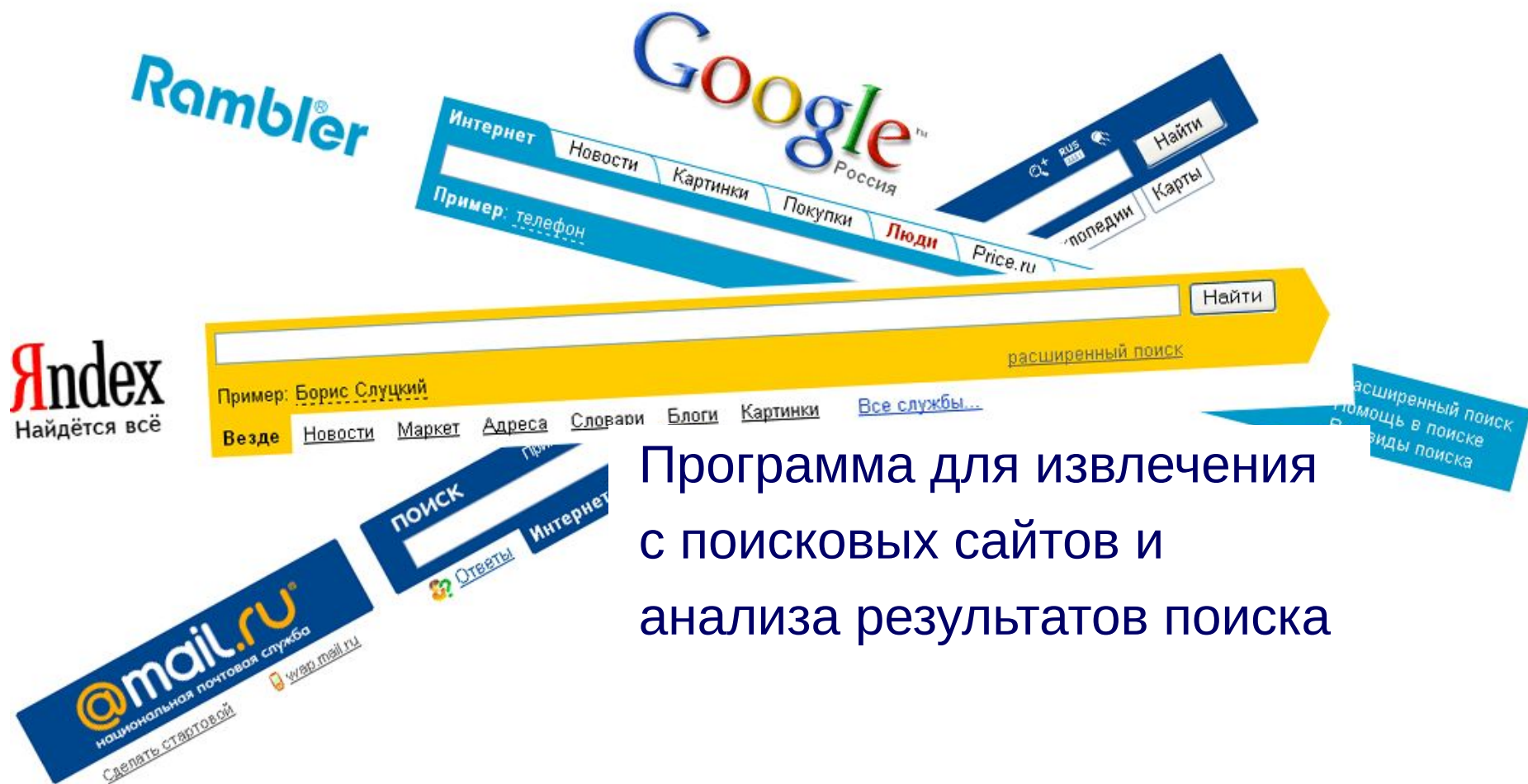


ПРОСТОЙ СОФТ: АНАЛИЗ РЕЗУЛЬТАТОВ ПОИСКА



Программа для извлечения
с поисковых сайтов и
анализа результатов поиска

Основные функции системы

- Извлечение результатов поиска с сайтов Google.ru, Yandex.ru, Rambler.ru и Mail.ru в локальную базу данных.
- Анализ результатов поиска.

Преимущества системы

- Автоматизация процесса извлечения результатов поиска с сайтов в локальную базу
- Эффективный анализ - удобные способы сортировки и фильтрации данных
- Возможность изменения структуры данных – добавления таблиц и полей

Составляющие системы

- Вкладка «**Google.ru**»
- Вкладка «**Yandex.ru**»
- Вкладка «**Rambler.ru**»
- Вкладка «**Mail.ru**»
- Вкладка «**Анализ**»

Рассмотрим подробнее процесс работы в системе

Общий порядок работы

- Вы определяете для себя, какие данные Вы будете искать на поисковых сайтах.
- При помощи вкладок для извлечения осуществляется извлечение данных в локальную базу.
- Когда данные уже извлечены, но Вы еще не отключились от Интернета, Вы можете заполнить поля GooglePageRank и Яндекс тИЦ . Это поможет Вам при дальнейшем анализе «нужности» и «полезности» каждого найденного сайта.
- Дальнейшая работа осуществляется во вкладке «Анализ», где данные из Интернета будут представлены в доступной для использования форме, с возможностью сортировки, фильтрации, а также скрытия/отображения различных параметров.
- Если необходимо, можно включить многопользовательский режим.

Извлечение данных

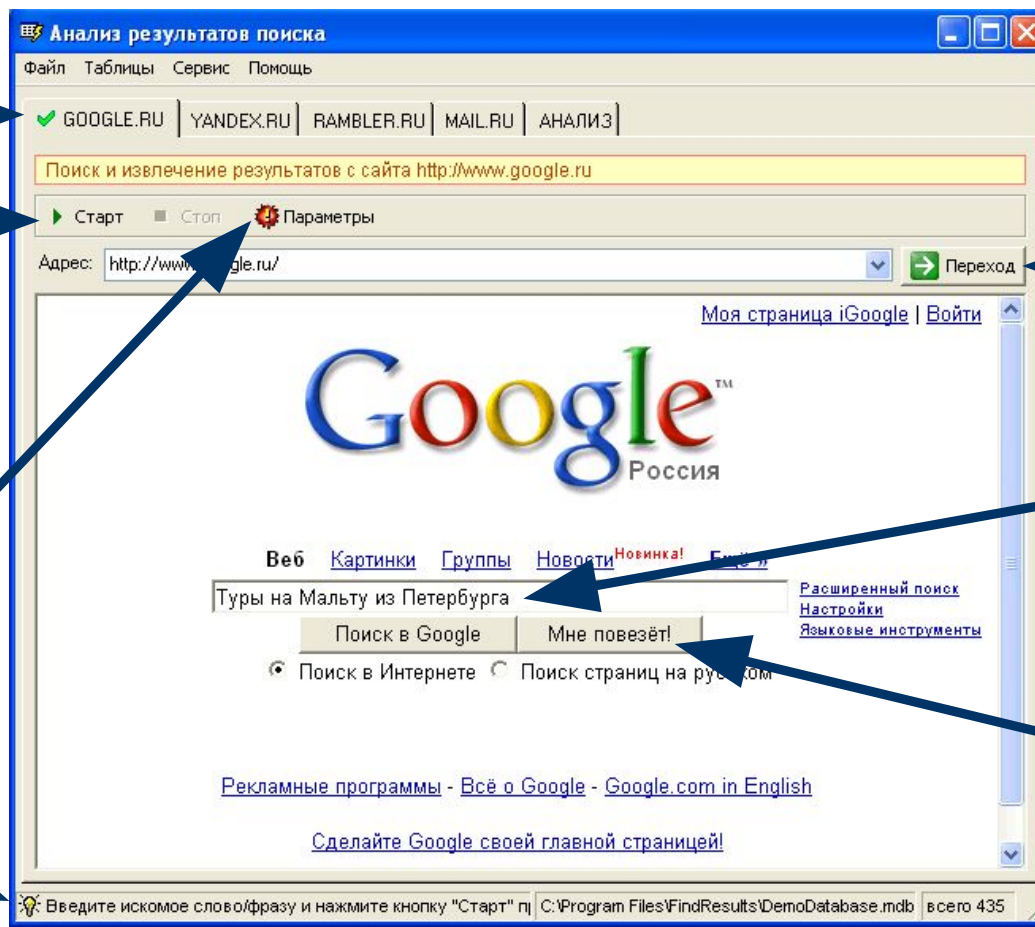
Принцип извлечения данных на всех вкладках поисковиков одинаков. Необходимо зайти на нужную вкладку, ввести интересующие параметры поиска (поисковый запрос), например, Туры на Мальту из Петербурга.

