***Поиск слова в текстовом документе***

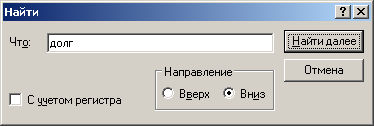
**Что нужно знать**:

* текстовые редакторы и текстовые процессоры имеют встроенную функцию поиска; большинство программ (*Блокнот, OpenOffice, LibreOffice*) просто ищут цепочку символов, то есть находят все формы данного слова
* в наиболее совершенных редакторах (*Microsoft Office*) есть возможность отметить режим *Только слово целиком*, при этом программа ищет только заданное слово именно в этой форме
* если нужно найти слова, начинающиеся только со строчной или только с заглавной буквы, нужно включить флажок *С учётом регистра*

**Пример 1 (демо-2021).** С помощью текстового редактора определите, сколько раз, не считая сносок, встречается слово «долг» или «Долг» в тексте романа в стихах А.С. Пушкина «Евгений Онегин» (файлы 10-0.docx, 10-0.txt). Другие формы слова «долг», такие как «долги», «долгами» и т.д., учитывать не следует. В ответе укажите только число.

**Решение (простейшие текстовые редакторы):**

1. Решение этой задачи существенно зависит от возможностей программы, которую вы используете. Рассмотрим сначала самый худший вариант, когда в вашем распоряжении только простейший текстовый редактор типа Блокнота.
2. Загрузите файл 10-0.txt в редактор.
3. С помощью комбинации клавиш *Ctrl+F* (или верхнего меню *Правка – Найти*) нужно вызвать окно поиска и ввести нужное слово:

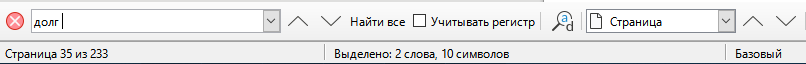


1. **Важно**! Поскольку нас интересует слово «долг», которое может начитаться как с заглавной, так и со строчной буквы, НЕ нужно включать флажок «С учетом регистра»
2. далее щелкаем по кнопке *Найти далее* (можно также щёлкнуть мышью на тексте и нажимать клавишу *F3*)
3. к сожалению, простые программы ищут только заданную цепочку символов и не умеют искать слово в одной заданной форме (то есть будут найдены также слова «долго», «долги», «долгий» и т.п., которые нас не интересуют); поэтому приходится просматривать все найденные слова и вручную считать, сколько раз встретится именно слово «долг», а не другие слова, содержащие эту цепочку символов
4. Ответ: 2.

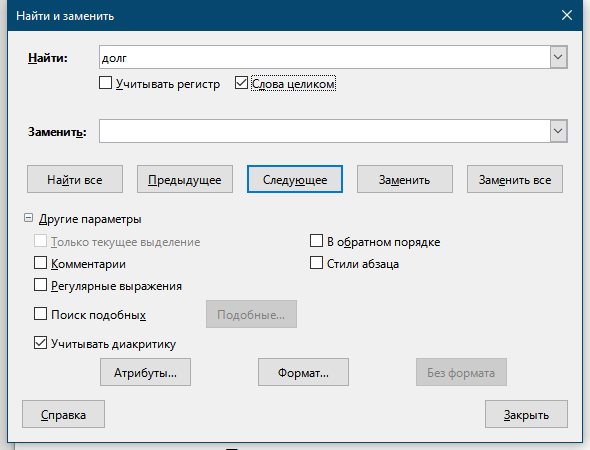
**Решение (LibreOffice, А.Т. Фомин, г. Ейск):**

**«Быстрый» поиск:**

1. Нажмите на клавиатуре сочетание клавиш Ctrl + F. В нижней части окна появляется панель поиска:
2. Введите слово для поиска. Чтобы исключить вхождение указанного слова как подстроки в других словах, введите после слова символ «пробел». Нажмите на кнопку «Найти все». Результат поиска появится в статус-баре ниже:



**«Расширенный» поиск**

1. Перейдите в меню программы Правка → Найти и заменить (сочетание клавиш Ctrl + H). Появится диалоговое окно:
2. Установите флажок «Слова целиком» и нажмите на кнопку «Найти все». Числовой результат поиска появится в статус-баре, в нижней части окна программы. Для перемещения в документе, по найденным словам, нажимайте кнопки «Предыдущее» (Alt + П) или «Следующее» (Alt + Ю).

