

# Классы символов в регулярных выражениях 1

## Категория Юникода или блок Юникода: `\p{}`

В стандарте Юникода каждому символу присваивается общая категория. Например, определенный символ может быть прописной буквой (категория `Lu`), десятичной цифрой (категория `Nd`), математическим символом (категория `Sm`) или разделителем абзацев (категория `Zl`). Определенные наборы символов стандарта Юникод также занимают некоторый диапазон или блок последовательных кодовых точек.

Например, базовая латинская кодировка находится в диапазоне от `\u0000` до `\u007F`, а арабская кодировка находится в диапазоне от `\u0600` до `\u06FF`.

Конструкция регулярного выражения

`\p{Имя}`

соответствует любым символам, которые относятся к общей категории Юникода или именованному блоку, где `name` — сокращенное название категории или имя блока.

## Отрицательная категория Юникода или блок Юникода:

`\P{}`

В стандарте Юникода каждому символу присваивается общая категория. Например, определенный символ может быть прописной буквой (категория `Lu`), десятичной цифрой (категория `Nd`), математическим символом (категория `Sm`) или разделителем абзацев (категория `Zl`). Определенные наборы символов стандарта Юникод также занимают некоторый диапазон или блок последовательных кодовых точек.

Например, базовая латинская кодировка находится в диапазоне от `\u0000` до `\u007F`, а арабская кодировка находится в диапазоне от `\u0600` до `\u06FF`.

Конструкция регулярного выражения

`\P{Имя}`

соответствует любым символам, которые не относятся к общей категории Юникода или именованному блоку, где `name` — сокращенное название категории или имя блока.

# Классы символов в регулярных выражениях 2

## Поддерживаемые общие категории

### Юникода

В Юникоде определяются общие категории, приведенные в следующей таблице. Дополнительные сведения см. в подразделах "Формат файла UCD" и "Общие значения категорий" в кодировке Юникод, сек. 5.7.1, таблица 12.

Категория	Описание
Lu	Буква: прописные буквы
Ll	Letter, Lowercase
Lt	Буква: заглавный регистр
Lm	Буква: модификатор
Lo	Буква: другие
L	Все буквенные символы. Включает символы Lu, Ll, Lt, Lm и Lo
Mn	Метка: не занимающая место
Mc	Метка: комбинированная
Me	Метка: вложенная
M	Все объединяющие знаки. Включает категории Mn, Mc и Me.
Nd	Число: десятичная цифра
Nl	Число: буква
No	Число: другое
N	Все числа. Включает категории Nd, Nl и No.
Pc	Пунктуация: соединитель
Pd	Пунктуация: тире
Ps	Пунктуация: открывающий знак пунктуации
Pe	Пунктуация: закрывающий знак пунктуации
Pi	Пунктуация: открывающая кавычка (в зависимости от использования поведение может быть аналогичным Ps или Pe)
Pf	Пунктуация: закрывающая кавычка (в зависимости от использования поведение

# Классы символов в регулярных выражениях 3

Категория	Описание
	может быть аналогичным Ps или Pe)
Po	Пунктуация: другие знаки пунктуации
P	Все знаки препинания. Включает категории Pc, Pd, Ps, Pe, Pi, Pf и Po.
Sm	Символ: математический
Sc	Символ денежной единицы
Sk	Символ: модификатор
So	Символ: другие
S	Все символы. Включает категории Sm, Sc, Sk и So.
Zs	Разделитель: пробел
Zl	Разделитель: строка
Zp	Разделитель: абзац
Z	Все знаки разделения. Включает категории Zs, Zl и Zp.
Cc	Другое: управляющий символ
Cf	Другое: формат
Cs	Другое: заменяющий символ
Co	Другое: индивидуальное использование
Cn	Другое, не назначенное или нехарактерное
C	Все остальные символы. Включает категории Cc, Cf, Cs, Co и Cn.

Можно определить категорию Юникода для любого отдельного символа, передав его в метод `GetUnicodeCategory`. В следующем примере метод `GetUnicodeCategory` используется для определения категории каждого элемента в массиве, содержащем выбранные латинские символы.