Подсистема аналитической обработки неструктурированной информации

Руководство пользователя

Москва, 2015 г.

ОГЛАВЛЕНИЕ

АННОТАЦИЯ ..... **- 4 -**

ВВЕДЕНИЕ ..... **- 5 -**

ГЛАВА 1. НАЗНАЧЕНИЕ И УСЛОВИЯ ПРИМЕНЕНИЯ СИСТЕМЫ ..... **- 7 -**

1.1. Общие сведения о Системе ..... - 7 -
1.2. Основные понятия. Терминология ..... - 8 -
1.3. Как работать с руководством ..... - 9 -

ГЛАВА 2. ПОДГОТОВКА К РАБОТЕ ..... **- 11 -**

2.1. Вход в Систему ..... - 11 -
2.2. Страница «Запрос» ..... - 13 -

ГЛАВА 3. ОПИСАНИЕ ОПЕРАЦИЙ, СВЯЗАННЫХ С ПРОВЕДЕНИЕМ ПОИСКА. ..... **- 19 -**

3.1. Что такое запрос? ..... - 19 -
3.2. Проведение запроса ..... - 20 -
3.3. Документ ..... - 21 -
3.4. Структура документа. Поля. ..... - 22 -
3.5. Результаты поиска: выборка документов. Навигация по документам. Страницы «Выборка» и «Документ». ..... - 23 -
3.6. Просмотр документов ..... - 27 -
3.7. Инфопортрет. Ведение диалога с Системой. Страница «Выборка» – «Инфопортет» ..... - 32 -

ГЛАВА 4. ОСНОВЫ ЭФФЕКТИВНОГО ПОИСКА ..... **- 37 -**

4.1. Уточнение запроса ..... - 37 -
4.2. Тематическое уточнение ..... - 42 -
4.3. Морфологический поиск ..... - 42 -
4.4. Страница «Выборка» – «Свойства» ..... - 43 -
4.5. Параллельная работа одного пользователя с несколькими выборками ..... - 46 -

ГЛАВА 5. ЯЗЫК ЗАПРОСОВ. СТРОГИЙ ЗАПРОС ..... **- 50 -**

5.1. Что такое язык запросов? ..... - 50 -
5.2. Запрос по полю документа ..... - 50 -
5.2.1. Общие правила составления запросов по полям документа ..... *- 52 -*
5.2.2. Запрос по дате создания документа (поле «date») ..... *- 52 -*
5.2.3. Запрос по номеру документа в базе (поле #) ..... *- 53 -*
5.2.4. Условные операторы ..... *- 53 -*
5.2.5. Запрос по источнику документа (поле «source») ..... *- 53 -*
5.2.6. Запрос по автору документа (поле «author») ..... *- 54 -*
5.2.7. Запрос по основному текстовому полю «message» (поле по умолчанию) ..... *- 55 -*
5.3. Позиционный запрос ..... - 55 -
5.4. Логические операторы ..... - 56 -
5.4.1. Логический оператор «И» ..... *- 56 -*
5.4.2. Логический оператор «ИЛИ» ..... *- 56 -*
5.4.3. Логическое отрицание «НЕ» ..... *- 57 -*
5.4.4. Логическое отрицание «И НЕ» ..... *- 57 -*
5.5. Запрос по части документа ..... - 57 -
5.6. Сложный запрос ..... - 58 -
5.7. Конструктор запросов. Страница «Запрос» – «Конструктор» ..... - 59 -
5.9 Синонимический поиск ..... - 67 -
5.10 Нечеткий поиск ..... - 68 -
5.11 Поиск телефонных номеров, фамилий, номеров документов ..... - 69 -
5.12 Поиск по объектам ..... - 73 -

ГЛАВА 6. МЕТОДИЧЕСКИЕ ПРИЕМЫ ПРОВЕДЕНИЯ ПОИСКА 77

6.1. Кластеризация 77
6.2. Папки. Страница «Запрос» – «Папка» 78
6.2.1. Назначение папок 78
6.2.2. Создание папок 79
6.2.3. Работа с папками 79
6.2.4. Сортировка выборки по папке или запросу 81

ГЛАВА 7. АНАЛИЗ 84

7.1. Общие положения. Структура аналитического исследования 84
7.2. Страница «Анализ» – «Таблица» 85
7.3. Работа с «важными» выборками 87
7.4. Сравнительный анализ. Страница «Анализ» – «Исследование» 89
7.5. «Публикационный рейтинг». Страница «Запрос» («Динамика по времени») 103
7.6. Динамический анализ главных тем. Страница «Анализ» («Динамика тем») 106
7.7. Разработка отчетного документа 108

ГЛАВА 8. РЕДАКТИРОВАНИЕ ДОКУМЕНТОВ 109

ГЛАВА 9. НАСТРОЙКИ 111

9.1. Основные настройки. Страница «Настройки» («Поиск») 111
9.2. Изменение настроек системы. Страница «Настройки» («Поиск») 112
9.3. Настройки параметров показа. Страница «Настройки» («Внешний вид») 113
9.4. Настройка статистики. Страница «Настройки» («Статистика») 119
9.5. Страница «Настройки» («Автоотсечение по рангам») 122

ГЛАВА 10. АВАРИЙНЫЕ СИТУАЦИИ 125

10.1. Включение ActiveX элементов 126

РЕКОМЕНДАЦИИ ПО ОСВОЕНИЮ 128

ПРИЛОЖЕНИЕ 143

АННОТАЦИЯ

Данный документ является **«Руководством пользователя»** и описывает работу ***RCO ZOOM v.4.0*** на уровне пользователя.

Он может быть использован как в качестве справочного руководства по данной системе, так и в качестве учебника.

Принцип, положенный в основу структурирования документа – описание системы от простого к сложному. В документе подробно рассматриваются возможности системы по поиску и анализу информации в текстовых базах через сети Intranet и Internet. В каждой главе содержатся примеры.

Алгоритмы функционирования системы, ее архитектура, требования к техническому оборудованию, а также ряд вопросов, связанных с загрузкой информации в систему и разграничением прав доступа, содержатся в **«Руководстве администратора»**.

ВВЕДЕНИЕ

В современных условиях увеличение интенсивности поступающих потоков информации и, как следствие, накопление огромных архивов текстовых документов требует применения новых средств организации доступа к данным. В больших информационных базах для ряда объектов возникает избыточность информации при попытках использования традиционных инструментов (рубрикаторы, поисковые системы).

Поэтому наибольшее предпочтение отдается системам, способным осуществлять поиск и обработку информации в считанные секунды.

Кроме того, главной потребностью пользователя является извлечение из информации пользы в виде знаний и точных решений.

К основным задачам пользователя относятся следующие:

1. Сбор информации об объектах (персонах, событиях, регионах, предприятиях, оргструктурах, социумах и пр.);
2. Определение свойств и функций с привязкой ко времени и месту, выделение главного;
3. Определение основных тенденций;
4. Определение связей объектов;
5. Выделение аномалий.

К вспомогательным (или инструментальным) задачам пользователя относятся:

1. Формулировка гипотез;
2. Проверка гипотез;
3. Разделение объектов;
4. Проверка решения на устойчивость, выделение инвариантов;
5. Работа в реальном времени. Ведь от скорости решения задач зависит актуальность рекомендаций, а, следовательно, и смысл самой работы по их решению.

Таким образом, одной из основных задач, возникающих при работе с большими текстовыми базами данных, является оперативный поиск необходимой информации. Существующие в настоящее время поисковые системы зачастую не обеспечивают требуемого качества и оперативности поиска.

Главная проблема качества поиска заключается в сложности подбора ключевых слов для формулировки запроса. Это может быть связано как с недостаточным владением терминологией

исследуемой предметной области, так и с наличием в языке многозначных и синонимичных слов. Кроме того, возможны орфографические ошибки в написании искомых слов.

Оперативность поиска зависит как от навыков и опыта самого пользователя, так и от сложности решаемой задачи.

Как отмечалось выше, существует широкий круг аналитических задач, требующих выявления скрытых взаимосвязей между объектами, определения ключевых факторов, влияющих на развитие ситуации, в том числе, задачи прогнозирования тенденций развития исследуемых процессов. Для таких задач необходима специализированная Система.

Система обладает уникальным инструментарием для проведения эффективного поиска и аналитических исследований информации в кратчайшие сроки.

Система позволяет работать с огромными массивами текстовой информации в реальном времени (объем базы – до сотен гигабайт, время обработки – секунды или минуты).

Средство отображения – *информационный портрет* дает возможность решать поставленные выше задачи (формулировать и проверять гипотезы, разделять объекты, выделять инварианты) в первом приближении.

При этом ситуация с информационной избыточностью полностью меняется: **чем больше информации, тем лучше**.

ГЛАВА 1. Назначение и условия применения Системы

1.1. Общие сведения о Системе

Система – это инструмент для сбора и создания корпоративного хранилища текстовой информации, обладающий уникальными возможностями для проведения оперативного поиска и аналитических исследований.

Система предназначена для работы в локальной сети организации на базе MS Windows и MS Internet Information Server. Работа пользователей осуществляется при помощи Web-браузера. Таким образом, может быть реализован удаленный доступ к Системе.

Система обеспечивает:

- ✓ *сбор* текстовой информации из открытых источников (электронных СМИ, печатных материалов, почтовых серверов и сайтов Internet) и интегрирование ее в БД организации;
- ✓ *структурирование* и *хранение* текстовой информации организации (создание унифицированной БД);
- ✓ эффективный *поиск* документов как в собственной базе данных организации, так и в сети Internet;
- ✓ автоматическая *обработка* документов (Система самостоятельно определяет ключевые темы документов, тем самым, предлагая варианты для проведения все более направленного поиска и уточнения предметной области);
- ✓ *анализ* собранной информации (определение тенденций развития исследуемых процессов, решение задач прогнозирования, многообъектного анализа);
- ✓ подготовка отчетов.

Система предусматривает широкие возможности для проведения поиска и анализа информации.

Поиск:

- по ключевым словам,

- по фрагменту документа,
- в рамках предложения или абзаца,
- с учетом морфологии,
- по дате или промежутку времени,
- по источнику документа,
- по номеру документа,
- по автору документа,
- по близости слов,
- по названию документа.

Анализ:

- формирование *«информационного портрета»* – множества значимых слов и словосочетаний, которые представляют собой главные темы совокупности найденных документов,
- создание автореферата документа – определение наиболее значимого фрагмента в тексте,
- сравнительный анализ – выявление связей между объектами, закономерностей их развития, а также определение индивидуальных особенностей объекта,
- динамический анализ – изучение развития объекта во времени по интересующим темам,
- объектный анализ – изучение нескольких объектов по одним и тем же темам.

Функциональные возможности Системы, оптимизирующие работу пользователей:

- автоматическое определение главных тем в документах,
- сортировка найденных документов по дате, значимости, в соответствии с заданным фильтром,
- авторубрикация – сортировка найденных документов по тематическим папкам или по «информационному портрету».

1.2. Основные понятия. Терминология

Перед началом работы с Системой необходимо изучить терминологию, используемую в Системе и настоящем руководстве.

Ключевым понятием Системы является понятие ***документа***.

В отличие от традиционного представления о бумажном документе, как о деловой бумаге, служащей подтверждением какого-либо факта или права на что-либо, в информационных технологиях под ***документом*** понимают единицу представления информации, создаваемую с помощью специальных программных средств.

Разумеется, это могут быть письма, отчеты, счета, схемы, книги и т. п., то есть все, что мы привыкли видеть на бумаге.

Документ **–** единица хранения и представления информации в Системе.

(Словарь терминов теории информации)

Информационная база (в дальнейшем просто *«база»*) — структурированное хранилище информации (документов), объемом свыше десятка гигабайт. Система способна работать с несколькими информационными базами, если они предварительно объединены в группу.

Выборка — набор (список) *документов*, соответствующих заданным критериям отбора (условиям запроса).

Главная тема — слово или словосочетание, значимое для данной *выборки*. *Выборка* может иметь множество главных тем. Каждая тема имеет свой *рейтинг*.

Запрос — логическая конструкция, записанная по специальным правилам языка запросов, задающая критерий отбора документов из базы. Запрос может представлять собой слово, последовательность слов или логическое выражение, составленное в соответствии с правилами языка запросов.

Значимость документа — величина, характеризующая типичность (представительность) *документа* в *выборке*. Более представительным (типичным) считается *документ*, наиболее соответствующий *информационному портрету* (инфопортрету) *выборки*.

Имя базы — название, которое присваивается администратором *базе* при ее создании.

Имя пользователя — название, которое присваивается пользователю или группе пользователей администратором для обеспечения доступа в Систему.

Информационный портрет — множество *главных тем* (слов и словосочетаний), отражающих информационное содержание объекта, т.е. характеризующих его.

Проведение запроса — получение *выборки* документов, удовлетворяющей условиям *запроса*.

Сортировка (ранжирование) — упорядочение списка *документов*, полученного по *запросу*, в соответствии с заданным условием.

Рейтинг – уровень *значимости* темы, представленный в виде числового значения.

1.3. Как работать с руководством

Руководство предназначено как для тех, кто впервые знакомится с Системой, так и для пользователей, желающих узнать о новых возможностях Системы. Прочитав руководство, Вы сможете извлечь максимальную пользу при работе с Системой.

Каждая последующая глава подготовлена на основе материала предыдущих. Кроме того, используется один и тот же пример, который постепенно и последовательно усложняется. Поэтому рекомендуем Вам придерживаться последовательности изложения, предложенной руководством.

Примечание

Все примеры, приведенные в руководстве, построены с использованием тестовой базы Системы. Тестовая база представляет собой базу деловых изданий российской прессы.

ГЛАВА 2. Подготовка к работе

В этой главе мы подробно опишем первые шаги в освоении Системы, детально рассмотрим окна, с которых всегда будет начинаться работа, а также основные операции, которые необходимо выполнять для входа в Систему и проведения запроса.

2.1. Вход в Систему

Доступ к Системе осуществляется через Web-браузер. Поэтому для того чтобы войти в Систему, необходимо сначала открыть окно Web-браузера. Для этого нажмите кнопку *«Пуск»* на панели задач, выберите *«Программы»*, затем, в открывшемся списке программ найдите *Internet Explorer* и щелкните на значке . Другой способ запуска браузера: найдите на рабочем столе ярлык *Internet Explorer* и дважды на нем щелкните. Откроется окно Web-браузера.

Примечание

Основным техническим требованием для обеспечения доступа к Системе является установка на рабочем месте Internet Explorer версии 6.0 или выше.

Затем в адресной строке введите соответствующий адрес и щелкните Enter.

Примечание

Уточните адрес у системного администратора Вашей организации.

Перед вами откроется станица *«Вход в систему ZOOM» (*Рисунок 1*)*

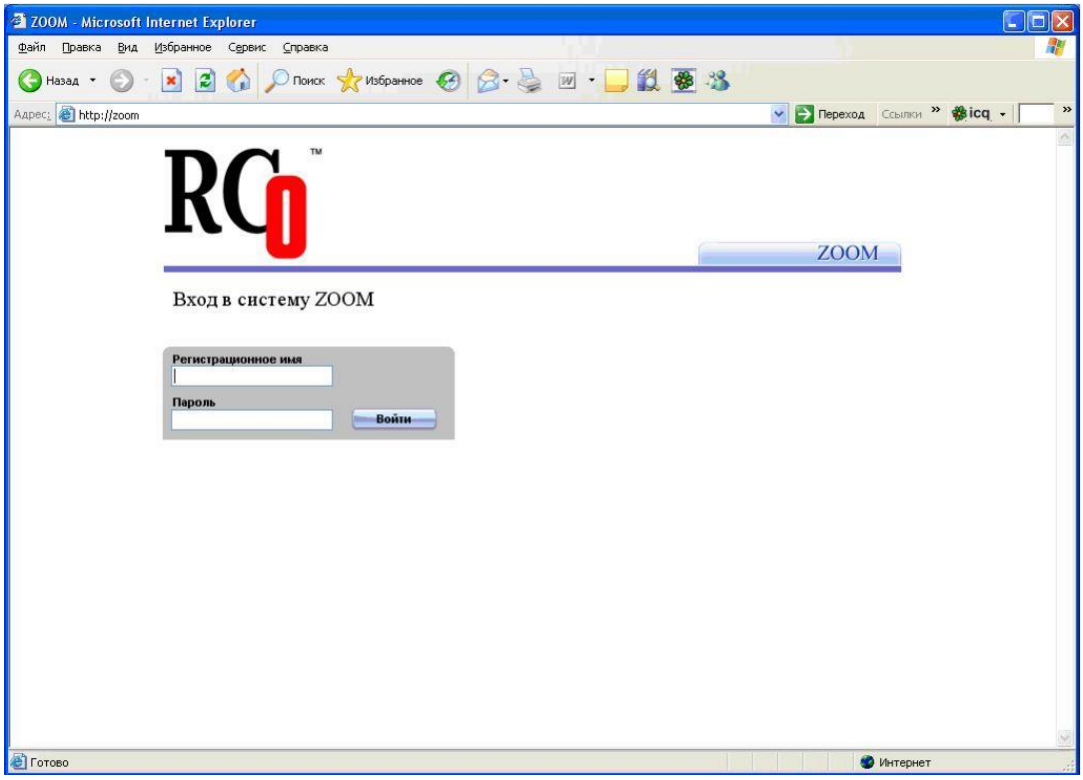


Рисунок 1. *Страница «Вход в систему ZOOM»*

В строке *«Регистрационное имя»* введите имя, присвоенное Вам администратором системы, а в строке *«Пароль»* — соответствующий пароль. Для первого входа в систему Вам потребуется только регистрационное имя.

Примечание

Регистрационное имя узнайте у администратора системы! Пароль заведите самостоятельно на странице *«Настройки» — «Внешний вид»* (см. ниже и п. 7.9.3).

Примечание

При первом входе в систему по регистрационному имени, Вам нужно завести свой пароль. Для этого щелкните по вкладке *«Настройки»*. Таким образом, Вы перейдете на страницу настроек. Вверху страницы располагается меню. Перейдите в раздел *«Внешний вид»*, щелкнув по соответствующему пункту меню. Внизу открывшегося раздела находится закладка *«Изменение пароля»*. Поле *«Старый пароль»* оставьте пустым, в поле *«Новый пароль»* введите придуманный Вами пароль для последующих входов в систему. Рекомендуем заводить пароли длиной не менее 6 символов, использовать цифры, знаки пунктуации и различный регистр букв. Не забудьте свой пароль, записывать пароли не рекомендуется. В поле *«Подтвердите новый пароль»* повторите ввод пароля и нажмите кнопку *«Изменить»*.

После того, как Вы ввели регистрационное имя и пароль (или только регистрационное имя при первичном входе в систему), нажмите кнопку *«Войти»*.

Если Вы сделали ошибку в регистрационном имени или пароле, на экране появится сообщение об ошибке (Рисунок 2). В этом случае повторите ввод еще раз.

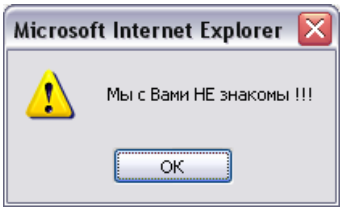


Рисунок 2. *Сообщение об ошибке*

Если вы забыли пароль, обратитесь к администратору системы, он обеспечит Вам доступ, затем Вы сможете завести новый пароль!

2.2. Страница «Запрос»

После успешной регистрации при входе в систему, перед Вами откроется страница *«Запрос»* (Рисунок 3). При повторном входе в систему открывается последний проведенный Вами запрос (Рисунок 4).

Кроме того, вверху страницы располагается навигационная панель для перехода на другие страницы.

Под меню и над словом «Запрос» для уведомления пользователей о наличии проблемных баз выводится сообщение «Внимание! Следующие базы недоступны для поиска:», под которым следует список проблемных баз. Если проблемных баз нет, то сообщение не выводится.

Рассмотрим каждую из вкладок навигационной панели:

- ✓ ***«Запрос»*** — с данной страницы проводятся все поисковые запросы;
- ✓ ***«Выборка»*** — страница предназначена для просмотра текущей выборки документов, полученной в результате проведения запроса;
- ✓ ***«Документ»*** — страница предназначена для просмотра текущего документа (при щелчке мышью по ссылке на документ система автоматически перемещается на данную страницу);

Сразу после входа в систему вкладки *«Выборка»* и *«Документ»* оказываются недоступными. Для того чтобы они стали доступны, необходимо провести запрос.

- ✓ ***«Анализ»*** — на данной странице осуществляются операции над выборками;

- ✓ ***«Настройки»*** — на данной странице можно устанавливать индивидуальные параметры поиска, интерфейса, а также параметры расчета статистики в зависимости от требуемой точности;
- ✓ ***«Помощь»*** — на данной странице пользователь может получить необходимую справочную информацию для работы.

