

文字化け対策の裏技

古くから使われてきた文字化け対策に、`<!--京-->`を挿入するという方法がある。文字コードがEUC-JPのページは、これで文字化けを防ぐことができる。

なぜこの`<!--京-->`^{※1}が文字化け対策になるかという、Shift_JISコードの特性に関する。

▶ Shift_JISには存在しない文字

日本語は2バイト文字と言って、コンピュータ内では2つの値の組み合わせで文字を表現している。EUC-JPで、「京」という文字は後半の2バイト目が「0xFE」という値になっている。

では、Shift_JISではどうかというと、2バイト目に「0xFE」という値が使われる文字は存在しない。

つまり、EUC-JPの「京」をページ内を書いておけば、ブラウザは勝手に「Shift_JISにない文字を使っているからEUC-JPだ!」と判断してくれるわけだ。

もちろんShift_JISのページであれば「京」の文字はShift_JISのコードになっているので、特に問題が起こるわけではない。あくまでEUC-JPの文字化け対策に使えるというだけだ。

▶ イマドキではないテクニック

この方法、2007年末まではYahoo Japanでも使われていた方法だが、今となってはさすがにちょっと古い。そもそもEUC-JPを指定してWebページ制作をするケースは現在少なくなりつつあり、UTF-8へと移行してきている。また`<meta>`タグ内で文字コード指定をしておけば通常文字化けを起こすこともほぼなくなっている、今後こういった方法を使うことはなくなるだろう。

CSSの記述などにも言えることだが、裏技的な手法はメンテナンス性も含めて今ではあまりメリットがない。その昔はブラウザの仕様などもまちまちで、正しいページ作りをしても表示が崩れたり文字化けすることも多かった。

現在、OSやブラウザは昔に比べ大きく進化した。今後はいかにスマートな方法でページを作るかが重要になる。

! EUC-JP文字化け対策単語

「京」意外にも、EUC-JP用の文字化け対策単語はいくつかある。
 ・雀の往来
 ・有朋自遠方来
 ・美乳
 などが有名だ。

※1 `<!--京-->`のようにコメント扱いにするのは、ブラウザで表示されないようにするため。これをHTMLの最上部付近に書くことでEUC-JPと判断される。

適正な文字コードは

4つある文字コードのうち、どれを採用するかは迷うところだ。それぞれの文字コードにはメリットとデメリットがあるので、慎重に検討することをおすすめする。

▶ それぞれの特徴

各文字コードの特徴はNo.059でもまとめているが、改めて整理すると以下のようになる。

① Shift_JIS

Shift_JISは文字化けしにくく、携帯電話の文字コードとしても使われている。ただし使える文字数が少ないので、幅広い文字を使う場合には対応が難しくなる。

② EUC-JP

EUC-JPはデータベースで使われていた文字コードなので、データベースと連動したコンテンツ作りには使いやすい。ただ近年ではMySQLなどのWebデータベースもUTF-8に対応している上、データベース内の文字化け問題も解消されていることを考えると、EUC-JPを選択するメリットはあまりないかもしれない。

③ ISO-2022-JP

ISO-2022-JPは通常では特にメリッ

トがないのでWebページで使うことは少ない。

④ UTF-8

UTF-8は使える文字数が格段に多く、さらに多言語対応のため、1つのページに日本語と中国語を表示することなども可能だ^{※1}。大手のWebサイトもUTF-8へ移行してきている例が多い。

▶ できるだけUTF-8で

UTF-8は最も新しい文字コードで、多言語対応の面からも今後はUTF-8が標準になっていくと考えられる。そのため、特に支障がなければUTF-8を選択した方がよい。一時はBlogのトラックバックが文字化けする問題などもあったが、Movable TypeなどのBlogツールは標準がUTF-8のため、それ以外の文字コードを使っている場合に文字化けが起こっていた。これらはUTF-8に統一することで回避できる。

Webサイトは、一度制作するとメンテナンスを繰り返しながら長い期間運用していくことが多い。そのため将来性のある設定作りを心がけた方がよい。文字コードについても、将来を見越してUTF-8を選択することが理想的だと言えるだろう。

※1 中国語フォントがインストールされていないと文字化けする。