**«Сложный» поиск**

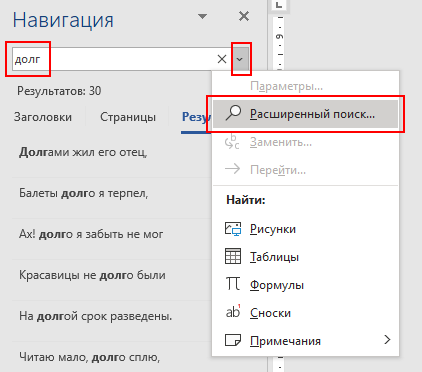
1. Для поиска подстроки по определенным («сложным») критериям используйте «Регулярные выражения». Установите флажок «Регулярные выражения» в диалоговом окне «Найти и заменить». Теперь в строке поиска можно использовать регулярные выражения. Например, давайте найдем в тексте все слова, которые начинаются на долг, а заканчиваются любым одним алфавитным символом (от «а» до «я»). Для этого введите следующее регулярное выражение: [:space:]долг[а-я][:space:]

Количество таких слов в романе будет 14 («долго» и «долги»).

**Некоторые регулярные выражения:**

|  |  |
| --- | --- |
| **Символ** | **Результат/Использование** |
| Любой символ | Представление данного символа, если не указано иное. |
| . | Представление любого отдельного символа, кроме разрыва строки или конца абзаца. Например, условие поиска "п.ск" возвращает и "пуск", и "писк". |
| \* | Поиск нулевого или большего числа символов перед "\*". Например, при поиске "Аб\*в" будут найдены "Ав", "Абв", "Аббв", "Абббв" и т. д. |
| + | Поиск одного или большего числа символов перед "+". Например, при поиске "AX.+4" будет найдено "AXx4", но не "AX4".  Всегда будет найдена самая длинная возможная строка, соответствующая данному искомому элементу в абзаце. Если в абзаце содержится строка "AX 4 AX4", выделяется весь фрагмент. |
| ? | Поиск нулевого числа или одного символа перед "?". Например, при поиске термина "Тексты?" будут найдены "Текст" и "Тексты", а при поиске "x(ab|c)?y" будут найдены "xy", "xaby" или "xcy". |
| \ | Поиск воспринимает специальные символы после "\" как обычные символы, а не как регулярное выражение (за исключением сочетаний \n, \t, \> и \<). Например, "текст\." находит "текст.", а не "тексты" или "тексту". |
| \> | Поиск элемента, находящегося в конце слова. Например, при поиске "бор\>" будет найдено "выбор", но не "борьба". |
| \< | Поиск элемента, находящегося в начале слова. Например, при поиске "\<бор" будет найдено "борьба", но не "выбор". |
| [абв123] | Представление одного из символов, находящихся в скобках. |
| [а-д] | Соответствует любому из символов между "a" и "e", включая начальный и конечный символы. Символы упорядочены по кодовым значениям. |
| [а-де-х] | Представление любого символа, находящегося между "а" и "д" или между "е" и "х". |
| [^а-с] | Соответствует любому символу, кроме символов между "a" и "s". |
| | | Поиск терминов, которые встречаются до "|" и после "|". Например, результатом поиска "этот|тот" будет "этот" и "тот". |
| {2} | Определяет допустимое количество вхождений данного символа перед открывающей скобкой. Например, результатом поиска "сине{2}" также будет "синева". |
| {1,2} | Определяет максимальное и минимальное допустимое количество вхождений данного символа перед открывающей скобкой. Например, результатом поиска "сине{1,2}" также будет "синее" и "синева". |
| {1,} | Этот параметр определяет, какое минимальное число раз может встречаться символ перед открывающей скобкой. Например, при поиске "сине{2,}" будет найдено "синее", "синеее" и "синеееее". |
| [:alpha:] | Представление алфавитного символа. [:alpha:]+ позволяет найти один из них. |
| [:digit:] | Представление десятичного знака. [:digit:]+ позволяет найти один из них. |
| [:alnum:] | Представление алфавитно-цифрового символ ([:alpha:] и [:digit:]). |
| [:space:] | Представляет символ пробела (но не другие пробельные символы). |

