

# Элементы языка регулярных выражений — краткий справочник

Регулярное выражение — это шаблон, который обработчик регулярных выражений пытается сопоставить с введенным текстом. Шаблон состоит из односимвольных или многосимвольных литералов, операторов или конструкций.

## Escape-знаки

Обратная косая черта (\) в регулярных выражениях указывает, что следующий за ней символ либо является специальным знаком (как показано в следующей таблице), либо должен интерпретироваться буквально. Дополнительные сведения см. в разделе [Escape-символы](#).

Escape-символ	Описание:	Шаблон	Число соответствий
\a	Соответствует знаку колокольчика, \u0007.	\a	"\u0007" в "Ошибка!" + "\u0007"
\b	В классе символов соответствует знаку BACKSPACE, \u0008.	[b]{3,}	"\b\b\b\b" в "\b\b\b\b"
\t	Соответствует знаку табуляции, \u0009.	(w+)\t	"элемент1\t", "элемент2\t" в "элемент1\tэлемент2\t"
\r	Соответствует знаку возврата каретки, \u000D. (\r не эквивалентен знаку начала новой строки, \n).	\r\n(w+)	"\r\nЗдесь" в "\r\nЗдесь имеется\nдве строки."
\v	Соответствует знаку вертикальной табуляции, \u000B.	[v]{2,}	"\v\v\v" в "\v\v\v"
\f	Соответствует знаку перевода страницы, \u000C.	[f]{2,}	"\f\f" в "\f\f"
\n	Соответствует знаку новой строки, \u000A.	\r\n(w+)	"\r\nЗдесь" в "\r\nЗдесь имеется\nдве строки."
\e	Соответствует escape-знаку, \u001B.	\e	"\x001B" в "\x001B"
\ nnn	Использует восьмеричное представление для указания символа (nnn состоит из двух или трех цифр).	\w\040\w	"a b", "c d" в  "a bc d"
\x nn	Использует шестнадцатеричное представление для указания символа (nn состоит ровно из двух цифр).	\w\x20\w	"a b", "c d" в  "a bc d"
\c X \c x	Соответствует управляющему символу ASCII, который задан как X или x, где X или x является буквой управляющего символа.	\cC	"\x0003" в "\x0003" (Ctrl-C)
\unnnn	Совпадение со знаком Юникода в шестнадцатеричном представлении (строго четыре цифры, представленные как nnnn).	\w\u0020\w	"a b", "c d" в  "a bc d"
\	Если за этим знаком следует символ, не распознанный как escape-символ из этой и других таблиц данной темы, то соответствует в точности этому символу. Например, \* — это то же самое, что и \x2A, а \. — то же самое, что	\d+[+~x\*]\d+	"2+2" и "3*9" в "(2+2) * 3*9"

	и \x2E. Это позволяет обработчику регулярных выражений распознавать языковые элементы (такие как *или ?) и символьные литералы (представленные как \* или \?).		
--	--	--	--

## Классы символов

Класс символов соответствует какому-либо одному набору символов. Классы символов состоят из языковых элементов, приведенных в следующей таблице. Дополнительные сведения см. в разделе [Character Classes](#).

Класс знаков	Описание:	Шаблон	Число соответствий
[ <i>character_group</i> ]	Соответствует любому одиночному символу, входящему в <i>character_group</i> . По умолчанию при сопоставлении учитывается регистр.	[ae]	"a" в "gray"  "a", "e" в "lane"
[^ <i>character_group</i> ]	Отрицание: соответствует любому одиночному символу, не входящему в <i>character_group</i> . По умолчанию символы в <i>character_group</i> чувствительны к регистру.	[^aei]	"r", "g", "n" в "reign"
[ <i>first-last</i> ]	Диапазон символов: соответствует одному символу в диапазоне от <i>first</i> до <i>last</i> .	[A-Z]	"A", "B" в "AB123"
.	Подстановочный знак: соответствует какому-либо одному знаку, кроме "\n". Для сопоставления символа точки (. или \u002E) перед ней нужно поставить дополнительную обратную косую черту (\.).	a.e	"ave" в "nave"  "ate" в "water"
\p{ <i>имя</i> }	Соответствует любому одиночному символу в общей категории Юникода или в именованном блоке, указанном в параметре <i>name</i> .	\p{Lu}  \p{IsCyrillic}	"C", "L" в "City Lights" "Д", "Ж" в "ДЖем"
\P{ <i>имя</i> }	Соответствует любому одиночному символу, не входящему в общую категорию Юникода или в именованный блок, указанный в параметре <i>name</i> .	\P{Lu}  \P{IsCyrillic}	"i", "t", "y" в "City" "e", "m" в "ДЖем"
\w	Соответствует любому алфавитно-цифровому знаку.	\w	"I", "D", "A", "1", "3" в "ID A1.3"
\W	Соответствует любому символу, который не является буквенно-цифровым знаком.	\W	" ", "." в "ID A1.3"
\s	Соответствует любому знаку пробела.	\w\s	"D " в "ID A1.3"
\S	Соответствует любому знаку, не являющемуся пробелом.	\s\S	" _ " в "int __ctr"
\d	Соответствует любой десятичной цифре.	\d	"4" в "4 = IV"
\D	Соответствует любому символу, не являющемуся десятичной цифрой.	\D	" ", "=", " ", "I", "V" в "4 = IV"

