

Работа № 1 – Поиск информации в сети Интернет. Поиск информации в текстовом процессоре Ms Word на основе регулярных выражений

1. Цель занятия:

Изучить возможности поиска информации в сети Интернет, научить студентов системе рационального поиска определённой (нужной) информации, в том числе на основе составления сложных поисковых запросов; научить выполнять анализ и отбор полученной информации.

Изучить возможности поиска и замены информации в текстовом процессоре Ms Word с использованием регулярных выражений; дать навыки продвинутого поиска и замены информации в текстовых документах.

2. Теоретические сведения

2.1. Поиск информации в сети Интернет

Поиск информации – одна из самых востребованных на практике задач, которую приходится решать любому пользователю сети Интернет. Для того чтобы найти информацию в сети Интернет, ему необходимо воспользоваться специальной программой или поисковой системой.

Основные виды и типы поисковых систем

К наиболее популярным на сегодняшний день поисковым системам можно отнести:

1). **Яндекс** — российская поисковая система, создана в 1997 году; является четвёртой среди поисковых систем мира по количеству обрабатываемых запросов. Обеспечивает средство полнотекстового поиска информации с учетом морфологии русского языка; поиск по сайтам, своему каталогу, новостям, товарам, географическим картам, словарям, блогам, картинкам, документам.

2). **Google** — крупнейшая в мире и первая по популярности поисковая система интернета. Основана в 1998 году. Обеспечивает поиск по сайтам; поиск документов в разных форматах; поиск картинок, видео, новостей, поиск по картам; возможность поиска по русскоязычным и российским сайтам; есть система перевода текста на др. языки (49 языков).

3). **Rambler** — российская поисковая система, создана в 1996 году; осуществляет поиск по сайтам с учетом морфологии русского и английского языков.

4). **Bing** — поисковая система, разработанная международной корпорацией Microsoft в 2009 году, занимает 2-е место в списке самых популярных поисковых сайтов по объёму трафика; осуществляет поиск по сайтам, изображениям и видео; обеспечивает возможность поиска по русскоязычным и российским сайтам; есть система перевода текста на др. языки (22 языка)

5). **Yahoo!** — американская поисковая система, создана в 1994 году, пятый по посещаемости веб-сайт в сети Интернет; осуществляет поиск по сайтам, адресам E-mail и др.

Выделяют два основных типа поисковых систем Интернет: *индексные* и *каталоговые*.

Индексные поисковые системы (Яндекс, Rambler, Google, Bing и др.) собирают информацию в Интернет в автоматическом режиме с помощью специальных программ-роботов, посещающих веб-страницы; индексируют информацию, содержащуюся на страницах, внося в свои базы данных.

Каталоговые системы поиска (Rambler, Yahoo! и др.) содержат тематически структурированный каталог серверов, чаще всего пополняются вручную. В каталоговой системе поиска можно начинать поиск с более крупной тематической рубрики, постепенно спускаясь дальше по рубрикам, прийти к ссылке на нужный сайт.

В одних случаях бывает удобней воспользоваться первым типом поисковых систем, в других — вторым. Есть поисковые системы, сочетающие в себе оба принципа работы.

Виды запросов, зависимость запросов от населенных пунктов

Поисковый запрос – это слово или словосочетание, которое вводит пользователь поисковой системы, желая получить определенную информацию. После получения запроса поисковая система автоматически выводит наиболее подходящие (по её мнению) страницы. Все запросы можно условно разделить на несколько видов и типов.

Информационные. Наиболее популярный в поисковых системах тип. Отвечает на вопросы «как», «для чего», «каким образом» и т. д. Например, «куда пойти учиться», «что делать, если ...».

Навигационные. С их помощью могут разыскиваться определенные сайты, в том случае если пользователь забыл точный адрес или сомневается в варианте написания URL.

Транзакционные. К данному типу относятся те, что предполагают совершение какого-либо действия. Например, «купить компьютер» или «заказать такси».

Мультимедийные. Предполагают поиск мультимедийной информации: музыки, фильмов, изображений и т.д.

Общие. К данному типу относятся запросы, которые сложно однозначно отнести к вышеописанным типам. Часто они состоят из 1–2 слов, например, «мышка», «машина» и т.д.

Выделяют также *геозависимые* и *геонезависимые* запросы.

Геозависимые. Осуществляют выдачу информации по географическому расположению. Например, запрос «заказать пиццу» выдаст результаты, которые относятся исключительно к вашему региону. Таким образом, поисковая выдача для разных регионов будет существенно отличаться.