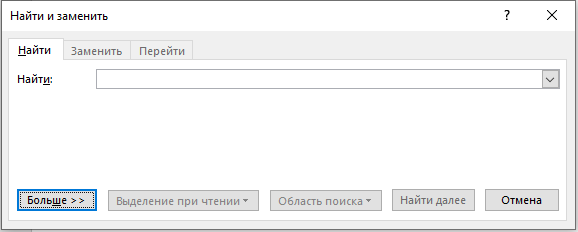
1. Более подробную информацию по использованию регулярных выражений см. на сайте справки к программе LibreOffice Writer:

<https://help.libreoffice.org/3.5/Common/List_of_Regular_Expressions/ru>

1. (**С. Кабанов**) можно использовать и такой приём: заменить искомое слово любым символом; при этом будет выдано количество замен, которое является ответом

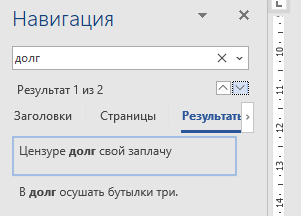
**Решение (Microsoft Word):**

1. в программе *Word* открываем окно поиска (*Ctrl+F*) и вводим нужное слово; получаем те же результаты, что и в *Блокноте*:
2. справа от поля ввода открываем список стрелкой, выбираем *Расширенный поиск*



и нажимаем кнопку *Больше >>*.

1. тут есть возможность выбрать режим поиска «*Только слово целиком*»:
2. после этого программа будет искать только то, что нам нужно, причём в современных версиях *Word* на панели *Навигация* слева от документа сразу видно, сколько слов найдено и в каких предложениях:



1. Ответ: 2.

У Word есть мощные возможности поиска, которые позволяют искать вам текст, числа, форматы, параграфы, разрывы страниц, использовать подстановочные символы, коды полей и многое другое. Используя подстановочные символы, вы можете искать просто всё что угодно в вашем документе. Поскольку функция поиска совмещена с заменой, то вы можете также выполнять весьма замысловатые преобразования текста.

***Регулярное выражение****— это условное обозначение критериев, которым должна соответствовать искомая строка. С помощью регулярных выражений можно найти множество строк или слов, соответствующих заданным условиям.*

***Подстановочные символы (wildcards)****— это****\*****(звёздочка),****.****(точка) и****?****(знак вопроса), которые имеют в регулярных выражениях специальное значение. Например, символ****\*****(звёздочка) обозначает любое количество любых символов, а****?****(знак вопроса) означает любой один символ.*

