

Метасимволы в Regex

- []** — используется для указания набора или диапазона символов
- ** — указывает на начало последовательности или экранирует служебные символы
- .** — выбирает любой символ, кроме новой строки `\n`
- ^** — проверяет, начинается ли строка с определенного символа / слова / набора символов
- \$** — проверяет, заканчивается ли строка в соответствии с шаблоном `r'До свиданья.$'`
- *** — ноль или больше совпадений с шаблоном
- +** — одно и более совпадений `r'к.+ператив'`
- ?** — ноль или одно совпадение `r'ф.?нтастика'`. Также, нейтрализует «жадность» выражений, которые используют `.`, `*`, `+` для выбора любых символов.
- {}** — точное число совпадений `r'Интерсте.{2}ар'`
- |** — любой из двух вариантов `r'уйду|останусь'`
- ()** — захватывает группу для дальнейших манипуляций — `re.sub(r'(www)', r'1.', "wwwgear.com")`
- <>** — создает именованную группу — `re.search('(P<группа1>(w+),(?P<группа2>(w+),(?P<группа3>(w+))', 'дом,улица,фонарь')`

Наборы и диапазоны символов

- [есн]** — проверит, есть ли в строке любой из указанных символов `е`, `с` или `н`. Наличие любой цифры из набора проверяется так же — `[0169]`
- [a-e]** — вернет совпадения по каждому символу из алфавитного диапазона `re.findall("[a-e]", "здесь есть символы из диапазона")`. Так же возвращает совпадения по диапазону цифр — `[5-9]`
- [^абвгд]** — проверит наличие в строке символов, кроме указанных в наборе — `re.findall("[^абвгд]", "АБВГДейка — детская передача", re.I)`
- [0-5][0-9]** — возвращает совпадения по двузначным цифрам от 00 до 59 — `re.findall("[0-5][0-9]", "будильник сработает в 07:45")`

Последовательности

- ** — специфическая последовательность символов
- \A** — проверяет, начинается ли строка с определенной последовательности символов. `re.findall(r"\ADом", txt)`, проверит, начинается ли предложение со слова «Дом»
- \b** — возвращает совпадение, если слово начинается или заканчивается нужной последовательностью символов `re.findall(r".com\b", s)` проверит, есть ли в строке хотя бы одно доменное имя зоны `.com`
- \B** — возвращает совпадение, если определенные символы есть в строке, но не в начале или не в конце слова `re.findall(r"\Bро", 'розовая от мороза')`, `re.findall(r'ин\B', 'синий апельсин')`
- \d** — проверяет, что в строке есть цифры от 0 до 9 `re.findall("\d", 'при пожаре звоните 112')`
- \D** — удостоверяет, что цифр в строке нет `re.findall("\D", 'цифр нет')`
- \s** — проверяет наличие пробелов в строке `re.findall("\s", "один пробел")`
- \S** — возвращает совпадение, если в строке есть любые символы, кроме пробелов — `re.findall("\S", "непустая строка")`.
- \w** — проверяет, есть ли в строке «словесные» символы — знак нижнего подчеркивания, цифры и буквы — `re.findall(r"\w", "_\\")`
- \W** — возвращает совпадение по каждому «несловесному» символу — `re.findall("\W", "здесь есть такие символы!")`
- \Z** — проверит, заканчивается ли строка нужной последовательностью символов `re.findall("конец\Z", "это конец")`

Флаги в Regex

Функциональность регулярных выражений расширяется за счет флагов

Краткий синтаксис	Полный синтаксис	Назначение
re.A	re.ASCII	Возвращает совпадения только по ASCII-символам вместо всей таблицы Unicode
re.I	re.IGNORECASE	Игнорирует регистр символов
re.M	re.MULTILINE	Используется совместно с метасимволами <code>^</code> и <code>\$</code> . В первом случае возвращает совпадения в начале каждой новой строки <code>\n</code> , во втором – в конце <code>\n</code>
re.S	re.DOTALL	Заставляет метасимвол <code>.</code> возвращать совпадения по абсолютно всем символам, включая <code>\n</code> . Без этого флага точка <code>.</code> соответствует любому символу, кроме <code>\n</code>
re.X	re.VERBOSE	Разрешает комментарии в Regex-выражениях.
re.L	re.LOCALE	Учитывает региональные настройки при использовании <code>\w</code> , <code>\W</code> , <code>\b</code> , <code>\B</code> , <code>\s</code> и <code>\S</code> . Используется только при работе с байтовыми строками, не совместим с re.ASCII.

[Proglib.io](https://proglib.io) 

[Proglib.academy](https://proglib.academy) 