

REGULAR EXPRESSIONS

LINGUISTICS 384

Regular expression queries provide a powerful way of searching texts, allowing for higher-level descriptions than just literal words or characters. For example, to search for either “car” or “cars”, use the pattern `cars?`. The `?` allows the previous element (“s”) to occur either 0 or 1 time.

Pattern	Description	Example query
<code>.</code>	Match any character	<code>b.t</code>
<code>?</code>	Match the previous element zero or one time	<code>car.?</code>
<code>*</code>	Match the previous element zero or more times	<code>car.*</code>
<code>+</code>	Match the previous element one or more times	<code>car.+</code>
<code>{3,7}</code>	Match the previous element between 3 and 7 times	<code>.{10,12}</code>
<code>{3,}</code>	Match the previous element at least 3 times	<code>.{14,}</code>
<code>{,7}</code>	Match the previous element no more than 7 times (including 0)	<code>.{,2}</code>
<code> </code>	Match the previous element or the subsequent element	<code>ab cd</code>
<code>(...)</code>	Group the characters within the parentheses	<code>talk(ed ing)</code>
<code>[ptk]</code>	Match either p, t, or k	<code>[mb]et</code>
<code>[^ptk]</code>	Match anything except p, t, or k	<code>[^mb]et</code>
<code>^a</code>	Match a only at the beginning of an entire query	<code>^met</code>
<code>a\$</code>	Match a only at the end of an entire query	<code>met\$</code>
<code>\w</code>	Match any alphanumeric character (<u>w</u> ord-like)	<code>\w+</code>
<code>\W</code>	Match anything except alphanumeric characters	<code>\W+</code>
<code>\d</code>	Match any number (<u>d</u> igit)	<code>\w+\.\d+</code>
<code>\D</code>	Match anything except numbers	
<code>\s</code>	Match any space character (eg. space, tab)	
<code>\S</code>	Match anything except space characters	
<code>\b</code>	Word boundary	<code>\bhello\b</code>

1 Examples

Find either Meghan or Megan: `Megh?an`

Find either Katelyn or Catelyn: `[CK]atelyn`

Find either Ben or Benjamin: `Ben(jamin)?`

Find either Emily or Emilie: `Emil(ie|y)`

Find a local phone number: `\d{3}-\d{4}`

Find an OSU id: `\w+\.\d+`