



Регулярні вирази

Як перевірити, чи в тексті є хоча б
одна цифра?

Або чи в тексті є 4 букви підряд?

Або чи містить текст пробіли?

Або що текст є коректною датою?

Можна писати складні цикли і
перевіряти кожен символ тексту.

А можна використати регулярні
вирази.

Регулярні вирази в будь-якій мові програмування використовуються для того, щоб перевірити, чи якийсь текст відповідає заданому шаблону (зразку).

У RHP є функція

`preg_match(шаблон, текст)`

яка перевіряє, чи текст відповідає шаблону.

Повертає `true/false`.

Приклад: перевірити, чи в тексті є хоча б одна цифра.

```
if(preg_match("/[0-9]/", $text))
```

Перевірити, чи в тексті є 4 малі латинські літери підряд.

```
if(preg_match("/[a-z]{4}/", $text))
```

Перевірити, чи в тексті є 4 малі або великі латинські літери підряд:

```
if(preg_match("/[a-zA-Z]{4}/", $text))
```

або

```
if(preg_match("/[a-z]{4}/i", $text))
```

Тут **i** означає case insensitive (нечутливість до регістру).

Перевірити, чи в тексті є 5 або більше будь-яких символів.

```
if(preg_match("/.{5,}/", $text))
```

Тут крапка означає будь-який символ.

Перевірити, чи в тексті є хоча б
одна крапка.

```
if(preg_match("/\./", $text))
```

Символ крапки в регулярних
виразах позначається \.

Перевірити, чи в тексті є хоча б
одна мала українська літера.

```
if(preg_match("/[a-яієїґ]/", $text))
```

Перевірити, чи в тексті є хоча б одна літера будь-якої мови.

```
if(preg_match("/p{L}/u", $text))
```

Тут **L** означає letter, **u** означає Unicode.

Квантифікатори (кількість повторень)

$\{4\}$ — те, що зліва, має повторюватись 4 рази

$\{4,\}$ — 4 або більше разів

$\{2,4\}$ — від 2 до 4 разів

$+$ те саме, що $\{1,\}$

$*$ те саме, що $\{0,\}$

$?$ те саме, що $\{0,1\}$

Попередні регулярні вирази перевіряли, чи є в тексті хоча б 1 або кілька певних символів.

А як перевірити, чи весь текст (а не частина) має 4 цифри?

```
if(preg_match("/^[0-9]{4}$/", $text))
```

^ означає, що лівіше більше нічого не може бути (це початок тексту)

\$ означає, що правіше більше нічого не може бути (це кінець тексту)

Приклад:

```
preg_match("/[0-9]{4}/", "x1234x")  
// true
```

```
preg_match("/^[0-9]{4}$/", "x1234x")  
// false
```


Якщо треба перевірити наявність символів **^** або **\$** у тексті, то пишеться **\^** і **\\$** відповідно.

```
preg_match("/^\$[0-9]+(\.[0-9]+)?$/",  
            '$123.99')  
            // true
```

Набір символів можна об'єднувати круглими дужками.

Перевірити, чи в тексті є слово
hello або world.

```
if(preg_match("/hello|world/",  
    "hello friend"))
```

Символ | означає **або**.
Він частіше використовується для
пошуку слів, а не поодиноких
символів.

Регулярний вираз для перевірки коректності e-mail

```
if(preg_match('/^[a-zA-Z0-9\.\-\_]{2,}@[a-zA-Z0-9\-\_]+\.[a-z]{2,3}$/',  
$text))
```

Жадібний пошук (Greedy search)

```
preg_match("/h.*o/", "hello world  
hello there")
```

Шукає якомога більше (до останньої появи літери **o**).

Лінивий пошук (Lazy search)

```
preg_match("/h.*?o/", "hello world  
hello there")
```

Шукає якомога менше (до першої появи літери o).

Корисні сайти для тестування
регулярних виразів:

<https://regex101.com>

<https://regexr.com>