1. Заходим на вкладку

5. Жмем кнопку "Старт" – начинается

4. Вводим параметры извлечения: таймаут загрузки страниц и количество обрабатываемых страниц

Следуем указаниям в строке-статусе



2. Жмем кнопку "переход"

3. Вводим поисковый запрос

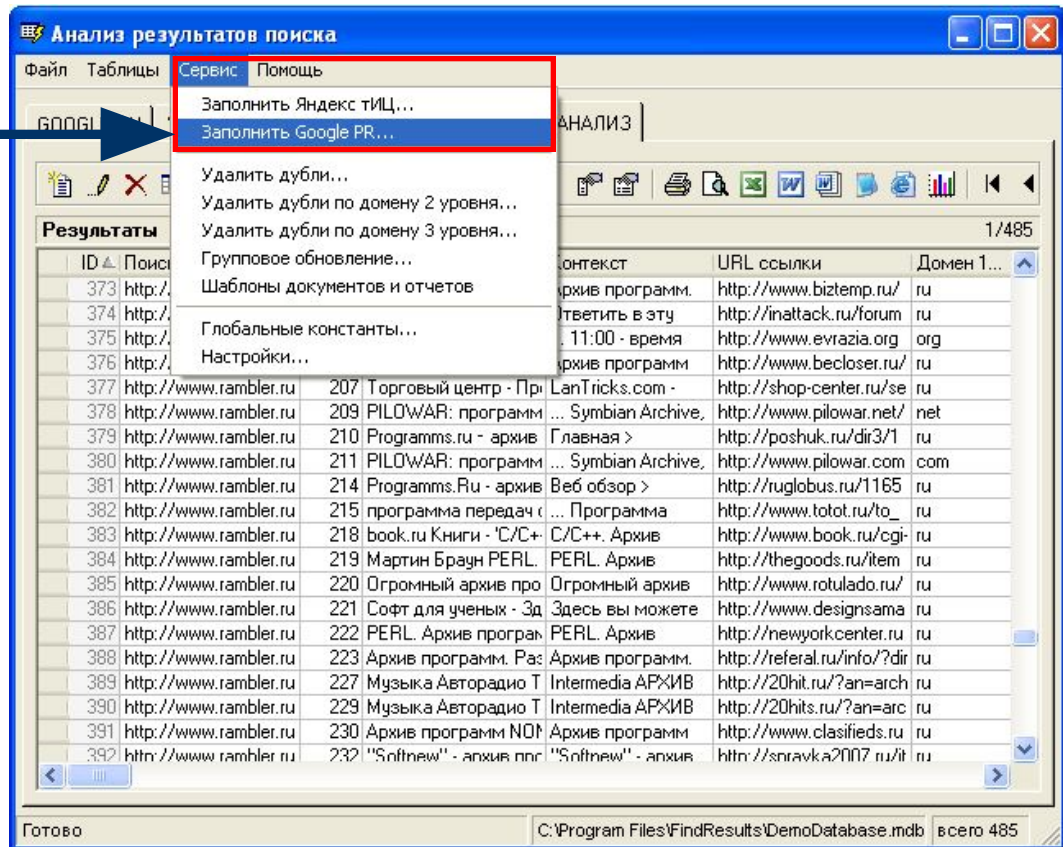
Кнопку "поиск" на странице нажимать НЕ надо

Извлечение данных – Google PR и Yandex тИЦ

После того, как результаты поиска извлечены, но Вы еще подключены к Интернет, можно выбрать пункт меню **“Заполнить Google PR”**.

Программа пройдет по всем найденным поисковиками сайтам и извлечет для каждого значение параметра Google PR (Google Page Rank)- индекса популярности по оценкам Google.

Аналогично можно поступить с параметром **Yandex тИЦ** (тематический Индекс Цитирования).



Построение аналитических полей

По извлеченным результатам поиска строятся аналитические поля. Программа как бы "читает" результат поиска (ссылку на ресурс) и раскладывает все "по полочкам". Вы получаете строго структурированную информацию.

Программа построила аналитические поля. Правый щелчок мыши по названию любого поля выдает полный перечень

всех полей.

Анализ результатов поиска

| ID | Поисковый сайт | Название ссылки | Контекст | URL ссылки | Домен... | Домен 2 уровня | Домен 3... |
|-----|----------------------|------------------------|----------------------|---------------------------|----------|-------------------|------------|
| 466 | http://www.google.ru | Мальта, отдых на Ма | Новый аттракцион | http://malta.megatis.ru/ | ru | megatis.ru | malta.meg |
| 467 | http://www.google.ru | Туры на Мальту прод | В самом конце мая | http://www.votpusk.ru/ | ru | votpusk.ru | |
| 468 | http://www.google.ru | В Москве и Петербу | 27 октября в отеле | http://www.travel.ru/ne | ru | travel.ru | |
| 469 | http://www.google.ru | Экскурсии по Санкт-И | Групповые туры из | http://www.vt-service.ru | ru | vt-service.ru | |
| 470 | http://www.google.ru | туры на Мальту | туры на Мальту. | http://www.paneuro.ru/ | ru | paneuro.ru | |
| 471 | http://www.google.ru | Мальта - Поезд Пете | Путешествия и туры | http://maltaru.ru/news_ | ru | maltaru.ru | |
| 472 | http://www.google.ru | Апорт-Каталог: По ст | Он-лайн поиск туров. | http://catalog.aport.ru/i | ru | aport.ru | catalog.ap |
| 473 | http://www.google.ru | Мальта. Отдых на Ма | Города и курорты. | http://www.aurore.ru/c | ru | aurore.ru | |
| 474 | http://www.google.ru | Отдых и туры на Мал | Отдых и туры на | http://www.maktr.ru/rus | ru | maktr.ru | |
| 475 | http://www.google.ru | Отель милано две ма | Мальта Отель Бей | http://pekacexatiwigico. | ru | land.ru | pekacexa |
| 476 | http://www.google.ru | БЕЗОПАСНОСТЬ и О | Туры на Мальту | http://www.bezopasno.r | ru | bezopasno.ru | |
| 477 | http://www.google.ru | Мальта - Туристичес | ... FL503 | http://www.ice-nut.ru/m | ru | ice-nut.ru | |
| 478 | http://www.google.ru | Туры на Мальту Ту | Золотое Кольцо и | http://www.vipvoyage.r | ru | vipvoyage.ru | |
| 479 | http://www.google.ru | Туры в Санкт-Петерб | Туры на | http://www.vipvoyage.r | ru | vipvoyage.ru | |
| 480 | http://www.google.ru | Австрия - отдых в Ав | Санкт-Петербург | http://www.online-austri | ru | online-austria.ru | |
| 481 | http://www.google.ru | ТУРФИРМА "Конкор | Туры в Китай из | http://www.concorde-tr | ru | net.ru | concorde- |
| 482 | http://www.google.ru | ТУРФИРМА "Конкор | Туры в Австрию из | http://www.concorde-tr | ru | net.ru | concorde- |
| 483 | http://www.google.ru | Refer.Ru / Страны, Р | ... туроператор по | http://www.refer.ru/648 | ru | refer.ru | |
| 484 | http://www.google.ru | Туры в Италию из Се | Сайт турфирмы | http://www.classic-tour. | ru | classic-tour.ru | |
| 485 | http://www.google.ru | Туры VIP - туры Ну | Туры.VIP - туры. | http://www.santour.ru/c | ru | santour.ru | |

- ID
 - Поисковый сайт
 - Позиция
 - Название ссылки
 - Контекст
 - URL ссылки
 - Домен 1 уровня
 - Домен 2 уровня
 - Домен 3 уровня
 - Страна домена
 - GooglePR
 - YandexCY
 - Рубрика
 - Регион
 - Размер страницы Kb
 - Дата публикации
 - Время извлечения
 - Заметки
 - Избранное
 - Неинтересное
- Добавить поле...
- Удалить поле "Домен 3 уровня"...
- Свойства поля "Домен 3 уровня"...

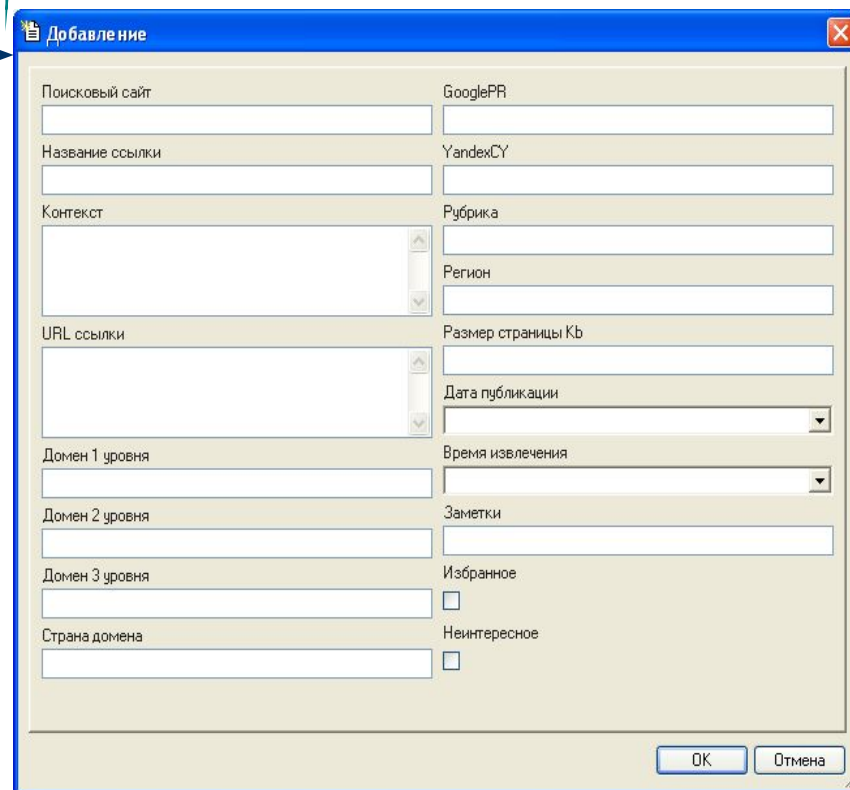
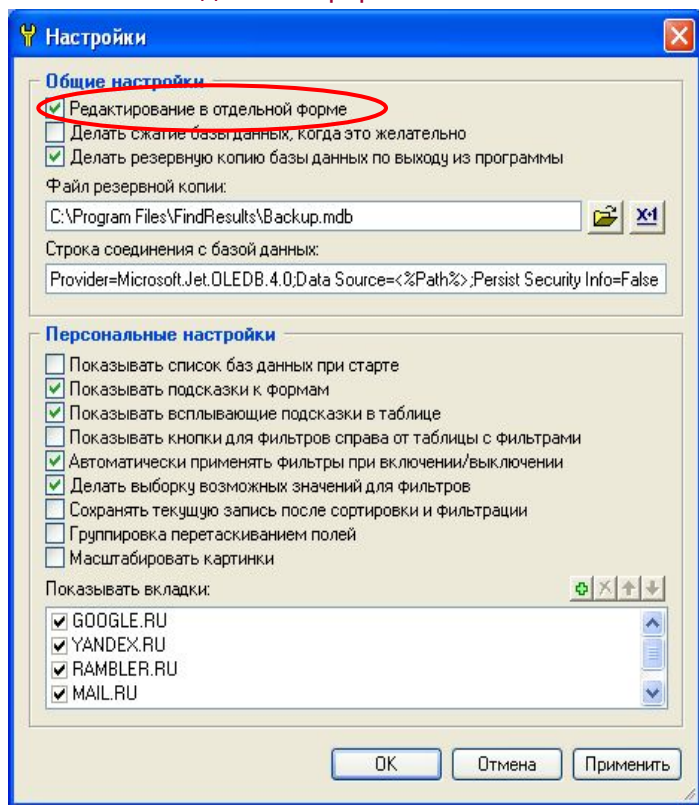
Занесение данных вручную

Можно просто вести базу данных, вводя информацию о сайтах с клавиатуры или копируя их через буфер обмена.

Задаем редактирование в отдельной форме

Жмем “Новая запись”

Заносим информацию о сайте



Фильтрация данных

Для того чтобы показывать не все записи таблицы, а только удовлетворяющие определенным критериям, используются **фильтры**.

Допустим нас интересуют только сайты с доменом 1 уровня ru, значением GooglePR больше или равно 3, занимающие в результатах поиска позиции не ниже 30-ой.