**Как использовать подстановочные символы в продвинутом поиске**

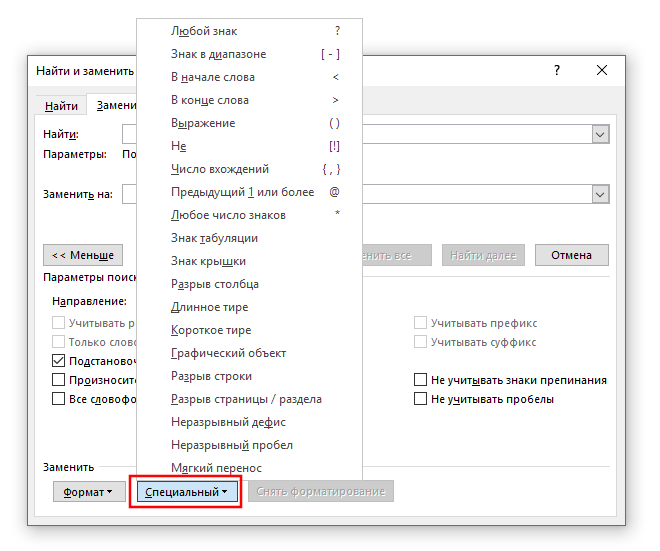
В ленте Word переключитесь на вкладку «Главная» и нажмите кнопку «Заменить»:

В окне «Найти и заменить» кликните «Больше >>», чтобы развернуть диалоговое окно и увидеть дополнительные опции. Если вы увидели кнопку «<< Меньше», значит всё прошло удачно.

В раскрывшемся окне для показа опций поиска, включите флажок «Подстановочные знаки».

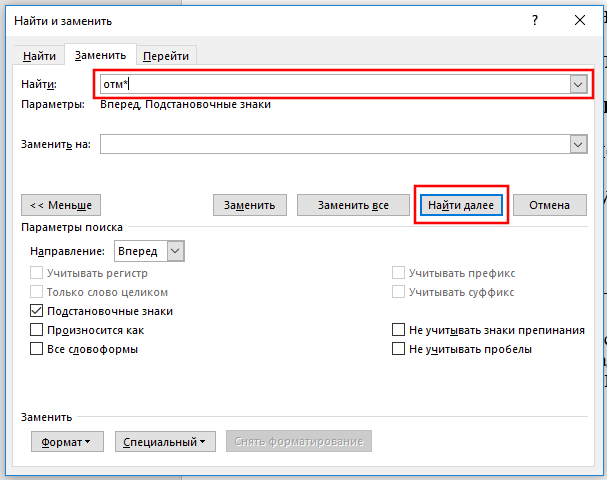
Обратите внимание, что после того, как вы включили опцию «Подстановочные знаки», Word сразу под полем «Найти:» показывает, что эта опция включена. Также когда выбран флажок «Подстановочные знаки», некоторые опции становятся недоступными для включения, а именно: «Учитывать регистр», «Только слово целиком», «Учитывать префикс», «Учитывать суффикс».

Теперь нажмите кнопку «Специальный» для просмотра списка подстановочных знаков.

[](https://zawindows.ru/wp-content/uploads/2018/11/wildcards-word5.png)

В Word доступны следующие подстановочные символы:

| Символ | Значение |
| --- | --- |
| ? | Любой знак |
| [-] | Символ в диапазоне |
| < | В начале слова |
| > | В конце слова |
| () | Выражение - единая последовательность символов. Также применяется для обратных ссылок |
| [!] | Не |
| {;} | Число вхождений |
| @ | Предыдущий 1 или более |
| \* | Любое число знаков |
| ^t | Знак табуляции |
| ^^ | Знак крышки |
| ^n | Знак столбца |
| ^+ | Длинное тире |
| ^= | Короткое тире |
| ^g | Графический объект |
| ^l | Разрыв строки |
| ^m | Разрыв страницы / раздела |
| ^~ | Неразрывный дефис |
| ^s | Неразрывный пробел |
| ^- | Мягкий перенос |

