

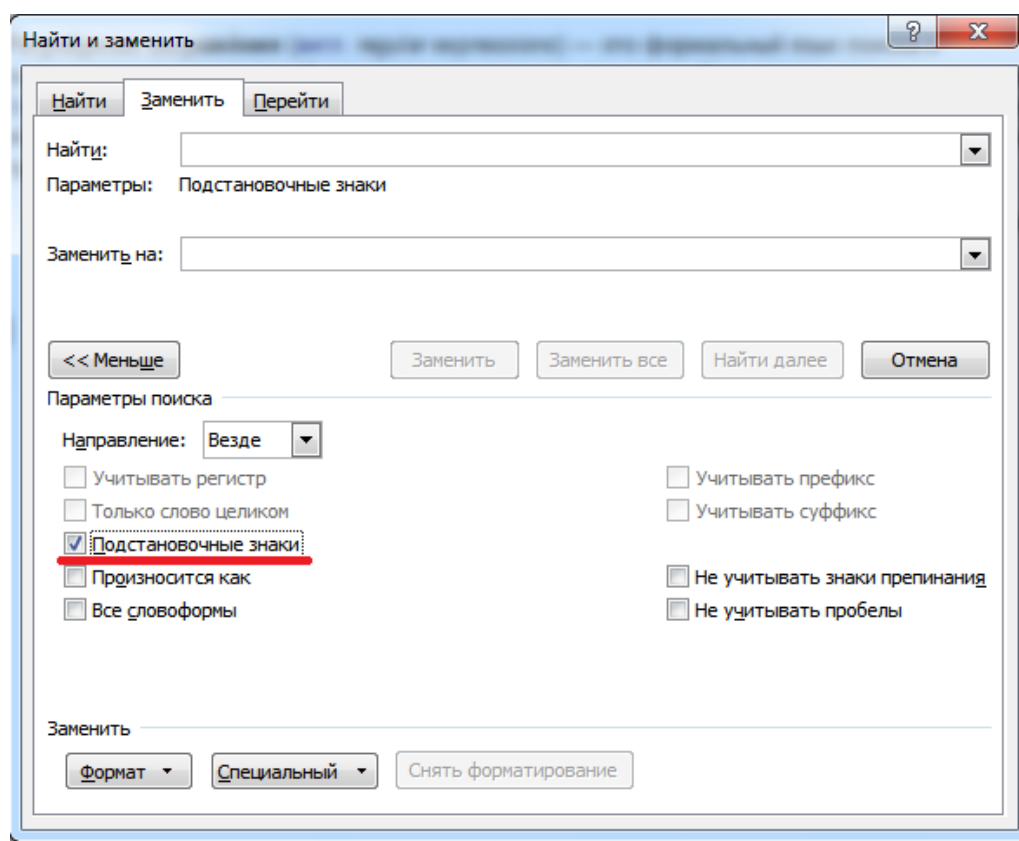
## Регулярные выражения MS-Word

Регулярные выражения(англ. *regular expressions*)— это формальный язык поиска и осуществления манипуляций с подстроками в тексте, основанный на использовании метасимволов. По сути это строка-образец (англ. *pattern*, по-русски её часто называют «шаблоном», «маской»), состоящая из символов и метасимволов и задающая правило поиска.

