

# Unicode Escape in PHP

PHP requires to implement a simple function to escape unicode. Some servers send unicode encoded string for supporting multi byte characters. Its encoded result will be something like `\uBBF8\uC158 \uC774\uC2A4\uD0C4\uBD88 4: \uC775\uC2A4\uD2B8\uB9BC \uB370\uC774`

function `unicode_escape()` can be implemented as following:

```
function unicode_escape($str, $encoding=null) {
    if (is_null($encoding)) $encoding = ini_get('mbstring.internal_encoding');
    return preg_replace_callback('/\\\\\\u([0-9a-fA-F]{4})/u', create_function('$match', 'return
mb_convert_encoding(pack("H*", $match[1]), ".var_export($encoding, true).", "UTF-16BE");'), $str);
}
```

Below is an example to check the source string by decoding possible options:

```
#!/usr/bin/php
<?php
/**
 * Test Code
 *
 * Author: Chun Kang
 * Date: 2021.11.02
 */

echo "Please type what you want to decode:\n";
$src = readline();

if (!strlen($src))
{
    $src = "\\uBBF8\\uC158 \\uC774\\uC2A4\\uD0C4\\uBD88 4: \\uC775\\uC2A4\\uD2B8\\uB9BC \\uB370\\uC774";
}

echo "\n#Source: {$src}\n\n";

echo "#Investigation Result\n";

$res = urldecode($src);
echo "case 1) urldecode: {$res}\n";

$res = rawurldecode($src);
echo "case 2) rawurldecode: {$res}\n";

$res = utf8_decode($src);
echo "case 3) utf8_decode: {$res}\n";

function unicode_escape($str, $encoding=null) {
    if (is_null($encoding)) $encoding = ini_get('mbstring.internal_encoding');
    return preg_replace_callback('/\\\\\\u([0-9a-fA-F]{4})/u', create_function('$match', 'return
mb_convert_encoding(pack("H*", $match[1]), ".var_export($encoding, true).", "UTF-16BE");'), $str);
}

$res = unicode_escape( $src);
echo "case 4) unicode_escape: {$res}\n";

$res = preg_replace_callback('/\\\\\\u([0-9a-fA-F]{4})/', function ($match) {
    return mb_convert_encoding(pack('H*', $match[1]), 'UTF-8', 'UTF-16BE');
}, $str);

echo "case 5) mb_convert_encoding: {$res}\n";

echo "\n\n";
```

Its result will be like below:

Please type what you want to decode:

\uBBF8\uC158 \uC774\uC2A4\uD0C4\uB888 4: \uC775\uC2A4\uD2B8\uB9BC \uB370\uC774

#Source: \uBBF8\uC158 \uC774\uC2A4\uD0C4\uB888 4: \uC775\uC2A4\uD2B8\uB9BC \uB370\uC774

#Investigation Result

case 1) urldecode: \uBBF8\uC158 \uC774\uC2A4\uD0C4\uB888 4: \uC775\uC2A4\uD2B8\uB9BC \uB370\uC774

case 2) rawurldecode: \uBBF8\uC158 \uC774\uC2A4\uD0C4\uB888 4: \uC775\uC2A4\uD2B8\uB9BC \uB370\uC774

case 3) utf8\_decode: \uBBF8\uC158 \uC774\uC2A4\uD0C4\uB888 4: \uC775\uC2A4\uD2B8\uB9BC \uB370\uC774

case 4) unicode\_escape: 4:

case 5) mb\_convert\_encoding: