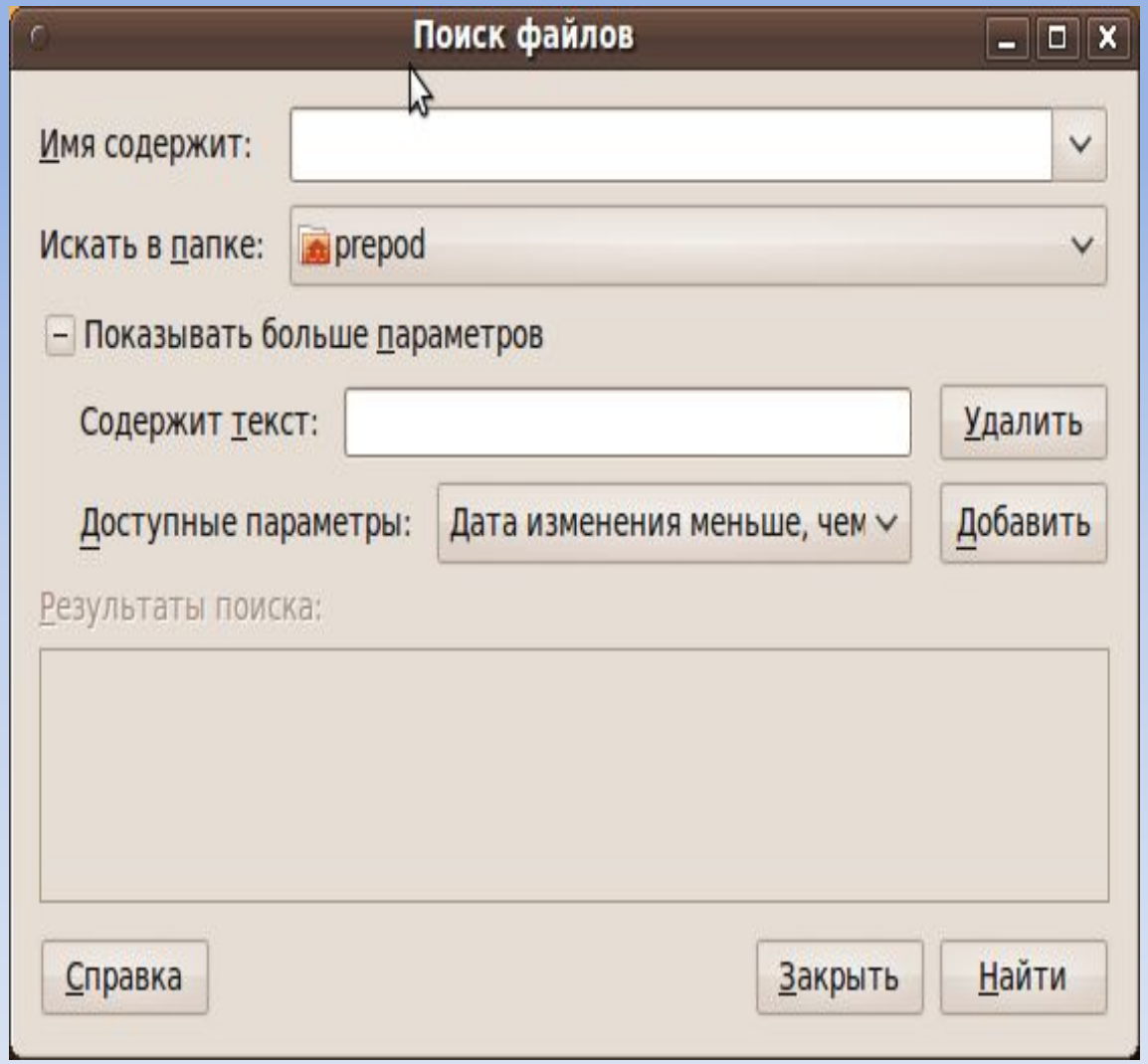
Поиск информации с использованием компьютера.
Программные поисковые серверы.
Использование ключевых слов, фраз для поиска информации.
Комбинации условия поиска.

Поиск информации с использованием компьютера

После некоторого времени работы за компьютером, вы обнаружите, что не всегда можете вспомнить место, где расположен тот или иной нужный вам документ. В таком случае вам следует воспользоваться средствами поиска информации, включенными в операционную систему. При этом следует помнить, что для успешного поиска файла необходимо правильно указать критерии поиска.

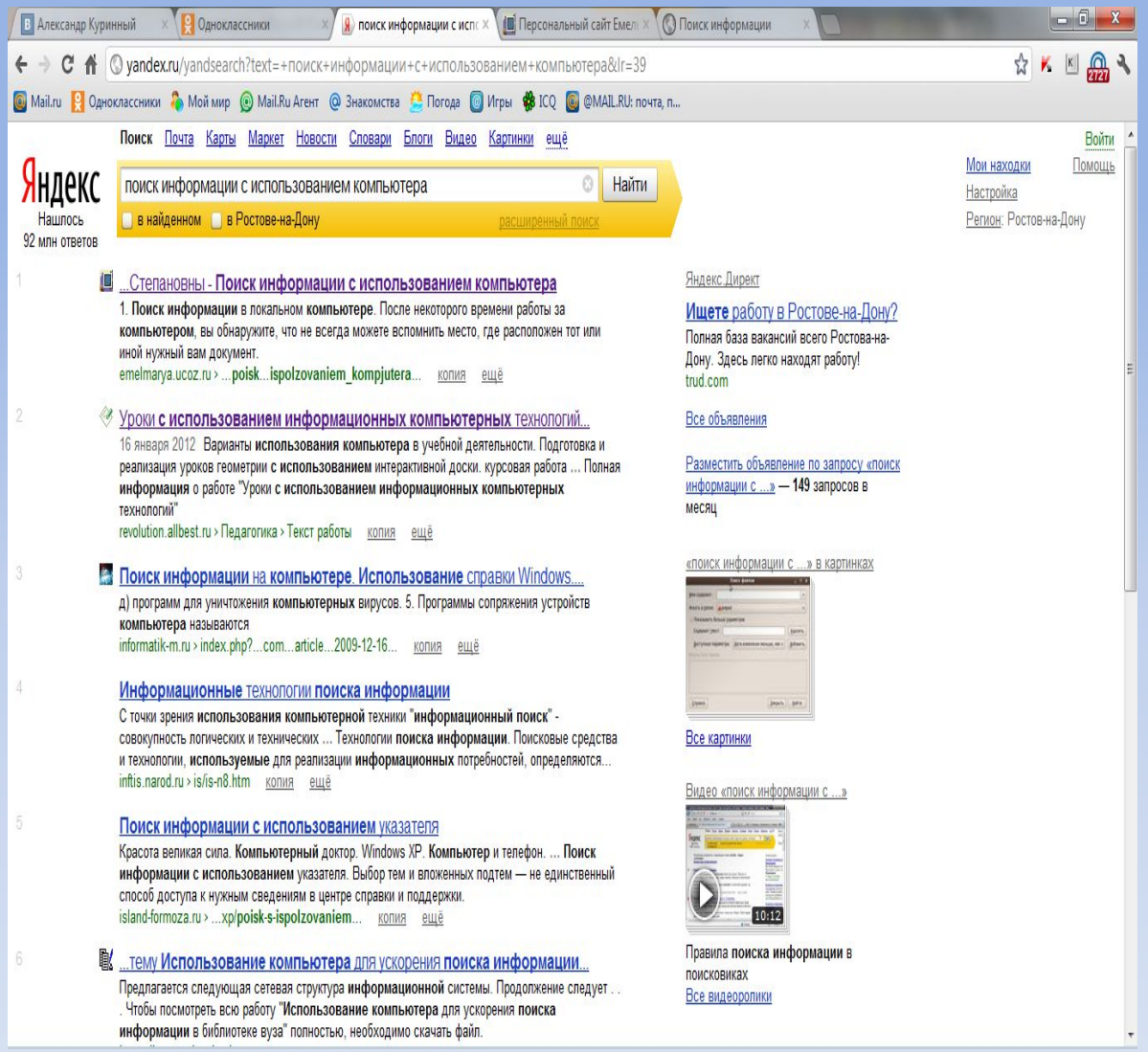
В общем случае поиск информации в Linux Ubuntu удобно проводить, выбрав команду главного меню **Искать файлы** из меню **Стандартные**. С их помощью вы можете задать критерии поиска и дать команду начала процедуры поиска нужного файла. Если пользователь не помнит названия файла, но знает содержимое, используется дополнительные параметры поиска.



Поиск информации в сети Интернет

С каждым годом объемы Интернета увеличиваются в разы, поэтому вероятность найти необходимую информацию резко возрастает. Интернет объединяет миллионы компьютеров, множество разных сетей, число пользователей увеличивается на 15-80% ежегодно. И, тем не менее, все чаще при обращении к Интернет основной проблемой оказывается не отсутствие искомой информации, а возможность ее найти. Как правило, обычный человек в силу разных обстоятельств не может или не хочет тратить на поиск нужного ему ответа больше 15-20 минут. Поэтому особенно актуально правильно и грамотно научиться, казалось бы, простой вещи – где и как искать, чтобы получать ЖЕЛАЕМЫЕ ответы.

Чтобы найти нужную информацию, необходимо найти её адрес. Для этого существуют специализированные поисковые сервера (роботы индексов (поисковые системы), тематические Интернет-каталоги, системы мета-поиска, службы поиска людей и т.д.). В данном мастер-классе раскрываются основные технологии поиска информации в Интернет, предоставляются общие черты поисковых инструментов, рассматриваются структуры поисковых запросов для наиболее популярных русскоязычных и англоязычных поисковых систем.



Пограммные поисковые серверы

Основу, «каркас» Internet составляют около 300 миллионов серверов (Netcraft, март 2011), постоянно подключенных к сети.
К серверам Internet
могут подключаться сотни миллионов пользоват елей Internet (2 млрд. -
генсек Международного союза электросвязи ОО Н 27 января 2011).