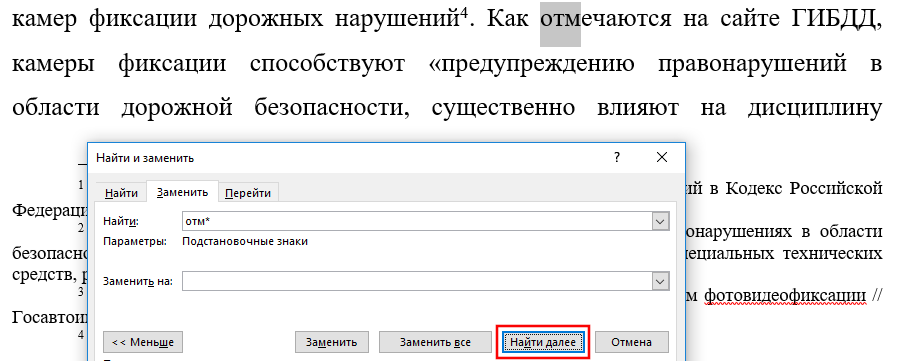
**[](https://zawindows.ru/wp-content/uploads/2018/11/wildcards-word6.png)Для чего используются подстановочные символы?**

**Использование звёздочки для указания любого количества символов**

Скорее всего, самым часто используемым подстановочным символом для вас станет звёздочка. Она означает, что вы хотите найти любое количество символов. Например, чтобы найти все слова, начинающиеся с «отм», напечатайте «отм\*» в строке поиска и затем кликните кнопку «Найти далее». Наше регулярное выражение означает любое количество букв (\* звёздочка), следующих после «отм».

В качестве альтернативы ввода символа звёздочки с клавиатуры, вы можете использовать выбор специального символа из списка подстановочных знаком. Вначале наберите «отм» в строке «Найти». Поставьте галочку «Подстановочные знаки». Затем кликните кнопку «Специальный» и выберите «Любое число символов». После этого нажмите кнопку «Найти далее»:

Word оценивает поиск и показывает вам первое вхождение, которое он найдёт в документе. Продолжайте кликать «Найти далее», чтобы найти все части текста, которые соответствуют вашему поисковому термину.

[](https://zawindows.ru/wp-content/uploads/2018/11/wildcards-word8.png)

Вы должны помнить важную вещь: при включении подстановочных знаков, поиск автоматически становится чувствительным к регистру (такова особенность регулярных выражений, частью которых являются подстановочные символы). Поэтому поиск «**отм\***» и «**Отм\***» даст различные результаты.

**Использование знака вопроса для поиска определённого количества символов**

В то время как звёздочка означает любое количество символов, знак вопроса в регулярном выражении означает единичный (один) символ. Например, если «**отм\***» будет искать строки начинающиеся с «отм» за которыми идёт любое количество символов, то «**отм?**» будет искать строки, начинающиеся с «отм», за которой идёт только один символ.

Как и звёздочку, знак вопроса можно использовать в любой части слова — в том числе в начале и в середине.

Также можно использовать вместе несколько знаков вопроса вместе, тогда они будут обозначать несколько букв.

Например, регулярное выражение для поиска «**о?о?о**» оно означает букву «о», за которой идёт любой символ, затем снова идёт буква «о», затем опять любой символ и затем опять буква «о» найдёт следующие слова:

* п**отоко**в
* мн**огопо**лосных
* мног**ополо**сных
* к**оторо**е
* прав**ового**
* так**ового**
* так**ого о**снования

Можно найти слова с четырьмя буквами «о», шаблон «**о?о?о?о**»:

* к**оторого**
* мн**огополо**сных

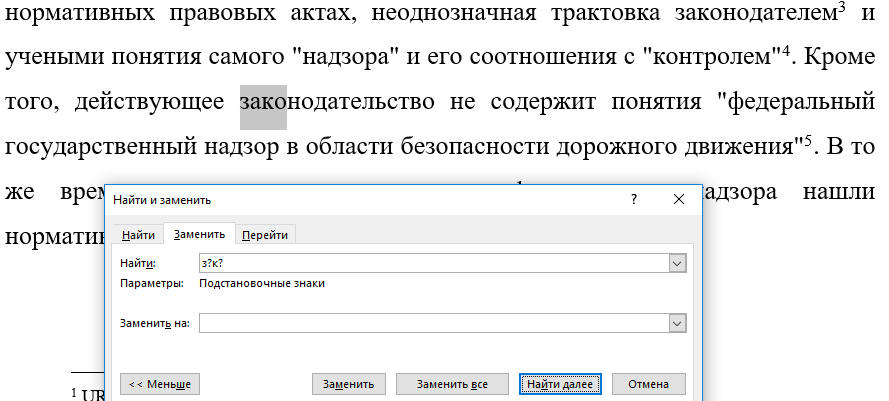
Или с тремя буквами «а», шаблон «**а?а?а**»:

* н**аказа**ния
* з**адача**
* апп**арата**
* выск**азана**
* н**ачала**х

Необязательно использовать одинаковые буквы — составляйте выражения под ваши задачи.

Например, чтобы найти слова, в которых первая буква «з», затем идёт любой другой символ, а затем буква «к» и вновь любой символ, шаблон для поиска «**з?к?**» найдёт:

* **зако**номерности
* **зако**нодательно

[](https://zawindows.ru/wp-content/uploads/2018/11/wildcards-word9.png)

**Использование знака собачка (@) и фигурных скобок ({ and}) для поиска вхождений предыдущего символа**

Вы можете использовать знак собачка (**@**) для указания одного или более вхождения предыдущего символа. Например, «**ro@t**» найдёт все слова, которые начинаются на «ro» и заканчиваются на «t» и которые имеют любое количество букв «o» между этими частями. Поэтому по этим условиям поиска будут найдены слова «rot», «root» и даже «roooooot».

Для большего контроля поиска предыдущих символов, вы можете использовать фигурные скобки, внутри которые укажите точное число вхождений предыдущего символа, который вы хотите найти. Например, поиск «**ro{2}t**» найдёт «root», но не найдёт «rot» или «roooooot».

Также поддерживает синтаксис вида **{n;}**- означает искать количество вхождений символа более n раз; **{;m}** - означает искать количество вхождений символа менее m раз; **{n;m}**- означает искать количество вхождений символа более n раз, но менее m раз.

Чтобы показать более практический пример, немного забежим вперёд, следующий поиск использует набор символов (будут рассмотрены в этой статье чуть ниже), в результате, будут найдены все слова, в которых подряд идут четыре согласных буквы:

|  |  |
| --- | --- |
| 1 | [бвгджзйклмнпрстфхцчшщ]{4} |

**Использование угловых скобок (< и >) для обозначения начала и конца слова**

Думаю, вы заметили, особенно когда составляли поисковые запросы со звёздочкой, что пробел считается за обычный символ и могут быть найдены довольно неожиданные фрагменты большого размера, либо фрагменты, состоящие из двух слов. Вы можете использовать угловые скобки (символы «больше чем» и «меньше чем») для обозначения начала и конца слова поиска. Например, вы можете искать «**<но>**» и Word найдёт все вхождения «но», но не найдёт слова вроде «новости», «законодатель».

Это довольно полезно, но эта техника становится более мощной, когда вы комбинируете её с другими подстановочными символами. Например, с помощью «**<з????>**» вы можете найти все слова, которые начинаются на «з» и состоят ровно из пяти букв.

Вам необязательно использовать обе угловые скобки в паре. Вы можете обозначить просто только начало или конец слова, используя только одну соответствующую скобку. Например «**ство>**» найдёт слова

* руковод**ство**
* множе**ство**
* сред**ство**
* количе**ство**

**Использование квадратных скобок ([ и ]) для поиска определённых символов или диапазонов символов**

Вы можете использовать квадратные скобки для указания любых символов или диапазонов символов. Например «**[а]**» будет искать любые вхождения буквы «а».

