

ПОИСК ИНФОРМАЦИИ В СЕТИ ИНТЕРНЕТ



2015 г

Краткое содержание:

- [Проблема поиска](#) • [Поисковые системы](#) • [Поиск по рубриктору поисковой системы](#) • [Поиск по ключевым словам](#) • [Правила формирования запросов в поисковой системе Яндекс](#)

Поиск информации в сети Интернет

КРАТКОЕ СОДЕРЖАНИЕ:

ПРОБЛЕМА ПОИСКА

Говорят, что в Интернете есть все. На самом деле – это не так. Материалы для размещения в сети готовят живые люди, и потому там можно найти лишь то, что они сочли нужным опубликовать. Впрочем, река питается ручьями, и, благодаря их творчеству сегодня в Интернете уже образовалось около двух миллиардов Web-страниц. В результате каталогизация имеющихся в сети ресурсов стала серьезной проблемой. Несмотря на то, что проблемой занимаются тысячи организаций, она не только не приближается к разрешению, но и становится острее. Процент каталогизированных ресурсов неуклонно падает. В последние годы это падение стало катастрофическим. Так, если в 1999 году процент каталогизированных ресурсов приближался к 40%, то всего лишь за один следующий год он опустился до 25 %.

Вывод простой: пространство Web быстрее наполняется, чем систематизируется.

К сожалению, у нас нет оснований предполагать, что в ближайшее время что-нибудь может измениться к лучшему. В итоге поиск информации в Интернет можно считать одной из самых трудных задач.

ПОИСКОВЫЕ СИСТЕМЫ

Для поиска интересующей вас информации необходимо указать браузеру адрес Web-страницы, на которой она находится. Это самый быстрый и надежный вид поиска. Для быстрого доступа к ресурсу достаточно запустить браузер и набрать адрес страницы в строке адреса.



Адреса Web-страниц приводятся в специальных справочниках, печатных изданиях, звучат в эфире популярных радиостанций и с экранов телевизора.

Если вы не знаете адреса, то для поиска информации в сети Интернет существуют **поисковые системы**, которые содержат информацию о ресурсах Интернета.

Каждая поисковая система – это большая база ключевых слов, связанных с Web-страницами, на которых они встретились. Для поиска адреса сервера с интересующей вас информацией надо ввести в поле поисковой системы ключевое слово, несколько слов или фразу. Тем самым вы посылаете поисковой системе запрос. Результаты поиска выдаются в виде списка адресов Web-страниц, на которых встретились эти слова.

Как правило, поисковые системы состоят из трех частей: **робота, индекса и программы обработки запроса.**

Робот (Spider, Robot или Bot) - это программа, которая посещает веб-страницы и считывает (полностью или частично) их содержимое.

Роботы поисковых систем различаются индивидуальной схемой анализа содержимого веб-страницы.

Индекс - это хранилище данных, в котором сосредоточены копии всех посещенных роботами страниц.

Индексы в каждой поисковой системе различаются по объему и способу организации хранимой информации. Базы данных ведущих поисковых машин хранят сведения о десятках миллионов документов, а объемы их индекса составляют сотни гигабайт. Индексы периодически обновляются и дополняются, поэтому результаты работы одной поисковой машины с одним и тем же запросом могут различаться, если поиск производился в разное время.

Программа обработки запроса - это программа, которая в соответствии с запросом пользователя «просматривает» индекс на предмет наличия нужной информации и возвращает ссылки на найденные документы.

Множество ссылок на выходе системы распределяется программой в порядке убывания от наибольшей степени соответствия ссылки запросу к наименьшей.

В России наиболее распространенными поисковыми системами являются:

- [Рамблер \(www.rambler.ru\)](http://www.rambler.ru);
- [Яндекс \(www.yandex.ru\)](http://www.yandex.ru);
- [Мэйл \(www.mail.ru\)](http://www.mail.ru).

За рубежом поисковых систем гораздо больше. Самыми популярными являются:

- [Alta Vista \(www.altavista.com\)](http://www.altavista.com);
- [Fast Search \(www.alltheweb.com\)](http://www.alltheweb.com);
- [Northern Light \(www.northernlight.com\)](http://www.northernlight.com).

Яндекс является, пожалуй, наилучшей поисковой системой в российском Интернете. Эта база данных содержит около 200 000 серверов и до 30 миллионов документов, которые система просматривает в течение нескольких секунд. На примере этой системы покажем как осуществляется поиск информации.

Поиск информации задается введением ключевого слова в специальную рамку и нажатием кнопки «Найти», справа от рамки.



Результаты поиска появляются в течение нескольких секунд, причем ранжированные по значимости – наиболее важные документы размещаются в начале списка. При этом ранг найденного документа определяется тем, в каком месте документа находится ключевое слово (в

заглавии документа важнее, чем в любом другом месте) и числом упоминаний ключевого слова (чем больше упоминаний, тем ранг выше).

The screenshot shows the Yandex search interface. The search bar contains the word "аквариум". Below the search bar, there are checkboxes for "в найденном" and "в регионе: Петербург". A yellow bar highlights the search results, showing the number of pages (3 740 620), sites (3 932), and catalogs (92). Below this, there is a section titled "Каталог: всё на тему «Аквариум»" with two search results listed. The first result is "Aquarium's Very Own Home Page" with a description, URL, and file size. The second result is "AQUARIUM-PRO. Аквариумы: продажа аквариумов, оформление дизайн аквариумов ..." with a description, URL, and file size.

Результат поиска: страниц — 3 740 620, сайтов — не менее 3 932, в каталоге — 92
Запросов за месяц аквариум — 52 586. [Купить это слово.](#)

Каталог: всё на тему «Аквариум»

1. [Aquarium's Very Own Home Page](#)
текст ссылок: Группа **Аквариум...** борис гребенщиков и **аквариум...**
[www.aquarium.ru:8083](#) · 2 КБ — найден по ссылке
[Еще с сайта 367](#) · Рубрика: [Рок-музыка](#)
2. [AQUARIUM-PRO. Аквариумы: продажа аквариумов, оформление дизайн аквариумов ...](#)
профессиональное оформление пресноводных и морских **аквариумов**
Пресноводные **аквариумы** - это очень элегантные **аквариумы**, которые подойдут для дизайна интерьера любого офиса.
[www.aquarium-pro.ru](#) · 10 КБ · 25.01.2007
[Сохраненная копия](#) · [Еще с сайта 50](#)

Таким образом, сайты, расположенные на первых местах в списке, являются ведущими не с содержательной точки зрения, а практически, по отношению к частоте упоминания ключевого слова. В связи с этим, не следует ограничиваться просмотром первого десятка предложенных поисковой системой сайтов.

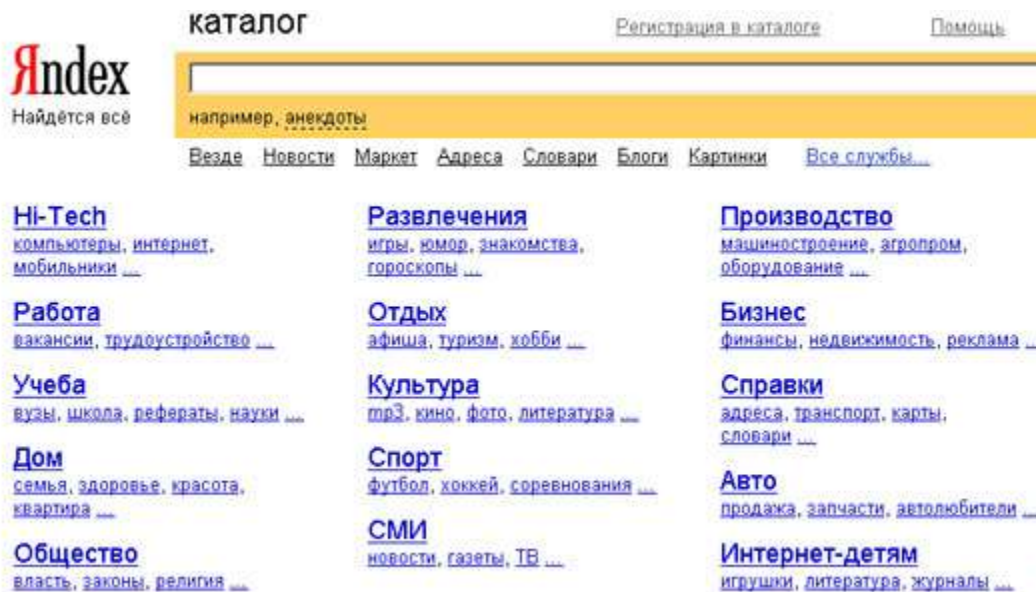
Содержательную часть сайта можно косвенно оценить по краткому его описанию, размещаемому Поисковой системой под адресом Сайта. Некоторые недобросовестные авторы сайтов, для того, чтобы повысить вероятность появления своей Web-страницы на первых местах Поисковой системы, умышленно включают в документ бессмысленные повторы ключевого слова. Но как только поисковая система обнаруживает такой «замусоренный» документ, она автоматически исключает его из своей базы данных.

Даже ранжированный список документов, предлагаемый поисковой системой в ответ на ключевую фразу или слово, может оказаться практически необозримым. В связи с этим в Яндекс (как и других мощных Поисковых Машинах) предоставлена возможность в рамках первого списка, выбрать документы, которые точнее отражают цель поиска, то есть уточнить или улучшить результаты поиска. Например, на ключевое слово список из 34 899 Веб-страниц. После ввода в команду «Искать в найденном» уточняющего ключевого слова список сокращается до 750 страниц, а после ввода в эту команду еще одного уточняющего слова этот список сокращается до 130 Веб-страниц.

Поиск по рубрикатору поисковой системы

Поисковые каталоги представляют собой систематизированную коллекцию (подборку) ссылок на другие ресурсы Интернета. Ссылки организованы в виде **тематического рубрикатора**, представляющего собой иерархическую структуру, перемещаясь по которой, можно найти нужную информацию.

Приведем в качестве примера структуру поискового интернет-каталога Яндекс.



Это каталог общего назначения, так как в нем представлены ссылки на ресурсы Интернета практически по всем возможным направлениям. В этом каталоге выделены следующие **темы**:

Бизнес и экономика;
Общество и политика;
Наука и образование;
Компьютеры и связь;
Справочники и ссылки;
Дом и семья;
Развлечения и отдых;
Культура и искусство.

Каждая **тема** включает множество **подразделов**, а они, в свою очередь, содержат **рубрики** и т. д.

Предположим, вы готовите мероприятие ко Дню победы и хотите найти в Интернете слова известной военной песни Булата Окуджавы «Вы слышите, грохочут сапоги». Поиск можно организовать следующим образом:

Яндекс > Каталог > Культура и искусство > Музыка > Авторская песня

Такой способ поиска является достаточно быстрым и эффективным. В конце вам предлагается всего несколько ссылок, среди которых есть ссылки на сайты с песнями известных бардов. Остается только найти на сайте архив с текстами песен Б. Окуджавы и выбрать в нем нужный текст.

Другой пример. Предположим, вы собираетесь приобрести мобильный телефон и хотите сравнить характеристики аппаратов разных фирм. Поиск мог бы вестись по следующим рубрикам каталога:

Яндекс > Каталог > Компьютеры и связь > Мобильная связь > Мобильные телефоны

Получив ограниченное количество ссылок, можно достаточно оперативно их просмотреть и выбрать телефон, исследовав характеристики по фирмам и модификациям аппаратов.

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 Найти.



Поисковая система найдет в своей базе и покажет документы, содержащие эти слова. Таких документов может оказаться множество, но много в данном случае не обязательно означает хорошо.

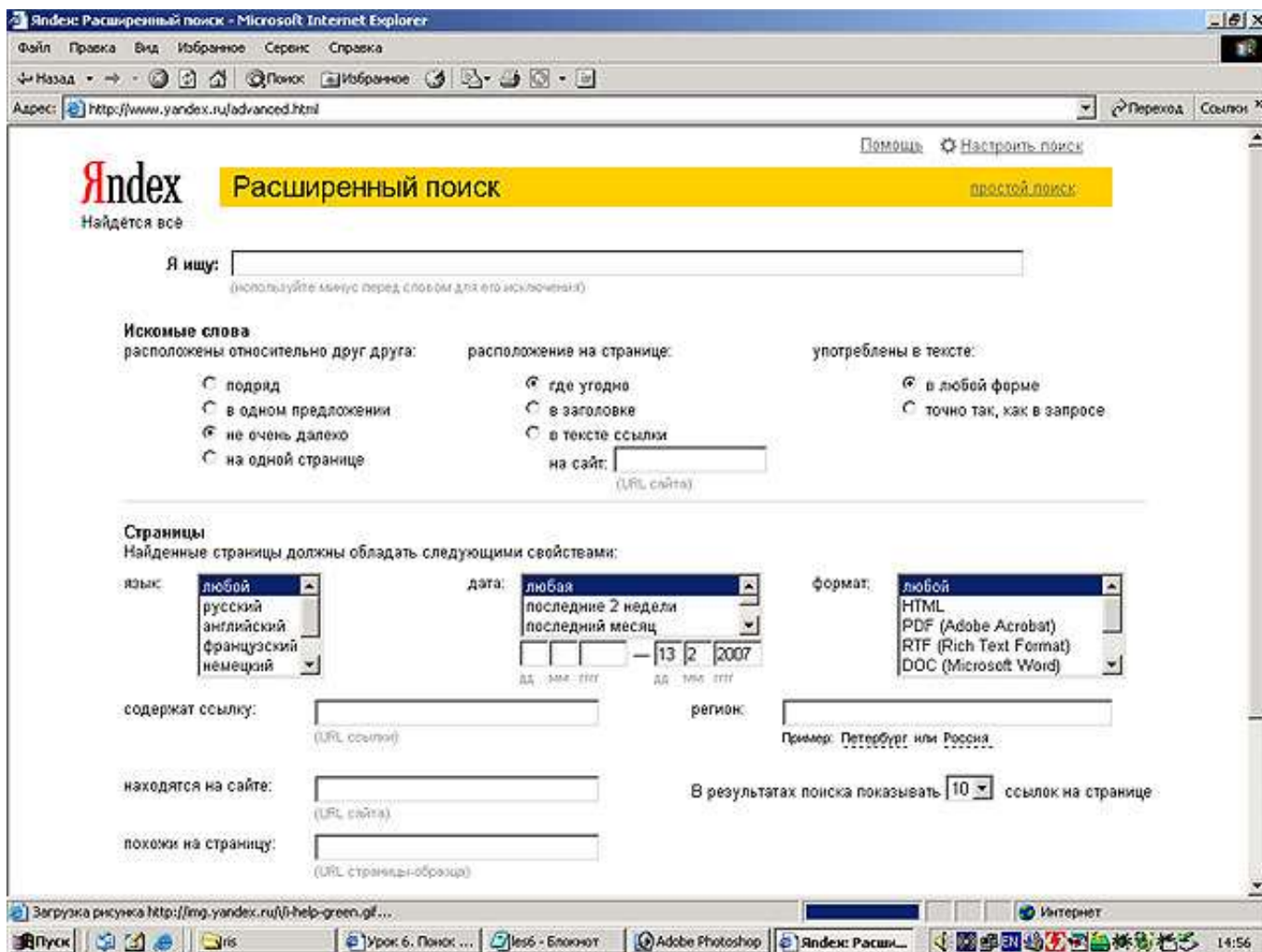
Проведем несколько экспериментов с любой из поисковых систем. Предположим, что мы решили завести аквариум и нас интересует любая информация по данной теме. На первый взгляд самое простое — это поиск по слову «аквариум». Проверим это, например, в поисковой системе Яндекс. Результатом поиска будет огромное количество страниц - огромное количество ссылок. Причем, если посмотреть внимательнее, среди них окажутся сайты, упоминающие группу Б. Гребенщикова «Аквариум», торговые центры и неформальные объединения с таким же названием, и многое другое, не имеющее отношения к аквариумным рыбкам.

Нетрудно догадаться, что такой поиск не может удовлетворить даже непритязательного пользователя. Слишком много времени придется потратить на то, чтобы отобрать среди всех предложенных документов те, что касаются нужного нам предмета, и уж тем более на то, чтобы ознакомиться с их содержанием.

Можно сразу сделать вывод, что вести поиск по одному слову, как правило, нецелесообразно, ведь по одному слову очень сложно определить тему, которой посвящен документ, веб-страница или сайт. Исключение составляют редкие слова и термины, которые практически никогда не используются вне своей тематической области.

Поиск информации в сети Интернет

Имея определенный набор наиболее употребительных терминов в нужной области, можно использовать расширенный поиск. На рис. показано окно расширенного поиска в поисковой системе Яндекс. В этом режиме возможности языка запросов реализованы в виде формы. Подобный сервис, включающий словарные фильтры, предлагается почти всеми поисковыми системами.



Попробуем уточнить условия поиска и введем словосочетание «аквариумные рыбки». Количество ссылок уменьшится более чем в 20 раз. Этот результат нас устраивает больше, но все равно среди предложенных ссылок могут встретиться, например, русские сувенирные наборы спичечных этикеток с изображениями рыбок, и коллекции заставок для Рабочего стола компьютера, и каталоги аквариумных рыбок с фотографиями, и магазины аквариумных аксессуаров. Очевидно, что следует продолжить движение в направлении уточнения условий поиска.

Для того чтобы сделать поиск более продуктивным, во всех поисковых системах существует специальный **язык формирования запросов** со своим синтаксисом. Эти языки во многом похожи. Изучить их все достаточно сложно, но любая поисковая машина имеет справочную систему, которая позволит вам освоить нужный язык.

Несколько простых правил формирования запроса в поисковой системе Яндекс.

Ключевые слова в запросе следует писать строчными (маленькими) буквами.

Это обеспечит поиск всех ключевых слов, а не только тех, которые начинаются с прописной буквы.

При поиске учитываются все формы слова по правилам русского языка, независимо от формы слова в запросе.

Например, если в запросе было указано слово «знаю», то условию поиска будут удовлетворять и слова «знаем», «знаете» и т. п.

Для поиска устойчивого словосочетания следует заключить слова в кавычки.

Например, «фарфоровая посуда».

Для поиска по точной словоформе перед словом надо поставить восклицательный знак.

Например, для поиска слова «сентябрь» в родительном падеже следует написать «!сентября».

Для поиска внутри одного предложения слова в запросе разделяют пробелом или знаком &

Например, «приключенческий роман» или «приключенческий&роман». Несколько набранных в запросе слов, разделенных пробелами, означают, что все они должны входить в одно предложение искомого документа.

Если вы хотите, чтобы были отобраны только те документы, в которых встретилось каждое слово, указанное в запросе, поставьте перед каждым из них знак плюс «+». Если вы, наоборот, хотите исключить какие-либо слова из результата поиска, поставьте перед этим словом минус «-». **Знаки «+» и «-» надо писать через пробел от предыдущего и слитно со следующим словом.**

Например, по запросу «Волга -автомобиль» будут найдены документы, в которых есть слово «Волга» и нет слова «автомобиль».

При поиске синонимов или близких по значению слов между словами можно поставить вертикальную черту «|».

Например, по запросу «ребенок | малыш | младенец» будут найдены документы с любым из этих слов.

Вместо одного слова в запросе можно подставить целое выражение. Для этого его надо взять в скобки

Например, «(ребенок | малыш | дети | младенец) +(уход | воспитание)».

Знак «~» (тильда) позволяет найти документы с предложением, содержащим первое слово, но не содержащим второе.

Например, по запросу «книги ~ магазин» будут найдены все документы, содержащие слово «книги», рядом с которым (в пределах предложения) нет слова «магазин».

Если *оператор* повторяется *один раз* (например, & или ~), поиск производится в пределах предложения. *Двойной оператор* (&&,~~) задает поиск в пределах документа.

Например, по запросу «рак ~~ астрология» будут найдены документы со словом «рак», не относящиеся к астрологии.

Вернемся к примеру с аквариумными рыбками. После прочтения нескольких предлагаемых поисковой системой документов становится понятно, что поиск информации в Интернете следует начинать не с выбора аквариумных рыбок. Аквариум - сложная биологическая система, создание и поддержание которой требует специальных знаний, времени и серьезных капиталовложений.

На основании полученной информации человек, производящий поиск в Интернете, может кардинально изменить стратегию дальнейшего поиска, приняв решение изучить специальную литературу, относящуюся к исследуемому вопросу.

Для поиска литературы или полнотекстовых документов возможен следующий запрос:

«+(аквариум | аквариумист | аквариумистика) +начинающим +(советы | литература) +(статья | тезис | полнотекстовый) -(цена | магазин | доставка | каталог)».

После обработки запроса поисковой машиной результат оказался весьма успешным. Уже первые ссылки приводят к искомым документам.

Теперь можно подытожить результаты поиска, сделать определенные выводы и принять решение о возможных действиях:

- Прекратить дальнейший поиск, так как в силу различных причин содержание аквариума вам не под силу.
- Прочитать предлагаемые статьи и приступить к устройству аквариума.
- Поискать материалы о хомячках или волнистых попугайчиках.