*Table 4-1. Regular expression metacharacter syntax*

|  |  |  |
| --- | --- | --- |
| **Subexpression** | **Matches** | **Notes** |
| **General** | | |
| \^ | Start of line/string |  |
| $ | End of line/string |  |
| \b | Word boundary |  |
| \B | Not a word boundary |  |
| \A | Beginning of entire string |  |
| \z | End of entire string |  |
| \Z | End of entire string (except allowable final line terminator) |  |
| . | Any one character (except line terminator) |  |
| […] | “Character class”; any one character from those listed |  |
| [\^…] | Any one character not from those listed |  |
| **Alternation and Grouping** | | |
| (…) | Grouping (capture groups) |  |
| | | Alternation |  |
| (?:\_re\_ ) | Noncapturing parenthesis |  |
| \G | End of the previous match |  |
| \ n | Back-reference to capture group number "n" |  |
| **Normal (greedy) quantifiers** | | |
| { m,n } | Quantifier for “from m to n repetitions” |  |
| { m ,} | Quantifier for "m or more repetitions” |  |
| { m } | Quantifier for “exactly m repetitions” |  |
| {,n } | Quantifier for 0 up to n repetitions |  |
| \\* | Quantifier for 0 or more repetitions | Short for {0,} |
| + | Quantifier for 1 or more repetitions | Short for {1,} |
| ? | Quantifier for 0 or 1 repetitions (i.e., present exactly once, or not at all) | Short for {0,1} |
| **Reluctant (non-greedy) quantifiers** | | |
| { m,n }? | Reluctant quantifier for “from m to n repetitions” |  |
| { m ,}? | Reluctant quantifier for "m or more repetitions” |  |
| {,n }? | Reluctant quantifier for 0 up to n repetitions |  |
| \\*? | Reluctant quantifier: 0 or more |  |
| +? | Reluctant quantifier: 1 or more |  |
| ?? | Reluctant quantifier: 0 or 1 times |  |
| **Possessive (very greedy) quantifiers** | | |
| { m,n }+ | Possessive quantifier for “from m to n repetitions” |  |
| { m ,}+ | Possessive quantifier for "m or more repetitions” |  |
| {,n }+ | Possessive quantifier for 0 up to n repetitions |  |
| \\*+ | Possessive quantifier: 0 or more |  |
| ++ | Possessive quantifier: 1 or more |  |
| ?+ | Possessive quantifier: 0 or 1 times |  |
| **Escapes and shorthands** | | |
| \ | Escape (quote) character: turns most metacharacters off; turns subsequent alphabetic into metacharacters |  |
| \Q | Escape (quote) all characters up to \E |  |
| \E | Ends quoting begun with \Q |  |
| \t | Tab character |  |
| \r | Return (carriage return) character |  |
| \n | Newline character |  |
| \f | Form feed |  |
| \w | Character in a word | Use \w+ for a word |
| \W | A nonword character |  |
| \d | Numeric digit | Use \d+ for an integer |
| \D | A nondigit character |  |
| \s | Whitespace | Space, tab, etc., as determined by java.lang.Character.isWhitespace() |
| \S | A nonwhitespace character |  |
| **Unicode blocks (representative samples)** | | |
| \p{InGreek} | A character in the Greek block | (Simple block) |
| \P{InGreek} | Any character not in the Greek block |  |
| \p{Lu} | An uppercase letter | (Simple category |
| \p{Sc} | A currency symbol |  |
| **POSIX-style character classes (defined only for US-ASCII)** | | |
| \p{Alnum} | Alphanumeric characters | [A-Za-z0-9] |
| \p{Alpha} | Alphabetic characters | [A-Za-z] |
| \p{ASCII} | Any ASCII character | [\x00-\x7F] |
| \p{Blank} | Space and tab characters |  |
| \p{Space} | Space characters | [ \t\n\x0B\f\r] |
| \p{Cntrl} | Control characters | [\x00-\x1F\x7F] |
| \p{Digit} | Numeric digit characters | [0-9] |
| \p{Graph} | Printable and visible characters (not spaces or control characters) |  |
| \p{Print} | Printable characters |  |
| \p{Punct} | Punctuation characters | One of !"#$%&'()\\* +,-./:;<=>?@[]\^\_`{|} \~ |
| \p{Lower} | Lowercase characters | [a-z] |
| \p{Upper} | Uppercase characters | [A-Z] |
| \p{XDigit} | Hexadecimal digit characters | [0-9a-fA-F] |