Геонезависимые. Данный вид запросов не привязывается к географическому расположению человека, запрашивающего информацию. Поисковая выдача будет с большой долей вероятности одинакова для разных стран и регионов.

Поиск информации по ключевым словам. Язык запросов поисковой системы

Поисковая система обычно имеет *поле для ввода ключевых слов*, по которым она находит документы, содержащие эти ключевые слова. Чем лучше (качественнее) составлен *запрос* к поисковой машине, тем меньше общее количество найденных документов.

Используя *язык запроса*, содержащий логические операции можно значительно сузить область поиска и более качественно составить запрос к поисковому серверу.

Следует учитывать, что разные поисковые системы могут иметь как различающиеся языки

запросов, так и очень похожие. На сайтах поисковых систем обычно есть ссылки на справочную информацию по использованию поисковых сервисов, языков запросов, рекомендации и т. д.

Некоторые поисковые системы дополнительно имеют функцию *расширенного поиска*, позволяющую пользователю более точно описать то, что ему конкретно нужно и без знания языка запроса. Функцию расширенного поиска имеют, например, поисковые системы Яндекс и Google.

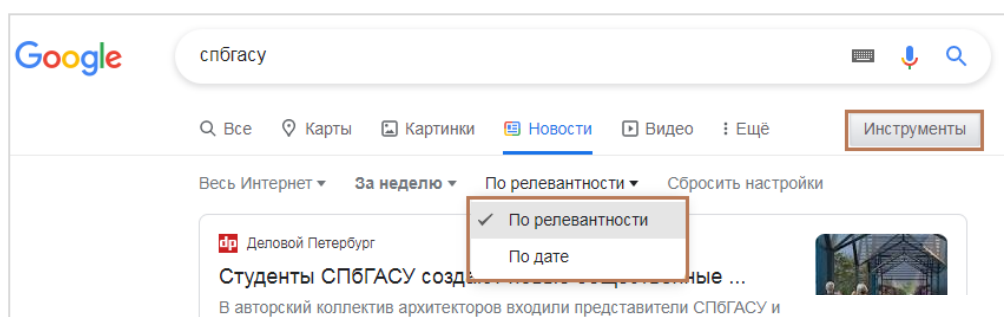
Поисковая система Google

Приведем часто встречаемые операторы запросов, используемые при поиске в Google.

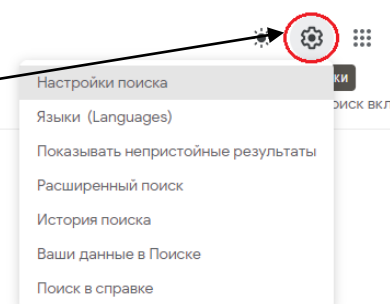
Оператор	Применение
” ”	Находит <i>точное совпадение</i> введенной <i>фразы</i> или <i>слова</i> . Оператор фиксирует количество слов в запросе, но не фиксирует словоформы (изменения падежа, числа и т.д.)
!	оставляет слово в <i>неизменной словоформе</i>
+	<i>делает акцент</i> на одно или несколько слов, помогает системе увидеть, какие из ключевых слов наиболее важные и сформулировать результаты поиска более точно
-	<i>исключает слова</i> из результатов поиска
*	ищет ключи, между которыми есть одно или несколько слов
OR или ()	ищет контент, где используется <i>один</i> из указанных в запросе ключей
AND	отображает результаты всех ключей (их может быть 2 и больше), которые объединены этим оператором
()	<i>Группирует</i> несколько запросов или операторов
site:	ищет страницы определенного <i>сайта</i>
intitle:	ищет страницы, в <i>заголовке</i> (title) которых находится указанные символы
filetype:	ищет страницы определенного <i>типа</i> (например, doc)


Основную часть любой поисковой выдачи составляют *органические (естественные) поисковые результаты* – список документов, найденных поисковой системой, на ранжирование и показ которых не *влияет продаваемая поисковой системой реклама*.

