

iranphp articles

تبدیل windows-1256 به utf-8

.....

.....

.....

عنوان مقاله :

نگارنده :

آدرس پست الکترونیک :

تاریخ نگارش :

تبدیل windows-1256 به utf-8 :

```
<?php

function convert_w1256_to_utf_8($text) {

    $converted_text=preg_replace_callback("/(:\&\#([0-9]{1,4})\;|([\x80-
\xFF]))/", "replace_content", $text);

    return $converted_text;
}

function replace_content($matches_content) {

    //windows-1256 character set

    $windows1256 = array(
        '€',      # 0x20AC EURO SIGN
        '• #',    'پ'x067E ARABIC LETTER PEH
        '‚',      # 0x201A SINGLE LOW-9 QUOTATION MARK
        'ƒ',      # 0x0192 LATIN SMALL LETTER F WITH HOOK
        '„',      # 0x201E DOUBLE LOW-9 QUOTATION MARK
        '…',      # 0x2026 HORIZONTAL ELLIPSIS
        '†',      # 0x2020 DAGGER
        '‡',      # 0x2021 DOUBLE DAGGER
        'ˆ',      # 0x02C6 MODIFIER LETTER CIRCUMFLEX ACCENT
        '‰',      # 0x2030 PER MILLE SIGN
        '• #',    'ٹ'x0679 ARABIC LETTER TTEH
        '‹',      # 0x2039 SINGLE LEFT-POINTING ANGLE QUOTATION MARK
        'Œ',      # 0x0152 LATIN CAPITAL LIGATURE OE
        '• #',    'ج'x0686 ARABIC LETTER TCHEH
        '• #',    'ز'x0698 ARABIC LETTER JEH
        '• #',    'ڙ'x0688 ARABIC LETTER DDAL
        '• #',    'گ'x06AF ARABIC LETTER GAF
        '‘',      # 0x2018 LEFT SINGLE QUOTATION MARK
        '’',      # 0x2019 RIGHT SINGLE QUOTATION MARK
        '‚',      # 0x201C LEFT DOUBLE QUOTATION MARK
        '”',      # 0x201D RIGHT DOUBLE QUOTATION MARK
        '•',      # 0x2022 BULLET
        '–',      # 0x2013 EN DASH
        '—',      # 0x2014 EM DASH
        '• #',    'ک'x06A9 ARABIC LETTER KEHEH
        '™',      # 0x2122 TRADE MARK SIGN
        '• #',    'ر'x0691 ARABIC LETTER RREH
        '›',      # 0x203A SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
        'œ',      # 0x0153 LATIN SMALL LIGATURE OE
        ' ',      # 0x200C ZERO WIDTH NON-JOINER
        ' ',      # 0x200D ZERO WIDTH JOINER
        '• #',    'و'x06BA ARABIC LETTER NOON GHUNNA
        ' ',      # 0x00A0 NO-BREAK SPACE
        '٫',      # 0x060C ARABIC COMMA
        '¢',      # 0x00A2 CENT SIGN
        '£',      # 0x00A3 POUND SIGN
        '¤',      # 0x00A4 CURRENCY SIGN
        '¥',      # 0x00A5 YEN SIGN
        '¦',      # 0x00A6 BROKEN BAR
```

```
'§',      # 0x00A7 SECTION SIGN
'¨',      # 0x00A8 DIAERESIS
'©',      # 0x00A9 COPYRIGHT SIGN
'•',      # 0x00AA ARABIC LETTER HEH DOACHASHMEE
'«',      # 0x00AB LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
'¬',      # 0x00AC NOT SIGN
'–',      # 0x00AD SOFT HYPHEN
'®',      # 0x00AE REGISTERED SIGN
'¯',      # 0x00AF MACRON
'°',      # 0x00B0 DEGREE SIGN
'±',      # 0x00B1 PLUS-MINUS SIGN
'²',      # 0x00B2 SUPERSCRIPIT TWO
'³',      # 0x00B3 SUPERSCRIPIT THREE
'´',      # 0x00B4 ACUTE ACCENT
'µ',      # 0x00B5 MICRO SIGN
'¶',      # 0x00B6 PILCROW SIGN
'·',      # 0x00B7 MIDDLE DOT
'¸',      # 0x00B8 CEDILLA