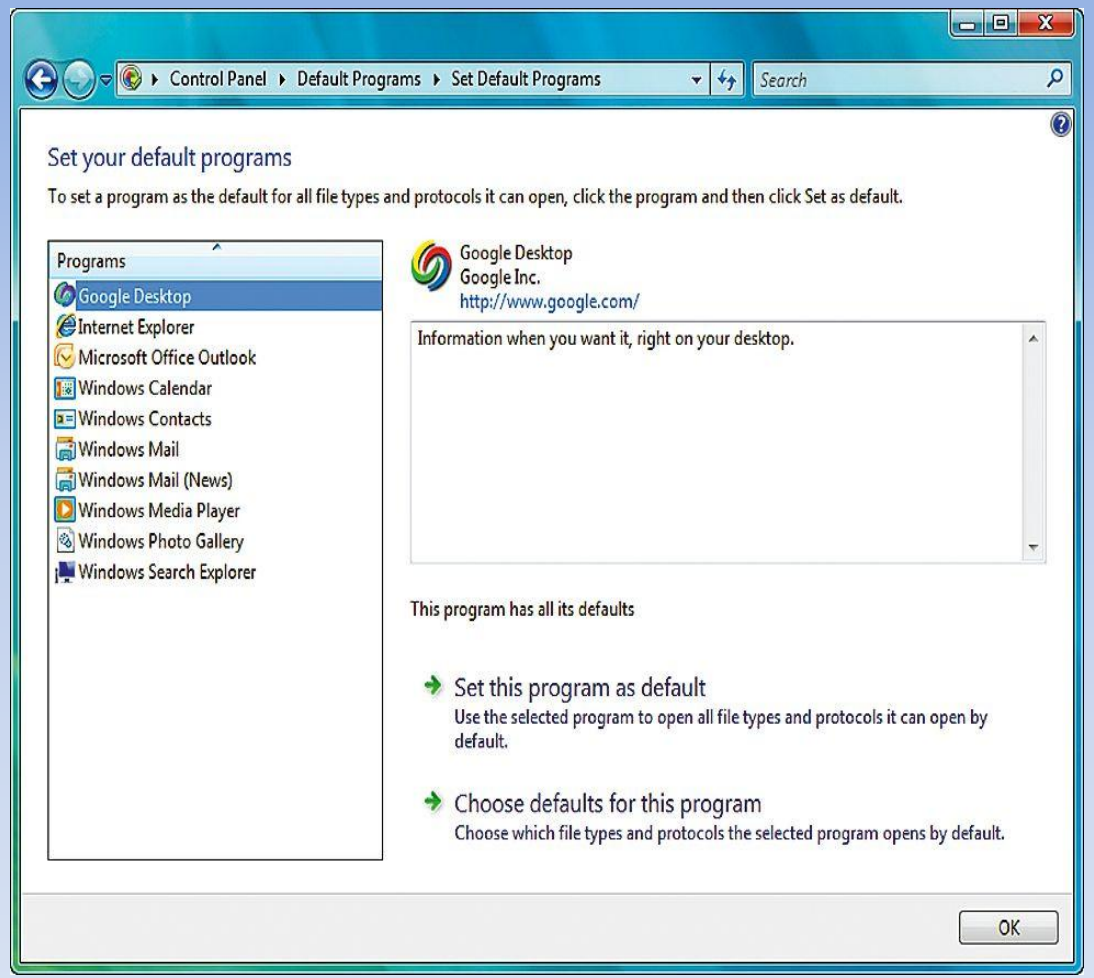
Сѐрвер (англ. server от англ. to serve — служить) (множественное число сѐрверы)
— в информационных технологиях — программн ый компонент вычислительной системы, выполняющий сервисные (обслуживающие) функции по запросу клиента,
предоставляя ему доступ к определённым ресур сам или услугам. Как правило,
каждый сервер обслуживает один (или несколько схожих)
протоколов и серверы можно классифицировать по типу услуг которые они предоставляют.

Файл-серверы
представляют собой серверы для обеспечения д оступа к файлам на диске сервера.
Прежде всего это серверы передачи файлов по заказу, по протоколам FTP, TFTP, SFTP и HTTP.
Протокол HTTP
ориентирован на передачу текстовых файлов, но серверы могут отдавать в качестве запрошен ных файлов и произвольные данные,
например динамически созданные веб-страницы, картинки, музыку и т. п.

На ноябрь 2009
года двумя наиболее распространёнными веб-серверами, вместе занимающими около 68 % рынка[1], являются:

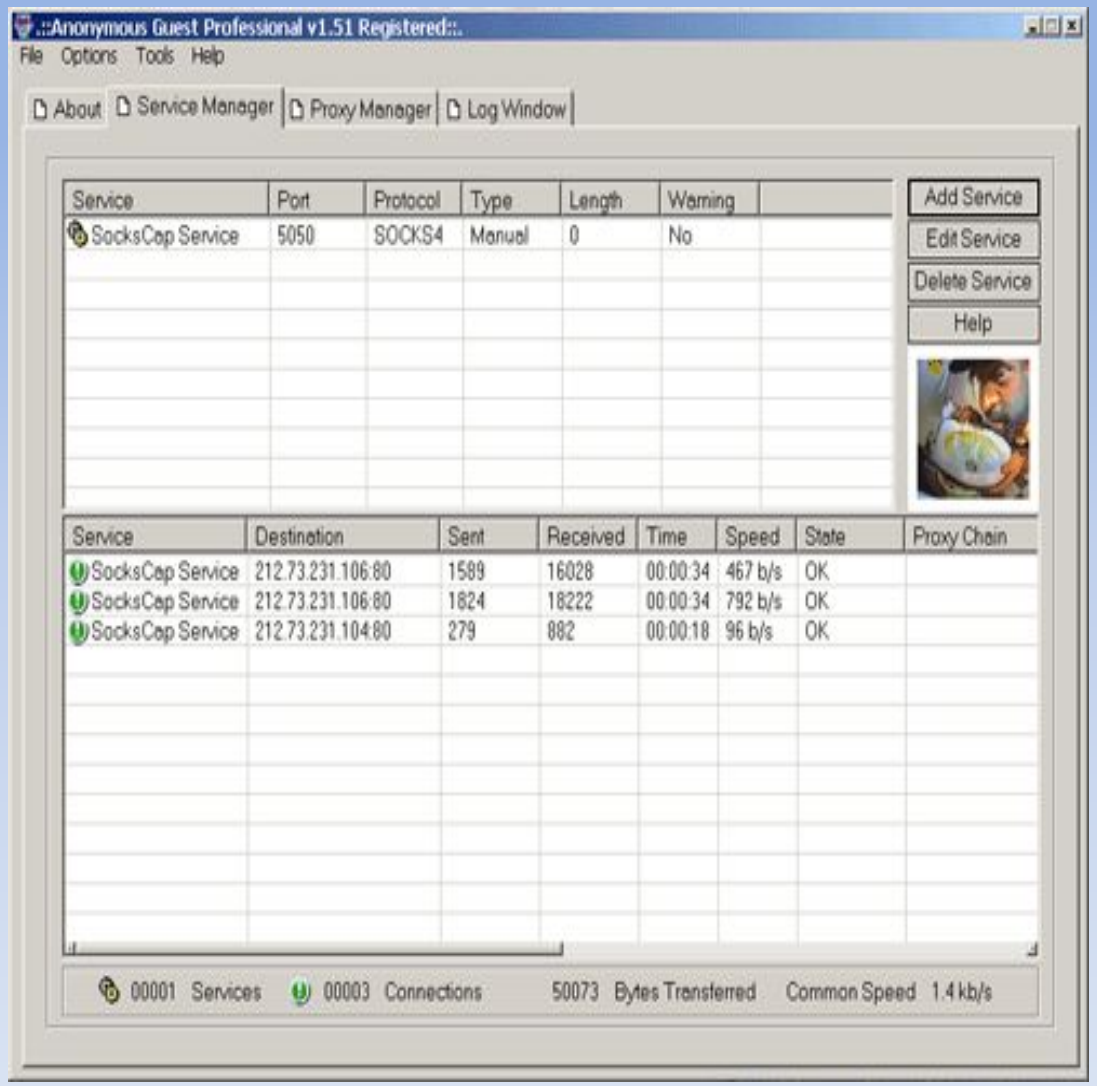
Apache — свободный веб-сервер,
наиболее часто используемый в Unix-подобных операционных системах;