- Устанавливаем фильтры, в соответствии с оговоренными условиями. Особую ценность представляют параметры GooglePR и Yandex тИЦ. По их значению можно отбирать действительно значимые сайты.
- В программе есть возможность установки фильтров по всем операторам языка SQL
- Находим 12 сайтов, удовлетворяющих нашему запросу

Анализ результатов поиска

Файл Таблицы Сервис Помощь

GOOGLE.RU | YANDEX.RU | RAMBLER.RU | MAIL.RU | АНАЛИЗ

Фильтры: всего 3 (включено 3) Описание таблицы:

| Включен | Связь | Поле | Условие | Значение |
|-------------------------------------|-------|----------------|---------|----------|
| <input checked="" type="checkbox"/> | | Домен 1 уровня | = | ru |
| <input checked="" type="checkbox"/> | И | GooglePR | >= | 3 |
| <input checked="" type="checkbox"/> | И | Позиция | <= | 30 |
| <input type="checkbox"/> | | * | | |

Результаты (представление) 1/12 (отфильтровано) всего 485

| ID | Поисковый сайт | Поз... | Название ссылки | Контакт... | URL ссылки | Домен 1... |
|-----|----------------------|--------|-----------------------|-------------------|--------------------------|------------|
| 9 | http://www.google.ru | 10 | freeSOFT - скачать... | Каталог программ | http://www.freesoft.ru | ru |
| 10 | http://www.google.ru | 12 | Programms - архив | Каталог | http://www.programms.r | ru |
| 11 | http://www.google.ru | 13 | Каталог программ / | Сайт представляет | http://klerk.ru/soft/dir | ru |
| 12 | http://www.google.ru | 14 | Каталог программ. с | Софт каталог | http://www.vidon.ru | ru |
| 13 | http://www.google.ru | 15 | Каталог программ и | Бесплатные free и | http://www.homesoft.ru | ru |
| 14 | http://www.google.ru | 17 | Каталог программ :: | Каталог | http://www.mega-soft.ru | ru |
| 18 | http://www.google.ru | 22 | Скачать бесплатные | Бесплатные и | http://www.izone.ru | ru |
| 20 | http://www.google.ru | 26 | Компьютерные прог | Каталог | http://www.przone.ru | ru |
| 21 | http://www.google.ru | 27 | Центр креативных те | Главное. Методы | http://www.inventech.r | ru |
| 22 | http://www.google.ru | 28 | Винград - работа для | о проекте. Каждый | http://vingrad.ru | ru |
| 194 | http://www.yandex.ru | 8 | Электронный порта | Избранные схемы | http://kazus.ru | ru |
| 199 | http://www.yandex.ru | 27 | Общие св-ва выбо | Каталог программ | http://www.kls.ru/shp | ru |

Готово C:\Program Files\FindResults\DemoDatabase.mdb всего 485

Свойства таблицы

Настроить интерфейс вкладки

Объявления можно при помощи

опции Свойства таблицы:

- удобная настройка полей, которые надо отображать.
- возможность настроить цвет, которым будут выделяться строки с определенными значениями определенных полей
- выбор полей, по которым будет отображаться дерево
- настройка подчиненных вкладок с подчиненными таблицами

Внутреннее имя: qdfMain

Тип: представление (VIEW)

Наименование таблицы (подпись): Результат

Показывать описание таблицы: Показывать описание таблицы:
Общие принципы работы:

Поля

Показывать поля:

| | |
|--|---|
| <input checked="" type="checkbox"/> ID | <input type="checkbox"/> ID |
| <input checked="" type="checkbox"/> Поисковый сайт | <input type="checkbox"/> Поисковый сайт |
| <input type="checkbox"/> Позиция | <input type="checkbox"/> Позиция |

Объединять одинаковые значения:

Цветовыделение

Выделение строк заданным цветом при условии:

| Цвет | Поле | Условие | Значение |
|------|--------------|---------|----------|
| ... | Избранное | = | Да |
| ... | Неинтересное | = | Да |

Итоги

Поля, по которым считать итоги:

| Поле | Групповая операция |
|------|--------------------|
| * | |

Дерево

Поля, по которым строить дерево:

| Уровень | Поле | Сортировка веток |
|-----------|----------------|------------------|
| Уровень 1 | Домен 1 уровня | По возрастанию |
| * | | |

Подчиненные таблицы

Привязка таблиц, которые отображать как подчиненные:

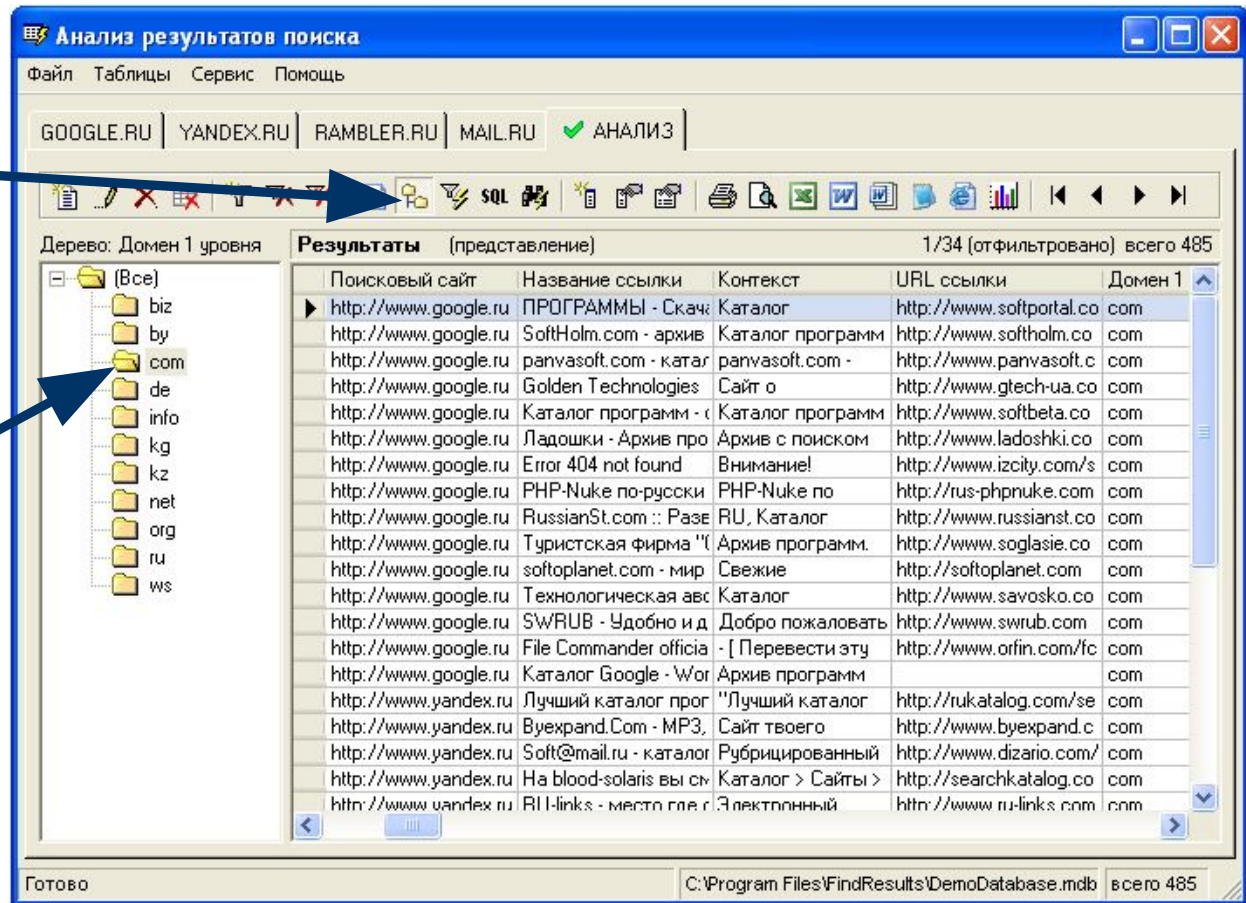
| Таблица | Тип связи | Связанное поле | Поле главной... | Доп. условия |
|---------|-----------|----------------|-----------------|--------------|
| * | | | | |

OK Отмена

Дерево

Когда необходимо отобразить данные с одинаковыми значениями полей, вместо таблицы с фильтрами удобнее использовать **дерево**. Предварительно в свойствах таблицы, в таком случае, надо установить, по каким полям дерево будет строиться.