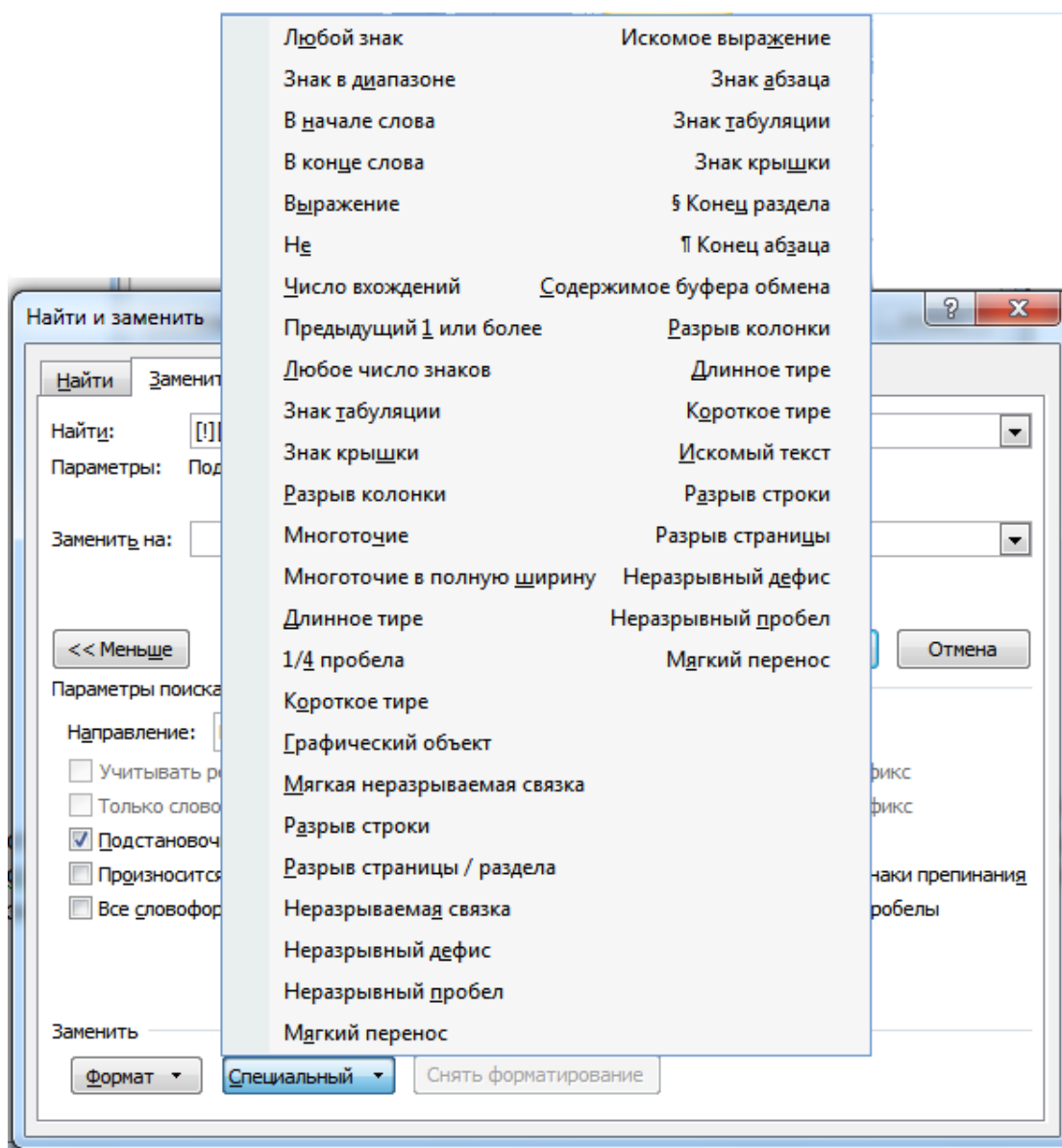
(Википедия)

Т.е. регулярным выражением задается шаблон, по которому происходит поиск в документе. Например, нужно найти только пятизначное число, значение которого заранее неизвестно, либо найти время, которое задано форматом ЧЧ:ММ (ЧЧ – часы: 00, 01, 02, ... 23; ММ – минуты: 00, 01, 02, ...). Регулярные выражения позволяют искать последовательности символов, которые имеют определенный формат, но не имеют заранее известного значения.

## Как включить регулярные выражения в MS-word?



Для того чтобы MS-WORD воспринимал регулярные выражения нужно поставить отметку «Постановочные знаки». После чего в полях «Найти» и «Заменить на» служебные символы регулярных выражений будут восприниматься как подстановочные знаки, а не как символы, которые нужно найти.



Как подсказку можно использовать список символов, который находится в выпадающем списке «Специальный».

## Краткое введение

### ? и \*

Два самых основных подстановочных знака это «?» и «\*». «?» используется для представления одного любого символа, а «\*» для представления любого количества любых символов.

S?t должно найти "Set", "Sat", "Sit", и т.д.

S\*t должно найти слова: "Secret", "Serpent", "Sailing boat".

### @

Указывает, что предыдущий символ должен встретиться один или более раз. Sel@ найдет Sell, Sellll и т.д.



Указывают на начало и конец слов. Например, <S\*t> найдет “Serpent” и “Sailing boat”, но НЕ “Sailing boats”(т.к. в boats еще есть s, которая является окончанием слова)



В таких скобках перечисляются символы, которые могут стоять на данной позиции на позиции. [asd] означает, что на данной позиции может стоять один из символов. Например, S[aeo]x найдет слова Sax, Sex, Sox.

К данным скобкам можно добавить символ @(повторение символа один или более раз). Например, если написать S[aeo]@x , то также найдем слова Saex, Soex, Saox(слова смысла не имеют и используются только для примера).

