<H2>を少し小さく表示する。

ただしこうしたブラウザがやってくれる「見た目」の処理は厳密にはHTMLの仕様には規定されていないため、各ブラウザが「これなら判りやすそうだ」と思う方法を用意している。H1 などのような Headings タグについてはほぼ表現方法が揃っているが、揃っておらず、ブラウザによって見え方が異なるタグもある。

つまり、HTML は文書の構造を記述するために作られており、見せ方は WWW ブラウザに任されている。ところが、HTML には..<（ボールド、太文字指定）や FONT タグのように、見た目をコントロールするためのタグも HTML には存在する。これは構造とは関係がない。

結論：HTML には文書構造を記述するタグと、見た目を指定するタグの両方が混在している。

これを理解しないで Web page を作ると、例えば以下のような間違いをしがちである。

- ・見た目のきれいさを過度に追求する
ブラウザによって見え方は異なるため、あまり意味がない。
- ・構造を記述するタグを見た目の調整のために使う
機械可読なデータを作っている意味が半減する

それぞれについて説明する。

見た目の違い

右は代表的なタグの一つである H1~H6 の表示結果が Netscape Communicator と Internet Explorer で微妙に異なることを示した例である。

画像の表示サイズはどちらのブラウザでも同じであるから、結果的に絵と文字の大きさのバランスを厳しく求め、調整しても、自分が使っているブラウザと違うブラウザで表示させた場合にその調整が崩れる場合が多い。

この例は共に Windows 環境で見ているが、Macintosh 上のブラウザではさらに違いがあるかもしれない。



他にも、これは HTML の問題ではないが、Netscape Communicator ではビットマップ形式の画像が表示されず、Internet Explorer ならば表示される、という問題を前の週で示した。

例えば右の例は IMG タグに、下記のように ALT パラメータを指定している。





```
<IMG ALT="ビットマップ" SRC="tori.bmp">
```

この結果、画像が表示されない場合には、そこに ALT タグで指定した文字が表示されるため、利用者は画像についてのヒントを得ることができる。こうした機能を利用して、より多くの環境の利用者にやさしい、表示が多少異なっても何とか意味は汲めるように心がけるのがよい。

場所: <http://www.kyoto-su.ac.jp/~yasuda/kiso2/2001tue12/11/graphics.html>

比較

Internet Explorer で表示結果がどのようになるか 見比べてみると良い。
では Netscape comm. は bitmap 形式を扱えない

			
(低圧縮) Bytes	JPEG (高圧縮) 2151 Bytes	PNG 4483 Bytes	BMP 11978 Bytes

スクリーンサイズもマシンによって大きく違うし、色が表示されるとも限らない。(極端な例として i-mode や l-mode を想像するとよい。)

そうした意味も含めて、「より多くの人にとって読みやすい」ページデザインを考えることが重要である。色、スクリーンサイズ、画像フォーマットなどによって見るができなくなる人が増えないよう配慮すること。例えば 10 号館の入り口には Netscape Navigator が動作する Macintosh がある。ここで自分のページを確認してみるのも効果的だろう。

何故構造を記述するのか

文書中に構造を記述することによって、自然にその文書が整形され見やすく表示される、というアイデアが HTML と WWW ブラウザの出発点である。しかし構造を示すのは表示のためだけではない。

- ・例えば <H1>などの Heading タグを自動的に抽出してやることで、簡単に目次情報を作ることができ、それをユーザーに見せることが可能になる(*1)。
- ・そして、WWW 文書を読むのは人間だけではない。例えば Google などのいわゆる検索エンジンの多くはコンピュータプログラムによって自動的に WWW 文書を見て回り、そこから情報を抽出してデータベースに記録する。その時検索にヒットした文字が本文中の一語なのか、見出しやタイトルの一語なのかによって、重要度は大きく違って来るだろう。

<H1>タグを使わず、ただタグを利用して文字サイズを大きくしたりするだけではこうしたことはできない。近い将来、<H1>などの Heading 指定に応じて「1.1, 1.2, 2.1」などと章節番号(セクション番号)を振っていく新しいブラウザが登場するかもしれない。そうした機能を持つソフトウェアが人気を博したとき、その普及速度は非常に速い。HTML のこうした特性を理解して記述しておくことが重要な理由である。

*1 HTML 4.01 の仕様書 「自動的に目次を生成するのに使われる可能性がある」

<http://www.w3.org/TR/html4/struct/global.html#edef-H1>

結論

- ・構造の記述はできている方がよい
- ・見た目が読みやすいに越したことはない
- ・ある程度これが両立する Web page を作ろう

HTML はまだ発展途上のものであり、規格も、使われ方も変わっていくだろう。基本的な WWW の精神や目的をわすれず、柔軟に変化に追随するように心がけて欲しい。ただ規格を守ればいいわけではない、ただきれいいに見えていれればいいわけでもない。立ち上がり時期で、変化の激しい今のインターネット環境の中で、自分の視点を高くもつことの重要さを理解して欲しい。