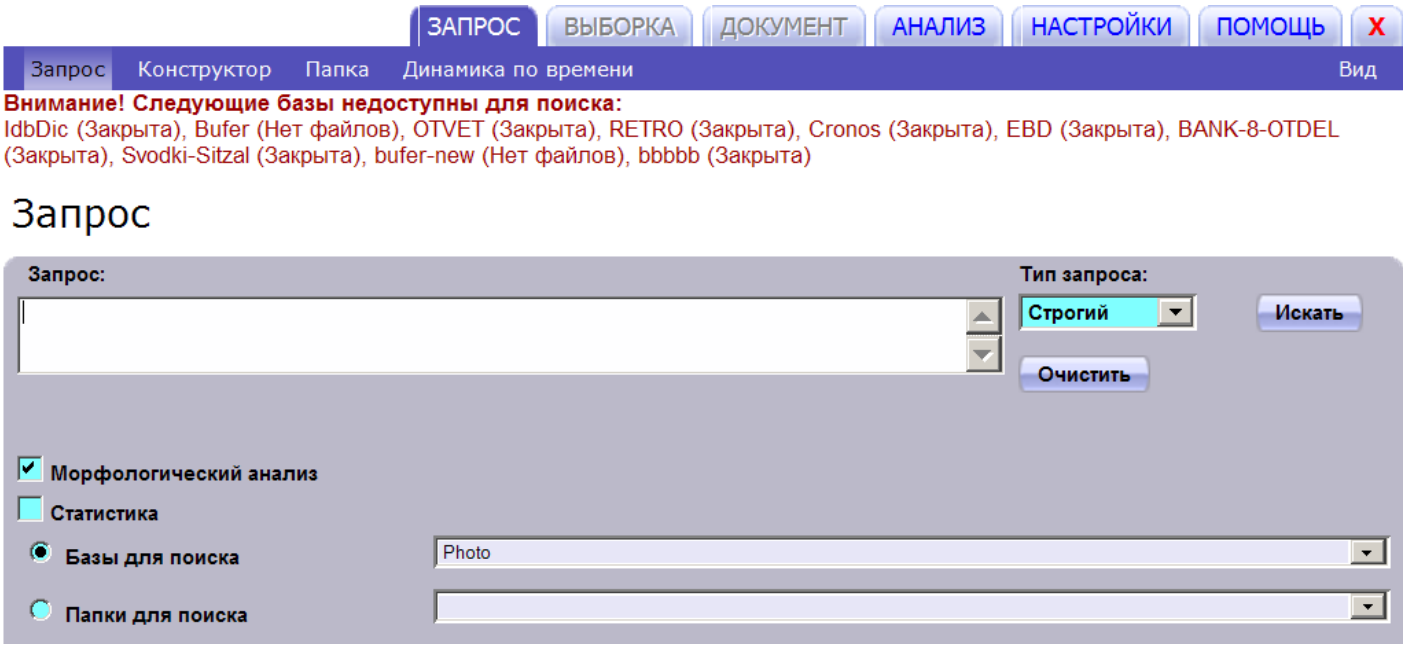


Рисунок 3. *Страница «Запрос»*

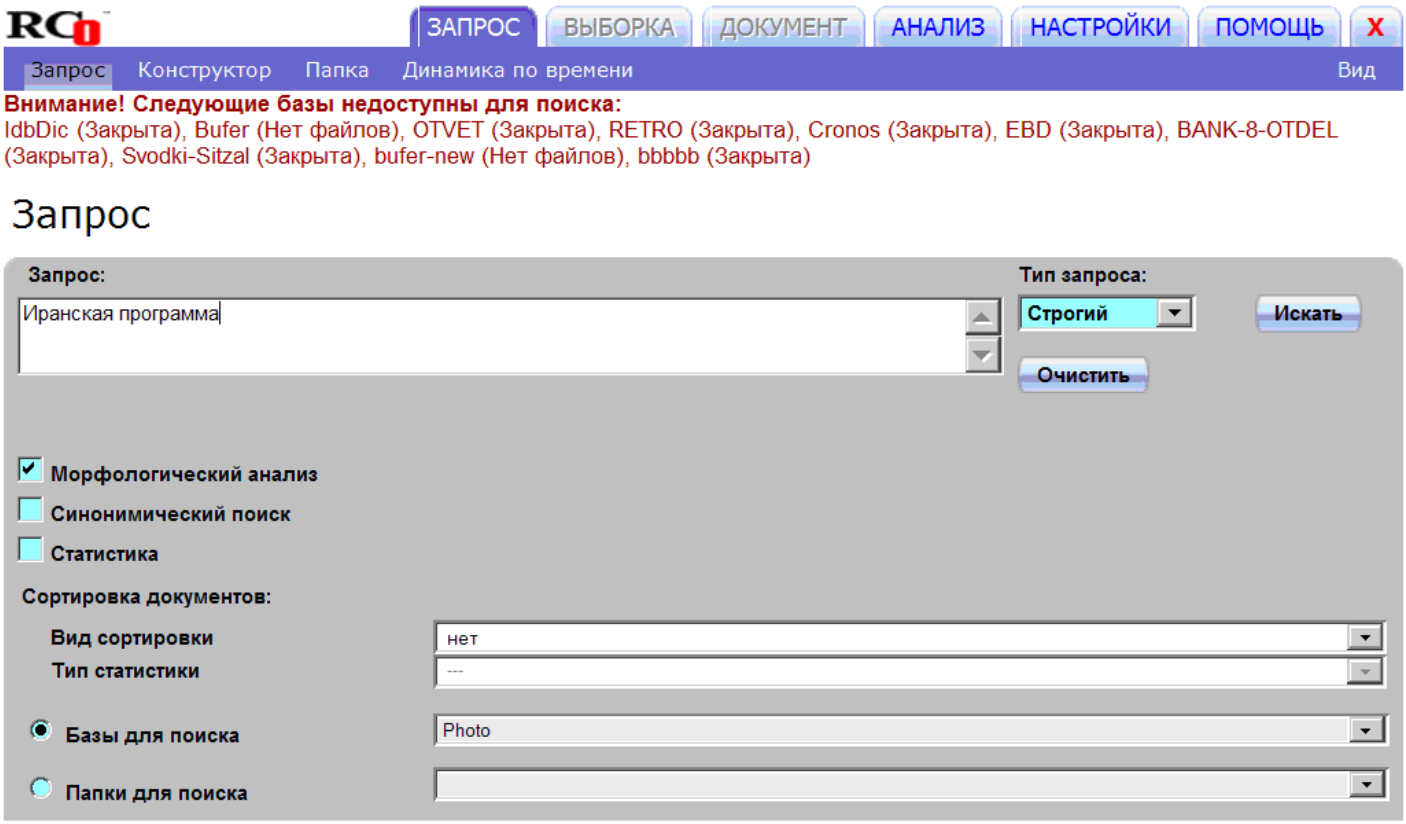


Рисунок 4. *При повторном входе в систему показывается последний запрос*

Итак, начинаем работу со страницы *«Запрос»* (Рисунок 3). Именно с этой страницы рекомендуем проводить все запросы.

Вверху страницы располагаются пункты меню:

- ✓ ***«Запрос»*** — в данном разделе проводятся поисковые запросы;
- ✓ ***«Конструктор»*** — раздел представляет собой упрощенную форму ввода сложного запроса;
- ✓ ***«Папка»*** — раздел для работы с папками;
- ✓ ***«Динамика по времени»*** — раздел для построения «публикационного рейтинга» объекта;
- ✓ ***«Вид»*** — с помощью данной формы можно устанавливать индивидуальные настройки интерфейса разделов.

Внизу страницы располагается дубликат навигационной панели для удобства работы.

Вообще вид страницы *«Запрос»* существенно зависит от установок в форме «Вид» (Рисунок 5). В минимальном варианте (все галочки сброшены) страница *«Запрос»* выглядит так (Рисунок 6)

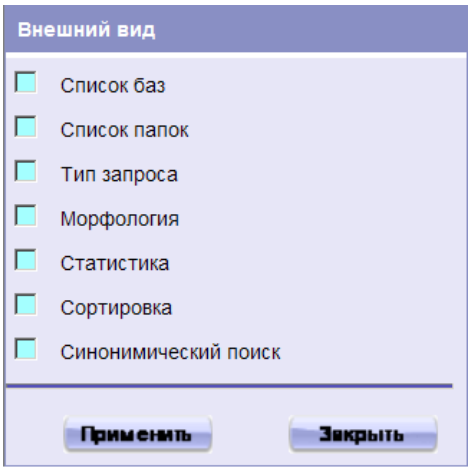


Рисунок 5. *Панель «Внешний вид»*

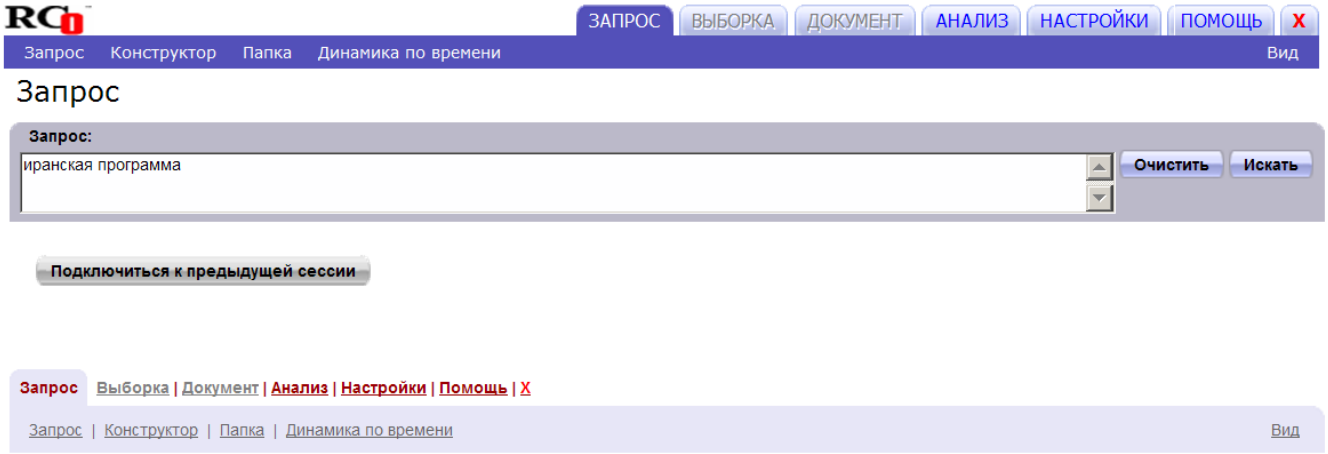


Рисунок 6. *Минимизированная панель «Запрос»*

Переходим в раздел *«Запрос»*. В центре страницы располагается поле, в которое нужно вводить текст *запроса* — слово, последовательность слов или логическое выражение, задающее критерий поиска.

Рассмотрим дополнительные элементы раздела *«Запрос»*, которые могут отображаться в окне, если установлены соответствующие параметры в разделе *«Вид»*.

Поле «Тип запроса»:

- ✓ ***строгий*** — для проведения сложного запроса с использованием элементов языка запросов (языку запросов посвящена «ГЛАВА 3. Описание операций, связанных с проведением поиска.»);
- ✓ ***простой*** — для проведения поиска по одному слову или нескольким словам, записанным через пробел.

Поле «Морфологический анализ»:

- ✓ ***вкл. (стоит флажок)*** – морфология включена, т.е. поиск будет осуществляться по всем возможным формам слов, указанным в запросе. Например, по запросу «банк» найдутся все его словоформы, в том числе «банка», «банки», «банками» и пр. Если флаг морфологии включен, он действует на все текстовые поля, указанные в запросе;
- ✓ ***выкл.*** – морфология выключена.

Поле «Синонимический поиск»:

- ✓ ***вкл. (стоит флажок)*** – поиск будет осуществляться по синонимам слов, указанным в запросе. Например, по запросу «хороший» найдутся все его синонимы, в том числе «лучше», «лучший», «добрый» и пр. Реально набор синонимов зависит от предметной области, к которой относится база а также от настроек, сделанных администратором в файле синонимов (подробнее см. *5.9 Синонимический поиск*);
- ✓ ***выкл.*** – поиск синонимов выключен.

Поле «Статистика»:

- ✓ ***вкл. (стоит флажок)*** – статистика включена, т.е. в процессе поиска будет проводиться анализ документов и строиться таблица – подсказка;
- ✓ ***выкл.*** – статистика выключена.

Поле «Вид сортировки»:

- ✓ ***нет*** – функция отключена, документы в списке будут расположены в том порядке, в котором они находятся в базе;
- ✓ ***по дате*** – в соответствии с датой документа, список найденных по запросу документов возглавляет самый новый документ;
- ✓ ***по значимости*** – документы в списке будут расположены в порядке уменьшения значимости, список найденных по запросу документов возглавит документ, наиболее близкий к информационному портрету выборки;
- ✓ ***по имени файла*** – документы в списке располагаются в алфавитном порядке имён исходных файлов.

Поле «Тип статистики»:

- ✓ ***слова*** – в списке Главных тем будут отображены значимые слова;

- ✓ ***слова и пары*** – в списке Главных тем будут отображены значимые слова и словосочетания.

Данное поле является доступным при установке ***«Вид сортировки» «по значимости»***.

Поле «Базы для поиска»: (например, тестовые базы СМИ)

- ✓ ***DemoBase***;
- ✓ ***SBase***.

Поле «Папки для поиска»: (Рисунок 7)

- ✓ Предъявляются папки, созданные пользователем;
- ✓ ***Просмотрено в сессии*** – папка, хранящая документы, просмотренные за текущую сессию.
- ✓ ***Просмотрено всего*** – папка, хранящая документы, просмотренные пользователем со времени создания папки (при первом входе создаётся автоматически).

Кнопка «Очистить» удаляет весь текст из окна запроса.

Кнопка «Искать» запускает поисковую машину Системы.

Рисунок 7. *При выборе папок для поиска предъявляются существующие папки*

ГЛАВА 3. Описание операций, связанных с проведением поиска.

Прочитав эту главу, Вы научитесь проводить запрос, просматривать документы, перемещаться по списку документов; узнаете, что такое информационный портрет объекта и как он помогает в поиске необходимой информации.

3.1. Что такое запрос?

Перед началом работы с Системой всегда необходимо четко поставить задачу, сформулировать цель. Иными словами, Вы всегда должны знать что искать и для чего Вам это нужно.

Составление запроса к базе является важнейшим моментом в процедуре поиска нужных документов. Орфографическая ошибка в запросе или его неверная формулировка приведет в конечном итоге к увеличению времени поиска информации. Для того чтобы проводить эффективный и оперативный поиск, необходимо сначала продумать структуру запроса, перечень ключевых слов, т.к. Система ориентируется, прежде всего, на текст запроса и осуществляет поиск в соответствии с ним.

Запрос – это «вопрос» к Системе, на который Система должна дать ответ в виде выборки документов, удовлетворяющих условиям данного запроса, и информационного портрета данной выборки.

Например, запрос *выборы* можно интерпретировать как вопрос к Системе: есть ли в базе документы, в тексте которых встречается слово *выборы*? Итак, Система проверяет (производит поиск) базу на наличие таких документов. В случае если такие документы в базе присутствуют, Система дает положительный ответ в виде *списка* соответствующих документов.

Запрос занимает ключевое место в общей структуре любого исследования, которое можно проводить с помощью Системы *(см. схему)*.

3.2. Проведение запроса

Рассмотрим наглядный пример того, как проводить поиск информации.

На странице *«Запрос»* (раздел *«Запрос»)* в поле *«Тип запроса»* установите *Простой*, это означает, что Ваш запрос будет состоять из одного или нескольких слов, записанных через пробел.

В Системе также существует *строгий* запрос *(«Тип запроса» – строгий*), составлений по специальным правилам *языка запросов*. (Об этом типе запроса более подробно будет рассказано ниже: ГЛАВА 5. Язык запросов. Строгий запрос).

Примечание

Система не различает строчных и прописных букв, запросы можно набирать в любом регистре.

Примечание

При проведении *простого* запроса символы ***~ ! @ # $ % ^ & * ( ) _ - + = { } [ ] : ; \ “ ‘ | \ ? / , . < >*** употреблять не следует, так как они заменяются на пробелы.

В поле *«Запрос»* введите текст запроса, например, *иранская программа*.

Мы будем исследовать информационный образ ядерной проблемы Ирана, развитие его во времени, образы персон, причастных к данной проблеме.

В поле *«Сортировка документов»* установите *«Вид сортировки» по значимости*, *«Тип статистики» – Слова и пары*, *«Морфологический анализ»* включен, *«Статистика»* включена (Рисунок 8).

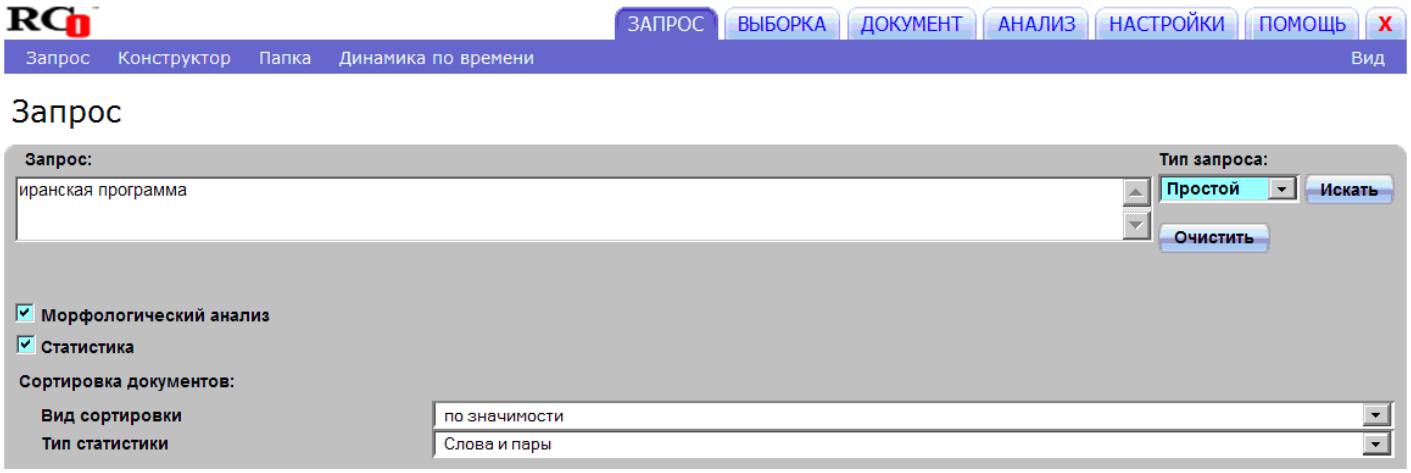


Рисунок 8. *Пример проведения поиска по запросу «Иранская программа»*

После установки основных параметров, нажмите кнопку *«Искать»*. Таким образом, Вы запустите поисковую машину Системы. Появится панель с вращающимся диском, символизирующая ход поиска (Рисунок 9).

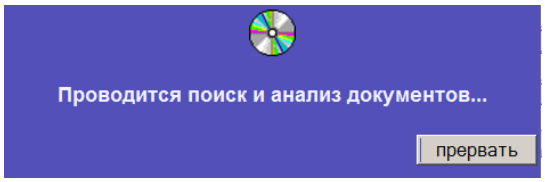


Рисунок 9. *Временная форма поиска*

Примечание

Если поиск длится очень долго (например, задан сложный запрос), есть возможность его прервать, для чего и служит кнопка прервать.

3.3. Документ

Поиск в базе осуществляется по документам. Документы бывают одностраничными и многостраничными. Вместе с документом может быть загружен исходный документ («оригинал»). В восстановленном документе, в отличие от оригинала, будет отсутствовать исходная разметка, но будут выделены цветом ключевые и значимые слова запроса. Просмотреть оригинал и восстановленный текст можно на странице «Документ» (вкладка «Показать текст», «Показать оригинал»). Если оригинал не был загружен, то внешний вид документа не меняется.

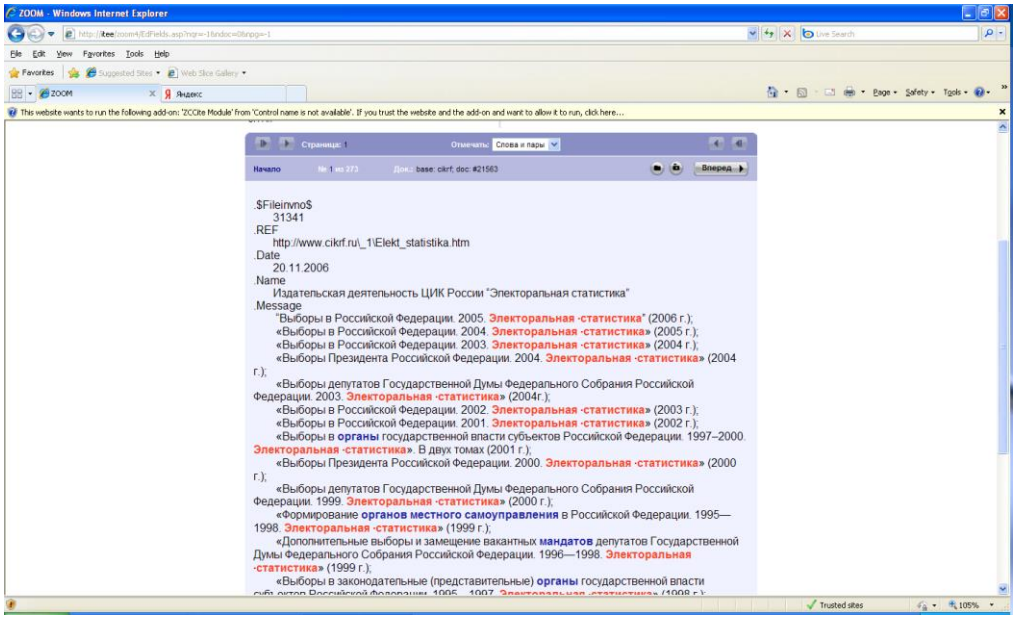


Рисунок 10. *Восстановленный текст*

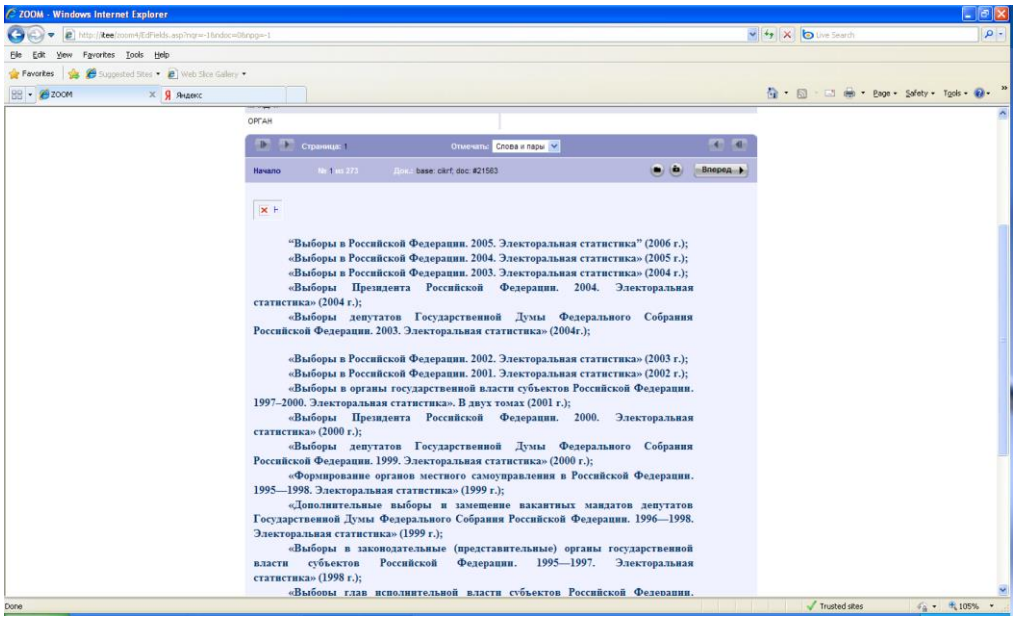


Рисунок 11. *Оригинал*

3.4. Структура документа. Поля.

Все документы базы имеют единую структуру. Они, как правило, разбиты на поля *(В таблице 1* – типичная структура для базы СМИ*)*. Каждое поле имеет текстовое имя и определенный тип хранящейся в нём информации.

В Системе возможны следующие типы полей:

Число — в данном поле можно хранить любые числа;

Дата — поле даты, чаще всего в данном поле хранятся даты создания документов;

Текст — содержательная часть, поле автора и т.д.

Поле	Имя поля	Тип поля
уникальный номер документа	doc: #	число
поле даты	Date	дата
поле автора	author	текст
поле источника	Source	текст
основное - текстовое поле (поле по умолчанию)	message	текст, число

Таблица 1. Поля документа

При этом некоторые поля в отдельных документах могут быть не заполнены.

Каждый элемент запроса относится к конкретному полю. Если же поле не указано, то этот элемент относится к *полю по умолчанию*, т.е. к основному текстовому полю.

Примечание

Создание списка полей базы документов и регистрация *«поля по умолчанию»* производятся администратором Системы при формировании базы.

3.5. Результаты поиска: выборка документов. Навигация по документам. Страницы «Выборка» и «Документ».

После проведения поиска система автоматически перешла на страницу *«Выборка»* (раздел *«Документы»*). Таким образом, мы получили выборку документов, удовлетворяющую условиям запроса *иранская программа*. Это означает, что во всех документах выборки в тексте содержатся слова *иранская программа*, причем они могут встречаться не только вместе, но и порознь (Рисунок 12, Рисунок 13)

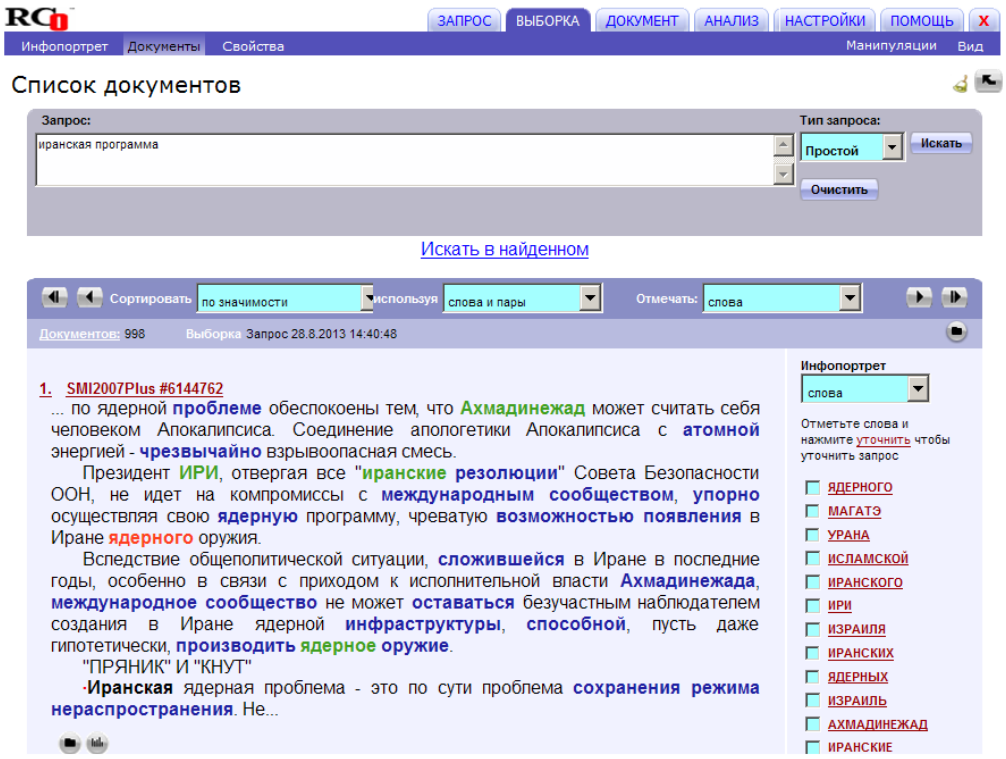


Рисунок 12. *Результаты поиска по запросу «Иранская программа»*

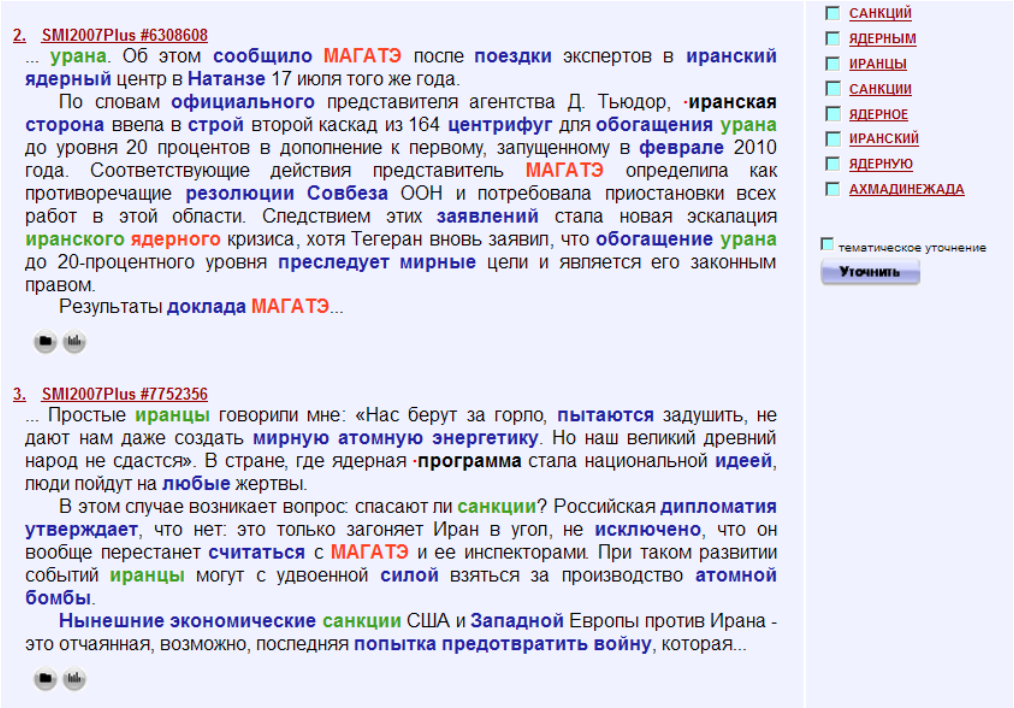


Рисунок 13. *Результаты поиска по запросу «Иранская программа»*

На странице *«Выборка»* над списком документов находится панель, на которой располагаются следующие поля:

- *«сортировать» – по дате, по значимости, по папкам, по имени файла* – в данном поле можно установить требуемый способ сортировки документов;
- *«используя» – слова, слова и пары* – поле доступно в случае, если выбран способ сортировки документов по значимости или по папкам. При установке параметра в данном поле сортировка документов будет проводиться по степени соответствия *слов* или *слов и пар слов* инфопортрету выборки (инфопортрету папки);
- *«отмечать» – слова, пары слов, слова и пары слов* – в данном поле можно установить требуемый параметр для разметки значимых *слов*/*пар слов*/*слов и пар слов* в документе.

Кроме того, на данной панели содержится информация о текущей выборке: количество документов, название выборки, а также кнопки навигации . С помощью кнопок навигации можно перемещаться по выборке документов.

– в начало/конец выборки;

– показать предыдущую/следующую страницу. При этом количество авторефератов документов, отображаемых на одной странице, устанавливается на странице *«Настройки» – «Внешний вид»*.

Для просмотра документа щелкните по ссылке, в которой указаны источник и номер публикации, система автоматически перейдет на страницу *«Документ»* – раздел *«Страница»* – откроется восстановленный документ (первая страница документа) *(*Рисунок 19*)*. Если нажать на ссылку с номером документа, он откроется для просмотра в отдельном окне.

В списке документов выборки под каждой ссылкой располагается наиболее значимый по смыслу фрагмент документа, который Система определяет автоматически. Таким образом, Система позволяет быстро ориентироваться в большом массиве найденных по запросу документов.

Под фрагментом документа в выборке располагается ряд кнопок (Рисунок 14). Количество кнопок зависит от характеристик документа, из которого взят вышележащий фрагмент.

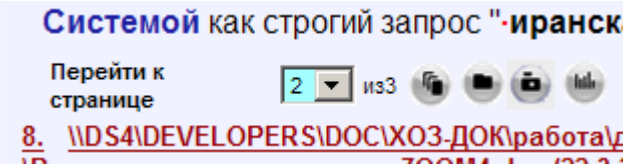


Рисунок 14. *Кнопки после фрагмента документа*

Рассмотрим их подробнее:

– позволяет перейти к указанной странице многостраничного документа. Для этого выбираем нужную нам страницу и нажимаем на ссылку заголовка документа. Кнопка появляется после многостраничных документов;

– позволяет просмотреть многостраничный документ целиком. Кнопка появляется после многостраничных документов;

– помещает документ в папку, выбранную из появляющегося при нажатии списка;

– показывает восстановленный документ. Кнопка появляется, если в архиве есть исходный документ в формате, отображаемом Системой;

– показывает ранг (число, связанное с порядком сортировки; чем он больше, тем раньше появляется документ) в выборке.

Главное меню страницы *«Выборка»* содержит следующие разделы:

1. *«Инфопортрет»* – переход на страницу просмотра инфопортрета (см. п. 3.7 *Инфопортрет. Ведение диалога с Системой. Страница «Выборка» – «Инфопортет»*);
2. *«Документы»* – отображает список документов, найденных по запросу;
3. *«Свойства»* – переход на страницу свойств (см. п. 4.4 *Страница «Выборка» – «Свойства»*);
4. *«Манипуляции»* – содержит следующие функции (Рисунок 15):

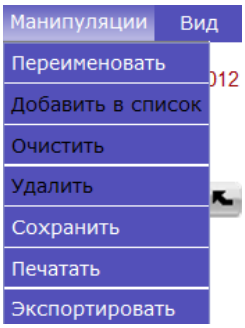


Рисунок 15. *Меню «Манипуляции»*

- ✓ ***Переименовать*** – позволяет переименовать выборку;
- ✓ ***Добавить в список*** – позволяет сделать временный запрос постоянным. Эта функция доступна для временных запросов;
- ✓ ***Очистить*** – удаляет все документы из папки (Доступно при работе с папками);
- ✓ ***Удалить*** – удаляет из папки отмеченные документы (Доступно при работе с папками);
- ✓ ***Сохранить*** – служит для сохранения документов. Появляется окно «Сохранение документов», в котором можно установить параметры сохранения:

- количество документов;
- оригинальный или восстановленный текст;
- каталог для сохранения (в том числе есть возможность создания нового каталога).

- ✓ ***Печатать*** – появляется окно «Печать документов» (Рисунок 16), в котором можно установить следующие параметры:

- количество документов для печати;
- оригинальный текст или список документов.

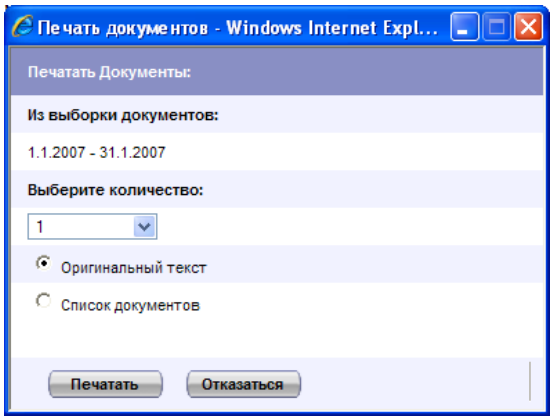


Рисунок 16. *Окно «Печать документов»*

- ✓ ***Экспортировать*** – появляется окно «Экспорт в …», позволяющее вывести документы на сервер или на машину пользователя.

5. *«Вид»* – показывает окно *«Внешний вид»* (Рисунок 18), содержащее следующие поля:

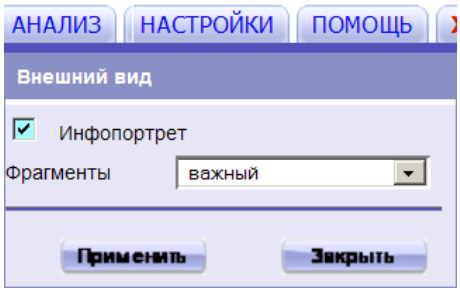


Рисунок 17. *Окно «Внешний вид»*

- ✓ ***Инфопортрет*** **–** инфопортрет отображается, если поставлена галочка. Вид инфопортрета зависит от того, как он был включён. Возможны варианты ():

  Не показывать – инфопортрет не отображается;

  Слова – инфопортрет будет состоять из отдельных слов;

  Пары слов – инфопортрет будет состоять из пар ключевых слов;

  Слова и пары – инфопортрет будет включать и отдельные слова и пары слов.

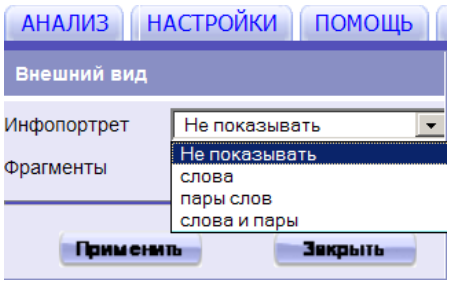


Рисунок 18. *Окно «Внешний вид»*

- ✓ ***Фрагменты*** – задаёт форму отображения выборки:

  --- – в списке документов отображаются только заголовки;

  важный – под заголовком будет отображаться наиболее значимый фрагмент документа;

  существенные – под заголовком будут отображаться несколько фрагментов, содержащих найденные слова (контекстные фрагменты).

После выбора параметров нажмите кнопку *«Применить»*.

3.6. Просмотр документов

Для перехода со страницы *«Выборка»* к просмотру документа щелкните по ссылке, в которой указаны источник и номер публикации, система автоматически перейдет на страницу *«Документ»* –

раздел *«Страница»* – откроется восстановленный документ (первая страница документа) (Рисунок 19). Если нажать на ссылку с номером документа, он откроется для просмотра в отдельном окне.

Рисунок 19. *Навигация по документам. Кнопки навигации. Список главных тем отключен*

Перед текстом расположен список *главных тем* данного документа, по которому можно определить, о чем пойдет речь в самом тексте документа (см Рисунок 21). Аналогично *автореферированию* (выделению значимых фрагментов), список главных тем помогает ориентироваться в тексте документа, минимизируя время на его прочтение. В разделе *«Вид»* можно подключить/отключить отображение списка главных тем документа (см. *п. 3.5*). После включения в разделе «Вид» опции «Показывать инфопортрет» под заголовком «Документ» появится строка, управляющая представлением инфопортрета (см. Рисунок 20). В окне «тип» выберите один из вариантов представления инфопортрета:

Рисунок 20. *Настройка представления инфопортрета*

Не показывать – означает отказ от предъявления инфопортрета;

Слова – в списке будут показаны ключевые слова;

Словосочетания – список будет состоять из ключевых словосочетаний;

Объединённый список – появятся и ключевые слова и словосочетания единым списком;

Отдельные списки – список слов будет дан в левом столбце, словосочетания – в правом (см. Рисунок 21).

Окошко «макс.стр.» отвечает за количество строк, занимаемых инфопортретом. Может принимать значения от 0 (инфопортрета не будет, только строка «Главные темы документа» (Рисунок 20)) до 25 строк.

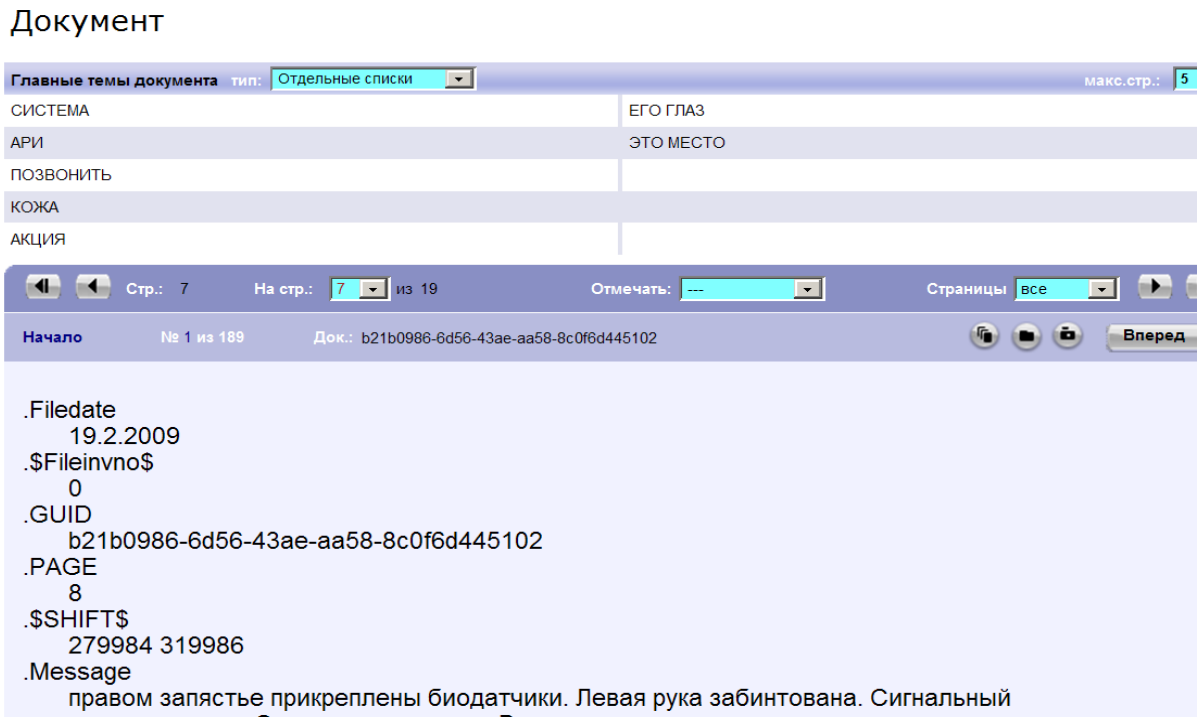


Рисунок 21. *Список главных тем документа*

Вверху и внизу страницы находятся кнопки навигации (перемещения по документам): Вперед (*Следующий документ)* и Назад (*Предыдущий документ)* (Рисунок 22). С помощью этих кнопок можно перемещаться по документам. При просмотре длинных документов или выборки можно вернуться к началу при помощи кнопки К началу ↑, возникающей слева/справа от текста (Рисунок 23). О навигации *по словам* см. в конце этого раздела.

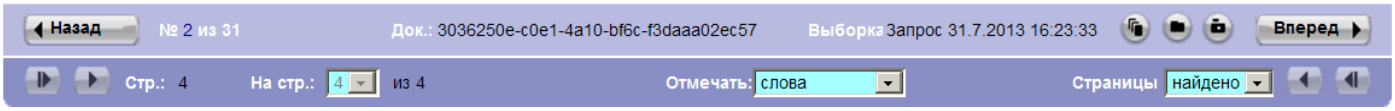


Рисунок 22. *Кнопки перемещения по документам*

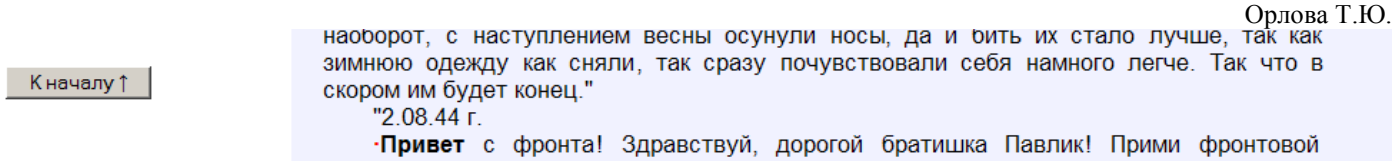


Рисунок 23. *Кнопка возврата к началу документа/выборки*

В разделе *«Весь документ»* содержится весь текст данного документа. Этот раздел является доступным в том случае, если документ многостраничный. Если в многостраничном документе искомое слово встречается на нескольких страницах, то на странице выборки предъявляется первая, в которой оно найдено, а под текстом документа появляется ссылка, в которой указано, что документ содержит искомое на других своих страницах (Рисунок 24).

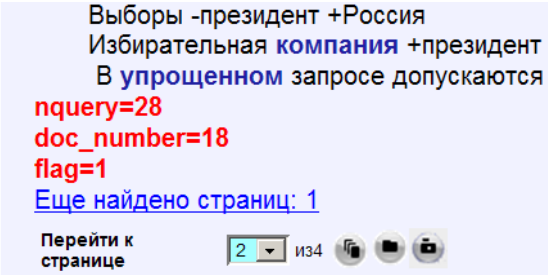


Рисунок 24. *Есть искомое слово на других страницах*

При нажатии на ссылку Еще найдено страниц открывается страница специального режима просмотра – отображение только найденных страниц. Отображаются сразу все страницы, содержащие поисковые слова, т.е. если всего в документе 20 страниц, а поисковые слова есть на 3 страницах, то будут показаны 3 страницы. Друг от друга страницы отделяются горизонтальными полосами с номером страницы (Рисунок 25).

Рисунок 25. *Разделитель страниц*

Раздел «Документ» – *«Показать текст»/«Показать оригинал»* переключает режим отображения документа в виде восстановленного текста или оригинала (см. п.3.3 *Документ*). Оригинал в основном окне отображается только в том случае, если он имеет соответствующий формат (текстовый, html или изображение). Для прочих документов (в том числе и не имеющих оригинала) отображается восстановленный текст.

Раздел *«Сохранить»* открывает форму для сохранения документа, в которой необходимо указать тип документа, выбрать каталог для сохранения и нажать кнопку *«Сохранить»*.

Раздел *«Печатать»* открывает форму для печати документа, в которой необходимо указать тип документа, отправляемого на печать и нажать кнопку *«Печатать»*.

Раздел *«Вид»* – позволяет устанавливать индивидуальные настройки раздела *«Страница»* – подключать/отключать список главных тем документа.

Под основным меню страницы *«Документ»* располагается панель, содержащая информацию о документе, а также поле *«Отмечать» – слова, пары слов, слова и пары слов* для разметки значимых слов в документе.

Кнопка

позволяет сохранить выделенный фрагмент документа в текстовый файл цитат, либо в документ Word (в зависимости от настроек системы, которые устанавливаются на странице *«Настройки» – «Внешний вид» – см. п. 7.9.3*). По щелчку на данной кнопке открывается форма, в которой необходимо указать каталог с файлом цитат и нажать «*Save»*. При необходимости можно очистить файл цитат, установив флажок в соответствующем поле.

Для того чтобы сохранить цитату в документ Word, на странице *«Настройки» – «Внешний вид»* должна быть указана соответствующая опция[1].

Также существует возможность перемещения по найденным словам. В этом случае возле найденного слова появляются стрелочки (Рисунок 26), являющиеся ссылкой на предыдущее и следующее найденные слова. При нажатии на стрелку после слова осуществляется переход к следующему найденному слову. Если это последнее в данном документе, произойдёт переход к следующему документу (к первому найденному в нём слову). При нажатии на стрелочку перед первым найденным словом, произойдёт переход к последнему найденному слову предыдущего документа. При переходе на другую страницу на короткое время появляется окошко «Листание на стр. …»

[1] Для того чтобы сохранить цитату в документ Word, на странице *«Настройки»* – раздел *«Внешний вид»* нажмите , затем кнопку *«Установить»*, далее перейдите на страницу *«Документ»* (откроется последний открытый документ), выделите фрагмент-цитату и щелкните по кнопке *«Цитировать»* – автоматически откроется новый документ Word с размещенной в нем цитатой (подробно страница *«Настройки»* описывается в следующих главах).

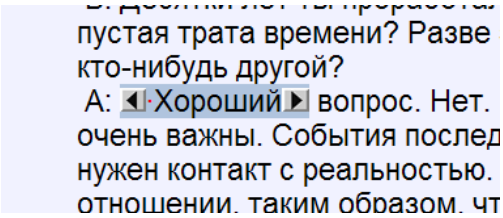


Рисунок 26. *Навигация по словам. Стрелки перехода к предыдущему и следующему ключевым словам*

3.7. Инфопортрет. Ведение диалога с Системой. Страница «Выборка» – «Инфопортет»

« “Информационный портрет” представляет собой как бы мозаику, состоящую из основных словоконструкций русского языка - ключевых слов и словосочетаний, присущих исследуемому объекту и расположенных в определенной последовательности. Например, тема “сахар”; ассоциации – “белый, сладкий, рафинад”, “пищевой продукт”…»

Кишинский Игорь Юрьевич,
руководитель департамента
специальных проектов

Информационный портрет – множество значимых слов и словосочетаний *(главных тем)*, отражающих информационное описание объекта.

Экспресс-анализ, который автоматически выполняется Системой, позволит получить представление об исследуемой предметной области, ознакомит Вас с используемой терминологией.

С помощью *информационного портрета* Вы можете вести так называемый ***диалог с Системой*** *(уточнять запрос)*. ***Уточнение запроса*** состоит в конкретизации предметной области. Вы можете выбрать интересующие темы из *инфопортрета* и уточнить ими запрос, таким образом, сужая границы поиска нужной информации. Подробно об уточнении запроса мы расскажем Вам в следующей главе. Сейчас мы остановимся на изучении *информационного портрета*.

При включенной статистике (поле *«Статистика»* – вкл. при проведении запроса) в списке документов на странице *«Выборка» – «Документы»* помимо фрагментов текста и кнопок навигации располагается *таблица-подсказка* (справа от списка). *Таблица-подсказка* представляет собой сокращенный список главных тем (сокращенный *инфопортрет*) *(см.* Рисунок 27*)*.

Для того чтобы посмотреть полный информационный портрет выборки перейдите в раздел *«Инфопортрет»*. Перед Вами откроется таблица, содержащая главные темы – описание объекта *(иранская программа)*. На панели, расположенной выше инфопортрета в поле *«Показать»* установите *слова и пары слов (см.* Рисунок 28*)*.

Слова и словосочетания, имеющие наивысшую степень значимости, как правило, отражают очевидные, общеизвестные вещи, характерные для объекта *(см.* Рисунок 29*)*.

Темы, имеющие среднюю степень значимости, связанны с особенностями объекта и явлениями, которые с ним происходят *(см.* Рисунок 30*)*.

Слова и словосочетания с наименьшей степенью значимости отображают редкие, малоизвестные, не малоизвестные, не типичные для данного объекта свойства *(см.* Рисунок 31*)*.

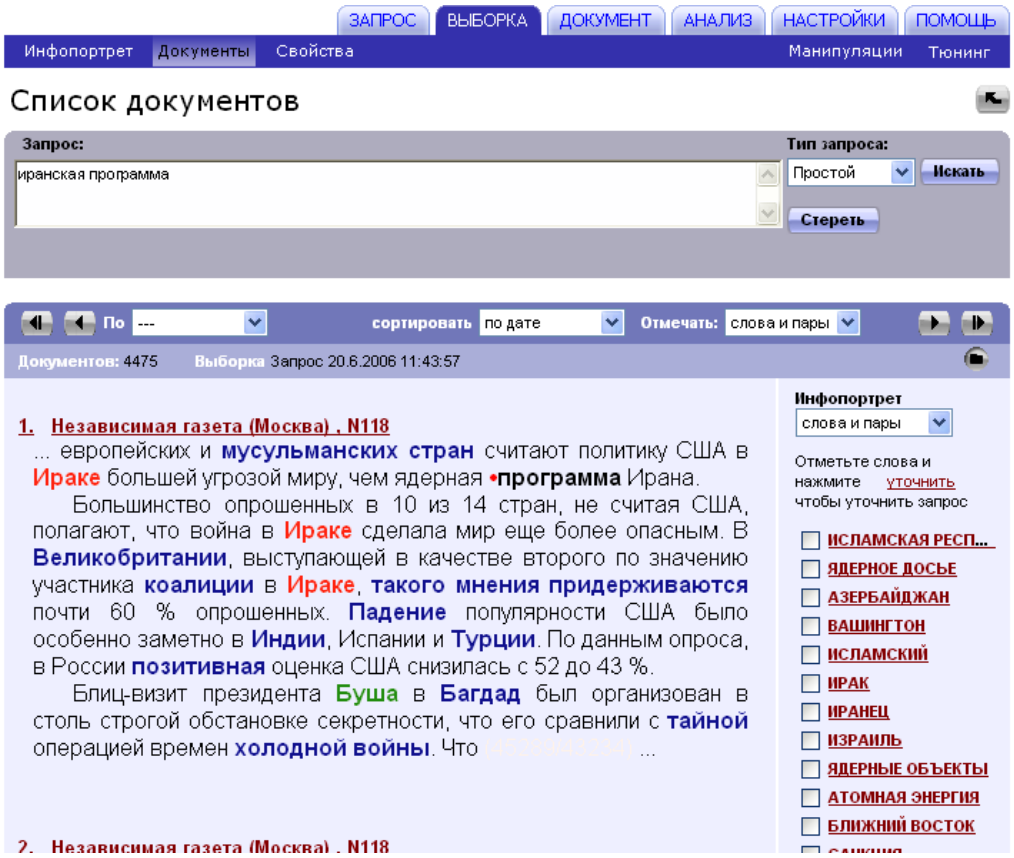


Рисунок 27. *Таблица – подсказка*

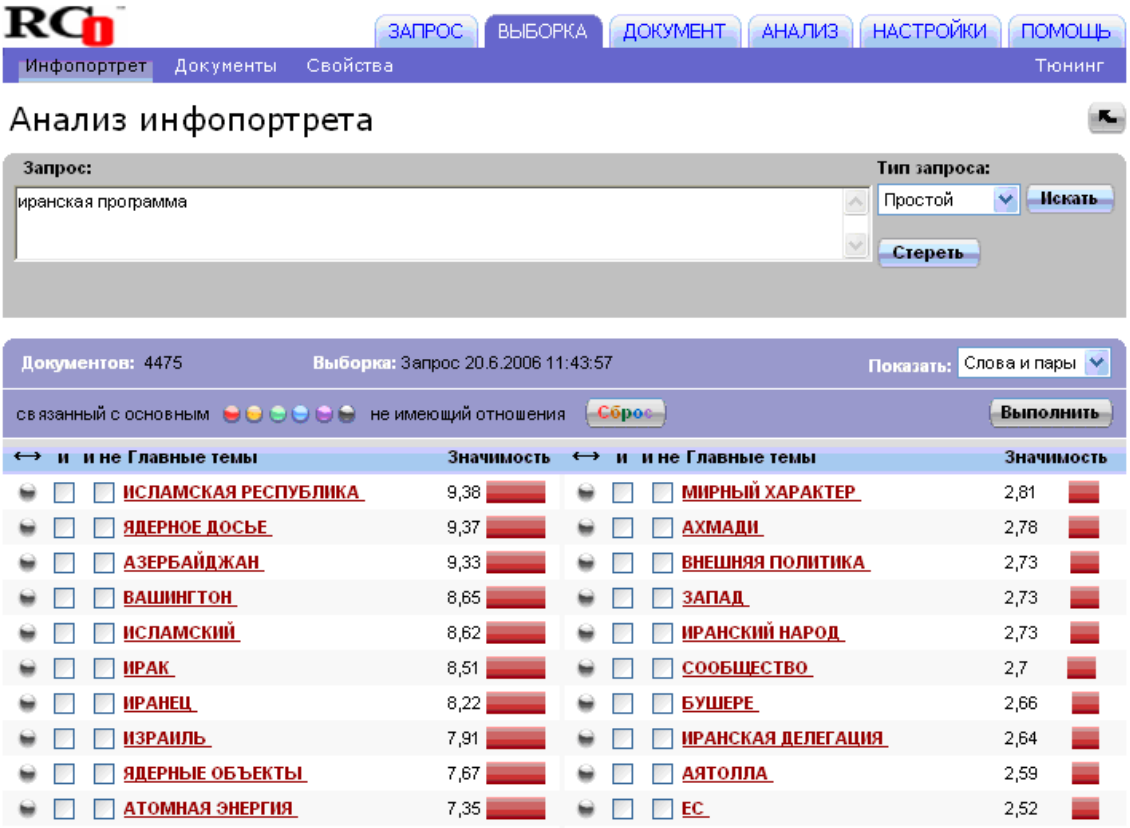


Рисунок 28. *Информационный портрет выборки «Иранская программа»*

ИСЛАМСКАЯ РЕСПУБЛИКА	9,38
ЯДЕРНОЕ ДОСЬЕ	9,37
АЗЕРБАЙДЖАН	9,33
ВАШИНГТОН	8,65
ИСЛАМСКИЙ	8,62
ИРАК	8,51
ИРАНЕЦ	8,22
ИЗРАИЛЬ	7,91
ЯДЕРНЫЕ ОБЪЕКТЫ	7,67
АТОМНАЯ ЭНЕРГИЯ	7,35
БЛИЖНИЙ ВОСТОК	6,07
САНКЦИЯ	5,99
ИРАНСКОЕ ДОСЬЕ	5,89
НЕФТЬ	5,66
БУШ	5,34
ИРАНСКИЕ ВЛАСТИ	5,23
ДОСЬЕ	5,17
ЯДЕРНОЕ ТОПЛИВО	5,01
ЭТА СТРАНА	4,98
ИНОСТРАННЫЕ ДЕЛА	4,92
МАХМУД	4,85
СОВБЕЗ	4,83
ЯДЕРНАЯ ПРОБЛЕМА	4,69
МЕЖДУНАРОДНОЕ АГЕНТСТВО	4,62

Рисунок 29. *Главные темы выборки «Иранская программа», имеющие высокую степень значимости*

МИРОВОЕ СООБЩЕСТВО	4,6
ЕВРОПЕЙСКАЯ ТРОЙКА	4,59
КИТАЙ	4,57
СОЕДИНЕННЫЕ ШТАТЫ	4,53
НЕРАСПРОСТРАНЕНИЕ	4,38
ИСЛАМСКАЯ РЕВОЛЮЦИЯ	4,35
ИРАНСКАЯ ПРОБЛЕМА	4,27
ИРАНСКИЙ ВОПРОС	4,17
РЕЗОЛЮЦИЯ	4,14
ПЕРСИДСКИЙ ЗАЛИВ	4,13
ГАЗ	4,07
ИРАНСКОЕ РУКОВОДСТВО	4,04
АРМЕНИЯ	3,78
ЯДЕРНЫЕ ТЕХНОЛОГИИ	3,76
МЕЖДУНАРОДНОЕ СООБЩЕСТВО	3,61
МИД	3,42
РАЙС	3,33
САУДОВСКАЯ АРАВИЯ	3,2
АТОМНАЯ ПРОГРАММА	3,17
АМЕРИКАНЕЦ	3,13
СБ	2,98
РОССИЙСКОЕ ПРЕДЛОЖЕНИЕ	2,95
ЯДЕРНЫЕ ИССЛЕДОВАНИЯ	2,94
ВИЗИТ	2,91
ИРАНСКАЯ СТОРОНА	2,82

Рисунок 30. *Главные темы выборки «Иранская программа», имеющие среднюю степень значимости*

ЭКОНОМИЧЕСКИЕ САНКЦИИ	2,23
АХМАДИНЕЖАД	2,23
ВОСТОК	2,2
ДИПЛОМАТИЧЕСКИЙ	2,2
СОЕДИНЕННЫЙ	2,17
БОМБА	2,16
ГОССЕКРЕТАРЬ	2,14
АХМАДИ-НЕЖАД	2,14
ИСФАХАНЕ	2,13
НАТО	2,13
ХАМАС	2,12
БЛИЖНИЙ	2,1
РАКЕТА	2,07
КОНДОЛИЗА	2,05
ЕВРОСОЮЗ	2,01
МИРНЫЙ АТОМ	2
КАСПИЙСКИЙ	1,99
ИРАНСКИЙ РЕЖИМ	1,95
ИРАНСКОЕ ПРАВИТЕЛЬСТВО	1,94
ФРАНЦИЯ	1,92
ИРИ	1,91
УРЕГУЛИРОВАНИЕ	1,9
ВОЕННАЯ ОПЕРАЦИЯ	1,89
ГРУЗИЯ	1,85
САУДОВСКИЙ	1,84

Рисунок 31. *Главные темы выборки «Иранская программа», имеющие низкую степень значимости*

Каждой *главной теме* (слову или словосочетанию) *инфопортрета* соответствует число (справа от главной темы, 5-ый столбец таблицы), которое характеризует степень значимости данного слова или словосочетания в выборке. Справа от чисел, располагаются столбцы столбчатой диаграммы, наглядно

иллюстрирующие значимость каждой конкретной темы. Чем больше число, тем больше значимость и соответствующий столбец диаграммы. Максимальная значимость присуща первому слову (словосочетанию) *инфопортрета,* минимальная – последнему.

В разделе *«Вид»* можно устанавливать параметры визуализации инфопортрета: отображение значимости главных тем, графика (столбчатой диаграммы), индикаторов качества выборки, кластеризации.

ГЛАВА 4. Основы эффективного поиска

В этой главе мы расскажем Вам о том, как наиболее рационально потратить время на поиск информации, не углубляясь в научное обоснование теории вероятностей и математической статистики. Вы узнаете основные приемы проведения поиска и отбора нужной информации.

4.1. Уточнение запроса

С помощью *инфопортрета* Вы можете уточнять (конкретизировать) запрос, сужая, таким образом, границы поиска. При уточнении запроса поиск осуществляется среди документов текущей выборки. Уточнить запрос можно тремя способами:

1. Используя функцию «И»

В *инфопортрете* выборки (столбец «И») отметьте, щелкнув по соответствующему полю, те слова и/или словосочетания, которые полностью соответствуют интересующей Вас предметной области. Тогда в результате уточнения в списке останутся только те документы, в которых упоминаются отмеченные Вами слова и/или словосочетания. Нажмите кнопку *«Выполнить»*, расположенную выше/ниже инфопортрета. При уточнении запроса по «И» в выборку попадут те документы, в которых присутствуют любые из отмеченных слов и/или словосочетаний.

2. Используя функцию «И НЕ»

В *инфопортрете* выборки (столбец «И НЕ») отметьте, щелкнув по соответствующему полю, те слова и/или словосочетания, которые не соответствуют исследуемой Вами предметной области. Тогда в результате такого уточнения в списке останутся только те документы, в которых указанные Вами слова и/или словосочетания отсутствуют. Нажмите кнопку *«Выполнить»*, расположенную выше/ниже инфопортрета. При уточнении запроса по «И НЕ» в выборку попадут только те документы, в которых отсутствуют любые из отмеченных слов и/или словосочетаний.

3. Щелкнув по слову или словосочетанию

В *инфопортрете* выборки щелкните по слову или словосочетанию, которым Вы хотите уточнить запрос. В этом случае, Вы получите *временную выборку*, в которой будет встречаться данное слово или словосочетание. *Временную выборку* нельзя в дальнейшем использовать для аналитических исследований.

Примечание

Уточнять запрос можно также с помощью *таблицы-подсказки* (сокращенного *инфопортрета*) (Рисунок 34), которая находится справа от списка документов. В этом случае применимы 1 и 3 способы уточнения.

В результате уточнения количество документов в списке сокращается, что позволяет направлять поиск и, тем самым, повышать его эффективность и оперативность.

Уточнение может быть многократным, количество последовательных уточнений не ограничено.

Примечание

Если нужно вернуться к запросу, который был до проведения уточнения, можно воспользоваться кнопкой «Откатить», появляющейся под кнопкой «Очистить» при уточнении (Рисунок 32).

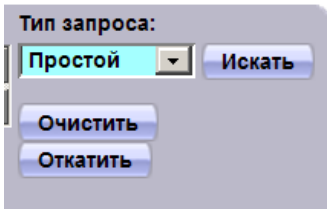


Рисунок 32. *После уточнения запроса появляется кнопка «Откатить»*

Итак, уточним запрос *иранская программа*, используя функцию «И», т.е. первым способом *(см.* Рисунок 33*)*. Возьмем в качестве уточняющего слова *главную тему Ядерная проблема*.

Таким образом, количество отобранных документов резко сократилось, информационный портрет выборки изменился *(см.* Рисунок 34, Рисунок 35*)*. Последовательно уточняя запрос, Вы получаете выборку, всё более соответствующую исследуемой предметной области и поставленной цели.

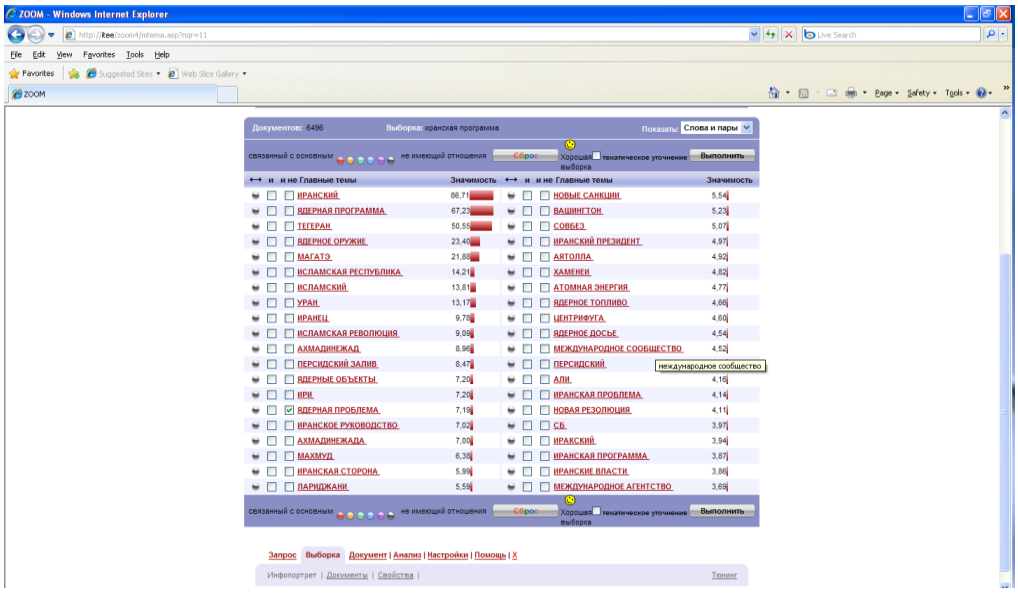


Рисунок 33. *Инфопортрет выборки «иранская программа»*

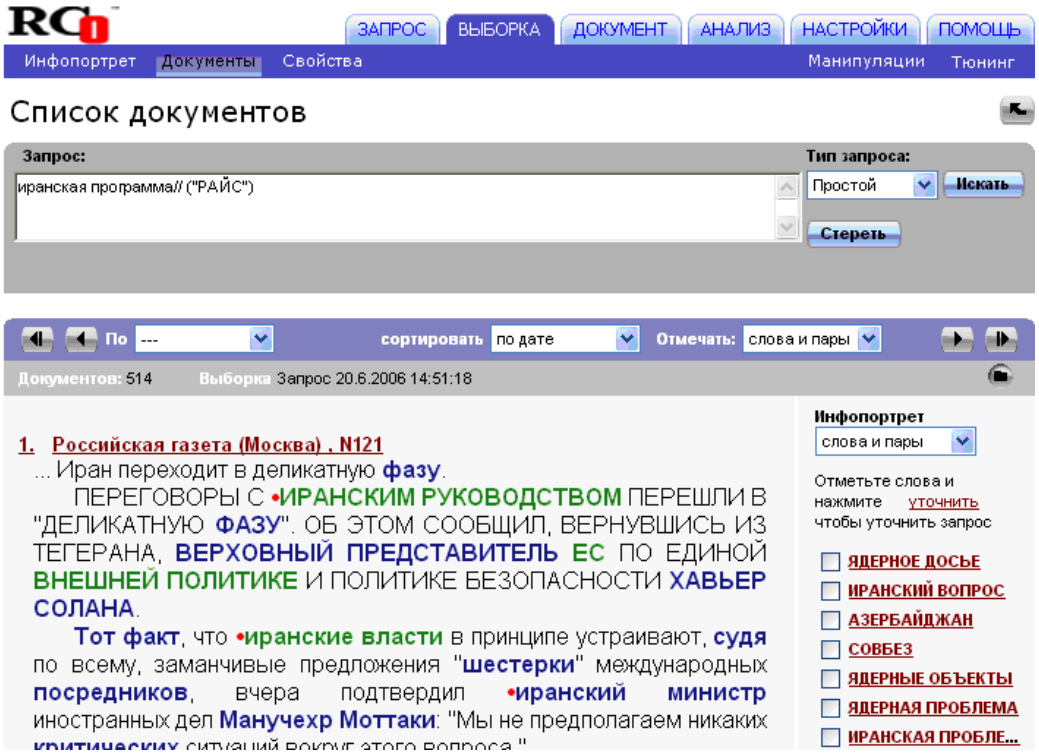


Рисунок 34. *Уточнение запроса «иранская программа» главной темой «Райс»*

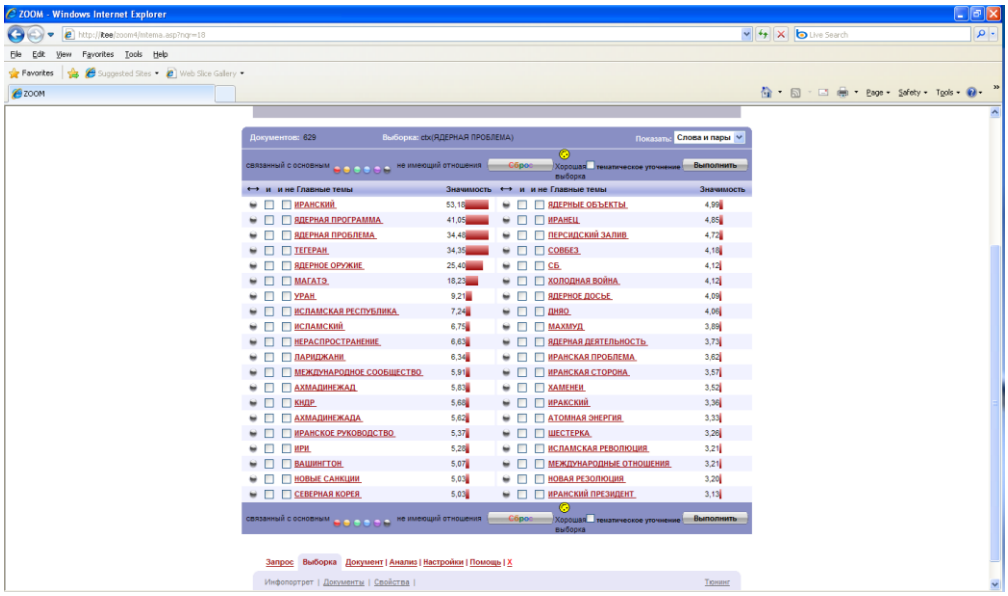


Рисунок 35. *Инфопортрет уточненной выборки «иранская программа» темой «Райс»*

При уточнении нажатием на слово мы можем вернуться из уточняющего запроса, являющегося временным, к основному запросу при условии, что основной запрос является постоянным. Для этого служит ссылка *«Возврат к основному запросу»* в правом верхнем углу под навигационной панелью (Рисунок 36).