## Привязки

Привязки, или атомарные утверждения нулевой ширины, приводят к успеху или сбою сопоставления, в зависимости от текущей позиции в строке, но не предписывают обработчику перемещаться по строке или обрабатывать символы. Метасимволы, приведенные в следующей таблице, являются привязками. Дополнительные сведения см. в разделе [Привязки](#).

Утверждение	Описание:	Шаблон	Число соответствий
<code>^</code>	По умолчанию соответствие должно начинаться в начале строки. В многострочном режиме соответствие должно начинаться в начале линии.	<code>^d{3}</code>	"901" в "901-333-"
<code>\$</code>	По умолчанию соответствие должно обнаруживаться в конце строки или перед символом <code>\n</code> в конце строки. В многострочном режиме соответствие должно обнаруживаться до конца линии или перед символом <code>\n</code> в конце линии.	<code>-d{3}\$</code>	"-333" в "-901-333"
<code>\A</code>	Соответствие должно обнаруживаться в начале строки.	<code>\Ad{3}</code>	"901" в "901-333-"
<code>\Z</code>	Соответствие должно обнаруживаться в конце строки или до символа <code>\n</code> в конце строки.	<code>-d{3}\Z</code>	"-333" в "-901-333"
<code>\z</code>	Соответствие должно обнаруживаться в конце строки.	<code>-d{3}\z</code>	"-333" в "-901-333"
<code>\G</code>	Соответствие должно обнаруживаться в той точке, где заканчивается предыдущее соответствие.	<code>\G(d)</code>	"(1)", "(3)", "(5)" in "(1)(3)(5)[7](9)"
<code>\b</code>	Соответствие должно обнаруживаться на границе между символом <code>\w</code> (алфавитно-цифровым) и символом <code>\W</code> (не алфавитно-цифровым).	<code>\bw+\sw+\b</code>	"them theme", "them them" в "them theme them them"
<code>\B</code>	Соответствие не должно обнаруживаться на границе <code>\b</code> .	<code>\Bend\w*\b</code>	"ends", "ender" в "end sends endure lender"

## Конструкции группирования

Конструкции группирования отображают части выражений регулярных выражений и обычно захватывают части строки входной строки. Конструкции группирования состоят из языковых элементов, приведенных в следующей таблице. Для получения дополнительной информации см. [Grouping Constructs](#).

Конструкция группирования	Описание:	Шаблон	Число соответствий
( <i>часть выражения</i> )	Захватывает соответствующую часть выражения и назначает ей порядковый номер,	<code>(\w)l</code>	"ee" в "deep"

	отсчитываемый от единицы.		
(?< имя > часть выражения )	Выделяет соответствующую часть выражения в именованную группу.	(?<double>\w)\k<double>	"ee" в "deep"
(?< имя1 - имя2 > часть выражения )	Задаёт сбалансированное определение группы. Дополнительные сведения см. в разделе "Сбалансированное определение группы" статьи <a href="#">Grouping Constructs</a> .	((('Open'()[^\()]*))+((?Close-Open')[^\()]*+))*((Open)(?!))\$	"((1-3)*(3-1))" в "3+2^((1-3)*(3-1))"
(?: часть выражения )	Определяет невыделяемую группу.	Write(?:Line)?	"WriteLine" в "Console.WriteLine()"  "Write" в "Console.Write(value)"
(?imnsx-имnsx: часть выражения )	Применяет или отключает указанные параметры в <i>subexpression</i> . Для получения дополнительной информации см. <a href="#">Regular Expression Options</a> .	A\d{2}(?i:\w+)\b	"A12xl", "A12XL" в "A12xl A12XL a12xl"
(?= часть выражения )	Утверждение положительного просмотра вперед нулевой ширины.	\w+(?=.)	"is", "ran" и "out" в "He is. The dog ran. The sun is out."
(?! часть выражения )	Утверждение отрицательного просмотра вперед нулевой ширины.	\b(?:un)\w+\b	"sure", "used" в "unsure sure unity used"
(?<= часть выражения )	Утверждение положительного просмотра назад нулевой ширины.	(?<=19)\d{2}\b	"99", "50", "05" в "1851 1999 1950 1905 2003"
(?<! часть выражения )	Утверждение отрицательного просмотра назад нулевой ширины.	(?<!19)\d{2}\b	"51", "03" в "1851 1999 1950 1905 2003"
(?> часть выражения )	Часть выражения поиска без возврата (или "жадного" поиска).	[13579](?>A+B+)	"1ABV", "3ABV" и "5AB" в "1ABV 3ABV 5AB 5AC"

