D

Oracle Regular Expression Support

Oracle's implementation of regular expressions conforms with the IEEE Portable Operating System Interface (POSIX) regular expression standard and to the Unicode Regular Expression Guidelines of the Unicode Consortium.

This appendix contains the following sections:

Multilingual Regular Expression Syntax

Table D-1 lists the full set of operators defined in the POSIX standard Extended Regular Expression (ERE) syntax. Oracle follows the exact syntax and matching semantics for these operators as defined in the POSIX standard for matching ASCII (English language) data. For more complete descriptions of the operators, examples of their use, and Oracle multilingual enhancements of the operators, refer to *Oracle Database Development Guide*. Notes following the table provide more complete descriptions of the operators and their functions, as well as Oracle multilingual enhancements of the operators. Table D-2 summarizes Oracle support for and multilingual enhancement of the POSIX operators.

Table D-1 Regular Expression Operators and Metasymbols

Operator	Description
\	The backslash character can have four different meanings depending on the context. It can: • Stand for itself • Quote the next character • Introduce an operator • Do nothing
*	Matches zero or more occurrences
+	Matches one or more occurrences
?	Matches zero or one occurrence
|	Alternation operator for specifying alternative matches
^	Matches the beginning of a string by default. In multiline mode, it matches the beginning of any line anywhere within the source string.
$	Matches the end of a string by default. In multiline mode, it matches the end of any line anywhere within the source string.
.	Matches any character in the supported character set except `NULL`

Table D-1 (Cont.) Regular Expression Operators and Metasymbols

Operator	Description
[ ]	Bracket expression for specifying a matching list that should match any one of the expressions represented in the list. A non-matching list expression begins with a circumflex (^) and specifies a list that matches any character except for the expressions represented in the list. To specify a right bracket (]) in the bracket expression, place it first in the list (after the initial circumflex (^), if any). To specify a hyphen in the bracket expression, place it first in the list (after the initial circumflex (^), if any), last in the list, or as an ending range point in a range expression.
( )	Grouping expression, treated as a single subexpression
{m}	Matches exactly m times
{m,}	Matches at least m times
{m,n}	Matches at least m times but no more than n times
\n	The backreference expression (n is a digit between 1 and 9) matches the n^{th} subexpression enclosed between '(' and ')' preceding the \n
[..]	Specifies one collation element, and can be a multicharacter element (for example, [.ch.] in Spanish)
[: :]	Specifies character classes (for example, [:alpha:]). It matches any character within the character class.
[==]	Specifies equivalence classes. For example, [=a=] matches all characters having base letter 'a'.

Regular Expression Operator Multilingual Enhancements

When applied to multilingual data, Oracle's implementation of the POSIX operators extends beyond the matching capabilities specified in the POSIX standard. Table D-2 shows the relationship of the operators in the context of the POSIX standard.

- The first column lists the supported operators.
- The second and third columns indicate whether the POSIX standard (Basic Regular Expression—BRE and Extended Regular Expression—ERE, respectively) defines the operator
- The fourth column indicates whether Oracle's implementation extends the operator's semantics for handling multilingual data.

Oracle lets you enter multibyte characters directly, if you have a direct input method, or you can use functions to compose the multibyte characters. You cannot use the Unicode hexadecimal encoding value of the form '`\xxxx`'. Oracle evaluates the characters based on the byte values used to encode the character, not the graphical representation of the character. All accented characters are considered word characters.

Table D-2 POSIX and Multilingual Operator Relationships

Operator	POSIX BRE syntax	POSIX ERE Syntax	Multilingual Enhancement
\	Yes	Yes	—

Table D-2 (Cont.) POSIX and Multilingual Operator Relationships

Operator	POSIX BRE syntax	POSIX ERE Syntax	Multilingual Enhancement
*	Yes	Yes	—
+	--	Yes	—
?	—	Yes	—
|	—	Yes	—
^	Yes	Yes	Yes
$	Yes	Yes	Yes
.	Yes	Yes	Yes
[ ]	Yes	Yes	Yes
( )	Yes	Yes	—
{m}	Yes	Yes	—
{m,}	Yes	Yes	—
{m,n}	Yes	Yes	—
\n	Yes	Yes	Yes
[..]	Yes	Yes	Yes
[::]	Yes	Yes	Yes
[==]	Yes	Yes	Yes

Perl-influenced Extensions in Oracle Regular Expressions

Oracle Database regular expression functions and conditions accept a number of Perl-influenced operators that are in common use, although not part of the POSIX standard. Table D-3 lists those operators. For more complete descriptions with examples, refer to *Oracle Database Development Guide*.

Table D-3 Perl-influenced Operators in Oracle Regular Expressions

Operator	Description
\d	A digit character.
\D	A nondigit character.
\w	A word character.
\W	A nonword character.
\s	A whitespace character.
\S	A non-whitespace character.
\A	Matches only at the beginning of a string, or before a newline character at the end of a string.
\Z	Matches only at the end of a string.
*?	Matches the preceding pattern element 0 or more times (nongreedy).
+?	Matches the preceding pattern element 1 or more times (nongreedy).
??	Matches the preceding pattern element 0 or 1 time (nongreedy).

Table D-3 (Cont.) Perl-influenced Operators in Oracle Regular Expressions

Operator	Description
{n}?	Matches the preceding pattern element exactly *n* times (nongreedy).
{n,}?	Matches the preceding pattern element at least *n* times (nongreedy).
{n,m}?	Matches the preceding pattern element at least *n* but not more than *m* times (nongreedy).