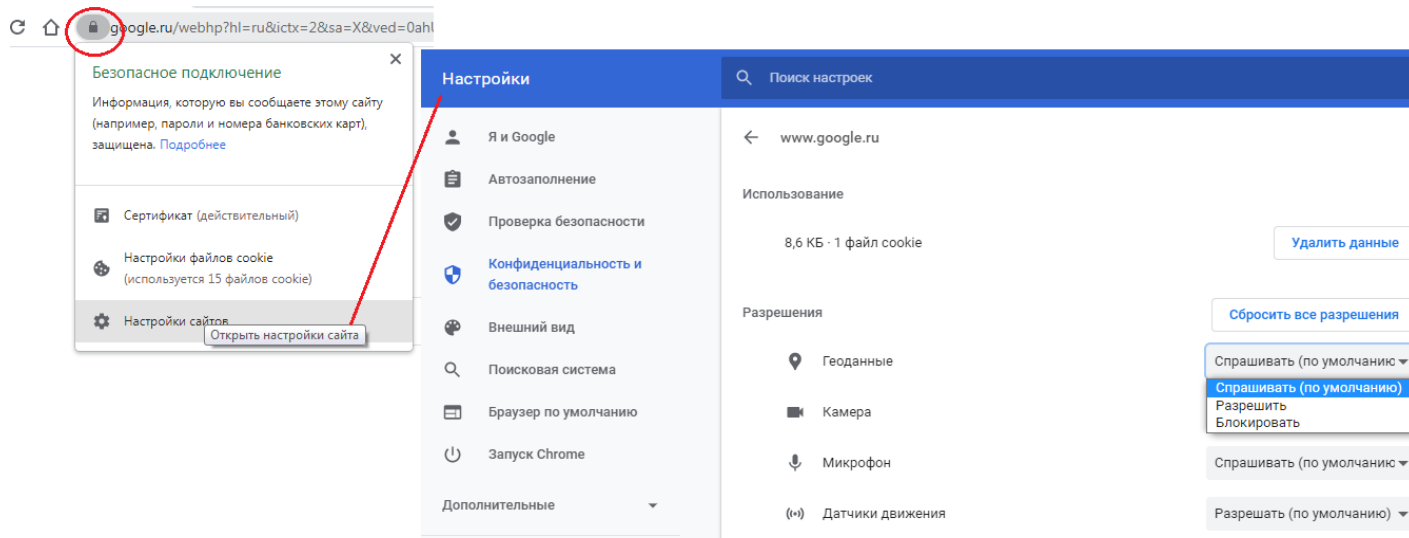
Обычно получаемый список документов бывает *упорядочен по убыванию релевантности (степени соответствия) документов поисковому запросу*, но во многих системах предусмотрены также и другие виды сортировки, например, по дате документов.



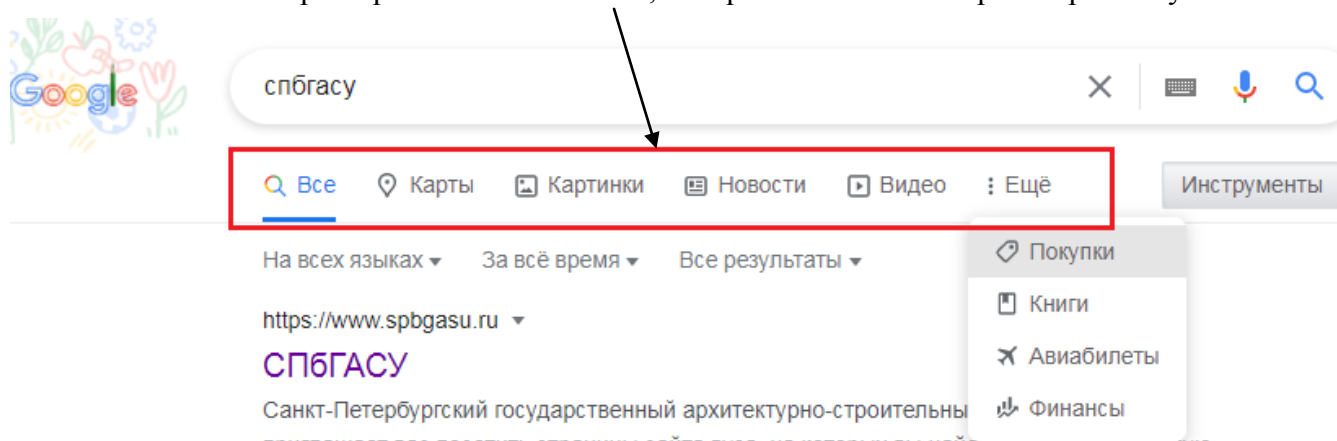
При работе с поисковой системой возможно *настроить параметры поиска*, кликнув на значок *Настройки* и выбрав пункт *Настройки поиска*. Помимо этого окно *Настройки* предоставляет доступ к использованию *Расширенного поиска* и др.