Можно не перечислять все символы, а задать их интервалом.

[A-Z] означает, любой заглавный символ английского алфавита.

[0-9] – любая цифра.

[A-z] – любая заглавная или строчная буква английского алфавита.

[0-9A-z] – любая буква английского алфавита или цифра.



Символ экранирования.

**Экранирование символов**— замена в тексте управляющих символов на соответствующие текстовые подстановки.(Википедия).

Если нам нужно найти символ «?», то просто так вписать мы его не можем, т.к. он воспримется как «любой символ», такая же ситуация будет и с «\*», «@», «<», «>» и т.д.

Для того чтобы нам найти в тексте символ, который является служебным в регулярных выражениях нужно перед этим символом поставить знак экранирования «\»

Например, S[aeo\?]x - найдет Sax, S?x, Sox и т.д.

Этот символ нужно использовать перед символами, если их нужно найти в тексте – () [] <> - @ ? ! \* \ .



! используется как НЕ, и относится ко всем скобкам. [!o] означает любой символ кроме «o». [!A-Z] – любой символ кроме заглавных букв.



Указывает на количество повторений предыдущего символа(такой смысл как и у “@”)

{n} – символ должен повториться n раз.

A{3} должно найти AAA.

[asd]{2} должно найти: aa, as, ad, sd, da,sa,ds и т.д.

{n;m} – символ должен повториться от n до m раз.

D{3;6} – символ D может повторяться от 3 до 6 раз

{n;} – символ может повторяться от n раз и больше.



Эти скобки никак не влияют на поиск. Они используются для группирования результатов поиска. Например, нам нужно найти текст «ббб ааа» и эти два слова поменять местами. Для этого мы в найти пишем (ббб) (ааа) , в данном варианте скобки являются служебными символами и не ищутся в тексте. Если в тексте будит найден результат, то мы можем

обратиться к каждой скобке, написав «\2 \1» в строке замены мы поменяем местами слова(\1 – это первая скобка \2 – вторая и т.д. ).

^

Используется для обозначения непечатаемых символов. Например, ^13 обозначает конец строки(если не стоит отметка «Подстановочные знаки», то этот символ обозначат ^р). Если отмечено «Подстановочные знаки», то в строке поиска нужно ставить ^13, но в строку «заменить на» нужно писать ^р.

## Таблицы служебных знаков

В **Microsoft Word** существуют два типа специальных знаков — *специальные символы* и *подстановочные знаки*. Те и другие удобно вводить в строки поиска и замены используя кнопку **Special (Специальный)** в диалоге поиска и замены.

Специальными символами называют символы, которых нет на клавиатуре и/или которые невозможно непосредственно вписать в строку поиска или замены. Например: принудительный обрыв строки, мягкий перенос, неразрывный пробел и т. п.

Если в диалоге поиска и замены задействована опция **Подстановочные знаки**, то становится возможным формировать сложные условия поиска. Для этого используются подстановочные знаки. С их помощью можно задавать диапазоны символов и выражения, отвечающие ряду условий.

Некоторые символы и знаки можно использовать только в строке поиска, некоторые — только в строке замены. Возможность использования других зависит от того, задействована или нет опция **Подстановочные знаки**.

Ниже приведен полный список символов и знаков.

А еще ниже — примеры их использования.

### СТРОКА ПОИСКА

Специальные символы, которые можно использовать только в том случае, если опция **Use wildcards (Подстановочные знаки)** не задействована(т.е. отметка «Подстановочные знаки не стоит»).

Спец. Символы	Что обозначают	
^p	Paragraph mark (¶)	Символ конца абзаца (¶)
^?	Any character	Любой символ
^#	Any digit	Любая цифра

^\$	Any letter	Любая буква
^e	Endnote mark	Концевая сноска
^d	Field	Поле
^f	Footnote mark	Сноска
^b	Section break (===End of section===)	Обрыв секции (===End of section===)
^w	Whitespace	Любое число и любая комбинация обыкновенных и неразрываемых пробелов, табуляций и концов абзаца

Специальные символы, использование которых не зависит от опции **Use wildcards** (**Подстановочные знаки**).