'¹',      # 0x00B9 SUPERSCRIPIT ONE
'•',      # 0x00BA ARABIC SEMICOLON
'»',      # 0x00BB RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
'¼',      # 0x00BC VULGAR FRACTION ONE QUARTER
'½',      # 0x00BD VULGAR FRACTION ONE HALF
'¾',      # 0x00BE VULGAR FRACTION THREE QUARTERS
'؟',      # 0x00BF ARABIC QUESTION MARK
'ح',      # 0x00C1 ARABIC LETTER HEH GOAL
'ء',      # 0x00C2 ARABIC LETTER HAMZA
'آ',      # 0x00C3 ARABIC LETTER ALEF WITH MADDA ABOVE
'أ',      # 0x00C4 ARABIC LETTER ALEF WITH HAMZA ABOVE
'ؤ',      # 0x00C5 ARABIC LETTER WAW WITH HAMZA ABOVE
'إ',      # 0x00C6 ARABIC LETTER ALEF WITH HAMZA BELOW
'ئ',      # 0x00C7 ARABIC LETTER YEH WITH HAMZA ABOVE
'ا',      # 0x00C8 ARABIC LETTER ALEF
'ب',      # 0x00C9 ARABIC LETTER BEH
'ة',      # 0x00CA ARABIC LETTER TEH MARBUTA
'ت',      # 0x00CB ARABIC LETTER TEH
'ث',      # 0x00CC ARABIC LETTER THEH
'ج',      # 0x00CD ARABIC LETTER JEEM
'ح',      # 0x00CE ARABIC LETTER HAH
'خ',      # 0x00CF ARABIC LETTER KHAH
'د',      # 0x00D0 ARABIC LETTER DAL
'ذ',      # 0x00D1 ARABIC LETTER THAL
'ر',      # 0x00D2 ARABIC LETTER REH
'ز',      # 0x00D3 ARABIC LETTER ZAIN
'س',      # 0x00D4 ARABIC LETTER SEEN
'ش',      # 0x00D5 ARABIC LETTER SHEEN
'ص',      # 0x00D6 ARABIC LETTER SAD
'ض',      # 0x00D7 ARABIC LETTER DAD
'×',      # 0x00D8 MULTIPLICATION SIGN
'ط',      # 0x00D9 ARABIC LETTER TAH
'ظ',      # 0x00DA ARABIC LETTER ZAH
'ع',      # 0x00DB ARABIC LETTER AIN
'غ',      # 0x00DC ARABIC LETTER GHAIN
'ـ',      # 0x00DD ARABIC TATWEEL
'ف',      # 0x00DE ARABIC LETTER FEH
```

```
'° # 'قx0642 ARABIC LETTER QAF
'° # 'كx0643 ARABIC LETTER KAF
'â' , # 0x00E0 LATIN SMALL LETTER A WITH GRAVE
'° # 'لx0644 ARABIC LETTER LAM
'â' , # 0x00E2 LATIN SMALL LETTER A WITH CIRCUMFLEX
'° # 'مx0645 ARABIC LETTER MEEM
'° # 'نx0646 ARABIC LETTER NOON
'° # 'هx0647 ARABIC LETTER HEH
'° # 'وx0648 ARABIC LETTER WAW
'ç' , # 0x00E7 LATIN SMALL LETTER C WITH CEDILLA
'è' , # 0x00E8 LATIN SMALL LETTER E WITH GRAVE
'é' , # 0x00E9 LATIN SMALL LETTER E WITH ACUTE
'ê' , # 0x00EA LATIN SMALL LETTER E WITH CIRCUMFLEX
'ë' , # 0x00EB LATIN SMALL LETTER E WITH DIAERESIS
'° # 'ىx0649 ARABIC LETTER ALEF MAKSURA
'° # 'يx064A ARABIC LETTER YEH
'î' , # 0x00EE LATIN SMALL LETTER I WITH CIRCUMFLEX
'ï' , # 0x00EF LATIN SMALL LETTER I WITH DIAERESIS
'í' , # 0x064B ARABIC FATHATAN
'î' , # 0x064C ARABIC DAMMATAN
'ï' , # 0x064D ARABIC KASRATAN
'ì' , # 0x064E ARABIC FATHA
'ô' , # 0x00F4 LATIN SMALL LETTER O WITH CIRCUMFLEX
'í' , # 0x064F ARABIC DAMMA
'ï' , # 0x0650 ARABIC KASRA
'÷' , # 0x00F7 DIVISION SIGN
'ì' , # 0x0651 ARABIC SHADDA
'ù' , # 0x00F9 LATIN SMALL LETTER U WITH GRAVE
'í' , # 0x0652 ARABIC SUKUN
'û' , # 0x00FB LATIN SMALL LETTER U WITH CIRCUMFLEX
'ü' , # 0x00FC LATIN SMALL LETTER U WITH DIAERESIS
' ' , # 0x200E LEFT-TO-RIGHT MARK
'° # 'x200F RIGHT-TO-LEFT MARK
'° # 'عx06D2 ARABIC LETTER YEH BARREE
); #windows1256

return $windows1256[ord($matches_content[1])-0x80];
}
?>
```