

Web アプリケーションでの文字化けの対処

- ・ [Microsoft コードページ 932](#)
- ・ [Unicode](#)

Java

■ HttpServletRequest クラス

```
public void setCharacterEncoding(java.lang.String env)
    throws java.io.UnsupportedEncodingException
```

この要求の本体で使用される文字エンコーディングの名前をオーバーライドします。
このメソッドは、要求パラメータの読み取り、
または `getReader()` を使用した入力の読み取りに先立って呼び出す必要があります。

■ URLEncoder、URLDecoder クラス

[HTML](#) 形式をエンコード、デコードするためのユーティリティクラス。

`String` と `application/x-www-form-urlencoded` MIME 形式 をエンコード、デコードするための static メソッドを含む。

- ・ 「a」から「z」、「A」から「Z」、「0」から「9」、「-」、「_」、「.」、「*」の文字「%」は使用可能だが、特殊なエスケープシーケンスの開始と解釈
- ・ 英数字文字の「a」から「z」、「A」から「Z」および「0」から「9」は元のまま残す
- ・ 特殊文字の「.」、「-」、「*」および「_」は元のまま残す
- ・ プラス記号「+」を空白文字「 」に変換する
- ・ 「%xy」という形式のシーケンスは、バイトの表現として扱う

[JavaScript](#) の `encodeURIComponent`、`decodeURIComponent` に対応

JSP

■ <%@ page contentType="text/html; charset=Shift-JIS" %>

出力の文字エンコーディングを page 指示子を用いて設定します。

■ <%@ page pageEncoding="Shift-JIS" %>

取り込むファイルの文字エンコーディングを指定

Web サーバや Web ブラウザでは、charset は次のような優先順位で処理するように [HTML](#) 仕様 4.0 で規定されています。

1. Content-Type の charset パラメータ ([HTTP](#))
2. http-equiv のある META 宣言の Content-Type や charset の値 ([HTML](#))
3. 要素の charset 属性 ([HTML](#))

Servlet や [JSP](#) の Content-Type で "text/html; charset=Shift_JIS" のように

charset を指定した場合は (1) に相当します

ECMAScript(JavaScript)

■ encodeURIComponent, decodeURI, encodeURIComponent, decodeURIComponent

escape/unescape と違って RFC 準拠 ([RFC 2396 Uniform Resource Identifiers \(URI\): Generic Syntax](http://tools.ietf.org/html/rfc2396))

JavaScript では ECMA-262 3rd Edition に対応するバージョン 1.5 で、これらの関数に対応

JScript はバージョン 5.5 で対応 (MSIE 5.5)

- [encodeURIComponent](#)
- [decodeURI](#)

encodeURIComponent では、"&" や "+" や "=" などの文字をエンコードしないので、GET や POST メソッドで利用するときは、encodeURIComponent を使用する。こちらは、英数以外は、"-_!.~*'" のみがエスケープされない。

■ HTML 特殊文字の処理

■ javascript: を利用するとデコードされる

参照

<http://www.w3.org/TR/html40/appendix/notes.html#non-ascii-chars>

<http://www.cresc.co.jp/tech/java/URLEncoding/JavaScript.URLEncoding.htm>