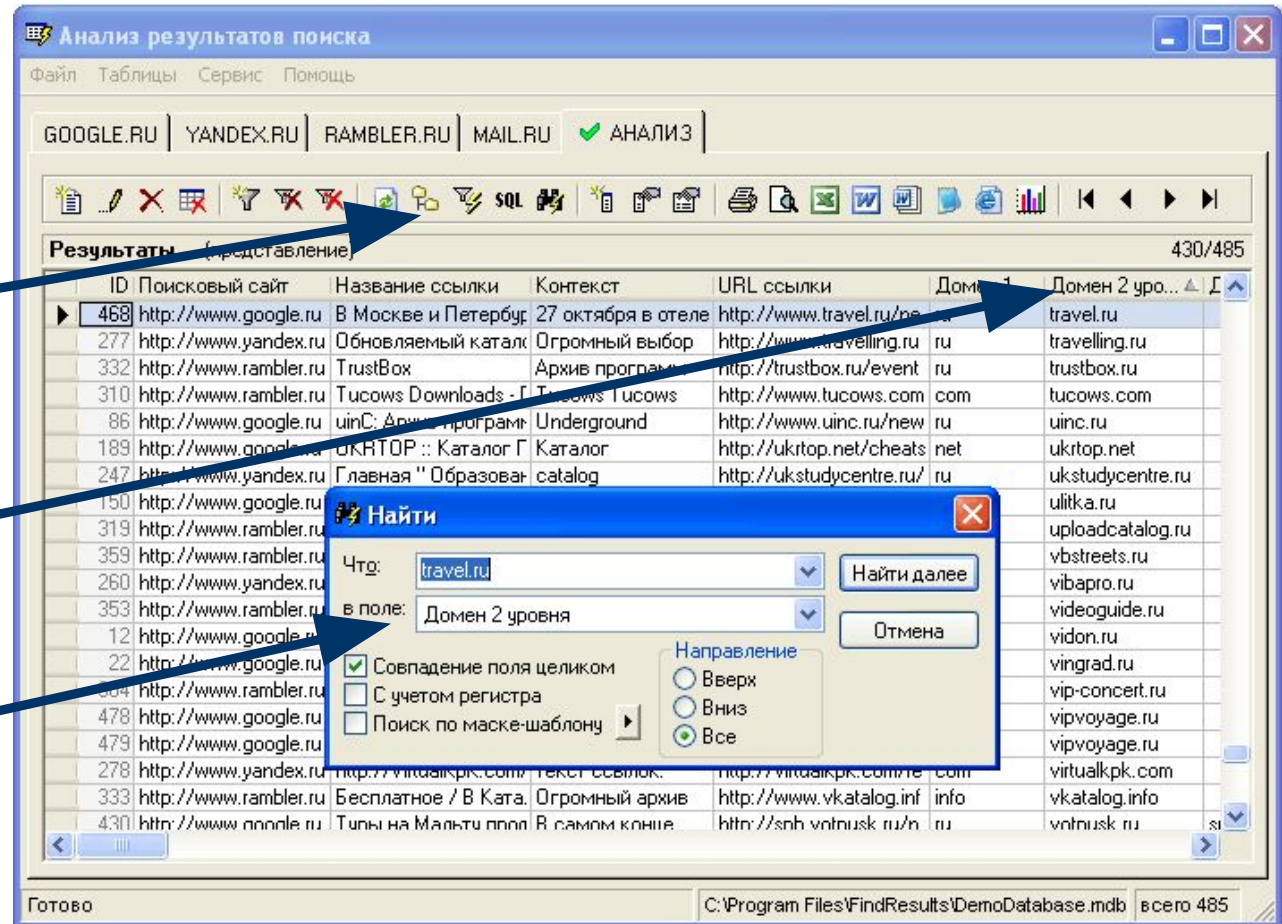
- отображаем дерево при помощи соответствующей кнопки панели инструментов
- выбираем нужную ветвь дерева, и отображаются записи только с выбранным значением поля, по которому построено дерево



Поиск по таблице

Допустим, Вы хотите найти определенный домен 2-го уровня, например, travel.ru

- Скрываем таблицу с фильтрами и дерево, нажав на соответствующие кнопки панели инструментов, чтобы больше данных было в области видимости
- Сортируем данные простым кликом по полю Домен 2 уровня
- Нажав кнопку поиска на панели инструментов ищем *travel.ru* в поле Домен 2 уровня



Изменение структуры

Возможно, для работы Вам понадобятся свои поля для работы с информацией о сайтах. Или, наоборот: некоторые поля, предусмотренные разработчиком, Вами точно использоваться не будут.

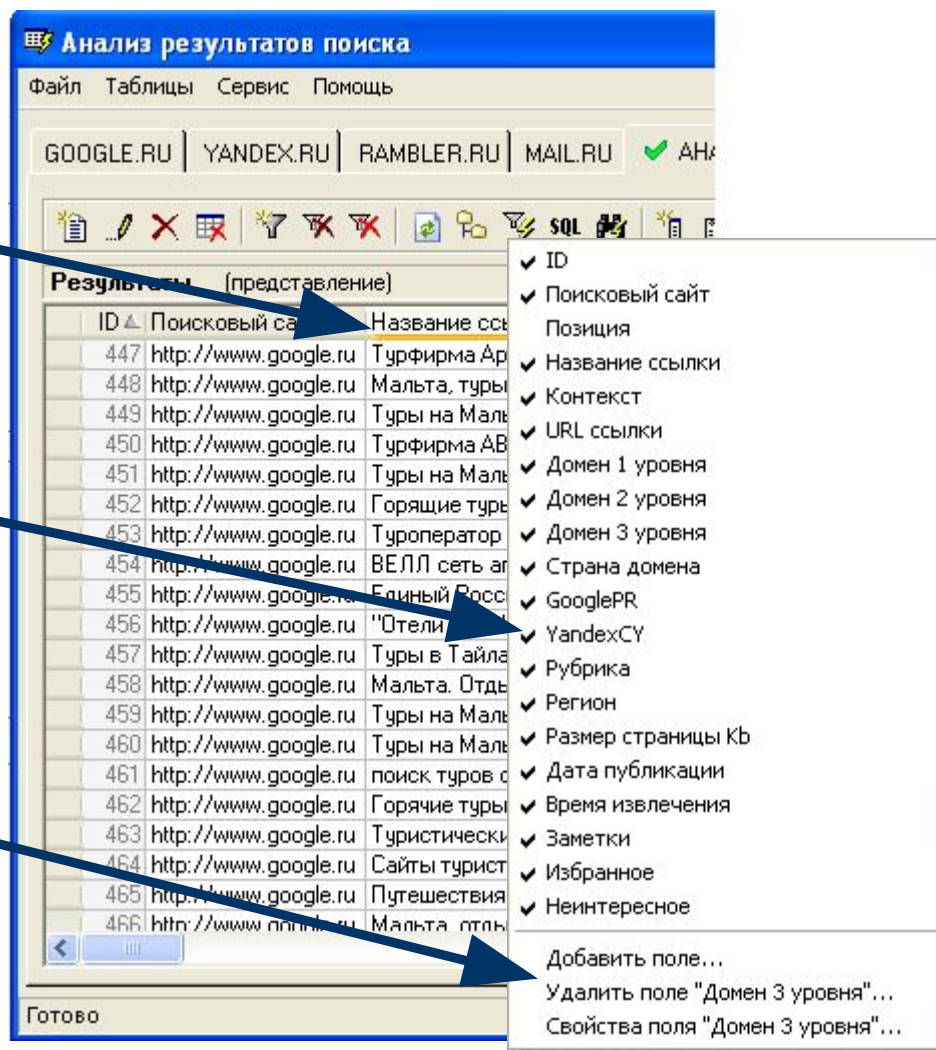
Большое преимущество программы состоит в том, что структуру данных можно легко изменять согласно требований пользователя.