<i>Спец. символы</i>	<i>Что обозначают</i>	
^t	Tab character (→)	Символ табуляции (→)
^a	Comment mark	Комментарий
^0nnn	ANSI or ASCII characters	Любой ANSI или ASCII символ с кодом nnn
^^	Caret character	Символ ^
^g	Graphic	Графика
^n	Column break (...Column Break...)	Обрыв колонки (...Column Break...)
^l	Manual line break (↵)	Принудительный обрыв строки (↵)
^m	Manual page break (---Page Break---)	Принудительный обрыв страницы (---Page Break---). Если опция Use wildcards задействована, то ищет и принудительный обрыв страницы, и обрыв секции.
^+	Em dash (—)	Длинное тире (—). Символ с кодом 0151.
^=	En dash (–)	Короткое тире (–). Символ с кодом 0150.
^s	Nonbreaking space (°)	Неразрываемый пробел (°)
^~	Nonbreaking hyphen (-)	Неразрываемый дефис (-)

^-	Optional hyphen (→)	Мягкий перенос (→)
----	---------------------	--------------------

Подстановочные знаки, которые можно использовать только в том случае, если опция **Use wildcards** (Подстановочные знаки) задействована (т.е. стоит отметка «Подстановочные знаки»).

Спец. знаки	Что обозначают	Пример строки поиска	Что будет находить
?	Один любой символ	б?к	бак, бок, бук, б5к, б¶к и т. п.
*	Любое число любых символов	б*к	бык, бардак, белый.¶ Полковник и т. п.
[]	Один из указанных символов	б[аоу]к	бак, бок, бук
[-]	Один символ из диапазона. Диапазон должен быть указан в порядке возрастания кодов символов.	[а-яё]	Любая строчная русская буква
		[А-ЯЁ]	Любая прописная русская буква
		[0-9]	Любая цифра
[!]	Один любой символ не указанный после восклицательного знака	б[!ы]к	бак, бок и т. п., но не бык
[!x-z]	Один любой символ, не входящий в диапазон указанный после восклицательного знака	[!а-яё]ок	Бок, Док и т. п., но не бок, док
		[!0-9]	Любой символ кроме цифр
{n}	Строго n штук предыдущего символа или выражения. Выражением является все то, что заключено в круглые скобки. Выражение может состоять как из конкретных символов, так и содержать спец. знаки.	10{3}	1000, но не 100, 10000
		10(20){2}	102020, но не 1020, 10202020
{n;}	n и более штук предыдущего символа или выражения	10{3;}	1000, 10000, 100000 и т. д., но не 100
{n;m}	От n до m штук предыдущего символа или выражения	10{3;4}	1000, 10000, но не 100, 100000
@	Ноль или более штук предыдущего символа или выражения	10@	10, 100, 1000, 10000 и т. д.

<	Начало слова	<бок	боксер, но не колобок
>	Конец слова	бок>	колобок, но не боксер

## СТРОКА ЗАМЕНЫ

Специальные символы, использование которых не зависит от опции **Use wildcards** (**Подстановочные знаки**).

<i>Спец. символы</i>	<i>Что обозначают</i>	
^p	Paragraph mark (¶)	Символ конца абзаца (¶)
^t	Tab character (→)	Символ табуляции (→)
^0nnn	ANSI or ASCII characters	Любой ANSI или ASCII символ с кодом nnn
^^	Caret character	Символ ^
^c	Clipboard contents	Содержимое буфера обмена
^&	Contents of the Find what box	Содержимое строки поиска (или то, что найдено)
^n	Column break (…Column Break…)	Обрыв колонки (…Column Break…)
^	Manual line break (↵)	Принудительный обрыв строки (↵)
^m	Manual page break (---Page Break---	Принудительный обрыв страницы (---Page Break---
^+	Em dash (—)	Длинное тире (—). Символ с кодом 0151.
^=	En dash (-)	Короткое тире (-). Символ с кодом 0150.
^s	Nonbreaking space (°)	Неразрываемый пробел (°)
^~	Nonbreaking hyphen (-)	Неразрываемый дефис (-)
^-	Optional hyphen (¬)	Мягкий перенос (¬)

Подстановочные знаки, которые можно использовать только в том случае, если опция **Use wildcards** (**Подстановочные знаки**) задействована.