Другим вариантом использования настроек в сети Интернет, является щелчок на значке  в адресной строке. Появится окно, изображенное на рис. ниже, в котором можно выполнить ряд настроек, в частности, касающихся использования ваших геоданных при поиске.



Ниже поисковой строки располагается меню, которое помогает конкретизировать условия поиска.



Поиск картинок

Google поддерживает «прямой» (на основе текстового запроса) и «обратный» (на основе загруженного файла) поиск изображений. Например, при «обратном» поиске изображений можно использовать один из алгоритмов действий:

1. Открыть браузер.
2. Перейти на сайт, где размещено изображение, по которому вы хотите выполнить поиск.
3. Нажать на нужное изображение *правой кнопкой мыши*.
4. Выбрать *Найти картинку* (Google).
5. Проанализировать результаты поиска.

или

1. Активировать меню «Картинки» ниже поисковой строки.
2. Нажать на иконку фотоаппарата.
3. Выбрать один из вариантов загрузки изображения (ввод адреса или загрузку файла).
4. Проанализировать результаты поиска.

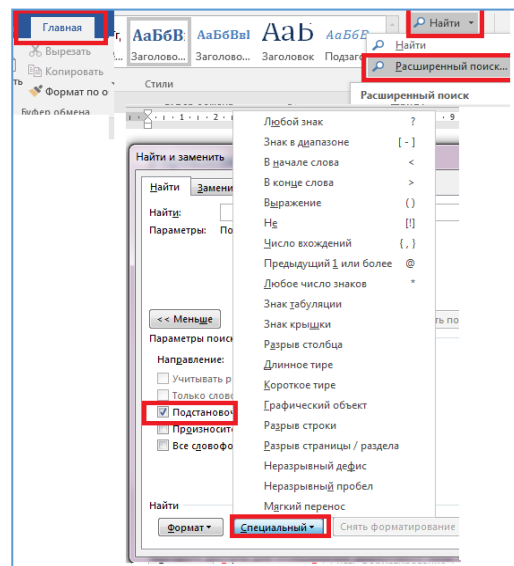
2.2. Поиск информации в документе Ms Word на основе регулярных выражений

Microsoft Word предоставляет необыкновенно мощный инструмент для поиска и замены. Умение грамотно им пользоваться существенно упрощает редактирование и правку документов.

Регулярные выражения – это формальный язык поиска и осуществления манипуляций с подстроками в тексте, основанный на использовании метасимволов. Регулярным выражением задается *шаблон*, по которому происходит поиск в документе. Регулярные выражения позволяют искать последовательности символов, которые имеют *определенный формат*, но не имеют заранее известного значения.

Для того чтобы *Ms Word* воспринимал регулярные выражения нужно отметить позицию «Подстановочные знаки» в окне «Найти и заменить». После чего в полях «Найти» и «Заменить на» служебные символы регулярных выражений будут восприниматься как подстановочные знаки, а не как символы, которые нужно найти (см. рис.).

Иногда приходится прибегать к *экранированию символов* – замену в тексте управляющих символов на соответствующие текстовые подстановки. Например, если требуется найти символ, который является служебным в регулярных выражениях, тогда необходимо перед этим символом поставить знак экранирования «\»



Например, $S[aeo\backslash?]x$ - найдет Sax , Sex , $S?x$, Sox . В таблице ниже приведены подстановочные знаки и примеры их использования.

Спец. знаки	Что обозначают	Пример строки поиска	Что будет находить
?	Один любой символ	б?к	бак, бок, бук и т. п.
*	Любое число любых символов	б*к	бык, бардак, балык 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 <i>строчная</i> русская буква
		[А-ЯЁ]	Любая <i>прописная</i> русская буква
		[0-9]	Любая <i>цифра</i>
[!]	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	[!Д-Р]ок	док , рок, но не Док, Рок
		[!0-9]	Любой символ кроме цифр
{n}	Строго n штук предыдущего символа или выражения. Выражением является все то, что заключено в <i>круглые скобки</i> . 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 символаили выражения	10@	10, 100, 1000, 10000 и т. д.
<	Начало слова	<бок	боксер, но не колобок
>	Конец слова	бок>	колобок, но не боксер

В *Microsoft Word* для поиска и замены применяют также *специальные символы*, используя кнопку *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.