Вы можете:

- Удалять ненужные поля и таблицы
- Скрывать/отображать поля и таблицы
- Создавать новые поля и таблицы

Удаление / скрывание полей

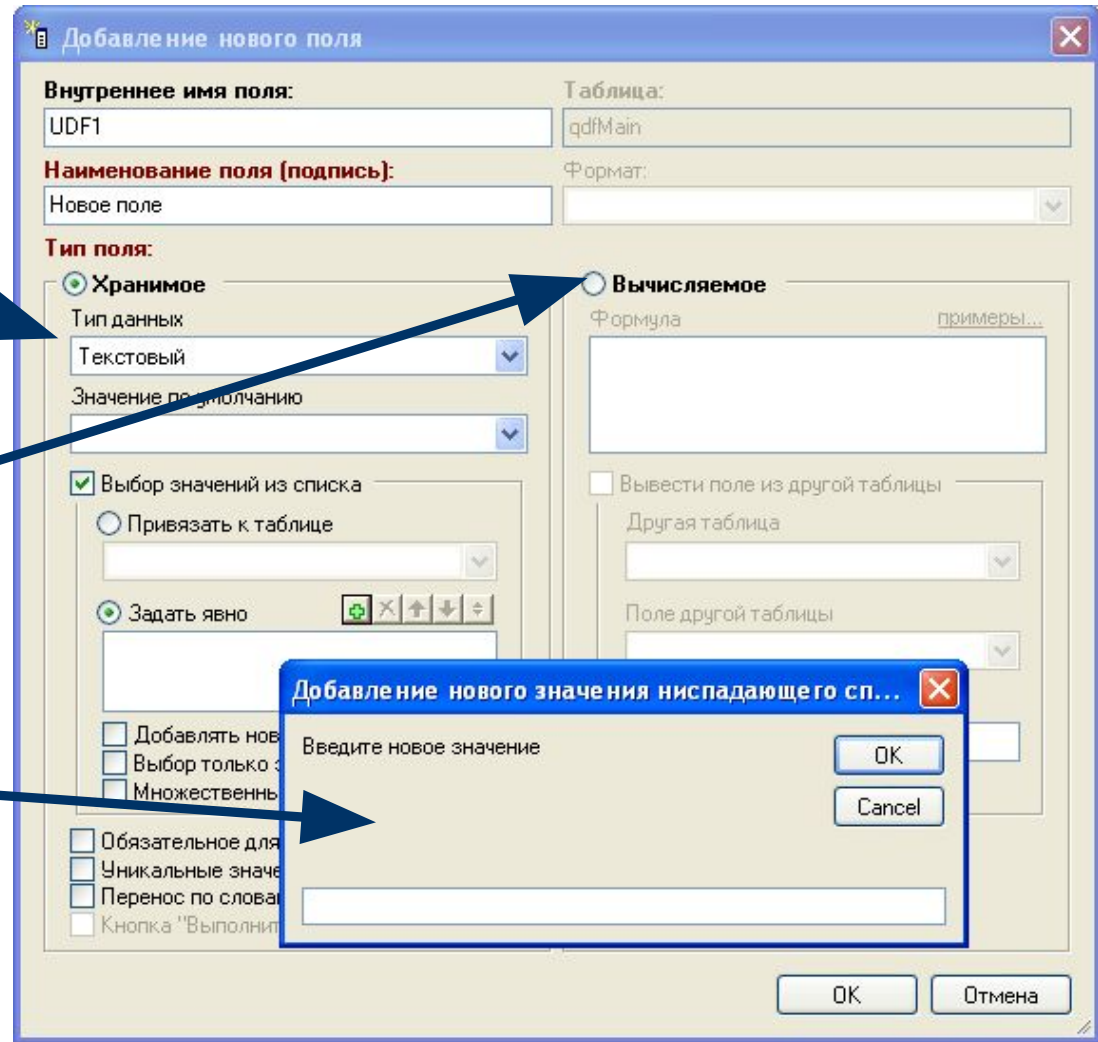
- Контекстное меню открывается простым щелчком правой кнопкой мыши по заголовкам полей
- Скрытие/отображение полей происходит путем проставления галочек напротив полей, которые должны быть видны
- Также при помощи данного контекстного меню можно добавить новое поле, удалить выбранное поле или посмотреть его свойства.



Добавление нового поля

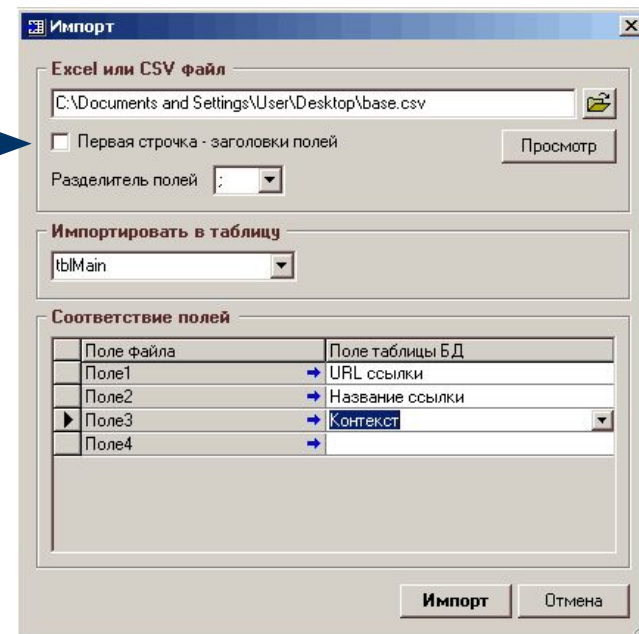
При добавлении нового поля предоставляются следующие возможности:

- возможность выбора типа данных
- возможность создания вычисляемых полей на основе sql (например, можно создать поле с формулой "[Размер страницы, Kb] / [GooglePageRank]" Чем меньше значение такого параметра, тем "лучше" сайт
- явное задание значений для выбора из выпадающего списка



Импорт данных

Вы можете **импортировать** объявления из каких-нибудь других источников. Формат - текстовый **CSV-файл** или файл **MS Excel**, можно задать разделитель полей. Импортировать можно любую информацию. Это могут быть данные из каких-либо каталогов сайтов и т.п. Вы сами выбираете нужную таблицу базы данных (сначала ее можно создать) и задаете соответствие полей откуда какую информацию копировать. Во время сопоставления необходимо помнить о типе полей.