## Квантификаторы

Квантор указывает количество вхождений предшествующего элемента (знака, группы или класса знаков), которое должно присутствовать во входной строке, чтобы было зафиксировано

соответствие. Кванторы состоят из языковых элементов, приведенных в следующей таблице. Для получения дополнительной информации см. [Quantifiers](#).

Квантификатор	Описание:	Шаблон	Число соответствий
*	Соответствует предыдущему элементу ноль или более раз.	\d*\d	".0", "19.9", "219.9"
+	Соответствует предыдущему элементу один или более раз.	"be+"	"bee" в "been", "be" в "bent"
?	Соответствует предыдущему элементу ноль или один раз.	"rai?n"	"ran", "rain"
{ <i>n</i> }	Предыдущий элемент повторяется ровно <i>n</i> раз.	",\d{3}"	",043" в "1,043.6", ",876", ",543" и ",210" в "9,876,543,210"
{ <i>n</i> , }	Предыдущий элемент повторяется как минимум <i>n</i> раз.	"\d{2,}"	"166", "29", "1930"
{ <i>n</i> , <i>m</i> }	Предыдущий элемент повторяется как минимум <i>n</i> раз, но не более чем <i>m</i> раз.	"\d{3,5}"	"166", "17668" "19302" в "193024"
*?	Предыдущий элемент не повторяется вообще или повторяется, но как можно меньшее число раз.	\d*?\d	".0", "19.9", "219.9"
+?	Предыдущий элемент повторяется один или несколько раз, но как можно меньшее число раз.	"be+?"	"be" в "been", "be" в "bent"
??	Предыдущий элемент не повторяется или повторяется один раз, но как можно меньшее число раз.	"rai??n"	"ran", "rain"
{ <i>n</i> }?	Предыдущий элемент повторяется ровно <i>n</i> раз.	",\d{3}?"	",043" в "1,043.6", ",876", ",543" и ",210" в "9,876,543,210"
{ <i>n</i> , }?	Предыдущий элемент повторяется как минимум <i>n</i> раз (как можно меньше).	"\d{2,}?"	"166", "29", "1930"
{ <i>n</i> , <i>m</i> }?	Предыдущий элемент повторяется не менее <i>n</i> и не более <i>m</i> раз (как можно меньше).	"\d{3,5}?"	"166", "17668" "193", "024" в "193024"

## Конструкции обратных ссылок

Обратная ссылка позволяет впоследствии идентифицировать ранее найденную соответствующую часть выражения в том же регулярном выражении. В следующей таблице перечислены конструкции обратных ссылок, поддерживаемые регулярными выражениями .NET. Для получения дополнительной информации см. [Backreference Constructs](#).

Конструкция обратных ссылок	Описание:	Шаблон	Число соответствий
\ <i>number</i>	Обратная ссылка. Соответствует значению нумерованной части выражения.	(\w)\1	"ee" в "seek"
\k< <i>имя</i> >	Именованная обратная ссылка. Соответствует значению именованного выражения.	(?<char>\w)\k<char>	"ee" в "seek"

## Конструкции чередования

Конструкции изменения модифицируют регулярное выражение, включая сопоставление по принципу "либо-либо". Такие конструкции состоят из языковых элементов, приведенных в следующей таблице. Дополнительные сведения см. в разделе [Alternation Constructs](#).

Конструкция изменения	Описание:	Шаблон	Число соответствий
	Соответствует любому элементу, разделенному вертикальной чертой ( ).	th(e is at)	"the", "this" в "this is the day."
(? <i>expression</i> ) <i>yes</i>   <i>no</i> )	Соответствует <i>yes</i> в случае соответствия шаблона регулярного выражения, определяемого выражением <i>expression</i> ; в противном случае соответствует дополнительной части <i>no</i> . <i>expression</i> интерпретируется как утверждение нулевой ширины.	(?(A)A\d{2}\b\b\d{3}\b)	"A10", "910" в "A10 C103 910"
(?( <i>name</i> ) <i>yes</i>   <i>no</i> )	Соответствует <i>yes</i> в случае соответствия именованной или нумерованной группы захвата <i>name</i> ; в противном случае соответствует дополнительному <i>no</i> .	(?(<quoted>")?(?(<quoted>).+)?"\S+s)	Dogs.jpg, "Yiska playing.jpg" в "Dogs.jpg "Yiska playing.jpg"

## Подстановки

Подстановки – это языковые элементы регулярных выражений, которые поддерживаются в шаблонах замены. Для получения дополнительной информации см. [Substitutions](#). Приведенные в следующей таблице метасимволы являются атомарными утверждениями нулевой ширины.