<i>Знаки</i>	<i>Что обозначают</i>

\n	Expression n	Выражение номер n из строки поиска
----	--------------	------------------------------------

Еще одна таблица

ЧТОБЫ НАЙТИ	ВВЕДИТЕ	ПРИМЕР
Один символ	?	Выражение "к?t" позволяет найти слова "кот" и "кит".
Любую последовательность символов	*	Выражение "в*d" позволяет найти слова "вход" и "выход".
Начало слова	<	Выражение "<(сол)" позволяет найти слова "солнце" и "солист", но не слово "фасоль".
Конец слова	>	Выражение "(ель)>" позволяет найти слова "ель" и "шрапнель", но не слово "ельник".
Один из указанных символов	[ ]	Выражение "кр[ea]н" позволяет найти слова "крен" и "кран".
Любой символ из указанного диапазона	[ - ]	Выражение "[з-ф]абор" позволяет найти слова "забор" и "табор". Диапазон должен быть указан в порядке возрастания.
Любой символ, кроме символов, которые находятся в диапазоне, указанном в скобках	[ !э-я ]	Выражение "к[!е-р]а" позволяет найти слова "кода" и "коса", но не слова "кожа" и "кора".
Предыдущий символ или выражение, повторенные n раз	{n}	Выражение "жарен{2}ый" позволяет найти слово "жаренный", но не слово "жареный".
Предыдущий символ или выражение, повторенные не менее n раз.	{n,}	Выражение "жарен{1,}ый" позволяет найти слова "жареный" и "жаренный".
Предыдущий символ или выражение, повторенные от n до m раз	{n,m}	Выражение "10{1,3}" позволяет найти числа 10, 100 и 1000.
Предыдущий символ или выражение, повторенные не менее одного раза	@	Выражение "больше@" позволяет найти слова "больше" и "большее".

## Примеры

Опция **Use wildcards (Подстановочные знаки)** должна быть задействована в диалоге поиска и замены. Это условие должно соблюдаться для всех приведенных ниже примеров.

### ПРИМЕРЫ ПОИСКА

Строка поиска	Что будет находить
[!^0013]^0013[!^0013]	Одиночный конец абзаца в окружении двух неконцов. (Подробный разбор см. ниже).
^0013{2;}	Два и более конца абзацев



<code>^0032{2;}</code>	Два и более пробела
<code>[.,:;!\\?]</code>	Знак препинания ( . , : ; ! ? )
<code>[0-9]^0032[0-9]</code>	Две цифры, разделенные пробелом
<code>[0-9]^0032[A-Za-zA-яЁё]</code>	Цифра и буква (английская или русская) разделенные пробелом. (Подробный разбор см. ниже).
<code>-[0-9]</code>	Дефис и цифра
<code>&lt;[0-9]@[0-9]&gt;</code>	Два целых числа, разделенные дефисом
<code>\([0-9]\)</code>	Цифра в круглых скобках
<code>&lt;[0-9]&gt;</code>	Целое число (точнее — непрерывная последовательность цифр)
<code>&lt;[0-9]@[0-9]&gt;</code>	Десятичная дробь, набранная через запятую
<code>&lt;[A-Za-z]&gt;</code>	Английское слово
<code>&lt;[A-яЁё]&gt;</code>	Русское слово
<code>&lt;[A-ЯЁ][a-яё]&gt;</code>	Русское слово, набранное строчными буквами, но с прописной. (Подробный разбор см. ниже).

#### ПОДРОБНЫЙ РАЗБОР НЕСКОЛЬКИХ СТРОК ПОИСКА

`[!^0013]^0013[!^0013]`

<i>1 символ строки поиска</i>	<i>2 символ</i>	<i>3 символ</i>
Любой символ, но не конец абзаца	Конец абзаца	Любой символ, но не конец абзаца
<code>[!^0013]</code>	<code>^0013</code>	<code>[!^0013]</code>
В квадратных скобках указан символ с кодом 0013 (конец абзаца). Знак «!» указывает на то, что в этой позиции может находиться любой символ кроме конца абзаца.	Конец абзаца (символ с кодом 0013).	В квадратных скобках указан символ с кодом 0013 (конец абзаца). Знак «!» указывает на то, что в этой позиции может находиться любой символ кроме конца абзаца.

`[0-9]^0032[A-Za-zA-яЁё]`

<i>1 символ строки поиска</i>	<i>2 символ</i>	<i>3 символ</i>
Любая цифра	Пробел	Любая буква (английская или русская)

[0-9]	^0032	[A-Za-zA-яЁё]
В квадратных скобках указан диапазон знаков от 0 до 9, т. е. цифры.	Пробел (символ с кодом 0032).	Допустимый символ определен тремя диапазонами и двумя конкретными символами. A-Z — английские прописные буквы; a-z — английские строчные буквы; А-я — русские прописные и строчные буквы; Ёё — прописная и строчная ё (чтобы охватить весь русский алфавит).