Специальные символы, используемые, если выключена опция *Подстановочные знаки*:

<i>Специальные символы</i>	<i>Что означают</i>
^p	Символ конца абзаца (¶)
^?	Любой символ
^#	Любая цифра
^\$	Любая буква
^e	Концевая сноска
^d	Поле
^f	Сноска
^b	Обрыв секции (---End of section---)
^f	Любое число и комбинация пробелов, табуляций и концов абзацев

Специальные символы не зависящие от опции *Подстановочные знаки*:

<i>Специальные символы</i>	<i>Что означают</i>
^t	Символ табуляции (→)
^a	Комментарий
^0nnn	Любой ANSI или ASCII символ с кодом nnn
^^	Символ ^
^g	Графика
^n	Обрыв колонки
^l	Принудительный обрыв строки
^m	Принудительный обрыв страницы
^+	Длинное тире, код 0151
^=	Короткое тире, код 0150
^s	Неразрываемый пробел (⁰)
^~	Неразрываемый дефис
^-	Мягкий перенос (¬)

Удобно использовать специальные регулярные выражения при замене текста. Причем, не можно принимать во внимание как окончания слов, так и менять форматирование.

Например, вы оформляете текст английскими терминами. Необходимо оформить все латинские буквы – курсивом. Для этого достаточно задать выбор всех латинских букв конструкцией [A-Za-z] и находясь в поле «Заменить на», нажать на кн. *Формат* и установить курсивное начертание.

Ниже приведены два варианта замены десятичной точки на запятую и наоборот:

(<[0-9]@).([0-9]@>)	\1,2	Десятичная точка заменена на запятую
(<[0-9]@),([0-9]@>)	\1.2	Десятичная запятая заменена на точку

Возможны также варианты с перестановками и заменой выражений. Отредактируем такое предложение: *Сережа ест грушу*.

<i>Строка поиска</i>	<i>Строка замены</i>	<i>Результат поиска и замены</i>
(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.

3. Задания

Задание 1. Изучите настройки браузера на вашем компьютере, определив какой браузер и какая поисковая система установлены у вас по умолчанию. Установите геоданные «по запросу».

Задание 2. Изучите *примеры поисковых запросов в Google в таблице 1. Заполните таблицу 2, сконструировав запросы, удовлетворяющие указанным условиям поиска.*

Таблица 1

Поисковая фраза (Google)	Описание поиска
обучение	ищется не обязательно точное совпадение – в результатах поиска могут быть слова в <i>другом падеже, числе</i> и т.д., а также страницы, содержащие <i>синонимы</i> слова. Также результаты поиска могут быть <i>персонализированы</i> – зависеть от вашего места <i>нахождения</i> и <i>интересов</i>
обучение программированию	использование двух и более слов выведет результаты, в которых используются все слова: между словами подразумевается логическое И . При этом в результатах поиска могут быть слова в другом падеже, числе, а также синонимы.
"обучение программированию онлайн"	при использовании « » будет выполнен поиск по <i>точному совпадению фразы</i> . Также в кавычках следует указывать специальные символы, которые в противном случае будут трактоваться иначе (например, / (слэш) без кавычек будет трактоваться в поисковом запросе как пробел)
tesla OR edison	указывать логическое ИЛИ полезно, когда оба термина встречаются редко
онлайн обучение (java OR php)	обязательно должно быть слово « онлайн » и слово « обучение » и обязательно должно быть одно из двух слов (или оба) « java » или « php »
онлайн обучение java OR php	это выражение означало бы: искать фразу в которой есть три слова « онлайн обучение java » или фразу « php » (без других уточняющих слов)
онлайн обучение -java -php	знак <i>минуса</i> должен быть в паре с <i>каждым исключаемым из поиска словом</i>
tesla -motors -"rock n roll"	можно исключить фразу из нескольких слов, используя знак минус (-), за которым следует фраза в кавычках
тесла AROUND(3) эдисон	находит слова, которые не просто присутствуют в одном документе, а находятся в <i>пределах X слов друг от друга</i>
"nikola tesla" AROUND(2) "thomas alva edison"	поиск по точно совпадающим фразам, стоящим в пределах определённого диапазона
site:vasexperts.ru pdf	будут показаны все упоминания pdf на сайте vasexperts.ru
nikola tesla (site:gov OR site:edu)	поиск материала на нескольких доменах верхнего уровня
недорогой "перелёт"	точное совпадение слова, если нужно исключить его варианты, а слово «недорогой» при этом может меняться на «дешевый» и т.д.
intitle:"практикум по поиску информации в интернет"	оператор "intitle:" с точно совпадающим заголовком (в кавычках) чтобы найти конкретный материал