Знак	Описание:	Шаблон	Шаблон замены	Входная строка	Результирующая строка
\$ <i>number</i>	Замещает часть строки, соответствующую группе <i>number</i> .	\b(\w+)(\s)(\w+)\b	\$3\$2\$1	"один два"	"два один"
\${ <i>имя</i> }	Замещает часть строки, соответствующую именованной группе <i>name</i> .	\b(?<word1>\w+)(\s)(?<word2>\w+)\b	\${ word 2} \${ word 1}	"один два"	"два один"
\$\$	Подставляет литерал "\$".	\b(\d+)\s?USD	\$\$\$1	"103 USD"	"\$103"

\$&	Замещает копией полного соответствия.	\\$?\d*\.?\d+	**\$&**	"\$1.30"	***\$1.30***
\$</code> <span data-ttu-id="69291-387">Замещает весь текст входной строки до соответствия.</span><span class="sxs-lookup"><span data-stu-id="69291-387">Substitutes all the text of the input string before the match.</span></span> B+ <code>\$	"AABBCC"	"AAAACC"			
\$'	Замещает весь текст входной строки после соответствия.	B+	\$'	"AABBCC"	"AACCCC"
\$+	Замещает последнюю захваченную группу.	B+(C+)	\$+	"AABBCCDD"	AACCCDD
\$_	Замещает всю входную строку.	B+	\$_	"AABBCC"	"AAAABBC CCC"

## Параметры регулярных выражений

Можно определить параметры, управляющие интерпретацией шаблона регулярного выражения обработчиком регулярных выражений. Многие из этих параметров можно указать в шаблоне регулярного выражения либо в виде одной или нескольких констант [RegexOptions](#). Этот краткий справочник перечисляет только встраиваемые параметры. Дополнительные сведения о встроенных параметрах и параметрах [RegexOptions](#) см. в статье [Regular Expression Options](#).

Встроенный параметр можно задать двумя способами:

- С помощью [прочих конструкций](#) (?imnsx-imnsx), где минус (-) перед параметром или набором параметров отключает эти параметры. Например, (?i-mn) включает сопоставление без учета регистра (i), отключает многострочный режим (m) и отключает захват неименованных групп (n). Параметр применяется к шаблону регулярного выражения от точки, в которой определен параметр, и действует либо до конца шаблона, либо до точки, в которой другая конструкция отменяет параметр.
- С помощью [grouping construct](#)(?imnsx-imnsx:часть выражения), которая определяет параметры для только для указанной группы.



Механизм регулярных выражений .NET поддерживает следующие встроенные параметры.

Параметр	Описание:	Шаблон	Число соответствий
i	Использовать соответствие без учета регистра.	\b(?:i)a(?:-i)a\w+\b	"aardvark", "aaaAuto" в "aardvark AAAAuto aaaAuto Adam breakfast"
m	Использовать многострочный режим. ^ и \$ соответствуют началу и концу строки (line), а не началу и концу строки (string).	Пример см. в подразделе "Многострочный режим" раздела <a href="#">Regular Expression Options</a> .	
n	Не захватывать неименованные группы.	Пример см. в подразделе "Только явные захваты" раздела <a href="#">Regular Expression Options</a> .	
s	Использовать однострочный режим.	Пример см. в подразделе "Однострочный режим" раздела <a href="#">Regular Expression Options</a> .	
x	Игнорировать знаки пробела в шаблоне регулярного выражения, не преобразованные в escape-последовательность.	\b(?:x) \d+ \s \w+	"1 aardvark", "2 cats" в "1 aardvark 2 cats IV centurions"

## Прочие конструкции

Прочие конструкции либо изменяют шаблон регулярных выражений, либо предоставляют сведения о нем. В следующей таблице перечислены все прочие конструкции, поддерживаемые .NET. Для получения дополнительной информации см. [Miscellaneous Constructs](#).

Конструкция	Определение	Пример
(?imnsx-imnsx)	Устанавливает или отключает такие параметры, как учет регистра в середине шаблона. Дополнительные сведения см. в статье <a href="#">Параметры регулярных выражений</a> .	\bA(?:i)b\w+\b соответствует "ABA", "Able" в "ABA Able Act"
(?#комментарий)	Встроенное примечание. Примечание заканчивается первой закрывающей скобкой.	\bA(?:#Matches words starting with A)\w+\b
# [до конца строки]	Комментарий режима X. Примечание начинается от знака # без обратной косой черты и продолжается до конца строки.	(?x)\bA\w+\b#Matches words starting with A