IIS от компании Microsoft,
распространяемый с ОС семейства Windows NT.



Назначение WEB

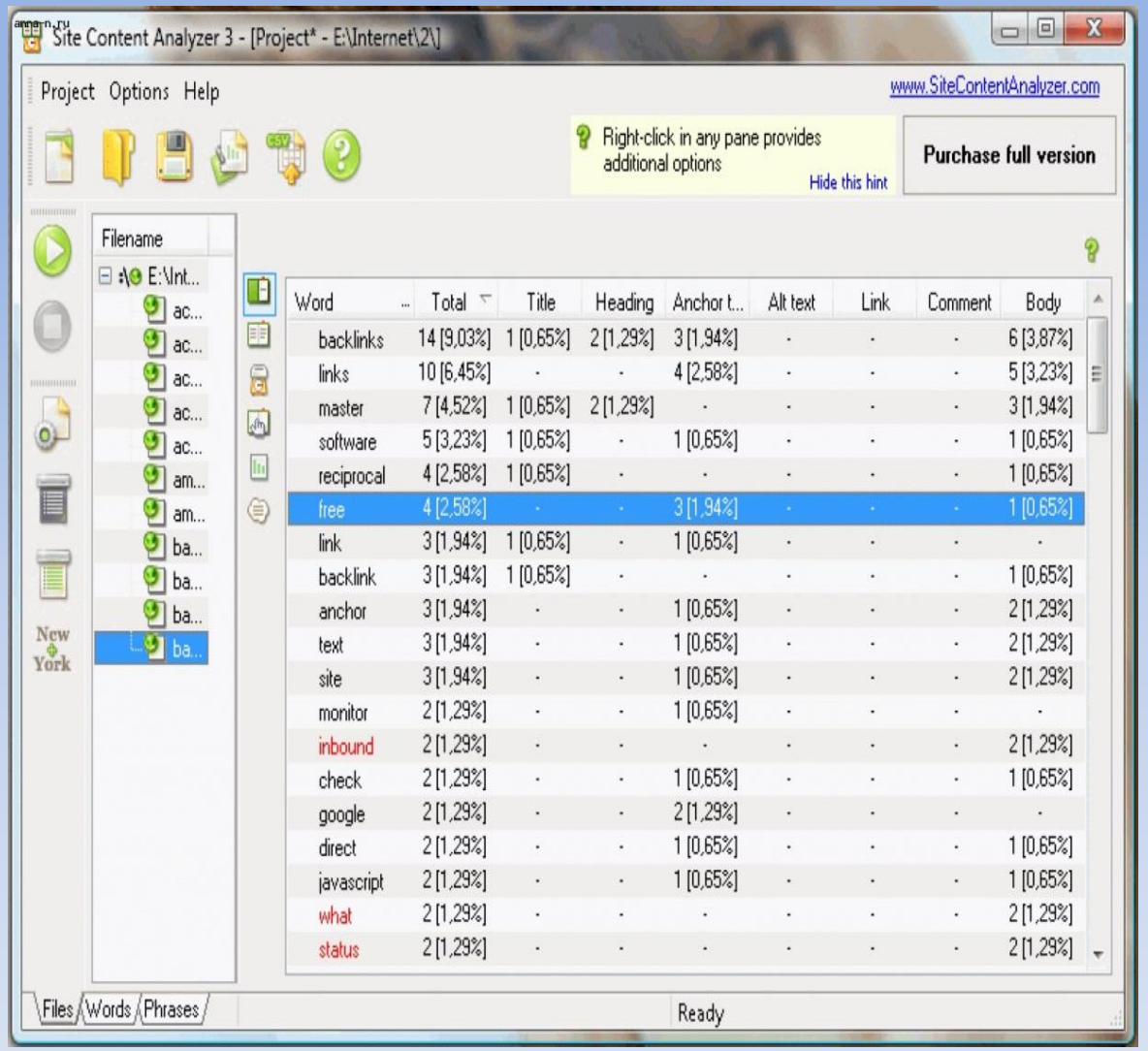
Веб-сервер является программой, запускаемой на подключённом к сети компьютере и использующей протокол HTTP для передачи данных.
В простейшем виде такая программа получает по сети HTTP-запрос на определённый ресурс, находит соответствующий файл на локальном жёстком диске и отправляет его по сети запросившему компьютеру.
Более сложные веб-серверы способны динамически распределять ресурсы в ответ на HTTP-запрос.
Для идентификации ресурсов (зачастую файлов или их частей) во Всемирной паутине используются единообразные идентификаторы ресурсов URI (англ. Uniform Resource Identifier).
Для определения местонахождения ресурсов в сети используются единообразные локаторы ресурсов URL (англ. Uniform Resource Locator). Такие URL-локаторы сочетают в себе технологию идентификации URI и систему доменных имён DNS (англ. Domain Name System) — доменное имя (или непосредственно IP-адрес в числовой записи) входит в состав URL для обозначения компьютера (точнее — одного из его сетевых интерфейсов), который исполняет код нужного веб-сервера.



Поиск по ключевым словам

Большинство поисковых машин имеют возможность поиска по ключевым словам. Это один из самых распространенных видов поиска. Для поиска по ключевым словам необходимо ввести в специальном окне слово или несколько слов, которые следует искать, и щелкнуть на кнопке **Поиск**. Поисковая система найдет в своей базе и покажет документы, содержащие эти слова.

Для того чтобы сделать поиск более продуктивным, во всех поисковых системах существует специальный язык формирования запросов со своим синтаксисом.



Использование ключевых фраз

Правила и советы при подборе ключевых фраз:

каждая ключевая фраза не более 5-ти слов, включая союзы и предлоги;

в словах допустимы только буквы, цифры и знак «-» ;

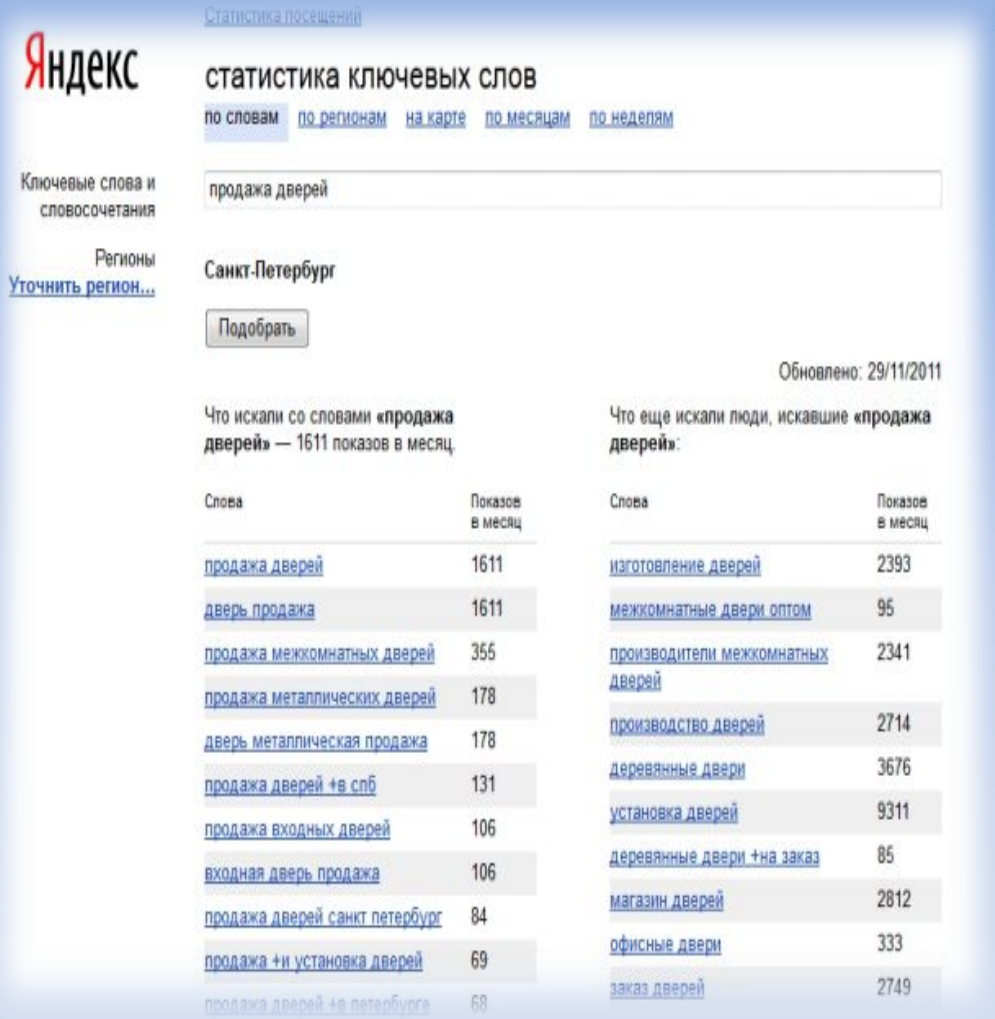
общая длина поля ключевых слов ограничена 4096 знаками;

подбираем максимально «широкие» в использовании фразы – лучше короткие, что бы было больше вариантов для использования с другими словами;

единственное или множественное число не имеет значения – выбираем что-то одно;

падеж не имеет значения – выбираем один из вариантов;

однокоренные глагол, существительное или прилагательное – это разные ключевые слова.

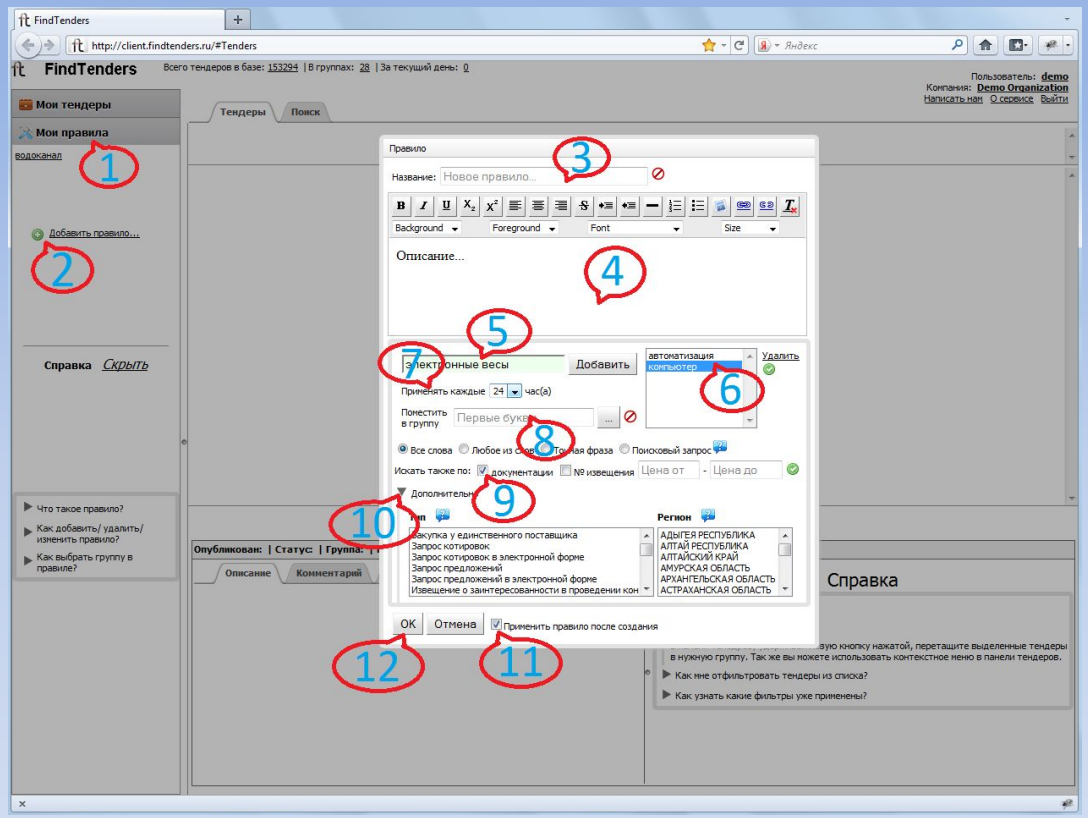


Поиск по ключевым фразам

Поиск ключевой фразы для микро-сайта я начинаю с использования программы **Market Samurai**. На мой взгляд, такой подход к поиску и анализу ключевой фразы является наиболее продуктивным для микро-сайта. Не следует забывать, что микро-сайт оптимизируется под одну единственную ключевую фразу. В дальнейшем, с развитием микро-сайта, возможно использование и других ключевых слов.

Market Samurai дает не совсем точную картину по выбранному ключу, однако, для поиска и анализа необходимой информации эта программа более, чем подходит.

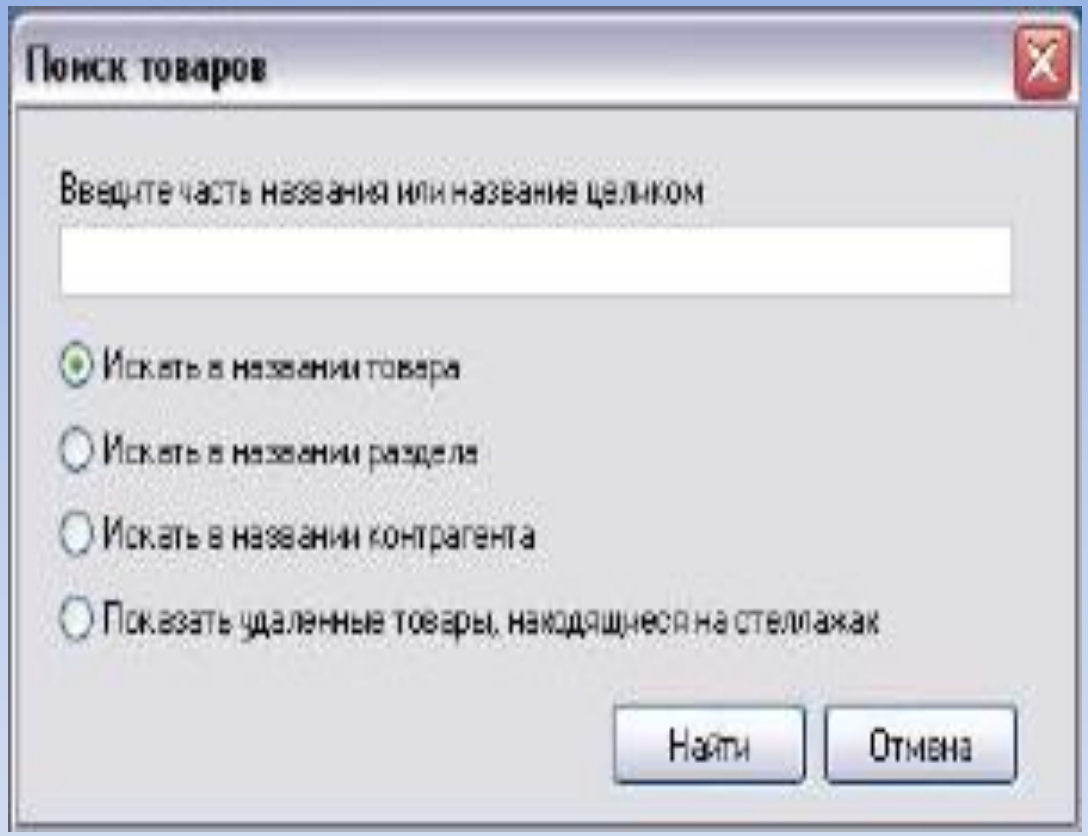
В подкасте я рассказываю о конкретном применении Market Samurai для определения ключевой фразы. Анализируются основные характеристики ключевых слов, производится оценка конкурентов и возможность выхода в топ-выдачу Google.



Комбинации условия поиска

Допускается создание запросов, включающих произвольное количество условий поиска, которые связываются любым числом операторов AND и OR. Запрос, содержащий комбинацию предложений AND и OR, может стать достаточно сложным. Поэтому полезно представить себе, как такие запросы интерпретируются при их выполнении и как они выглядят в режиме конструктора и в окне режима SQL.

Для наглядности в дальнейшем обсуждении предположим, что идет работа с таблицей employee, содержащей столбцы hire_date, job_lvl и status. В примерах предполагается, что требуется получить информацию, как долго сотрудник работал в компании (т.е. узнать его дату найма), какую работу он выполнял (уровень его ответственности), а также его статус (например, выход на пенсию).

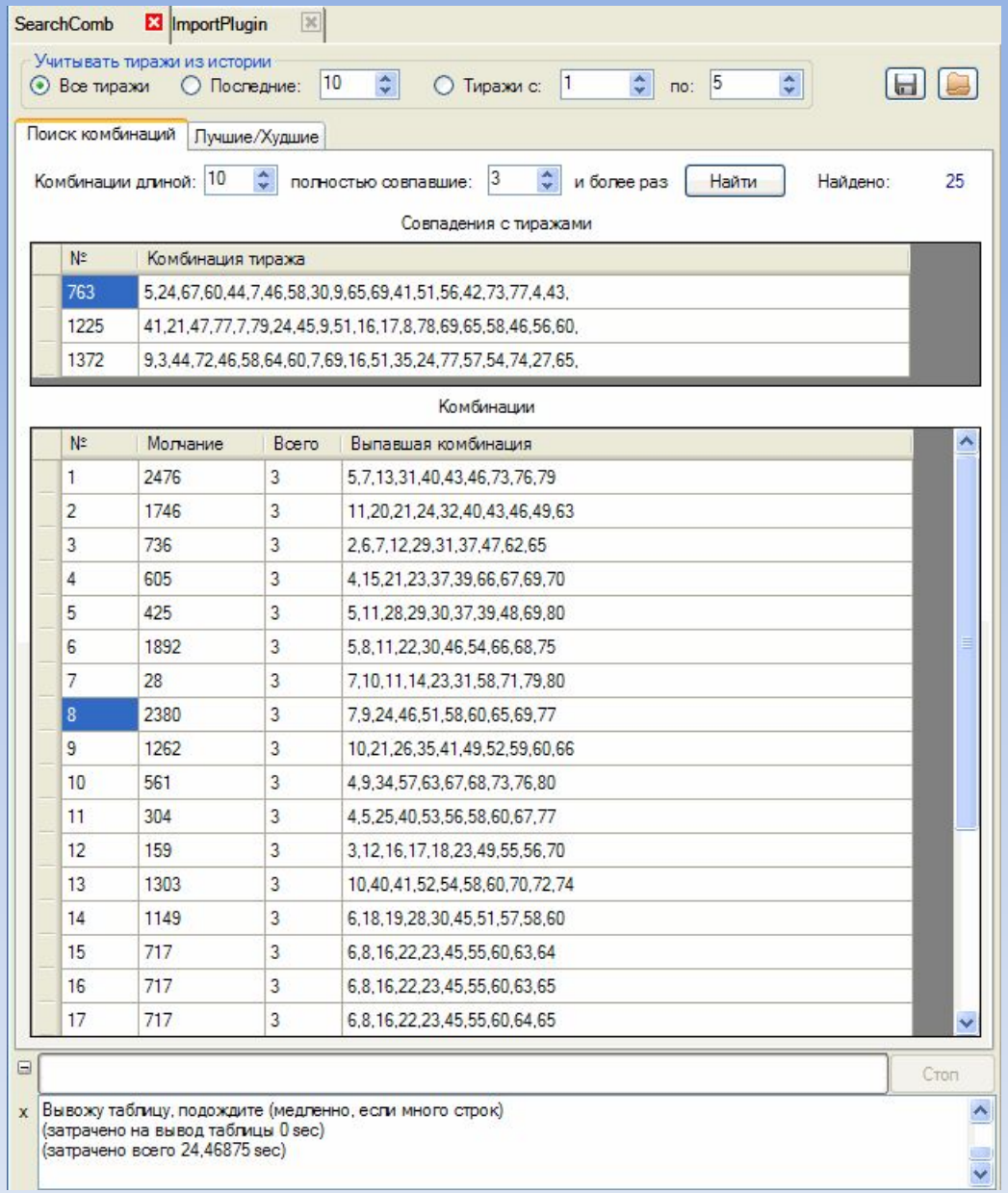


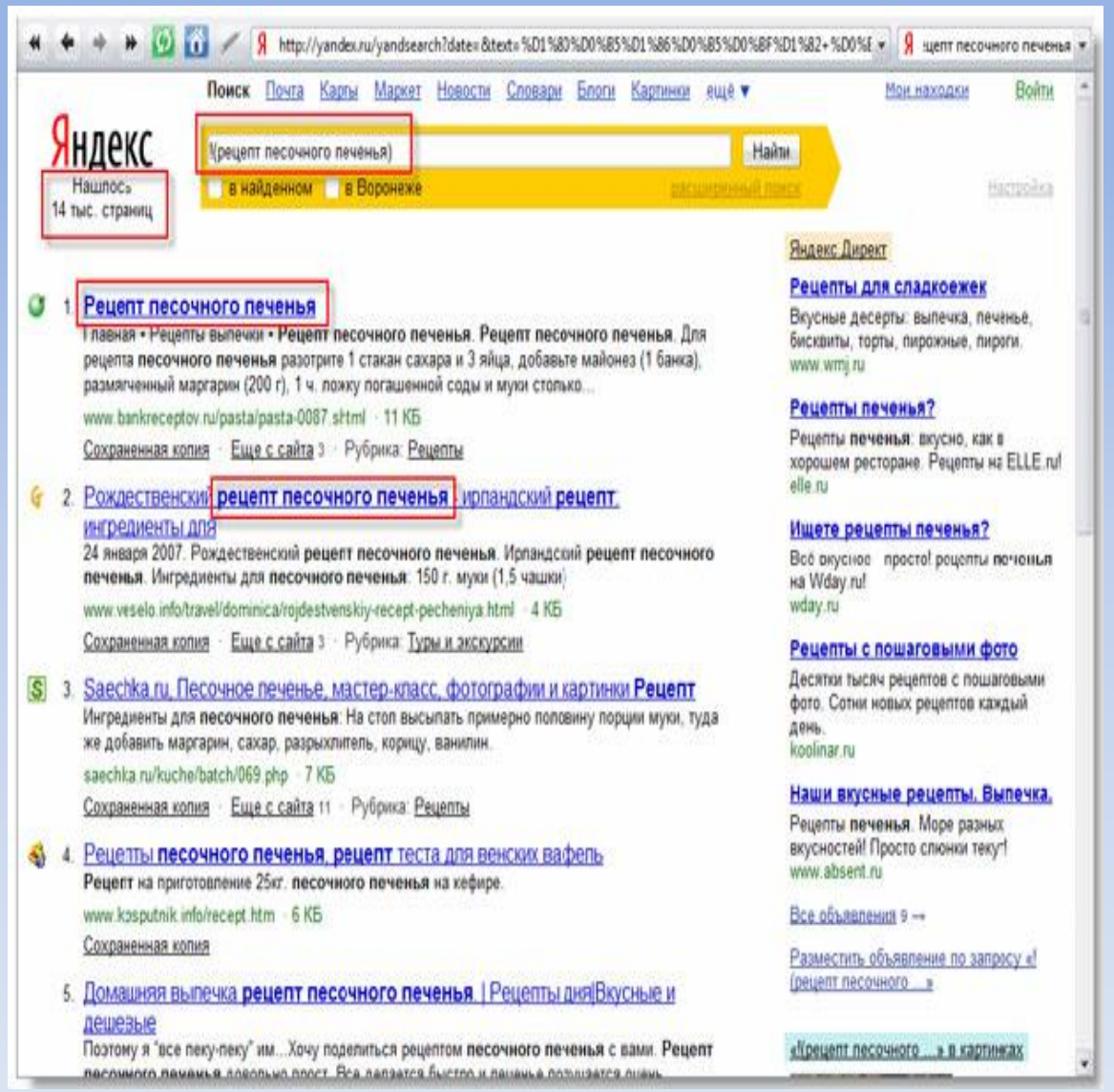
Комбинации условия поиска

Плагин "Поиск Комбинаций" (SearchComb) предназначен для поиска комбинаций в истории тиражей, удовлетворяющих тем или иным условиям.

Вверху задается диапазон учитываемых в поиске тиражей (все тиражи, последние или выборочно с .. по ...). Кнопки сохранения настроек и загрузки их.

Первая вкладка "Поиск комбинаций" предназначена для поиска в истории тиражей таких комбинаций, которые полностью (т.