

# Regex gyorsreferencia

Wettl Ferenc Ver. 0.1 2006-10-08

Illesztés egy karakterre (karakterosztályok)	
.	mindenre illeszkedik (kivéve az újsort: <code>\n</code> , ld. <code>s</code> opció).
<code>[abcde]</code>	minden felsorolt karakterre illeszkedik (karakterosztály)
<code>[^abcde]</code>	minden karakterre illeszkedik, kivéve a <code>^</code> után felsoroltakat (negált karakterosztály)
<code>[0-9a-zA-F]</code>	(-) tartományokat jelöl ki (kivéve az első és utolsó helyen)
<code>\p{név}</code>	illeszkedik a <i>név</i> nevű Unicode-os karakterosztály elemére ( <code>Basic_Latin</code> , <code>Cyrillic</code> )
<code>\P{name}</code>	nem illeszkedik a <i>név</i> nevű Unicode-os karakterosztály elemére
<code>\w</code>	egy szókarakterre illeszkedik (alap esetben <code>[a-zA-Z_0-9]</code> , de nemzeti beállítás is lehetséges).
<code>\W</code>	szókarakterre nem illeszkedik
<code>\s</code>	bármilyen „üres” (whitespace) karakter (szóköz, tabulátor, újsor...)
<code>\S</code>	nem üres karakter
<code>\d</code>	digitális számjegy ( <code>[0-9]</code> )
<code>\D</code>	nem digitális számjegy

Mennyiségjelzők (kvantorok)		Regex opciók	
<code>*</code>	0 vagy több illeszkedés (vele ekvivalens <code>{0,}</code> )	<code>i</code> case-insensitive	nem különbözteti meg a kis- és nagybetűt
<code>+</code>	1 vagy több illeszkedés (vele ekvivalens <code>{1,}</code> )	<code>m</code> multiline	<code>^</code> és <code>\$</code> a sor (és nem a karakterlánc) elejére ill. végére illeszkedik
<code>?</code>	1 vagy illeszkedés (vele ekvivalens <code>{0,1}</code> )	<code>s</code> Singleline	a pont ( <code>.</code> ) minden karakterre, így <code>\n</code> -re is illeszkedik
<code>{n}</code>	<i>n</i> illeszkedés	<code>x</code>	a nem karakterrövidítéses üres karaktereket nem veszi figyelembe
<code>{n,}</code>	legalább <i>n</i> illeszkedés		
<code>{n,m}</code>	legalább <i>n</i> , legfeljebb <i>m</i> illeszkedés		
<code>*?</code>	a legkevesebb illeszkedés (lusta csillag)		
<code>+?</code>	a legkevesebb, de legalább 1 illeszkedés (lusta +)		
<code>??</code>	ha lehet, 0 illeszkedés, de legfeljebb 1 (lusta ?)		
<code>{n,}?</code>	a lehető legkevesebb, de legalább <i>n</i> illeszkedés		
<code>{n,m}?</code>	<i>n</i> és <i>m</i> közt a lehető legkevesebb illeszkedés		

0 szélességű illesztések (horgonyok)	
<code>^</code>	a sor (vagy a karaktrelánc) eleje (ld. <code>m</code> opció)
<code>\$</code>	a sor (vagy a karaktrelánc) vége (a <code>\n</code> előtt)
<code>\A</code>	a karakterlánc eleje (törli az <code>m</code> opció hatását)
<code>\Z, \z</code>	a karakterlánc vége (törli az <code>m</code> opció hatását)
<code>\G</code>	az előző illesztés vége
<code>\&lt;</code>	szó eleje (előtte <code>\W</code> , utána <code>\w</code> )
<code>\&gt;</code>	szó vége (előtte <code>\w</code> , utána <code>\W</code> )
<code>\b</code>	szóhatár ( <code>\w</code> és <code>\W</code> közt)
<code>\B</code>	nem szóhatár

Karakterreprezentációk (escaped characters)	
<code>\b</code>	visszalépés
<code>\t</code>	tab
<code>\r</code>	carriage return
<code>\v</code>	vertical tab
<code>\f</code>	form feed
<code>\n</code>	new line
<code>\e</code>	escape
<code>\040</code>	ASCII karakter októálsan (itt szóköz)
<code>\x20</code>	ASCII karakter hexadecimálisan
<code>\cC</code>	ASCII kontrol karakter <code>\cC</code> = ctrl-C
<code>\u0020</code>	Unicode karakter hexadecimálisan
<code>\</code>	az utána következő karaktert jelenti, ha azzal együtt nem karakterrövidítés

Metakarakterek	
BRE (basic regular expression)	<code>.</code> <code>^</code> <code>\$</code> <code>[</code> <code>\</code>
ERE (extended regex) és	<code>.</code> <code>^</code> <code>\$</code> <code>[</code> <code>\</code>
PCRE (Perl compatible regex)	<code>(</code> <code>)</code> <code>+</code> <code>?</code> <code>{</code> <code>}</code>

Választó mintaillesztés (alternation)	
<code> </code>	a függőleges vonallal elválasztottak egyike
<code>(?(részkifejezés) igen nem)</code>	ha a <i>részkifejezés</i> illeszkedik, az <i>igen</i> részre, egyébként a <i>nem</i> részre illeszkedik; a <i>részkifejezés</i> 0 szélességgel, azaz horgonyként viselkedik (ekvivalens alak: <code>(?(?=expression)yes no)</code> )

Csoportosítás, gyűjtés (grouping, capture) és visszahivatkozás	
<code>( )</code>	a bezárt részkifejezés automatikusan sorszámot kap a nyitó zárójelek sorrendje szerint
<code>(?: )</code>	nem gyűjtő (nem sorszámozódó) részkifejezés
<code>(?imsx-imsx: )</code>	opciók megadása regexen belül (pl. <code>(?i-s: )</code> az <code>i</code> opciót be, az <code>s</code> -t kikapcsolja
<code>(?= )</code>	0 szélességű előretekintés (pl. <code>\w+(?=\d)</code> illeszkedik <code>abc</code> -re az <code>abc1</code> sztringben, de nem illeszkedik <code>abcd</code> -ben)
<code>(?! )</code>	0 szélességű negatív előretekintés (pl. <code>\b(?!q)\w+\b</code> a nem <code>q</code> betűvel kezdődő szavakra illeszkedik)
<code>(?&lt;= )</code>	0 szélességű visszatekintés (pl. <code>(?&lt;=20)06</code> a <code>20</code> utáni <code>06</code> -ra illeszkedik)
<code>(?! )</code>	0 szélességű negatív visszatekintés (pl. <code>\w+(?!nak)\b</code> )
<code>\\$számjegy</code>	visszahivatkozás, pl. <code>(\w)\1</code> a két azonos részből álló szavakra illeszkedik, mint „gyöngyön”, „elemelem”...