Таблица 2

Описание поиска	Запрос
Данные о журнале Вокруг света или о журнале Хакер за август 2021, исключив синонимы к названиям журналов на страницах результатов поиска	
Данные о стипендиях СПбГАСУ, обеспечить выдачу результатов только с сайта вуза	
Недорогие точки питания рядом с СПбГАСУ, исключая шаверма (шаурма)	
Курсы программирования на языке Golang (Go) или Python	
Программы обзорных экскурсий по Петербургу на текущий месяц, вывод в формате pdf	

Задание 3. Найти информацию в сети Интернет согласно своего варианта, грамотно отформатировать текст, указать источники информации. Текст ответа должен кратко отображать данную тему, размер 1-2 страницы.

1	Статистика запросов Google, как вывести, привести примеры использования
2	Статистика запросов Yandex, как вывести, привести примеры использования
3	Сайты библиотечных ресурсов, полезные для студента. Какие ресурсы предлагает библиотека вуза
4	Библиотеки готовых студенческих работ и рефератов
5	Правила оформления списка литературы (актуальные)
6	Конструкторы сайтов, сравнить их характеристики
7	Что такое доменное имя, как и где получить доменное имя
8	Чем занимается интернет-маркетинг. Способы продвижения товаров в интернете
9	Понятие информации, ее отличие от данных и знаний. Привести примеры
10	Свойства информации, привести примеры на разные свойства (достоверность,...)
11	Понятие информационных технологий и современные тенденции их развития
12	Понятие компьютерной сети, виды сетей
13	Понятие информационных систем, классификация ИС
14	Сеть Интернет. История развития сети Интернет
15	Понятие электронного офиса, аппаратные средства электронного офиса
16	Защита информации, основные виды защиты информации
17	Компьютерные вирусы, их классификация, краткое описание
18	Этапы развития информационных технологий
19	Классификация компьютеров, суперкомпьютеры, виды и характеристики
20	Понятие электронного документооборота, системы управления электронным документооборотом
21	Понятие BIM технологий, основные пакеты программ, используемые в BIM моделировании
22	Найти данные о средних баллах ЕГЭ вашего региона за прошлый год
23	Найти законы, регулирующие защиту информационных ресурсов
24	Найти законы, регулирующие авторские права
25	Пакеты прикладных программ, применяемые в строительстве и архитектуре
26	Найти министерства, связанные с образовательной деятельностью в России, их руководители, подведомственные им организации
27	Понятие больших данных. Языки программирования, используемые при работе в Big Data
28	Пакеты прикладных программ, используемые в математике, их характеристика
29	Современные антивирусные программы, их характеристика
30	Справочно-правовые информационные системы

Задание 4. Заполните в таблице ниже столбец «Выражения», обеспечив нужный результат поиска.

№	Результат поиска	Поисковое выражение
1	"род"и "рад", но не “рада"	
2	"вход" и "выход"	
3	"солнце" и "солонка", но не "фасоль"	
4	"ель" и "панель", но не "ельник"	
5	"крен" и "кран"	

6	"выбор" и "табор"	
7	"кожа" и "кора", но не "коза" и "коса"	
8	"жаренный", но не "жареный"	
9	99, 999 и 999	
10	"меньше" и "меньшее"	

Задание 5. В приведенном тексте используя замену, оформите *английские* слова *красным курсивом*, *русские* – *синим* шрифтом. Укажите фразы «Найти» и «Заменить на» в ответе на задание (можно использовать скрин экрана).

Alpha(nu)meric *display* - буквенно-цифровой (символьный) дисплей.

Alpha(nu)meric *keyboard* - буквенно(алфавитно)-цифровая клавиатура

Alternate *key* - альтернативный ключ

Alternate *Mark Inversion* (AMI) - чередующаяся инверсия единиц

Задание 6. Выполните замены.

Результат	Выражение «Найти»	Выражение «Заменить на»
Замена пробела между цифрами на неразрывный пробел		
Замена дефиса перед цифрой на минус		
Замена круглых скобок вокруг цифры на квадратные		
Замена нескольких пробелов подряд на один пробел		
Заменить Иванов на Петров, Иванову на Петрову, Ивановой на Петровой (одной заменой)		