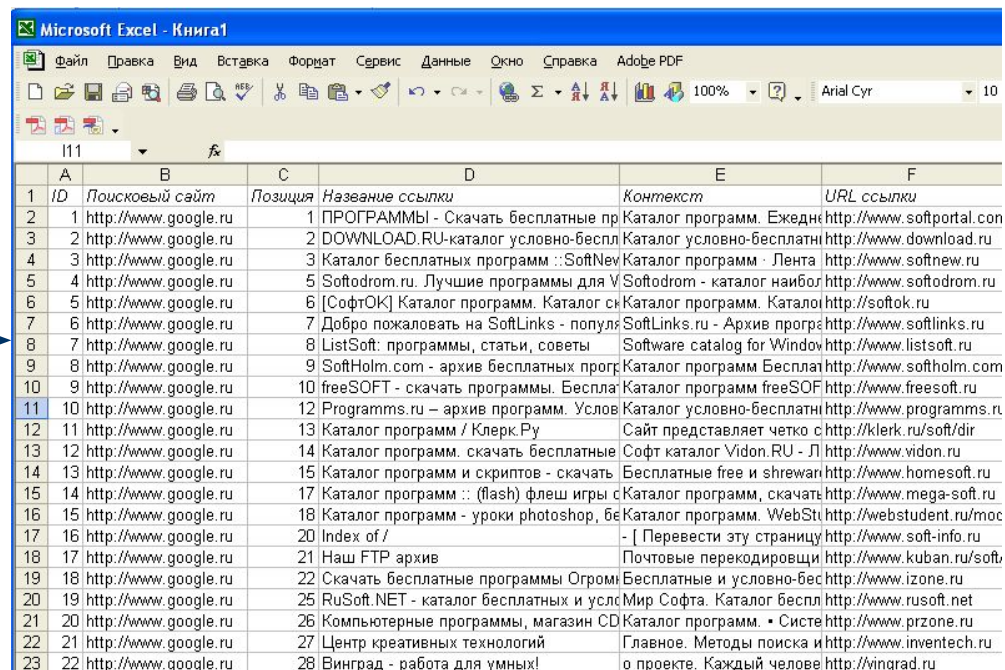


Экспорт данных

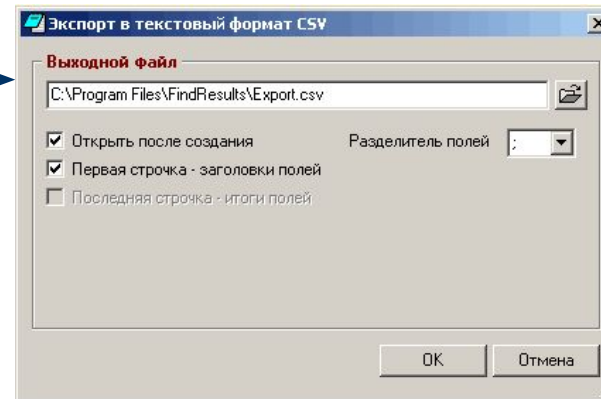
Экспортировать также можно абсолютно любую таблицу. Для этого нужно нажать одну из четырех кнопок для экспорта на панели инструментов (в Word, в Excel, в CSV или в HTML).

В MS Excel экспорт выполняется сразу без дополнительных вопросов.

Для экспорта в текстовый формат CSV выдается форма, где вы выбираете целевой файл, задаете разделитель полей. Можно задавать любой символ-разделитель. Также можно экспортировать названия (заголовки) полей первой строчкой. Экспортируются только видимые поля.



| ID | Поисковый сайт | Позиция | Название ссылки | Контекст | URL ссылка |
|----|----------------------|---------|---|-----------------------------|---------------------------|
| 1 | http://www.google.ru | 1 | ПРОГРАММЫ - Скачать бесплатные пр | Каталог программ. Ежедне | http://www.softportal.com |
| 2 | http://www.google.ru | 2 | DOWNLOAD.RU-каталог условно-беспл | Каталог условно-бесплатн | http://www.download.ru |
| 3 | http://www.google.ru | 3 | Каталог бесплатных программ :: SoftNew | Каталог программ - Лента | http://www.softnew.ru |
| 4 | http://www.google.ru | 5 | Softodrom.ru. Лучшие программы для V | Softodrom - каталог набол | http://www.softodrom.ru |
| 5 | http://www.google.ru | 6 | [СофтОК] Каталог программ. Каталог с | Каталог программ. Каталог | http://softok.ru |
| 6 | http://www.google.ru | 7 | Добро пожаловать на SoftLinks - популя | SoftLinks.ru - Архив progr | http://www.softlinks.ru |
| 7 | http://www.google.ru | 8 | ListSoft: программы, статьи, советы | Software catalog for Window | http://www.listsoft.ru |
| 8 | http://www.google.ru | 9 | SoftHolm.com - архив бесплатных прог | Каталог программ Бесплат | http://www.softholm.com |
| 9 | http://www.google.ru | 10 | freeSOFT - скачать программы. Беспла | Каталог программ freeSOF | http://www.freesoft.ru |
| 10 | http://www.google.ru | 12 | Programms.ru - архив программ. Услов | Каталог условно-бесплатн | http://www.programms.ru |
| 11 | http://www.google.ru | 13 | Каталог программ / Клерк.Ру | Сайт представляет четко c | http://klerk.ru/soft/dir |
| 12 | http://www.google.ru | 14 | Каталог программ. скачать бесплатные | Софт каталог Vidon.RU - Л | http://www.vidon.ru |
| 13 | http://www.google.ru | 15 | Каталог программ и скриптов - скачать | Бесплатные free и shrewari | http://www.homesoft.ru |
| 14 | http://www.google.ru | 17 | Каталог программ :: (flash) флеш игры с | Каталог программ, скачать | http://www.mega-soft.ru |
| 15 | http://www.google.ru | 18 | Каталог программ - уроки photoshop, б | Каталог программ. WebSt | http://webstudent.ru/moc |
| 16 | http://www.google.ru | 20 | Index of / | - [Перевести эту страницу | http://www.soft-info.ru |
| 17 | http://www.google.ru | 21 | Наш FTP архив | Почтовые перекодировщи | http://www.kuban.ru/soft |
| 18 | http://www.google.ru | 22 | Скачать бесплатные программы Огром | Бесплатные и условно-бес | http://www.izone.ru |
| 19 | http://www.google.ru | 25 | RuSoft.NET - каталог бесплатных и усл | Мир Софта. Каталог беспл | http://www.rusoft.net |
| 20 | http://www.google.ru | 26 | Компьютерные программы, магазин CD | Каталог программ. Систе | http://www.przone.ru |
| 21 | http://www.google.ru | 27 | Центр креативных технологий | Главное. Методы поиска и | http://www.inventech.ru |
| 22 | http://www.google.ru | 28 | Винград - работа для умных! | о проекте. Каждый челове | http://vingrad.ru |



Экспорт в текстовый формат CSV

Выходной файл
C:\Program Files\FindResults\Export.csv

Открыть после создания
 Первая строка - заголовки полей
 Последняя строка - итоги полей

Разделитель полей : ;

OK Отмена

Групповое обновление

По окончании извлечения и построения аналитических полей, чтобы не загромождать некоторые поля излишней информацией, можно воспользоваться опцией **Групповое обновление**.