<[А-ЯЁ] [а-яё]@>

1 позиция строки поиска	2 позиция
Любая русская прописная буква	Любые русские строчные буквы
<[А-ЯЁ]	[а-яё]@>
В квадратных скобках указан диапазон знаков от А до Я и Ё (т. е. любая русская прописная буква). Знак «<» указывает на то, что прописная буква должна быть в начале слова.	Диапазон в квадратных скобках определяет все русские строчные буквы. Знак «@» говорит о том, что число русских строчных букв может быть равным или больше нуля. Знак «>» указывает на конец слова.

#### ОБ ОПЕРАЦИЯХ С ВЫРАЖЕНИЯМИ

Для работы с выражениями задействуйте опцию **Use wildcards (Подстановочные знаки)**.

В строке поиска выражением является все то, что заключено в круглые скобки (которые в этом случае выступают в роли специальный знаков). Выражение может состоять как из конкретных символов, так и содержать подстановочные знаки или специальные символы. Порядок следования выражений определяет их нумерацию.

В строке замены выражения представляются в виде: \n, где n — номер выражения в строке поиска.

#### ПРОСТОЙ ПРИМЕР, ПОЯСНЯЮЩИЙ РАБОТУ С ВЫРАЖЕНИЯМИ

Будем редактировать предложение: Сережа ест грушу.

Строка поиска	Строка замены	Результат поиска и замены
(Сережа )(ест )(грушу).	\3 \2\1	грушу ест Сережа .
	\1\3 с хрустом \2	Сережа грушу с хрустом ест .
	\1и Катя едят \3	Сережа и Катя едят грушу .
	\1\2\3 и Катя \2\3	Сережа ест грушу и Катя ест грушу .

## ПРИМЕРЫ ПОИСКА И ЗАМЕНЫ

<i>Строка поиска</i>	<i>Строка замены</i>	<i>Результат поиска и замены</i>
<code>([!^0013]^0013)([!^0013])</code>	<code>\1^0013\2</code>	Вставка пустой строки между теми абзацами, между которыми пустой строки не было.
<code>(^0013){2;}</code>	<code>\1</code>	Удаление пустых строк
<code>^0032{2;}</code>	<code>^0009</code>	Замена двух или более пробелов на табуляцию
<code>^0032([.,;:\!?\])</code>	<code>\1</code>	Удаление пробела перед знаком препинания ( . , ; ! ? )
<code>([0-9])^0032([0-9])</code>	<code>\1^s\2</code>	Замена пробела между цифрами на неразбиваемый пробел
<code>([0-9])^0032([A-Za-zA-яЁё])</code>	<code>\1^s\2</code>	Замена пробела между цифрой и буквой (английской или русской) на неразбиваемый пробел
<code>-([0-9])</code>	<code>^0150\1</code>	Замена дефиса перед цифрой на минус
<code>\((([0-9])\)</code>	<code>[\1]</code>	Замена круглых скобок вокруг цифры на квадратные

## И ЕЩЕ ДВА ПРИМЕРА

### Пример 1

Замена в документе фамилии Иванов на фамилию Петров сразу во всех падежах:

<i>Строка поиска</i>	<i>Строка замены</i>	<i>Результат поиска и замены</i>
<code>Ивано([а-я]@&gt;)</code>	<code>Петро\1</code>	Замена фамилии Иванов на фамилию Петров сразу во всех падежах

Конечно, такая замена возможна только с фамилиями, у которых одинаковые падежные окончания.

### Пример 2

В документе очень много десятичных дробей. Часть дробей набрана через точку, часть — через запятую.

Требуется произвести такую замену, чтобы сделать написание всех дробей единообразным.

Вот два варианта этой замены:

<i>Строка поиска</i>	<i>Строка замены</i>	<i>Результат поиска и замены</i>
<code>(&lt;[0-9]@).([0-9]@&gt;)</code>	<code>\1,\2</code>	Десятичная точка будет заменена на запятую
<code>(&lt;[0-9]@),([0-9]@&gt;)</code>	<code>\1.\2</code>	Десятичная запятая будет заменена

		на точку
--	--	----------

## **Полезные ссылки**

1. [http://office.microsoft.com/ru-ru/word-help/HA102350661.aspx#\\_Toc286834987](http://office.microsoft.com/ru-ru/word-help/HA102350661.aspx#_Toc286834987)
2. [http://artefact.lib.ru/design/text\\_khozyainov.shtml](http://artefact.lib.ru/design/text_khozyainov.shtml)
3. <http://word.mvps.org/FAQs/General/UsingWildcards.htm> (англ.яз.)