В следующем примере, будет искаться строка, которая начинается на «р», затем идёт любая гласная, а затем снова буква «р»: «**р[аеиоуэюя]р**»

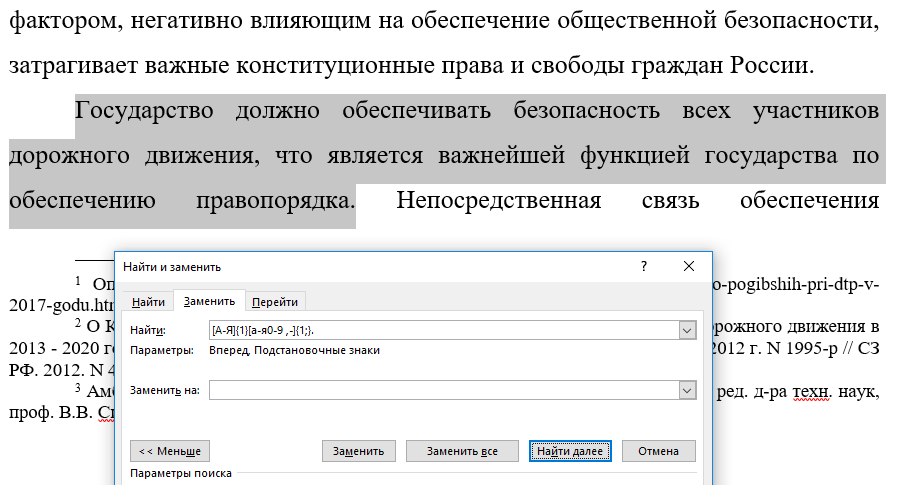
Далее аналогичный пример, но между буквами «р» должно быть две любых гласных: «**р[аеиоуэюя]{2}р**», будет найдено, к примеру, слово «приоритет».

Вы также можете искать квадратные скобки для поиска диапазонов символов, например «**[a-z]**» найдёт любую из этих букв в нижнем регистре. Поиск «**[0-9]**» найдёт любую из этих цифр.

Пример **р[а-и]{2}р** найдёт строку, которая начинается и заканчивается на букву «р» между которыми две любые буквы из указанного диапазона.

Следующий пример довольно сложный, но при этом и довольно интересный:

|  |  |
| --- | --- |
| 1 | [А-Я]{1}[а-я0-9 ,-]{1;}. |

[](https://zawindows.ru/wp-content/uploads/2018/11/wildcards-word10.png)

В начале строки должна стоят любая заглавная буква (**[А-Я]**) ровно один раз (**{1}**). Затем должны идти маленькие буквы, цифры, пробелы, запятые и тире (**[а-я0-9 ,-]**) хотя бы один раз и более (**{1;}**), в самом конце должна стоять точка (**.**)

Думаю вы догадались, что это регулярное выражение которое будет искать предложения. Перечислены не все возможные символы, поэтому не будут найдены предложения, содержащие кавычки и некоторые другие символы, а также вопросительные и восклицательные предложения.

То есть вы можете комбинировать подстановочные символы и дополнять их кванторами количества, а затем это всё комбинировать любое количество раз, пока не получите желаемый результат. На самом деле, это довольно сложно — почти как программирование, поскольку требует абстрактного мышления.

**Использование скобок для группировки поисковых терминов в последовательности**

Вы можете использовать круглые скобки в вашем поиске для группировки последовательностей символов.

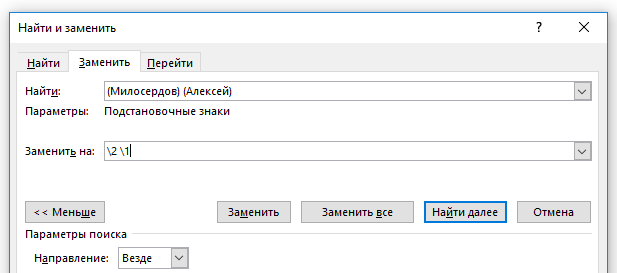
Можно использовать довольно простые шаблоны, например «**(го){3;}**» найдёт строки, в которых «го» встречается три и более раза подряд.

Но настоящую силу эта конструкция покажет при использовании в операциях поиска и замены.