е. совпали все числа комбинации с тем или иным тиражем) совпадали на протяжении анализируемой истории лотереи заданное число раз (или более). Например, на скриншоте выведены в таблицу комбинации длиной в 10 чисел, которые совпали со всеми тиражами лотереи Кено 20 из 80 (**Украинская УНЛ**) более трех раз. В нижней таблице выводятся найденные комбинации, в верхней, при позиционировании на какой-либо строке нижней, тиражи, в которых было найдено полное совпадение данной комбинации.



Комбинации условия поиска

Следует иметь в виду, что при некоторых условиях расчет, а особенно вывод в таблицу, могут занимать длительное время. Так например, миллион шестерок, выпадавших более 2-х раз за 3500 тиражей лотереи "Кено 20 из 80" (**Украинская Национальная Лотерея**), выводится на двухядерном Пентиуме 2.53 МГц примерно за три минуты (при том что расчет занимает полминуты). Но при этом программа начинает потреблять много памяти, и при нехватке оной возможно аварийное прекращение работы функции (не программы, а только данного расчета) по исключению OutOfMemoryException. В целях исключения такой ситуации в программе имеется параметр, ограничивающий число расчитываемых комбинаций. По-умолчанию он установлен в 500 000 записей (в конечном счете их может оказаться и меньше после дополнительной обработки). Если же у вас оперативной памяти достаточно и так необходимо получить несколько миллионов комбинаций (хотя и непонятна цель этого), в конфигурационном файле плагина (например, SearchComb_20x80.pcfg) можно задать требуемое значение параметра ResultLengthLimit.