NAME

PCRE2 - Perl-compatible regular expressions (revised API)

SIZE AND OTHER LIMITATIONS

There are some size limitations in PCRE2 but it is hoped that they will never in practice be relevant.

The maximum size of a compiled pattern is approximately 64 thousand code units for the 8-bit and 16-bit libraries if PCRE2 is compiled with the default internal linkage size, which is 2 bytes for these libraries. If you want to process regular expressions that are truly enormous, you can compile PCRE2 with an internal linkage size of 3 or 4 (when building the 16-bit library, 3 is rounded up to 4). See the **README** file in the source distribution and the **pcre2build** documentation for details. In these cases the limit is substantially larger. However, the speed of execution is slower. In the 32-bit library, the internal linkage size is always 4.

The maximum length of a source pattern string is essentially unlimited; it is the largest number a PCRE2_SIZE variable can hold. However, the program that calls **pcre2_compile()** can specify a smaller limit.

The maximum length (in code units) of a subject string is one less than the largest number a PCRE2_SIZE variable can hold. PCRE2_SIZE is an unsigned integer type, usually defined as size_t. Its maximum value (that is ~(PCRE2_SIZE)0) is reserved as a special indicator for zero-terminated strings and unset offsets.

All values in repeating quantifiers must be less than 65536.

The maximum length of a lookbehind assertion is 65535 characters.

There is no limit to the number of parenthesized groups, but there can be no more than 65535 capture groups, and there is a limit to the depth of nesting of parenthesized subpatterns of all kinds. This is imposed in order to limit the amount of system stack used at compile time. The default limit can be specified when PCRE2 is built; if not, the default is set to 250. An application can change this limit by calling pcre2_set_parens_nest_limit() to set the limit in a compile context.

The maximum length of name for a named capture group is 32 code units, and the maximum number of such groups is 10000.

The maximum length of a name in a (*MARK), (*PRUNE), (*SKIP), or (*THEN) verb is 255 code units for the 8-bit library and 65535 code units for the 16-bit and 32-bit libraries.

The maximum length of a string argument to a callout is the largest number a 32-bit unsigned integer can hold.

AUTHOR

Philip Hazel
University Computing Service
Cambridge, England.

REVISION

Last updated: 02 February 2019