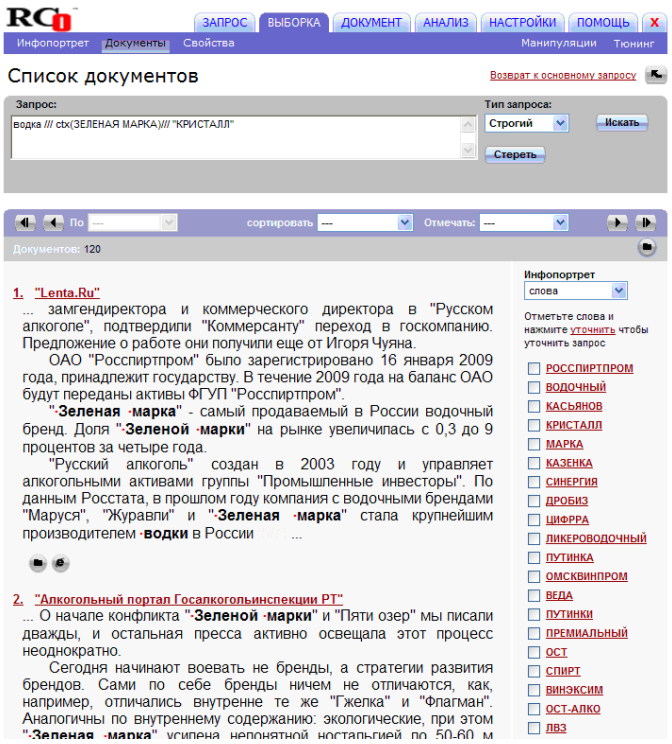


Рисунок 36. *Инфопортрет уточненной одним словом выборки. В правом верхнем углу ссылка «Возврат к основному запросу»*

Сделать временный запрос постоянным можно двумя способами:

- с помощью выпадающего меню *«Манипуляции» – «Добавить в список»*
- В окне *«Настройки» – «Внешний вид»* поставить галочку в пункте *«Сохранять новый запрос автоматически» в* разделе *«Анализ»*. В этом случае каждый новый запрос будет добавляться в список постоянных запросов.

Сортировка документов.

Часто возникает необходимость сортировки (ранжирования) документов тем или иным образом. Скажем, для того чтобы выбрать наиболее свежую и значимую информацию или, наоборот, исследовать ретроспективу.

Для этого в Системе предусмотрены следующие режимы сортировки документов:

- ***нет*** – режим отключен, тогда список документов, найденных по запросу, содержит документы в том порядке, в котором они были загружены в базу;
- ***по дате*** – в соответствии с датой создания документа, список возглавляет самый свежий документ;
- ***по значимости*** – документы располагаются в порядке уменьшения значимости, список найденных по запросу документов возглавит документ, наиболее близкий к информационному портрету выборки;
- ***по имени файла*** – в соответствии с алфавитным порядком имён файлов;
- ***по папке*** – документы в списке будут расположены в порядке уменьшения их соответствия инфопортрету папки.

Сортировка документов по значимости используется в тех случаях, когда нужно выделить перечень документов, в большей степени удовлетворяющих условиям запроса.

Первый способ сортировки документов: на странице *«Запрос»* – раздел *«Запрос»* в поле *«Вид сортировки»* установите нужный режим упорядочения документов, затем нажмите кнопку *«Искать»*.

Второй способ сортировки документов: на странице *«Выборка»* – раздел *«Документы»* в поле *«Сортировать»* установите вид сортировки – система автоматически отсортирует текущую выборку в соответствии с выбранным.

Третий способ включения сортировки: на странице *«Выборка»* – раздел *«Свойства»*) в поле *«Сортировка»* установите требуемый режим сортировки и нажмите *«Применить»*.

Режим сортировки можно также установить на странице *«Настройки»* – раздел *«Поиск»*, о которой речь пойдет в следующих главах.

4.2. Тематическое уточнение

Тематическое уточнение (ТУ) предназначено для уточнения списка документов, найденных по заданному предварительно простому и/или строгому запросу. В отличие от обычного уточнения, выборка будет включать документы, содержащие не только сами уточняющие слова, но и слова, связанные с ними. Тематических уточнений может быть несколько.

Активизировать тематическое уточнение можно, поставив галочку в соответствующем пункте в окне *«Инфопортрет»* (пункт расположен над инфопортретом) или *«Документы»* (пункт расположен под инфопортретом) (вкладка *«Выборка»*).

В строке запроса ТУ отделяется от запроса и/или другого уточнения тремя прямыми слешами. ТУ может быть использовано без предшествующего запроса.

Пример 1:

Польский премьер /// кнт(мировая война) – ТУ простого запроса;

кнт(польский премьер) /// футбол или матч /// скандал или болельщик – ТУ строгого запроса;

/// футбол или матч /// скандал – ТУ без предшествующего запроса.

Элементами ТУ могут быть слова или контексты, объединённые операцией «или». Для элементов ТУ нельзя использовать операцию логического отрицания (для отрицания надо использовать строгое уточнение).

Пример 2:

Польский премьер // не война – строгое уточнение простого запроса.

В остальном синтаксис ТУ ничем не отличается от синтаксиста строгого запроса, например, операция ***или*** может быть задана символами: «|», «или», «or».

Использование ТУ чаще всего уменьшает число найденных документов. При этом документы сортируются по тематике уточнения, т.е. в начало списка найденных документов поднимаются те из них, которые содержат больше элементов ТУ. Это справедливо как для простых, так и для строгих запросов.

4.3. Морфологический поиск

Режим морфологического поиска позволяет по каждому слову запроса искать все его словоформы, имеющиеся в морфологическом словаре Системы. Таким образом по запросу *найдет* будут найдены документы со словами *найти*, *нашел*, *нашедший*.

Допускается включение или отключение морфологического поиска отдельно для каждого слова. Для этого необходимо в конце слова дописать без пробелов %+ для включения или %– для отключения морфологического поиска. Таким образом, на слово банк будут найдены документы:

банк%- ищется только *"банк"*

банк%+ ищется *"банк"*, *"банка"*, *"банки"*, *"банками"* и т.д.

Для того чтобы включить морфологический поиск по всем словам запроса, на странице *«Запрос»* – раздел *«Запрос»* в поле *«Морфологических поиск»* установите флажок. Установка флажка влияет на все поля текстового типа.

Другой вариант включения морфологического поиска: на странице *«Настройки»* – раздел *«Поиск»* установите флажок в поле *«Морфологический поиск»* и нажмите *«Установить»*. Установка флажка также влияет на все поля текстового типа.

4.4. Страница «Выборка» – «Свойства»

Интерфейс раздела *«Свойства»* страницы *«Выборка»* (Рисунок 37 и Рисунок 38).

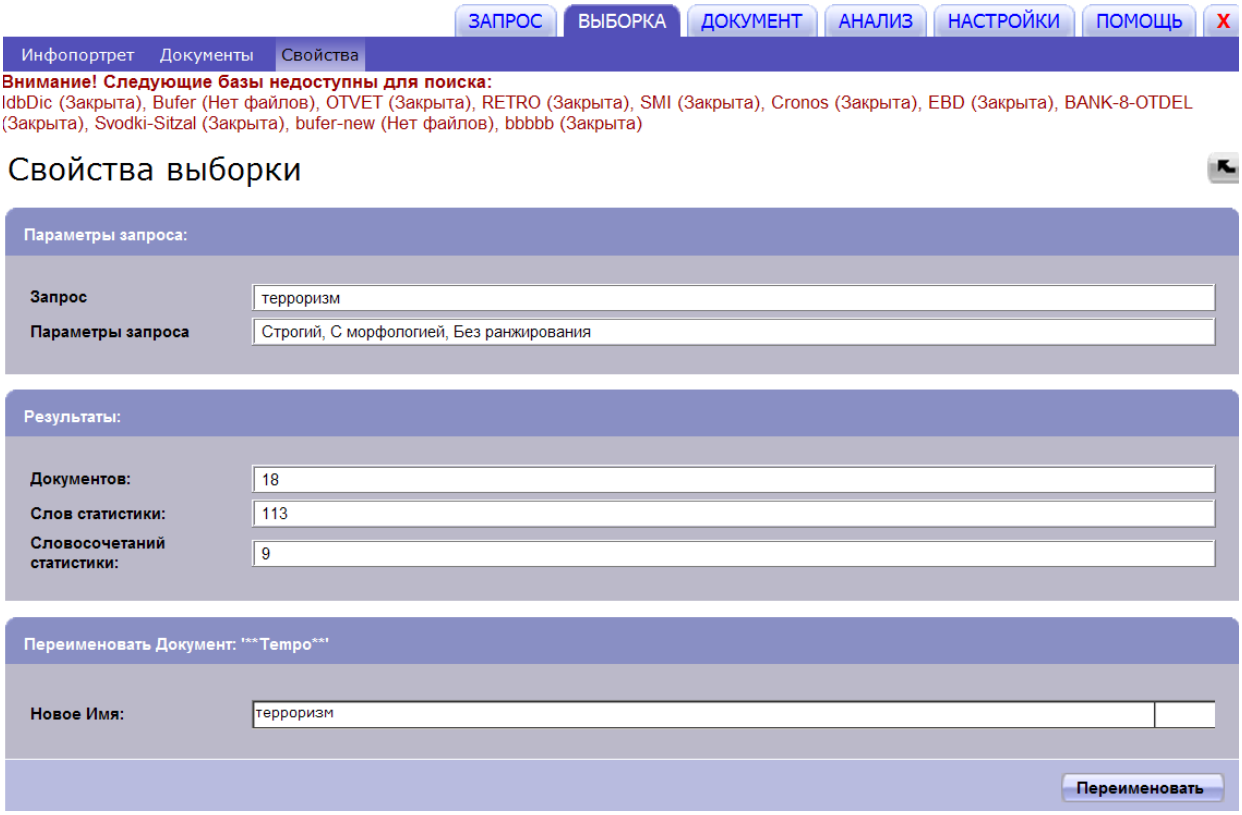


Рисунок 37. *Раздел «Свойства» страницы «Выборка»*

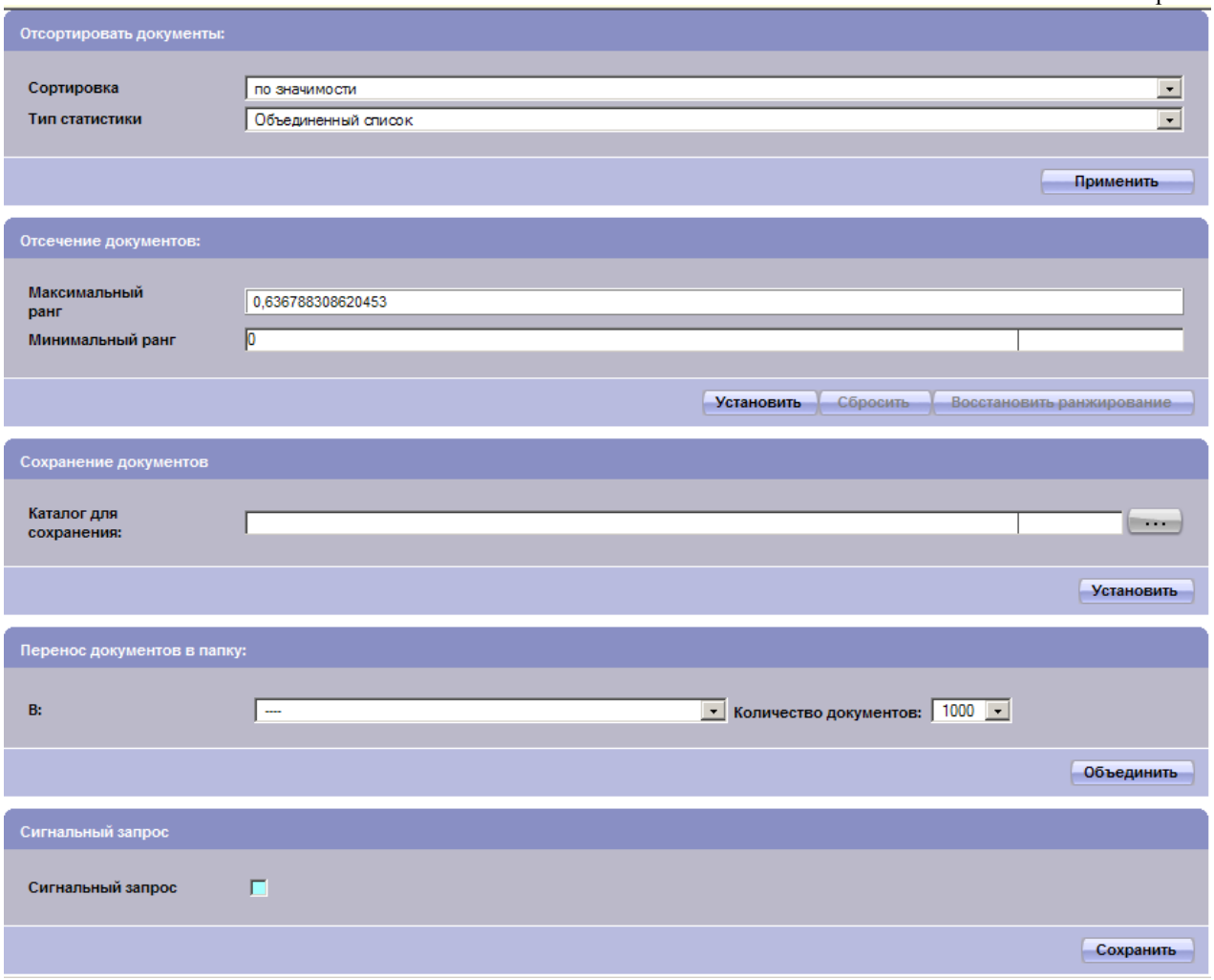


Рисунок 38. *Раздел «Свойства» страницы «Выборка»*

На этой странице располагается ряд панелей:

- ✓ ***Параметры запроса*** – содержит информацию о проведенном запросе, основные его параметры;
- ✓ ***Результаты*** – содержит статистические данные о запросе;
- ✓ ***Переименовать документ*** – дает возможность переименовать запрос. Для того чтобы переименовать запрос, удалите старое название, введите новое и нажмите кнопку *«Переименовать»*. Обратите внимание, что при этом новое название появится в заголовке блока *«Переименовать документ»*, а в строке ввода останется прежнее название (Рисунок 39);

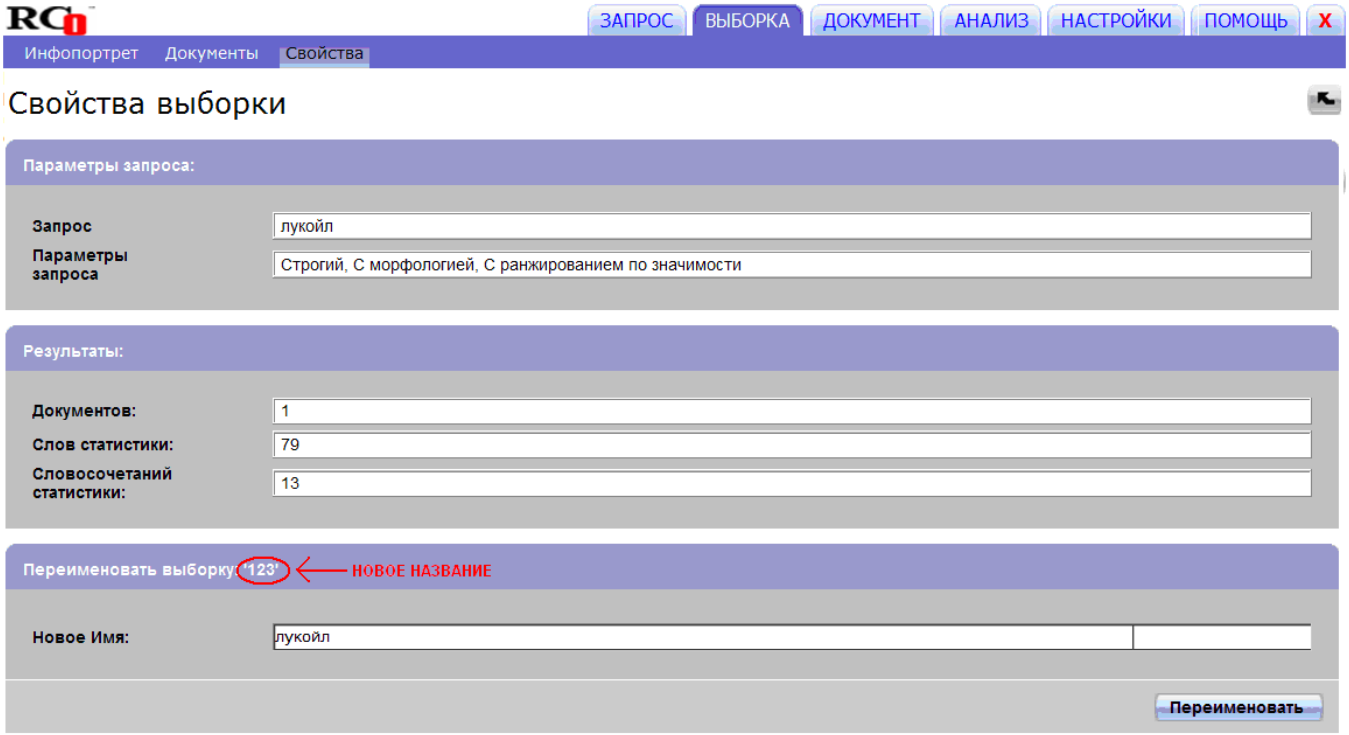


Рисунок 39. *Переименование запроса из «ЛУКОЙЛ» в «123»*

- ***Отсортировать документы*** – позволяет проводить сортировку найденных документов по значимости, дате или по папкам (Рисунок 38). Для сортировки документов выберите необходимые параметры и нажмите *«Применить»*;
- ***Отсечение документов*** – позволяет сократить количество документов путем удаления из выборки документов с рангом ниже установленного. Установить минимальное значение ранга можно в соответствующей строке панели *«Отсечение документов»*. Для выполнения отсечения нажмите кнопку *«Применить»*. Для отмены отсечения нажмите кнопку *«Сбросить»*.

  После отсечения оставшиеся в выборке документы можно отсортировать по другому параметру. Потом при желании можно восстановить сортировку, по которой проводилось отсечение (*«Выборка» – «Свойства» – «Отсечение документов» – «Восстановить ранжирование»*).

Примечание

Отсечение документов можно применить:

- к выборке при простом запросе (отсортированной по умолчанию или по значимости/ папкам);
- к выборке при строгом запросе (отсортированной по значимости/ папкам).

- ***Сохранение документов*** – дает возможность выбрать каталог для сохранения документов. Для того чтобы выбрать каталог, нажмите *«…»*, появится окно выбора каталога, затем выберите нужный каталог, нажмите ОК (если появляется ошибка «Невозможно загрузить компонент», см. гл. 9, п. 9.1). Затем нажмите кнопку *«Установить»*.
- ***Перенос документов в папку*** – позволяет перенести в заранее созданную папку (см. п. 6.2) определенное количество документов (например, первые 1000 документов). Для этого выберите папку и количество сохраняемых документов, нажмите кнопку *«Объединить»*.
- ***Сигнальный запрос*** – позволяет особо пометить запрос (допустимо только для сохранённых запросов). Каждый раз при пополнении базы проводится поиск по сигнальным запросам и, если появляются новые документы, соответствующие этому запросу, на всех страницах появится кнопка (колокольчик), показывающая наличие новых документов по данному запросу. При нажатии на кнопку появится окно со списком сработавших сигнальных запросов (Рисунок 40). В списке запросов сигнальные тоже помечаются «колокольчиком» .

Рисунок 40. *Список сработавших сигнальных запросов*

Примечания

Сигнальные запросы всегда проводятся по той же базе (базам), по которой проводился запрос, помеченный, как сигнальный.

Не следует заводить много сигнальных запросов, т.к. это снижает быстродействие системы. В отдельных случаях это может привести к проблемам с сохранением новых запросов, т.к. общее число сохраняемых запросов ограничено.

4.5. Параллельная работа одного пользователя с несколькими выборками

Для одновременной работы одного пользователя с несколькими выборками на одном компьютере, для начала необходимо в настройках сессии пользователя *(«Настройки» – «Внешний вид» – «Анализ»)* установить галочку *“Сохранять новый запрос автоматически”* (Рисунок 41)

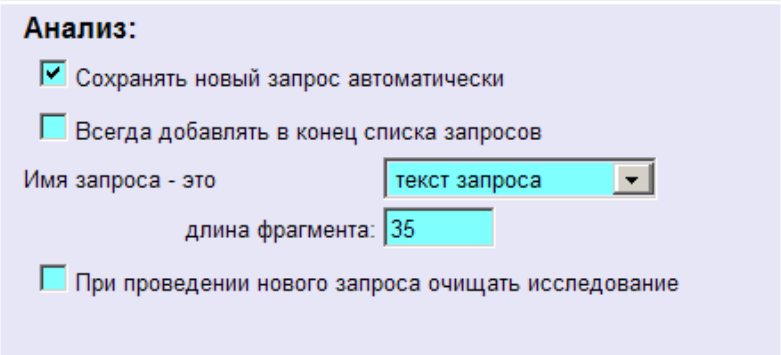


Рисунок 41. *Установка режима сохранения запроса*

Примечание: Настройка *«Имя запроса – это текст запроса»* не является обязательной, здесь она взята исключительно из соображений наглядности страницы *«Исследование»* (Рисунок 45).

Например, в одном окне был проведён запрос «Президент» (Рисунок 42)

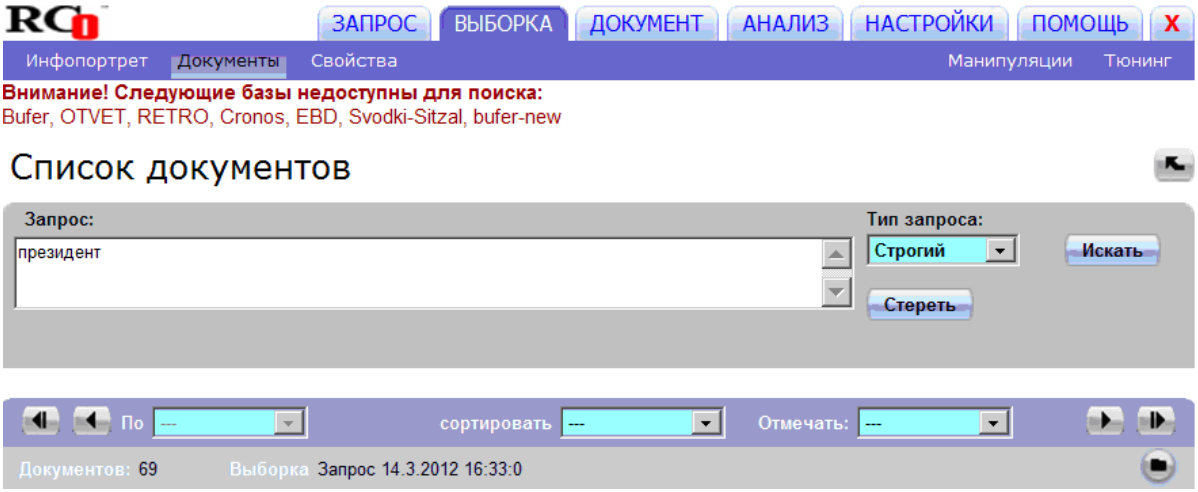


Рисунок 42. *Проведение запроса «Президент»*

Для проведения нового запроса нужно открыть новую вкладку/окно, зайти на сайт и провести запрос.