Что касается загромождения полей информацией, можно, например, в поле Поисковый сайт заменить названия сайтов с <http://www.yandex.ru> на YANDEX и т.д.

- Есть возможность не изменить выбранные значения, а удалить их
- Есть возможность посмотреть и изменить sql-инструкцию обновления
- Возможность задания нескольких условий отбора записей при помощи различных sql-операторов

Данная функциональность предназначена для проведения операций изменения или удаления сразу нескольких записей. Например, можно выставить поле "Статус" в "ок" для всех записей с пустым статусом. Будьте осторожны, поскольку после выполнения инструкции SQL изменения нельзя вернуть назад. Если нужно, сделайте резервную копию базы данных перед выполнением.

Выполнить Очистить Сохранить Загрузить SQL Инструкция

Таблица для обновления
tblMain

Тип обновления
 Удалить
 Изменить

| Поле | Присвоить | Значение |
|----------------|-----------|----------|
| Поисковый сайт | = | Yandex |
| * | | |

Диапазон записей
 Все
 Удовлетворяющие следующим условиям:

| Связь | Поле | Условие | Значение |
|-------|----------------|---------|----------------------|
| | Поисковый сайт | = | http://www.yandex.ru |
| * | | | |

Удаление дублей

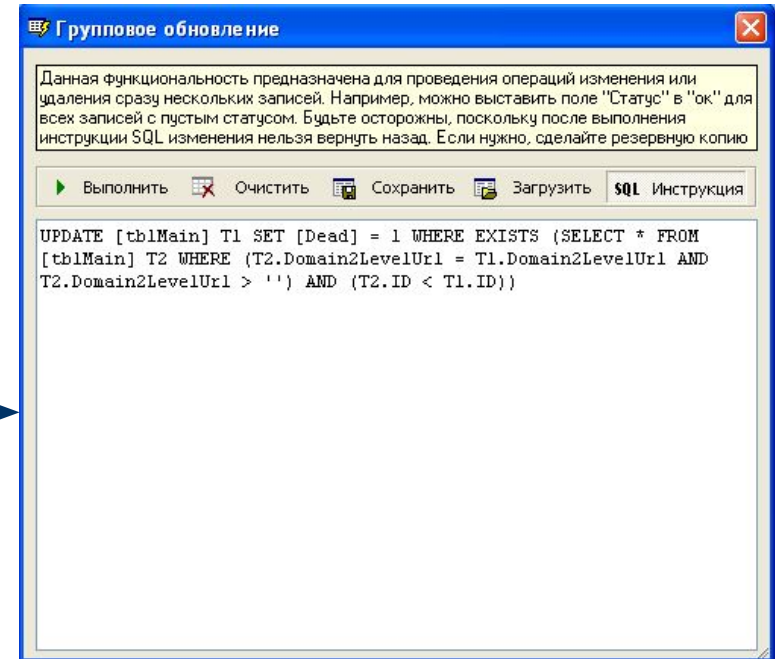
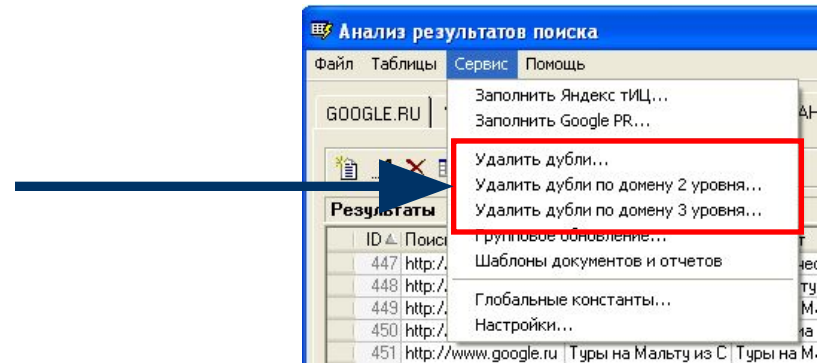
По окончании извлечения результатов будет полезно удалить дублированные результаты, оставляя в локальной базе данных только последнюю версию информации о сайте. Причем, есть возможность удаления дублей по полю **Домен 2 уровня** и по полю **Домен 3 уровня**.

Вы можете войти в **режим настройки инструкции SQL**, которая будет выполнена. Например, вы можете не удалять дублированные объявления, а только лишь отметить их как "Мертвые" (или как "Избранные"), вероятно, для последующего удаления вручную. В этом случае замените SQL-инструкцию с этой (здесь речь о доменах 2 уровня)

```
DELETE FROM [tblMain] T1 WHERE EXISTS (SELECT * FROM [tblMain] T2 WHERE (T2.Domain2LevelUrl = T1.Domain2LevelUrl AND T2.Domain2LevelUrl > ") AND (T2.ID < T1.ID))
```

на эту (легко заметить, что меняется только первая строчка инструкции)

```
UPDATE [tblMain] T1 SET [Dead] = 1 WHERE EXISTS (SELECT * FROM [tblMain] T2 WHERE (T2.Domain2LevelUrl = T1.Domain2LevelUrl AND T2.Domain2LevelUrl > ") AND (T2.ID < T1.ID))
```



**Если вы хотите установить
данную программу...**

Установка

Для установки программы на вашем компьютере необходимо запустить файл FindResults.msi, который входит в состав инсталляционного пакета.

Во время установки предлагается указать папку, в которую будет установлено приложение. По умолчанию это *C:\Program Files\FindResults*.

Минимальные требования к компьютеру:

- операционная система Windows 95, 98, NT, 2000, XP;
- процессор с частотой 100 MGz и более;
- оперативная память 16 Mb и более;
- жесткий диск (используется около 3Mb с пустой базой данных);

Техническая поддержка

Техническая поддержка осуществляется по электронной почте. Пожалуйста, пишите нам по адресу support@simple-soft.ru.

Полезным будет сообщить конфигурацию вашего компьютера и операционную систему (не забудьте указать язык), если у вас возникает системная ошибка, и другие уместные настройки.

Но, прежде всего, убедитесь, что у вас установлена последняя версия программы, посетив сайт программы <http://www.simple-soft.ru>

Ваше мнение может влиять на то, как будут выглядеть следующие версии программы. Если у вас есть идеи по улучшению программы, присылайте нам письма с предложениями.

С уважением,

Абрамов Иван Валерьевич
программист и руководитель компании "Простой софт"
Россия, Санкт-Петербург

Спасибо за внимание!