Encodings

Zeichenkodierungen und Escape-Funktionen.

- Encodings – Übersicht
- HtmlEncode
- HtmlMask
- ScriptEncode
- UrlEncode
- XmlEncode

Encodings – Übersicht

Bestimmte Tags unterstützen Encodings (Maskierungen) um die Ausgabe in bestimmten Situationen anders aussehen zu lassen. So werden z.B. für HTML spitze Klammern durch ihre entsprechenden HTML-Entites ersetzt oder für JavaScript Steuerzeichen escaped (ein Backslash vorangestellt).

Die meisten Tags unterstützen auch eine Angabe der gewünschten Kodierung via `encoding` Parameter, z.B.: <bx:recordfield.Name encode="..." /> Falls explizit kein Encoding angegeben wurde, versucht das Tag automatisch passen zu encoden (z.B. HtmlEncode bei Mime-Type text/html). Die folgenden Umwandlungen können angegeben werden:

- javascript
- url
- xml
- html
- sql
- sql-like
- sql-rlike
- htmltext
- plain

Informationen zu den einzelnen Encodings:

- HtmlEncode
- HtmlMask
- ScriptEncode
- UrlEncode
- XmlEncode

Auf jeder Seite steht ganz unten, welche Tags oder sonstigen Funktionen diese Umwandlung benutzen bzw. unterstützen.

HtmlEncode

Es werden folgende Ersetzungen vorgenommen:

Originalwert	ausgegebener Wert
`<`	`<`
`>`	`>`
`"`	`"`
`'`	`'`
`&`	`&`

HtmlMask

Hierbei werden sämtliche Zeichen durch HTML-Entities der Form `&#dd;` ersetzt, wobei `d` für eine dezimale Ziffer steht.

Aus `Hallo` wird somit `Hallo`.

ScriptEncode

Einige Sonderzeichen und besondere Zeichen werden durch JavaScript Escape-Sequenzen ersetzt.

Den folgenden Zeichen wird ein \ (Backslash) vorangestellt:

- Newline / Zeilenumbruch (`\n` / 0x0a)
- Carriage return / Wagenrücklauf (`\r` / 0x0d)
- Tab (`\t` / 0x09)
- Single-Quote / Apostroph (`'` / 0x27)
- Double-Quote / Anführungszeichen (`"` / 0x22)
- Backslash (`\` / 0x5c)

Strict / Non-Strict

Standardmäßig ist der Strict-Modus aktiviert, das heißt jeder Backslash wird verdoppelt.

Im Non-Strict-Modus werden Backslashes, die sich vor den oben aufgelisteten Zeichen (außer Single-Quote) befinden nicht verdoppelt (außer der Backslash steht ganz am Ende).

UrlEncode

Es wird die Java Klasse URLEncoder benutzt. Dabei werden "unsichere" Zeichen durch ein oder mehrere Blöcke `%hh` ersetzt, wobei `h` für eine hexadezimale Ziffer steht.

Je nach verwendetem Zeichensatz können die Hexadezimalen Blöcke anders aussehen, so wird ein kleines `ö` z.B. mit ISO-8859-1 zu `%f6`, mit UTF-8 allerdings wird es zu `%c3%b6`.

Es gelten folgende Regeln:

- Die Bereiche a-z, A-Z und 0-9 werden nicht ersetzt
- . (Punkt), - (Minus), * (Stern) und _ (Underscore) werden nicht ersetzt
- Leerzeichen wird durch + (Plus) ersetzt
- Alle anderen Zeichen sind "unsichere" Zeichen und werden in Hexadezimal dargestellt

XmlEncode

Es werden folgende Ersetzungen vorgenommen:

Originalwert	ausgegebener Wert
`<`	`<`
`>`	`>`
`"`	`"`
`&`	`&`
Zeichen zwischen dezimal 32 und 127 (beides inklusive)	*(unverändert)*
Alle restlichen Zeichen	`&#ddd;` d = dezimale Ziffer je nach Plattform, auf der das CMS läuft, können unterschiedliche Werte entstehen