Например, проведём в новом окне запрос «Выборы» (Рисунок 43)

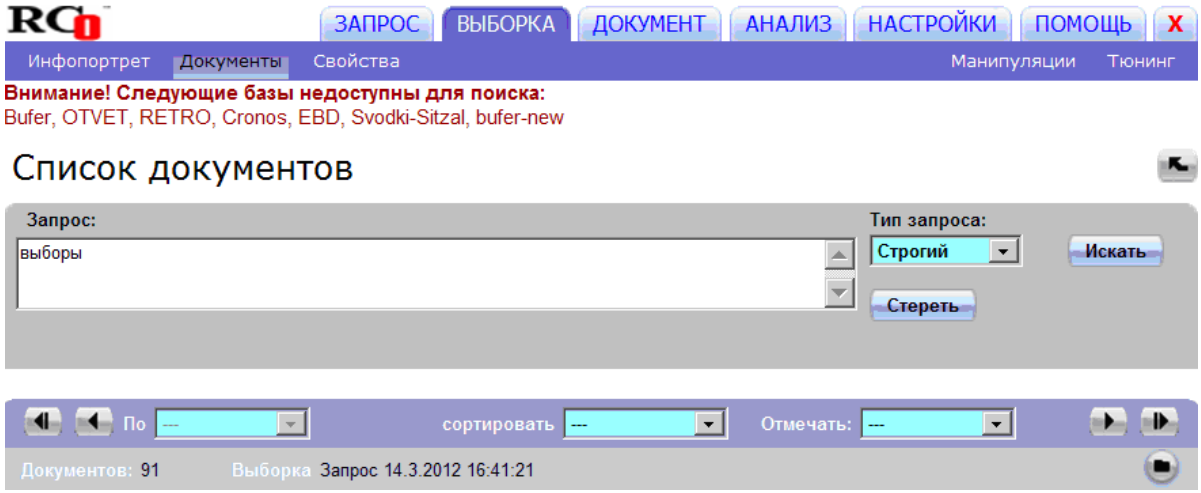


Рисунок 43. *Проведение запроса «Выборы»*

Таким образом, становится доступной одновременная работа с несколькими выборками (Рисунок 44)

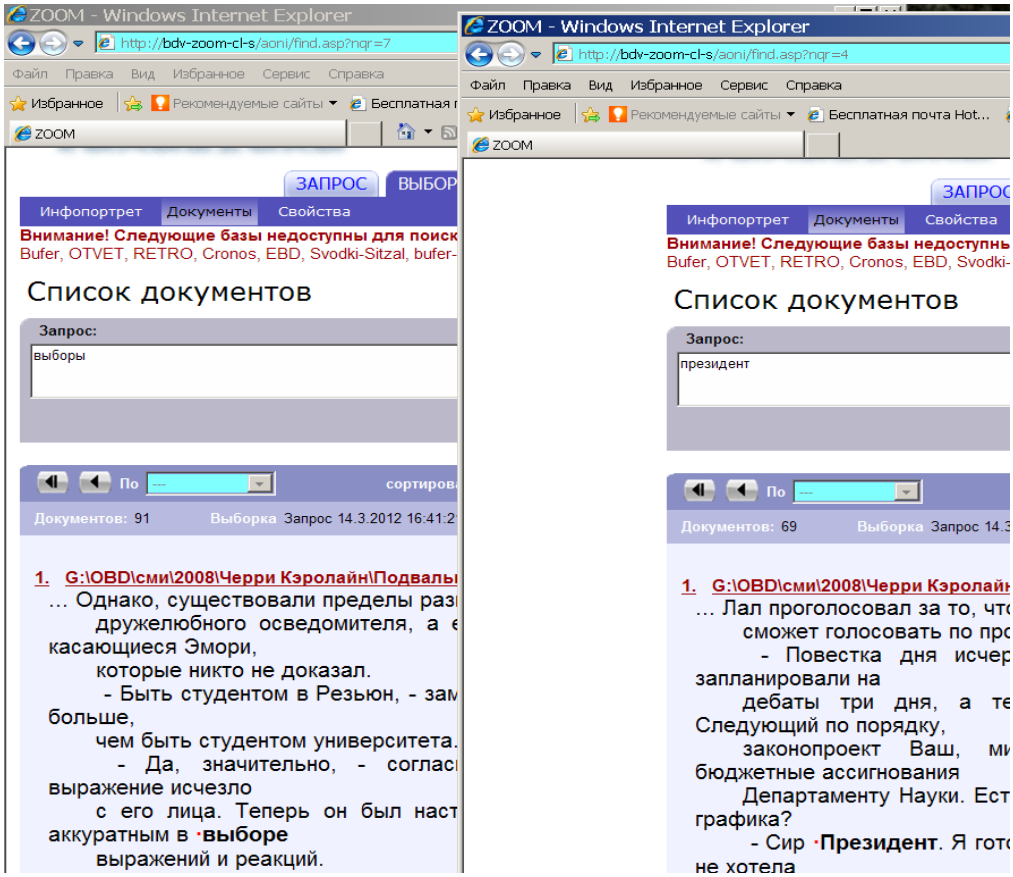


Рисунок 44. *Одновременная работа с несколькими окнами*

Переключение между выборками (если в ходе работы пользователь перешел на одной из вкладок/окон на вкладку запрос) возможно только через вкладку *«Анализ»* (Рисунок 45):нажатие на

пункт меню «выборка» в том окне, где перешли на запрос не обязательно приведет к исходной выборке.

Для этого нужно воспользоваться кнопкой напротив нужной выборки.

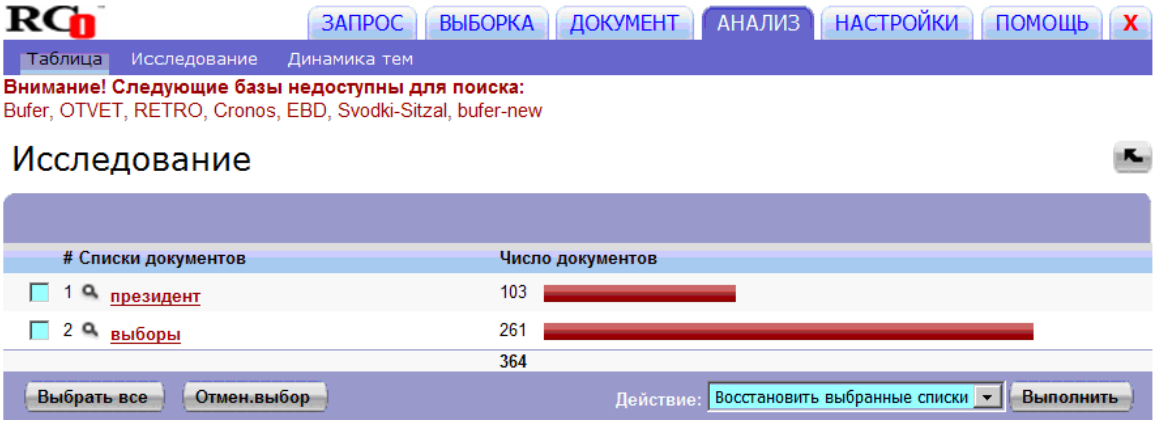


Рисунок 45. *Переключение между выборками во вкладке «Анализ»*

ГЛАВА 5. Язык запросов. Строгий запрос

В предыдущих главах мы рассмотрели пример *простого запроса* (по одному или нескольким словам, записанным через пробел). Но порой этого не достаточно для проведения эффективного поиска. Зачастую бывает необходимо отобрать документы либо за определенный промежуток времени, либо поступившие из конкретного источника, либо принадлежащие одному автору. Для этого следует получить представление о создании более сложных запросов по полям документа.

5.1. Что такое язык запросов?

В первой главе мы дали определение *запроса*, вернемся к этому определению.

Запрос — логическая конструкция, записанная по специальным правилам языка запросов, задающая критерий отбора документов из базы. Запрос может представлять собой слово, последовательность слов или логическое выражение, составленное в соответствии с правилами языка запросов.

Слово или последовательность слов, записанных через пробел – пример *простого* запроса, т.е. запроса, составленного на естественном языке без использования специальных правил. Примеры *простого* запроса: *«иранская программа»*, *«Райс»*, *«национальная безопасность США»*.

Используя только *простой* запрос, Вы не сможете в полной мере реализовать требования эффективности и оперативности поиска, т.к. Система ищет документы во всем объеме текстовой базы, по всем полям.

Для ускорения поиска составлять запрос рекомендуем по правилам *языка запросов*.

Язык запросов - формальный язык, включающий в себя набор символов, зарезервированных слов и правил (способов комбинирования символов и зарезервированных слов), в соответствии с которыми выстраивается логическая конструкция (запрос).

5.2. Запрос по полю документа

Для того чтобы ускорить поиск информации, особенно в тех случаях, когда необходимо найти документы за определенный промежуток времени, либо имеющих одного и того же автора, либо относящихся к определенному источнику, можно воспользоваться специальными правилами составления запросов по полям документа.

Примечание

Перед проведением запроса, составленного с использованием языка запросов, в поле *«Тип запроса»* (на странице *«Запрос»* – раздел *«Запрос»*) установите значение *Строгий*!

Обращаем Ваше внимание на то, что в предыдущих главах мы рассматривали примеры проведения *простых* запросов, представляющих собой слово или последовательность слов, записанных через пробел. *Простой* запрос (на странице *«Запрос»* – раздел *«Запрос»*, тип запроса – *простой*), в отличие от *строгого*, составляется на естественном языке, без использования специальных правил *языка запросов*.

Примечание

Ознакомиться со списком полей баз можно на странице *«Настройки»* – раздел *«Поиск»*. Если навести указатель мыши на символ «*i*», появится список полей (*см.* Рисунок 46) соответствующей базы.

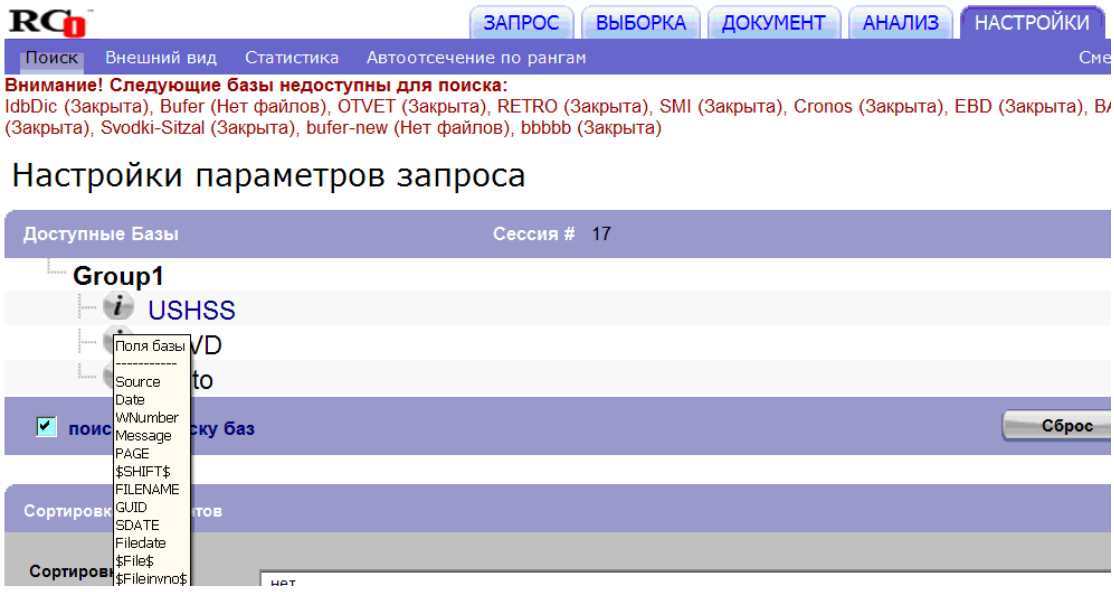


Рисунок 46. *Список полей базы на вкладке «Настройки»*

5.2.1. Общие правила составления запросов по полям документа

Общие правила составления запроса таковы:

Запрос должен быть адресован одному из полей документа. Имя поля, по которому ведется поиск, указывается всегда, кроме тех случаев, когда поиск ведется в *поле по умолчанию* (обычно, это основное текстовое поле). Так же, как и имя поля, функционируют знаки “#” и “**@**”.

Знак “**#**” используется при поиске по номеру документа.

Знак “**@**” употребляется при поиске во всех текстовых полях документа.

В общем виде запрос имеет следующий вид:

имя_поля: *(условие отбора)*

Что такое “*условие отбора документов*”?

Это – то условие, которому должны удовлетворять найденные по запросу документы. Оно представляет собой некоторые требования к элементам, заполняющим данное поле документа.

Если условие отбора не содержит *логических* и *контекстных* операторов, то скобки не обязательны (о логических и контекстных операторах мы расскажем далее).

Т.е. запрос начинается с указания имени поля, по которому ведется поиск (кроме *поля по умолчанию*), далее следует знак “:” (двоеточие), затем перечислены условия отбора документов по данному полю.

Двоеточие не ставится после знаков “#” и “@”.

Знак “*****” заменяет любое количество символов. Одиночная “*****” употребляется для поиска всех документов, у которых поле запроса непустое.

5.2.2. Запрос по дате создания документа (поле «date»)

Запрос по полю *«date»* следует использовать, если Вам необходимо отобрать документы за определенный промежуток времени, либо относящиеся к определенной дате.

Тип запроса зависит от типа поля. Для поля дат это дата в формате дд.мм.гггг, либо д.м.гггг.

Пример 1:

date: 07.11.2005

По такому запросу Система отберет все документы за 7 ноября 2005 года.

Пример 2:

date: 7.12.2004..14.1.2006

По такому запросу Система отберет все документы, относящиеся к периоду с 7 декабря 2004 года по 14 января 2006 года. Таким образом, можно задавать *диапазон* дат.

Диапазон задается указанием двух границ диапазона с разделителем в виде «..», либо «…» (многоточие).

5.2.3. Запрос по номеру документа в базе (поле #)

Каждому документу при поступлении его в базу присваивается уникальный номер. Если Вам требуется отобрать документы с определенными номерами, нужно указать их в запросе.

Тип поля номеров – числа.

Пример 1:

#1234

По такому запросу Система найдет документ с номером 1234.

Пример 2:

#638..1234

В этом случае Система отберет документы с номерами, лежащими в диапазоне с 638 по 1234.

Так же, как и в случае с датой, *диапазон* номеров задается указанием двух границ диапазона и разделителя в виде многоточия.

5.2.4. Условные операторы

К условным операторам относятся “больше”, “больше или равно”, “меньше”, “меньше или равно”. Для их обозначения в языке запросов используются знаки соответственно “>”, “>=”, “<”, “<=”.

Сюда же относится операция “равно”, обозначающаяся знаком “=”.

Пример 1:

date: >= 7.11.2005

По этому запросу будут найдены все документы от 7 ноября 2005 года и более поздние.

Пример 2:

#<1234

По такому запросу Система отберет все документы, номера которых меньше 1234.

5.2.5. Запрос по источнику документа (поле «source»)

Запрос по источнику документа необходим для того, чтобы отобрать документы, относящиеся к определенному источнику.

Тип поля *«source»* это слова либо *шаблоны слов* (см. приложение). Под словом здесь понимается последовательность букв и цифр. Название источника (слово) документа необходимо брать в кавычки.

Пример 1:

source: ‘ведомости’

Система отберет все документы (статьи) газеты «Ведомости».

Пример 2:

source: “мк”

Система отберет все документы (статьи) газеты «Московский комсомолец (МК)».

Для текстовых полей также можно использовать *условные операторы*. В этом случае поиск будет проводиться в соответствии с упорядоченным по алфавиту списком элементов поля (названий источников, имен авторов и пр.)

Пример 3:

source: >= 'итоги'

При таком запросе Система отберет все документы, поступившие из газеты «Итоги» и далее в алфавитном порядке изо всех остальных источников, стоящих в списке элементов поля «source» после элемента (источника) «Итоги». При этом результат запроса зависит от того, каким способом был загружен документ. Таким образом, источник «Аналитические итоги» попадёт по вышеуказанному запросу в выборку только в том случае, если поле source грузилось, как многословное. Если source загружалось как однословное – не попадает.

Для текстовых полей также можно задавать *диапазон* значений.

Пример 4:

source:(итоги...коммерсантъ)

В этом случае найдутся все документы, поступившие из газеты «Итоги» и далее в алфавитном порядке изо всех источников, стоящих в списке элементов поля *«source»* до элемента (источника) «Коммерсантъ», включая документы, поступившие из газеты «Коммерсантъ».

Однако хотим обратить Ваше внимание на то, что в отличие от задания диапазона дат, диапазон текстовых значений необходимо заключать в круглые скобки.

5.2.6. Запрос по автору документа (поле «author»)

Запрос по автору документа необходим для того, чтобы отобрать документы определенного автора.

Тип поля *«author»* – слова либо *шаблоны слов* (см. приложение), как для поля *«source»*.

Слова – имена и фамилии авторов документов необходимо брать в кавычки.

Пример:

author:’иванов’

По этому запросу Система отберет все документы автора Иванова.

author:*

По этому запросу Система отберет все документы с заполненным полем автора

Условные операторы и диапазоны используются так же, как и в запросе по источнику документа (см. предыдущий пункт).

5.2.7. Запрос по основному текстовому полю «message» (поле по умолчанию)

Основное текстовое поле документа обычно устанавливается администратором Системы в качестве поля по умолчанию. Для формирования запроса по такому полю нет необходимости указывать его название.

Примечание

Если в запросе имя поля не указано, то Система осуществляет поиск в *поле по умолчанию* (обычно, это поле основного текста документа).

5.3. Позиционный запрос

Позиционный запрос задаёт строгую последовательность слов, которая должна присутствовать в найденном документе. Минимальная последовательность содержит два слова.

В том случае, когда необходимо найти документы, содержащие слова в строго определенном порядке, используется оператор **ctx**. Если порядок слов может также быть обратным заданному, то в этом случае используется оператор **near**. Знак оператора ставится перед скобками, внутри которых описывается условие поиска – последовательность слов.

Обозначения операторов:

- ✓ **ctx**, либо **c**, либо **кнт**
- ✓ **near**, либо **n**, либо **рдм**

Пример 1:

кнт(Владимир Путин)

При таком запросе Система найдет все документы, в тексте (поле по умолчанию) которых встречается последовательность слов *Владимир Путин*.

Пример 2:

рдм(Владимир Путин)

В этом случае Система найдет документы, в тексте которых встречается последовательность слов *Владимир Путин*, а также документы, в тексте которых встречается последовательность *Путин Владимир*.

Примечание

Позиционные операторы можно также использовать в запросе по любому текстовому полю документа, в том числе полям *«author»* и *«source»*.

Пример 3:

author:**кнт**(Дмитрий Денисов)

В этом случае Система отберет документы, у которых в поле «author» указан автор Дмитрий Денисов.

Примечание

Если в запросе по полю документа используются позиционные операторы, то условие отбора (в примере это **кнт**(Дмитрий Денисов)) должно заключаться в круглые скобки.

5.4. Логические операторы

Рассмотрим все логические операторы на примерах.

5.4.1. Логический оператор «И»

Обозначается знаком **&**, либо **и**, либо **and**.

Пример: требуется найти все документы, в которых встречаются слова *Россия* и *Белоруссия*.

Запрос будет выглядеть следующим образом:

Россия **и** *Белоруссия*

В данном случае **и** является логическим оператором, соединяющим элементы запроса.

По первому элементу запроса Система выберет документы, содержащие в тексте (поле по умолчанию) слово *Россия*, по второму – слово *Белоруссия*. В результирующую выборку попадут документы, удовлетворяющие и первому, и второму элементу запроса, то есть содержащие в тексте (поле по умолчанию) оба слова *Россия* и *Белоруссия*.

5.4.2. Логический оператор «ИЛИ»

Обозначается знаками “**или**”, либо “**or**”.

Пример: требуется найти документы, где упоминается хотя бы одно из слов: *Россия* или *Белоруссия*.

Запрос будет выглядеть следующим образом:

Россия **или** *Белоруссия*

В данном случае **или** является логическим оператором, соединяющие элементы запроса.

В результирующую выборку попадут документы, удовлетворяющие хотя бы одному элементу запроса, т.е. документы, содержащие в тексте (поле по умолчанию) слово *Россия*, либо слово *Белоруссия*, либо то и другое.

5.4.3. Логическое отрицание «НЕ»

Обозначается знаками ^, либо **не**, либо **not**.

Пример: требуется найти документы, где не упоминается слово *Белоруссия*.

Запрос будет выглядеть следующим образом:

не *Белоруссия*

В данном случае **не** является логическим оператором, относящимся к элементу запроса.

В результирующую выборку попадут документы, в тексте (поле по умолчанию) которых не содержится слово *Белоруссия*.

5.4.4. Логическое отрицание «И НЕ»

Обозначается знаками **&^**, либо **и не**, либо **and not**.

Пример: требуется найти документы, где встречается слово *Россия*, но не упоминается *Белоруссия*.

Запрос будет выглядеть следующим образом:

Россия **и не** *Белоруссия*

В данной случае **и не** является логическим оператором, относящимся к элементам запроса.

В результирующую выборку попадут документы, в тексте (поле по умолчанию) которых содержится слово *Россия* и не упоминается *Белоруссия*.

Примечание

В предыдущих главах мы рассматривали примеры проведения *простых* запросов (слово/последовательность слов). В том случае, если *простой* запрос (на странице *«Запрос»* – раздел *«Запрос»*, тип запроса – *простой*) представляет собой последовательность слов, записанных через пробел, то Система автоматически вместо пробелов ставит оператор «***И***». Таким образом, *простой* запрос *«иранская программа»* интерпретируется Системой как *строгий* запрос *«иранская* **и** *программа» (см.п. 5.4.1.)*.

5.5. Запрос по части документа

В данном случае условие поиска должно быть реализовано в определенной *части документа* – в абзаце или предложении. В запросе *по части докум*ента не должны использоваться операторы позиционного запроса (**ctx** и **near**).

Если условие поиска должно быть реализовано в пределах абзаца, используется оператор **prgf**, либо **p**, либо **пргф**.

Если условие поиска должно быть реализовано в пределах предложения, используется оператор **Sent**, либо **s**, либо **пред**.

Знак оператора ставится перед скобками, внутри которых описывается условие поиска. Условие поиска может быть описано посредством логического выражения, элементами которого являются слова и *шаблоны слов* (см. приложение).

Пример 1:

пргф(Россия **и** Белоруссия)

По этому запросу будут найдены документы, в которых слова *Россия* и *Белоруссия* содержатся в одном абзаце.

Пример 2:

пред(Россия **и** Белоруссия)

По этому запросу будут найдены документы, в которых слова *Россия* и *Белоруссия* содержатся в одном предложении.

