

Regular Expressions Cheat Sheet

(PCRE-variant)

| | |
|--------|---------------------------------------------------------------------------------------------------------------------------------|
| abc | this matches the word 'abc' |
| . | this matches any character, but only one single character. |
| [agd] | this matches one of the 3 character: a, g or d. |
| [^agd] | this matches any character, except a, g or d. |
| [c-f] | the dash within the square brackets operates as a range. In this case it means either the letters c, d, e or f (same as [cdef]) |

Multipliers:

| | |
|---|----------------------------------------------------|
| * | the preceding character matches 0 or more times. |
| + | the preceding character matches 1 or more times. |
| ? | the preceding character matches 0 or 1 times only. |

| | |
|-------|-----------------------------------------------------------------------------|
| {n} | the preceding character matches exactly n times. |
| {n,m} | the preceding character matches at least n times and not more than m times. |
| {n,} | the preceding character matches at least n times. |

Boundaries

| | |
|----|------------------------------------|
| ^ | matches the beginning of the line. |
| \$ | matches the end of the line. |

| | |
|----|---------------------------------------------------------|
| \b | word <u>b</u> oundary (the beginning or end of a word). |
| \B | not a word <u>b</u> oundary. |

Grouping, backreferences, "OR" and "\":

| | |
|-------|-----------------------------------------------------------------|
| () | allows us to group several characters to behave as one. |
| \1,\2 | a (back)reference to what the ()'s matched earlier |
| | the logical OR operation. |
| \ | escape character: remove or add special meaning to a character. |

Speciale karakters

| | |
|----|----------------------------------------------------------------------------------------------------|
| \s | anything which is considered white <u>s</u> pace, same as [\t] (sometimes it can be [\t\r\n]) |
| \S | anything which is <i>not</i> considered whitespace. |

| | |
|----|---------------------------|
| \t | tab |
| \r | carriage return (CR) |
| \n | newline or line feed (LF) |

| | |
|----|--------------------------------------------------------------------|
| \d | a <u>d</u> igit, same as [0-9]. |
| \D | anything which is <i>not</i> a digit, same as [^0-9]. |
| \w | anything which is considered a <u>w</u> ord character [a-zA-Z0-9_] |
| \W | anything which is <i>not</i> considered a word character. |

Op <https://ryanstutorials.net/regular-expressions-tutorial/> staat er een goede regex-tutorial, met veel voorbeelden en een cheat sheet met extra info zoals "lookahead" en "lookbehind" operators.

Bij ERE hebben we naast "\b" ook "<" (begin v.e. woord) en ">" (einde v.e. woord).

Bij ERE komt "\w" overeen met [0-9A-Z_a-z³µªÀ-ÖØ-ßà-öø-ÿ] i.p.v. [0-9A-Z_]