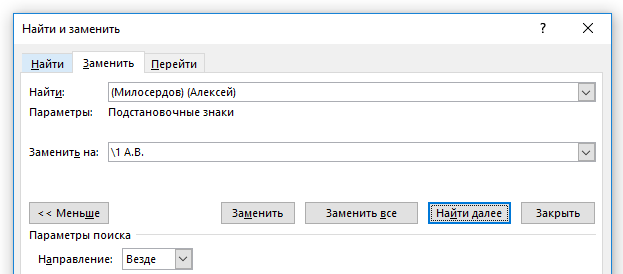
Показанная далее техника называется «обратные ссылки». Начнём с конкретного примера, чтобы было проще понять. Предположим, нам нужно во всём документе поменять местами два слова, допустим, имя и фамилию. К примеру, во всём тексте множество раз упоминается «Милосердов Алексей», а нам нужно, чтобы было «Алексей Милосердов».

Для этого в строке поиска мы вставляем «**(Милосердов) (Алексей)**», а в строке «Заменить на» пишем «**\2 \1**».

Скобки не участвуют в поиске, то есть в результате будет искаться фраза «Милосердов Алексей» При этом то, что было найдено в первых скобках, будет присвоено как значение «**\1**», а то, что было найдено во вторых скобках, станет значением «**\2**». При этом найденная строка заменится на «Алексей Милосердов».

[](https://zawindows.ru/wp-content/uploads/2018/11/wildcards-word11.png)

Можно сделать по-другому, допустим вместо фамилии и имени, я хочу заменить на фамилию и инициалы, тогда в строке поиска я ищу «(Милосердов) (Алексей)», а в строке «Заменить на» пишу «\1 А.В.».

[](https://zawindows.ru/wp-content/uploads/2018/11/wildcards-word12.png)

В скобках можно писать не только слова, там могут использоваться подстановочные символы в разных сочетаниях, а также кванторы количества. Когда находится целая фраза, то Word автоматически пронумеровывает эти группы слева направо (это происходит «под капотом» - мы это не видим), поэтому в поле «Заменить на» мы можем использовать эти группы по их номеру, перед которым нужно поставить обратный слэш.

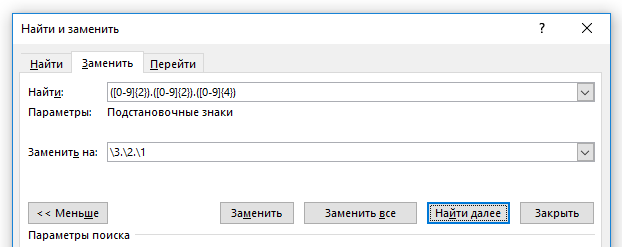
Рассмотрим более жизненный пример. Допустим, у нас по всему документу имеются даты вида 31.12.2019 (то есть в формате день.месяц.год), а мы хотим чтобы все эти даты были в формате 2019.12.31 (то есть год.месяц.день).

Тогда для поиска любых дат составляем регулярное выражение с подстановочными символами:

|  |  |
| --- | --- |
| 1 | «[0-9]{2}.[0-9]{2}.[0-9]{4}» |

«**[0-9]**» - это диапазон, обозначающий все цифры, «**{2}**» - это квантор количества, который говорит, что предыдущий символ должен встречаться ровно два раза. То есть будет искаться строка длиной ровно два символа, состоящая из цифр. Затем идёт точка, затем снова аналогичная строка и точка, и в конце строка из четырёх цифр.

Строка подходит для поиска, но чтобы были созданы обратные ссылки, мы заключаем нужные нам элементы в круглые скобки: «**([0-9]{2}).([0-9]{2}).([0-9]{4})**» - она будет работать точно также, как и предыдущая.

[](https://zawindows.ru/wp-content/uploads/2018/11/wildcards-word13.png)Теперь составляем строку «Заменить на». В начале идёт год, то есть третья группа, она обозначается как «**\3**», затем точка, затем вторая группа, затем опять точка и затем первая группа, получаем «**\3.\2.\1**».