Примечание

Запрос по части документа может быть адресован только основному текстовому полю (обычно, *поле по умолчанию*).

5.6. Сложный запрос

Позиционный запрос и запрос *по части документа* являются *контекстными* запросами (т.е. их можно объединить одним понятием *контекстного* запроса).

Любой запрос, *контекстный* или по определенному *полю* документа, может выступать как часть логического выражения, т.е. запросы могут соединяться между собой логическими операторами.

Пример 1:

date: 20.11.2004 и prgf(нато и косово) и sent(оон и гуманитарный) и не кнт (президент сша)

По этому запросу будут найдены документы за 20 ноября 2004 года, в которых есть параграф, содержащий оба слова “НАТО” и “Косово”, а также есть предложение, в котором содержатся слова ООН и гуманитарный, но во всем тексте документа не содержится последовательность слов *президент США*.

Пример 2:

date:14.6.2005 и author:кнт(Дмитрий Денисов) и (Самара или Москва)

По этому запросу будут найдены документы за 14 июля 2005 года, принадлежащие автору *Дмитрию Денисову*, и, содержащие в тексте (*поле по умолчанию*) слово *Самара*, либо слово *Москв*а, либо оба слова.

Пример 3:

#>1234 и не source:'ведомости' и рдм(В Путин)

По этому запросу будут найдены документы с номерами, большими 1234, из всех источников, кроме источника *«Ведомости»* и содержащие в тексте (*поле по умолчанию*) последовательность *В Путин*, либо *Путин В*.

Примечание

В этой главе мы описали основные приемы составления запросов, более подробно ознакомиться с языком запросов Вы сможете в ***Приложении*** к настоящему руководству.

5.7. Конструктор запросов. Страница «Запрос» – «Конструктор»

При расширенном поиске (с использованием языка запросов) для упрощения составления запросов можно использовать так называемый *конструктор запросов*.

Для перехода к конструктору запроса, необходимо на странице *«Запрос»* перейти в раздел *«Конструктор»* (см. Рисунок 47).

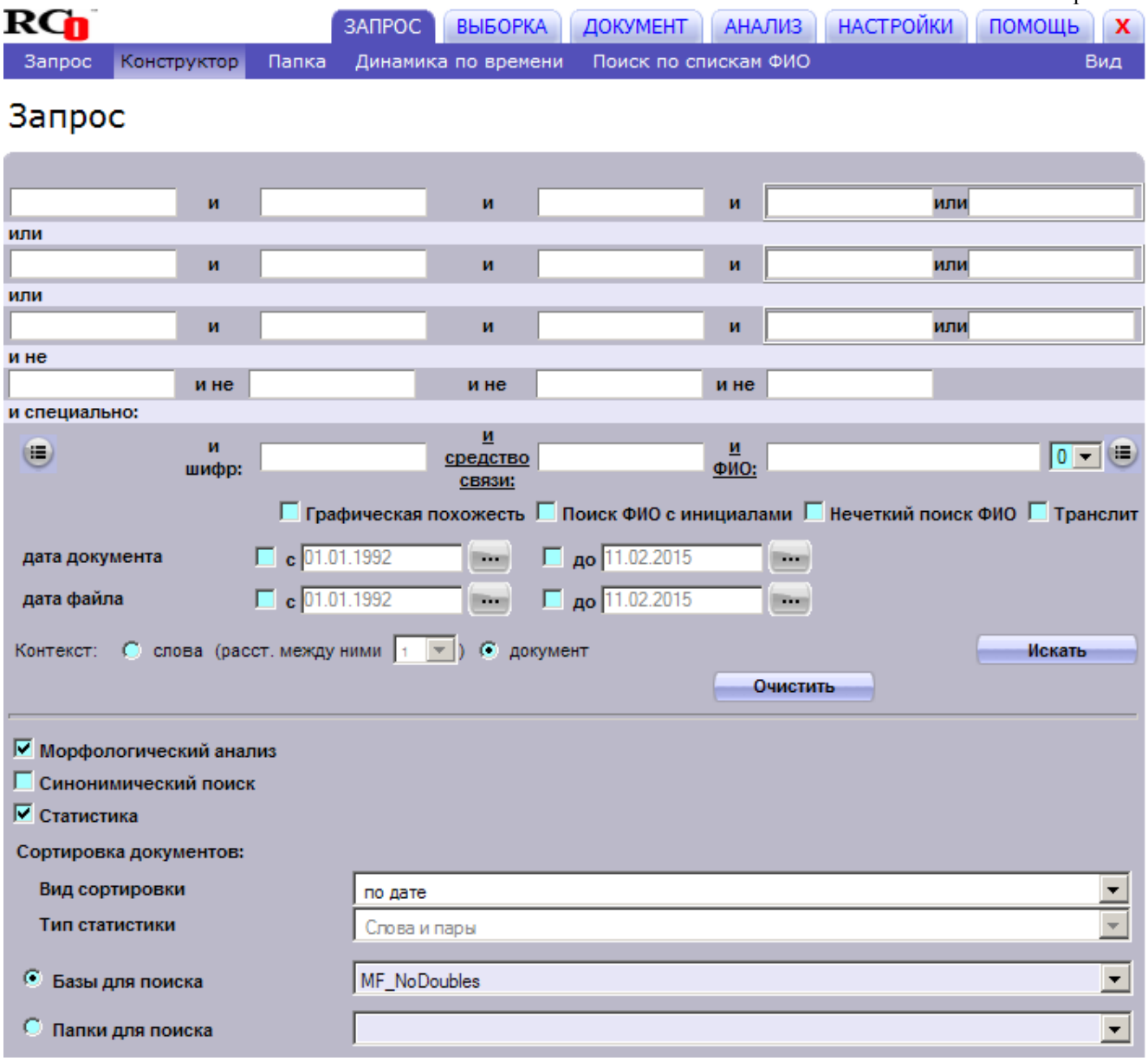


Рисунок 47. *Раздел «Конструктор»*

При составлении расширенного запроса слова записываются в соответствующие поля, соединенные между собой логическими операторами (см. пп. 5.4.1 – 5.4.4). Количество строк, объединяемых оператором «или» обычно ограничивается тремя (Рисунок 47). Если требуется иное, следует использовать поле «’или’ строк» (Рисунок 48) и выбрать из выпадающего меню количество в пределах от 1 до 25. На рисунке показано 5 строк ‘или’.

Примечание

Если на странице конструктора отсутствует поле «’или’ строк» (Рисунок 47), следует обратиться к администратору.

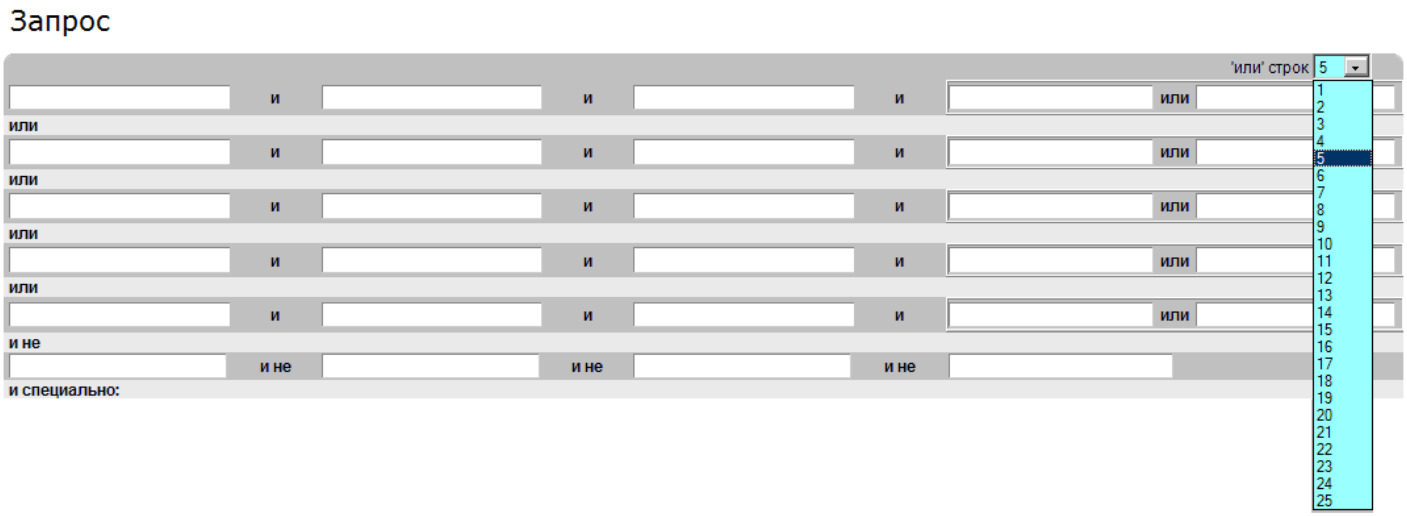


Рисунок 48. *Настройка количества строк «или»*

Дополнительные возможности расширения запроса предоставляются в средней части конструктора запроса под заголовком «и специально:». К таковым относятся возможности поиска телефонных номеров, фамилий, номеров документов (кодов, шифров) (см. п. 5.11). Кроме того, можно ограничить поиск временными рамками, основываясь на датах документов и файлов (что будет работать в случае, если база, в которой производится поиск, имеет поля даты документа/файла). В этой же части конструктора запроса можно задать расстояние между словами при контекстном поиске (кнопка «контекст» и выпадающее меню «слова (расстояние между ними)») или, если требуется поиск по всему документу, выбирается кнопка «документ». Для поиска по фамилиям может быть включён нечёткий поиск (см. п. 5.10).

На этой же странице можно включить режимы *«морфологический анализ», «синонимический поиск», «статистика»* и установить параметры *«сортировка документов», «тип статистики документов», «базы для поиска», «папки для поиска»*.

Поле «и шифр» (см. п. 5.11)***:***

- ✓ ***шифр***;
- ✓ ***номер паспорта***;
- ✓ ***номер закона/приказа, телеграммы***;
- ✓ ***адрес электронной почты***;
- ✓ ***ссылка интернета***.

Поле «и средство связи» (см. п. 5.11)***:***

- ✓ ***телефон*** – комбинация цифр, и других символов, составляющих телефонный номер;
- ✓ ***URL;***
- ✓ ***e-mail*** – адрес электронной почты;
- ✓ ***ники в соцсетях.***

Кроме того, существует возможность поиска документов, содержащих одновременно несколько средств связи. Подключение этой возможности осуществляется нажатием на надпись «и средство связи».

Примечание

Поля ввода списков средств связи и ФИО различны по цвету фона. Для списков осуществляется контроль допустимости вводимых символов.

Поле «и ФИО» (см. п. 5.11)***:***

- ✓ ***Фамилия, имя, отчество*** – комбинация фамилии, имени, отчества или инициалов искомого лица. Кроме того, существует возможность поиска документов, содержащих одновременно несколько фамилий. Подключение этой возможности осуществляется нажатием на надпись «и ФИО». Тот же эффект даёт кнопка , расположенная справа от поля ФИО.

Примечание

Поля ввода списков средств связи и ФИО различны по цвету фона. Для списков осуществляется контроль допустимости вводимых символов.

Поле «Дата документа»:

- ✓ ***вкл. (стоит флажок)*** – в процессе поиска будет проводиться анализ дат документов (если в базе, по которой ведётся поиск, есть поле даты документа);
- ✓ ***выкл.*** – поиск по дате документа выключен.

Поле «Дата файла»:

- ✓ ***вкл. (стоит флажок)*** – в процессе поиска будет проводиться анализ дат файлов (если в базе, по которой ведётся поиск, есть поле даты файла);
- ✓ ***выкл.*** – поиск по дате файла выключен.

Поля задания дат (расположены правее полей «Дата документа» и «Дата файла» и становятся активными при включении соответствующих им флажков):

- ✓ ***задана дата*** – в процессе поиска будет проводиться анализ дат документа/файла (если в базе, по которой ведётся поиск, есть поле даты документа/файла) и в выборку будут включены документы, чья дата больше или равна указанной;
- ✓ ***задана дата «до»*** – в процессе поиска будет проводиться анализ дат документа/файла (если в базе, по которой ведётся поиск, есть поле даты документа/файла) и в выборку будут включены документы, чья дата меньше или равна указанной.

Переключатель «контекст»/«документ»:

- ✓ ***в положении «контекст»*** – становится доступным выпадающее меню «слова (расстояние между ними)», позволяющее задавать расстояние между словами при контекстном поиске;
- ✓ ***в положении «документ»*** – поиск идёт по всему документу.

Поле «Графическая похожесть»:

- ✓ ***вкл. (стоит флажок)*** – в процессе поиска ФИО учитываются возможные опечатки (вместо русских букв в документах стоят похожие латинские буквы или цифры);
- ✓ ***выкл.*** – поиск по ФИО идёт, как введено в запросе (без учёта возможных опечаток).

Поле «Поиск ФИО с инициалами»:

- ✓ ***вкл. (стоит флажок)*** – в процессе поиска будет найдены фамилии, сопровождаемые как полными именами и отчествами, так и инициалами;
- ✓ ***выкл.*** – поиск по ФИО идёт, как введено в запросе (если в запросе написано полностью, ищется полностью, если в запросе введены инициалы, ищутся инициалы).

Поле «Нечеткий поиск ФИО»:

- ✓ ***вкл. (стоит флажок)*** – в процессе поиска будет найдены фамилии, близкие фонетически (по звучанию) или грамматически (по написанию) к заданной. Так, при нечётком поиске по фамилии Иванов будут найдены фамилии, написанные с ошибками: Ивонов, Иваноф, Ивановв;
- ✓ ***выкл.*** – поиск по ФИО идёт по установленным в системе общим правилам (как для остальных слов).

Поле «Транслит»:

- ✓ ***вкл. (стоит флажок)*** – в процессе поиска будет найдены русские фамилии, написанные латинскими буквами или наоборот (действует только в пределах поля ФИО). Так, при поиске со включённой опцией «Транслит» по ФИО Ivanov будут найдены документы, где упомянуты фамилии Ivanov и Иванов;
- ✓ ***выкл.*** – поиск по ФИО идёт по установленным в системе общим правилам (как для остальных слов).

Поле «Морфологический анализ»:

- ✓ ***вкл. (стоит флажок)*** – морфология включена, т.е. поиск будет осуществляться по всем возможным формам слов, указанным в запросе. Например, по запросу «банк» найдутся все его словоформы, в том числе «банка», «банки», «банками» и пр. Установка флажка влияет на все поля текстового типа;
- ✓ ***выкл.*** – морфология выключена.

Поле «Синонимический поиск»:

- ✓ ***вкл. (стоит флажок)*** – включен синонимический поиск, т.е. при указании в запросе одного из слов, имеющих синоним, автоматически выполняется поиск по всем синонимам (в том числе и многословным);
- ✓ ***выкл.*** – синонимический поиск выключен.

Поле «Статистика»:

- ✓ ***вкл. (стоит флажок)*** – статистика включена, т.е. в процессе поиска будет проводиться анализ документов и строиться таблица - подсказка;
- ✓ ***выкл.*** – статистика выключена.

Поле «Вид сортировки»:

- ✓ ***нет*** – функция отключена, документы в списке будут расположены в том порядке, в котором они находятся в базе
- ✓ ***по дате*** – в соответствии с датой документа, список найденных по запросу документов возглавляет самый свежий документ;
- ✓ ***по значимости*** – документы в списке будут расположены в порядке уменьшения значимости, список найденных по запросу документов возглавит документ, наиболее близкий к информационному портрету выборки;

- ✓ ***по имени файла*** – документы в списке располагаются в алфавитном порядке имён исходных файлов.

Поле «Тип статистики»:

- ✓ **---** – функция отключена;
- ✓ ***слова*** – в списке Главных тем будут отображены значимые слова;
- ✓ ***слова и пары*** – в списке Главных тем будут отображены значимые слова и пары слов.

Данное поле является доступным при установке ***«Вид сортировки» «по значимости»***.

Поле «Базы для поиска»: (например, тестовые базы СМИ)

- ✓ ***DemoBase***;
- ✓ ***SBase***.

Поле «Папки для поиска»:

- ✓ ***название папки, имеющейся у пользователя***.

Для того чтобы отобрать документы за определенную дату, установите флажок в поле *«дата документа»* и выберите из ниспадающих списков определенную дату. Ту же дату нужно выбрать в поле *«до»*. Если же нужно отобрать документы за определенный промежуток времени, установите флажки в полях: *«дата документа»* и *«до»*. Затем выберите из ниспадающих списков диапазон дат.

В том случае, если необходимо найти документы, в которых слова (элементы) запроса располагаются на определённом расстоянии друг от друга (например словосочетания), щелкните по полю *«контекст» «слова»*, а затем выберите, в пределах скольких слов могут располагаться элементы запроса (расстояние между словами). Например, если требуется отобрать документы, содержащие словосочетание «выборы президента», то установите флажок *«контекст»* и щелкните по полю *«расст. между ними»*, а затем выберите из списка 1 (т.е. включаем поиск словосочетания).

После составления запроса, выбора диапазона дат и установки необходимых параметров, щелкните по кнопке *«Искать»*. Система запустит поисковую машину.

Пример:

Необходимо отобрать все документы с 1 января 2005 года по 31 декабря 2005 года (диапазон – 1 год), в которых упоминается *генпрокурор Чайка* или *Устинов*, но отсутствует упоминание *Колесникова*.

Пример составления запроса: Рисунок 49.

RCI ЗАПРОС ВЫБОРКА ДОКУМЕНТ АНАЛИЗ НАСТРОЙКИ ПОМОЩЬ X

Запрос Конструктор Папка Динамика по времени Поиск по спискам ФИО Вид

Запрос

Генпрокурор и и и Чайка или Устинов
или
и и и или
или
и и и или
и не
Колесников и не и не и не
и специально:
и шифр: и средство связи: и ФИО: 0
Графическая похожесть Поиск ФИО с инициалами Нечеткий поиск ФИО Транслит
дата документа с 01.01.2005 до 11.12.2005
дата файла с 01.01.1992 до 11.02.2015
Контекст: слова (расст. между ними 1) документ Искать
Очистить
Морфологический анализ
Синонимический поиск
Статистика
Сортировка документов:
Вид сортировки нет
Тип статистики Слова
Базы для поиска USHSS
Папки для поиска

Рисунок 49. *Пример составления расширенного запроса*

5.9 Синонимический поиск

Программа предусматривает возможность синонимического поиска, т.е. поиска с учетом синонимов.

При указании в запросе одного из слов, имеющих синоним, автоматически выполняется поиск по синонимам (в том числе и многословным). При этом поддерживается выполнение как строгих, так и простых запросов.

Например:

Петроград;Ленинград;Санкт-Петербург;Петербург;Питер;Город на Неве
ДНК: дезоксирибонуклеиновая кислота

Включить синонимический поиск для всего запроса можно, поставив галочку в пункте *«Синонимический поиск» («Настройки» – «Поиск»)* (для всех запросов) или поставив галочку в поле *«Синонимический поиск»* (для текущего запроса). В простом запросе это единственный вариант.

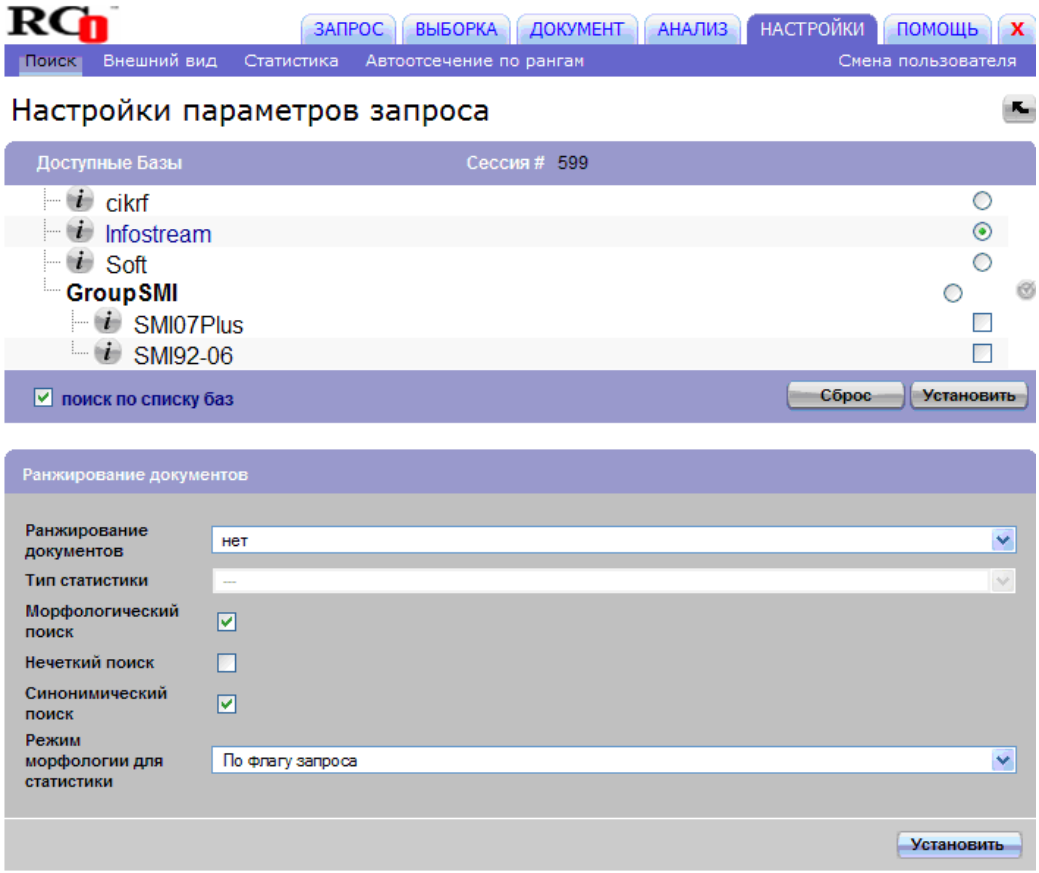


Рисунок 50. *Страница настроек*

В строгом запросе синонимический режим может быть включен/выключен индивидуально на каждом слове. Это осуществляется с помощью модификатора +(–)s – например, Ленинград%+s (включен синонимический поиск), ДНК%–s (выключен).

Для строгого запроса *Происшествия & Ленинград%+s* будут найдены те же документы, что и по запросу

Происшествия & (Ленинград%-s | Петроград%-s | Санкт-Петербург%-s | Петербург%-s | Питер%-s | ctx(Город%-s на%-s Неве%-s)

5.10 Нечеткий поиск

Программа предусматривает возможность поиска с учетом возможных орфографических ошибок в текстовых документах. Это так называемый нечеткий поиск. Включить этот режим можно, поставив галочку в пункте *«Нечеткий поиск»* (*«Настройки» – «Поиск» – «Сортировка документов»*).

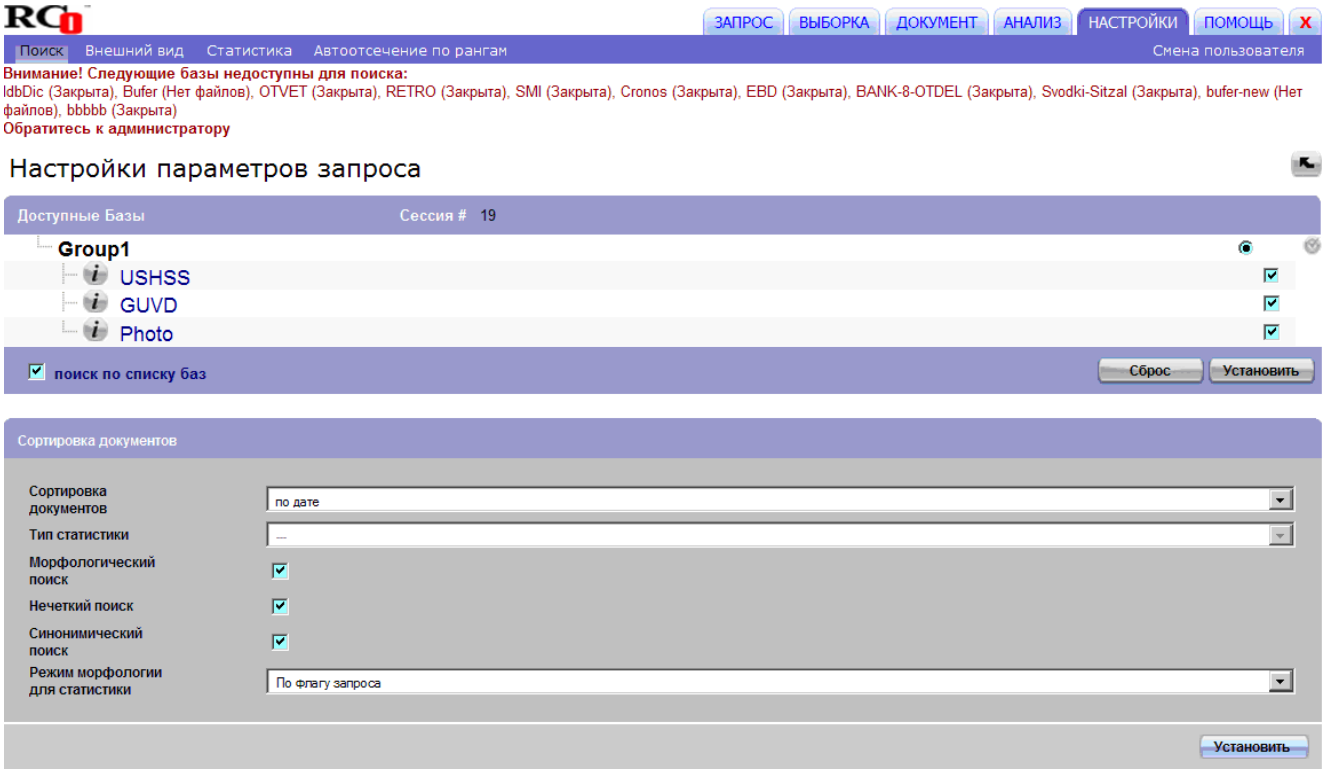


Рисунок 51. *Страница настроек*

Использование нечеткого поиска позволяет расширить количество найденных документов с учетом возможных допущенных в них ошибок.

Например, по запросу **Гвердцители** без подключения нечеткого поиска был найден 351 документ. Когда подключили нечеткий поиск, нашёлся 371 документ. Документов оказалось больше, т.к. теперь добавились случаи типа **Гвердтцители, Гвердцетели.**

5.11 Поиск телефонных номеров, фамилий, номеров документов

Поиск телефонных номеров. Для этого можно воспользоваться оператором *тел()/phone()* или соответствующим полем «и средство связи» в конструкторе запроса.

Например, по запросу **тел(“123-45-27”)** будет найдено существенно больше документов, содержащих интересующий нас номер. Это документы, в которых искомый номер представлен в формах:

123-45-67; 1234567; 495 123 45 67; 8 095 1 23 45 67; +7 (495) 123-45-67; 12-345-67; 123-4-567.

Можно усложнить условия поиска. Допустим, мы не помним какие-то цифры номера. В этом случае неизвестные цифры заменяются вопросительными знаками в запросе.

Например, по запросу **phone(“123 ?5 67”)** будут найдены документы, содержащие телефоны, отличающиеся только четвертой цифрой (причем в различных вариантах написания):

12-335-67; 8095 ·123-45-67; +7095 ·1232567; 8495 ·123-05-67; 123-55-67.

Поиск фамилий. Для поиска фамилий (а, при необходимости, имён и отчеств) в базе используется оператор *ФИО()/FIO()* или поле «ФИО» в конструкторе запроса. При этом все имена собственные вводятся в именительном падеже и единственном числе. Порядок задания слов в функции ФИО строгий: Фамилия Имя Отчество.

Например, по запросу **ФИО("Иванов")** в документах будет найдена только указанная фамилия, а имя «Иван», от которого она произошла, искаться не будет.

При поиске по фамилии и имени будут найдены документы, в которых имя встречается в указанном в запросе порядке, и документы, где имя стоит перед фамилией.

Например, по запросу **ФИО("Иванов Сергей")** будут найдены документы, содержащие следующие формулировки: «Иванов Сергей», «Иванов Сергей Алексеевич», «Сергей Иванов» и т.д.

Если нам нужно найти кого-то по имени и фамилии, но неизвестно, поименован этот человек в тексте полным именем или инициалами, можно найти все упоминания запросом **ФИО("Иванов П.~")** (символ "~" называется модификатором. Символ "~" можно отделять пробелом, но это не обязательно). В этом случае будут найдены:

Иванов П.И.; Петр Иванов; Иванов Павел; П.П. Иванов; Петр Петрович Иванов и т.д.

Заметим, что применение запроса **ФИО("Иванов Петр ~")** приведёт к такому же результату – будут найдены документы, как содержащие «Иванов Петр», так и «Иванов П.», т.е. по полному фамилии-имени будет искаться и сокращение (но, разумеется, при данном запросе Система уже не будет искать Иванова Павла).

Ещё более расширяет возможности поиска модификатор "?". Применять этот модификатор следует при поиске по имени-отчеству.

Например, по запросу **ФИО("Сидоров Петр ?")** будут найдены:
Петр Николаевич Сидоров, Петр П. Сидоров, Сидоров Петр, Сидоров Петр Романович и т.д.

Если есть необходимость одновременного поиска нескольких фамилий, в пределах функции ФИО можно перечислить их через вертикальную черту "|". В этом случае будет найдена любая из перечисленных фамилий.

Например, по запросу ФИО("Иванова|Петрова|Сидорова Мария Ивановна") будут найдены граждане, носящие имя и отчество «Мария Ивановна» и любую из перечисленных фамилий:

Иванова Мария Ивановна, Петрова Мария Ивановна, Сидорова Мария Ивановна

Такое использование функции ФИО может быть полезно при поиске по девичьей фамилии или если есть различные варианты написания имени.

Например, по запросу ФИО(«Алеутдинова Татьяна|Таньзиля») будут найдены:

Алеутдинова Таньзиля и Алеутдинова Татьяна

Поиск номеров документов (кодов, шифров). Не всегда очевидно, как машина разбивает на слова сложную конструкцию (номер паспорта, закона, приказа, телеграммы, адреса электронной почты или ссылки Интернета) на слова. Для поиска таких конструкций в базе используется функция *text()* или поле конструктора запроса «и шифр» (следует использовать строгую форму запроса):

text(www.v2004.ru)
text(vb2004@mail.ru)
text(№622-XII)
text(постановление N 12/81-3)

Использование специальных функций (поиск ФИО, кодов, дат) **в простом запросе.** Существует возможность подключения специальных функций даже в простом запросе. Для этого администратор должен соответствующим образом настроить Систему.

Дополнительные возможности поиска ФИО и телефонов.

Как уже говорилось выше (п.5.7) существует возможность поиска документов, содержащих одновременно несколько фамилий или нескольких телефонов. Подключение этой возможности для фамилий даёт кнопка , расположенная справа от поля ФИО (или кнопка *«и ФИО»*), дающая доступ к полю ввода списка фамилий. Для подключения списка телефонов нужно нажать кнопку *«и телефон»*. В обоих случаях появляется дополнительное поле списка (Рисунок 52). На рисунке показан случай со списком фамилий, что видно по синей рамочке, появившейся вокруг текста *«и ФИО»*.

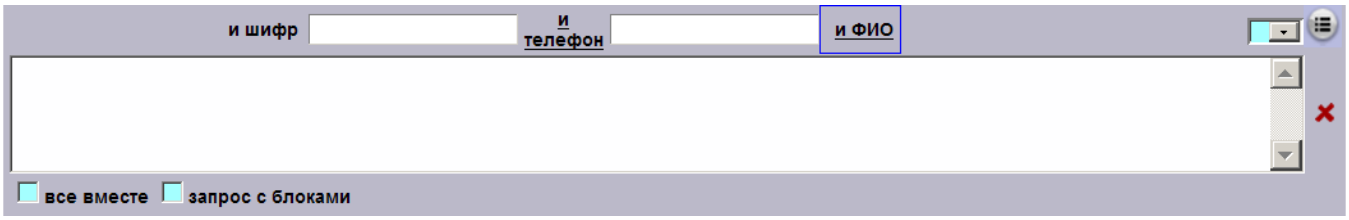


Рисунок 52. *Дополнительное поле для списка фамилий*

Поле «все вместе», появляющееся вместе с полем списка, влияет на представление результатов поиска. В случае, если в нём стоит галочка, происходит поиск документов, содержащих одновременно все фамилии (телефоны), перечисленные в списке (как если бы они были перечислены через оператор «и»). Иначе производится поиск документов, в которых фамилии (телефоны) встречаются по отдельности.

Примечание

Если установлена галочка в поле «все вместе», значение поля «запрос с блоками» игнорируется.

Поле «запрос с блоками». Если галочка не установлена, происходит обычный поиск фамилий (телефонов). Если галочка установлена, несколько видоизменяется способ представления выборки найденных документов. **Например,** рассмотрим запрос, состоящий из списка фамилий «Ельцин» «Гайдар» «Чубайс» (Рисунок 53)

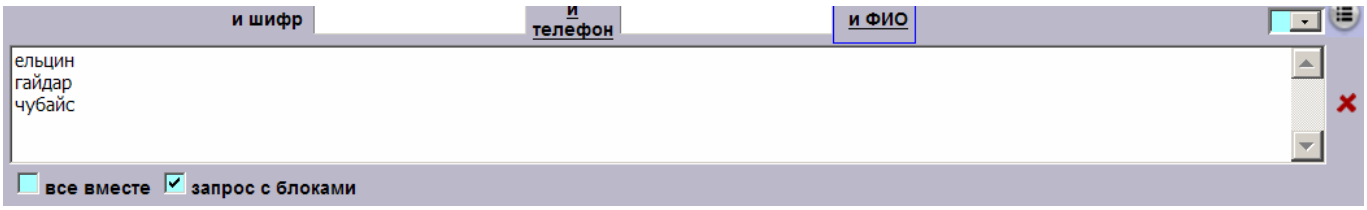


Рисунок 53. *Ищем по списку фамилий*

После проведения запроса на странице выборки появилась дополнительная кнопка Список блоков (Рисунок 54).

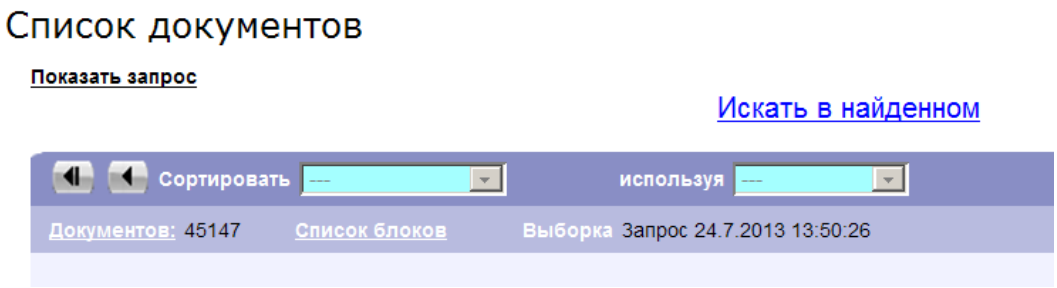


Рисунок 54. *Появился «Список блоков»*

Если нажать кнопку *«Список блоков»*, появится панель *«Навигация по блокам»* (Рисунок 55). На этой панели первая колонка – номер по списку, вторая – фамилия (телефон) из списка, третья колонка содержит количество документов, в которых найдены соответствующие фамилия (телефон). При этом, если документ содержит несколько элементов из списка, он будет учтён в каждом списке.

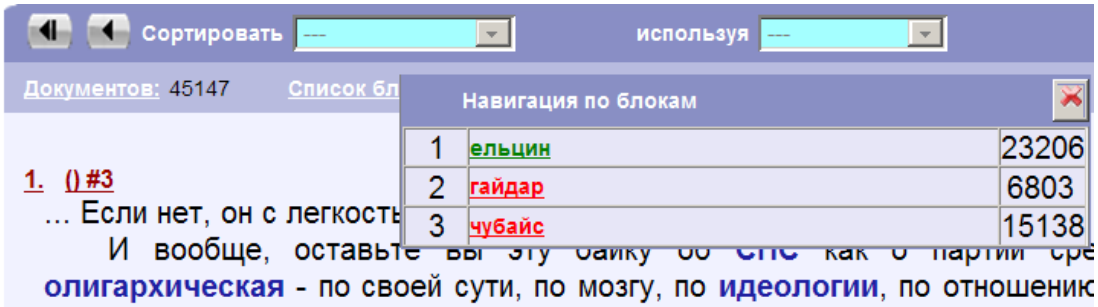


Рисунок 55. *Панель «Навигация по блокам»*

Цвет элемента списка во 2-й колонке зависит от того, какой блок найденных документов представлен в настоящий момент на странице выборки: предъявленные на странице окрашены зелёным, скрытые – красным. Если в блоке не содержится ни одного документа, соответствующий ему элемент списка будет серым.

При нажатии на другой элемент списка страница выборки обновляется. При этом выборка сдвигается к 1-го документу блока, в котором элемент найден. Порядок следования документов соответствует тому, в каком они были загружены в базу (сортировка другими способами не доступна).

5.12 Поиск по объектам

В некоторых документах содержится информация об объектах (источником такой информации может быть, например, обработка документов программой RCO FactExtractor), которая может быть использована при поиске. Объект – совокупность структурированных данных, связанных друг с другом, выделяемая из текста при загрузке (например, персона (перс), организация (орг), географическое название (гео)).

Пример: Если нам нужно найти человека по фамилии Путин, запрос будет выглядеть следующим образом:

OBJ(перс: Путин)

где OBJ – функция поиска по объектам, а перс – тип объекта (персона). Как уже подчёркивалось, искомый объект будет найден в тех документах, где он выделен до загрузки в Систему. При этом будут выделены не только те фрагменты текста, где фамилия Путин встречается напрямую, но и те, где упомянуты его имя и отчество, а возможно и «Президент Российской Федерации».

Например, в документе фигурирует не только фамилия, но и местоимения «он», «ему» (Рисунок 56).

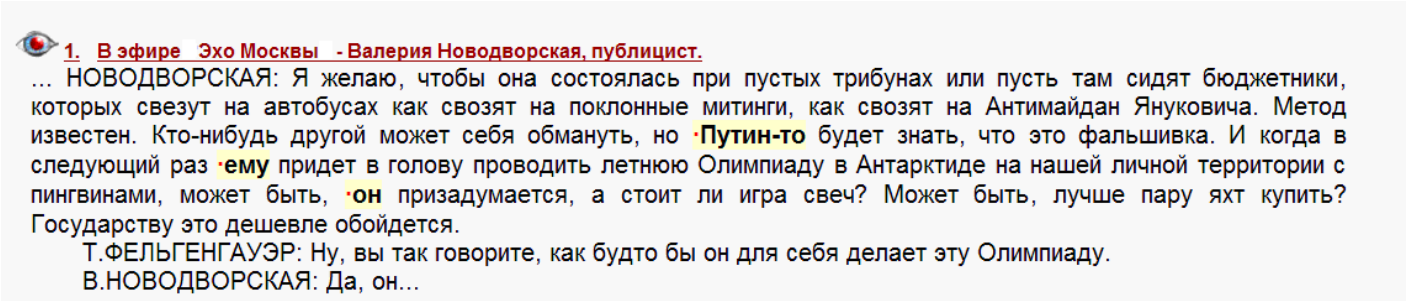

1. В эфире Эхо Москвы - Валерия Новодворская, публицист.
... НОВОДВОРСКАЯ: Я желаю, чтобы она состоялась при пустых трибунах или пусть там сидят бюджетники, которых свезут на автобусах как свозят на поклонные митинги, как свозят на Антимайдан Януковича. Метод известен. Кто-нибудь другой может себя обмануть, но ·Путин-то будет знать, что это фальшивка. И когда в следующий раз ·ему придет в голову проводить летнюю Олимпиаду в Антарктиде на нашей личной территории с пингвинами, может быть, ·он призадумается, а стоит ли игра свеч? Может быть, лучше пару яхт купить? Государству это дешевле обойдется.
Т.ФЕЛЬГЕНГАУЭР: Ну, вы так говорите, как будто бы он для себя делает эту Олимпиаду.
В.НОВОДВОРСКАЯ: Да, он...

Рисунок 56. *Документ с найденным объектом*

Для получения всех документов, в которых выделены персоны, нужно задать запрос со «звёздочкой» (таким образом будут найдены и подсвечены все объекты документов, имеющие тип «персона») (Рисунок 57)

Обк(перс:*)

Примечание:

Допускается (и рекомендуется) также аналогичный запрос без звёздочки. Его рекомендуется использовать, поскольку он выполняется быстрее

Обк(перс:)

13. Штурмовики майдана выходят в ночь

... за власть, теперь они вынуждены пожинать плоды такого сотрудничества. ·**Виталий** ·**Кличко** и ·**Арсений** ·**Яценюк** еще в июне прошлого года знали о том, что международный центр ·**Симона** ·**Визенталя** включил лидера "Свободы" ·**Олега** ·**Тягнибока** и другого видного деятеля партии ·**Игоря** ·**Мирошниченко** в список десяти топ-антисемитов мира в 2012 году. Они занимают пятое место вслед за египетскими "Братьями-мусульманами" и иранским режимом.

Центр обратился в госдепартамент США с рекомендацией запретить ·**Тягнибоку** и ·**Мирошниченко** пребывание на территории Штатов.

Выглядит парадоксальным, но стараниями все знающей, но закрывающей на неприглядные факты глаза либеральной оппозиции именно ·**Тягнибок** оказался на ·**майдане** среди ключевых борцов за европейские ценности. А сегодня...

Рисунок 57. *Документ с многочисленными объектами*

Особый интерес, однако, представляет не поиск отдельной персоны (её можно найти и просто по фамилии), а одного объекта во взаимодействии с другими.

В качестве примера рассмотрим запросы на совместный поиск объектов и текста. Их преимущество состоит в том, что заранее неизвестно, какой объект будет найден в результате:

near(obj(орг:) [5] (банкрот* | обанкрот*))

В результате мы получим информацию об обанкротившихся организациях (Рисунок 58)

9. Новый год может принести в Поднебесную дефолт

... Хубэй. Фонд был спасен с помощью гарантии от третьих лиц. Дефолтов такие фонды никогда не объявляли - власти Китая всегда делали все, чтобы не пострадали инвесторы.

На грани ·**банкротства** сейчас находится ·**фонд** ·**Jilin** ·**Trust**, который выдавал кредиты Liansheng Resources Group. Продавал этот продукт Строительный банк Китая, второй по величине активов банк страны. Liansheng стал банкротом. Вкладчики фонда, видимо, не получат свои деньги целиком.

Появилось сообщение о еще одном возможном ·**банкротстве**. ·**Инвестиционная** ·**компания** ·**Beijing** ·**Roll-In** ·**Investment** Co продавала свой продукт по управлению богатством. Средства от его продажи планировалось вложить в строительство нескольких жилых зданий при поддержке государства. Всего в фонде было порядка...

Рисунок 58. *Документ с банкротствах*

near(obj(орг:) [5] obj(перс:) [3] (директор | президент | председатель | управляющий | топ-менеджер | вице-президент | замгендиректора | замдиректора | замначальника | замруководителя | заместитель | председатель | главврач | шеф | начальник | руковод*))

Запрос позволит нам найти руководителей организаций-объектов (Рисунок 59)

1. Победителей премии "Марка №1 в России" наградили в Кремле <p> МАРКА №1В РОССИИ
... Сергея Трофимова, Жасмин.
ФОТО:
Александр Островский, основатель Независимой ·**лаборатории** "ИНВИТРО"
·**Владимир** ·**Черников**, ·**руководитель** ·**Департамента** средств массовой информации и рекламы города Москвы
·**Мария** ·**Гуржова**. ·**директор** по маркетингу ·**ТД** "·**Аскона**"
·**Александр** ·**Кузнецов**, исполнительный ·**директор** Северо-Западного ·**Регионального** ·**отделения** фабрики ·**Валерия**
Алексей Куон, ·**директор** по продажам и маркетингу ООО "Хенкель Баутехник"
Валерий ·**Леонтьев**
·**Темур** ·**Аминджанов**, ·**президент** ·**НПО** "·**СтарЛайт**"
·**Наталия** ·**Прокопьева**, ·**председатель** совета директоров ·**фармацевтической** ·**компании** "·**Эвалар**"
·**Антон** Касьяненко, ·**начальник** отдела маркетинга департамента безрецептурных препаратов ·**компании** "·**ШТАДА** ·**Маркетинг**"
Мина ·**Хачатрян**, ·**директор** по маркетингу ·**компании** "·**Голдер** - ·**Электроникс**" (бренды VITEK, Maxwell, Rondell)
·**Алексей** ·**Абрамейцев**, ·**председатель** правления АКБ "·**ЗАО** "·**РУССЛАВ-БАНК**"
Лев Лещенко
Владимир Белинский, бренд-менеджер компании "Русскии Алкоголь...

Рисунок 59. *Руководители организаций*

near(obj(гео:) [5] наводнение)

near(obj(гео:) [5] пожар)

near(obj(гео:) [5] взрыв)

Эти три запроса показывают нам документы, рассказывающие о различных бедствиях, с привязкой к географическим объектам.

Допускается применение логических выражений над объектами, например запрос

рдм((obj('орг:' и не 'орг:РИА*' и не 'орг:МЧС*')) [5] пожар)

найдёт документы про пожары, случившиеся в любой организации, за исключением тех документов, в которых рядом (в пределах 5 слов) с названием организации упоминаются МЧС и РИА (Рисунок 60)

17. Пожар парализовал работу воронежского мясокомбината Talex как минимум на год
... ·**Пожар** парализовал работу ·**воронежского** ·**мясокомбината** ·**Talex** как минимум на год
Воронеж. 27.01.2014. ABIREG.RU - Произошедший 25 января в Воронежской области ·**пожар** на ·**мясокомбинате** "·**Верхнехавский**" (входит в ГК Talex) полностью парализовал работу предприятия как минимум на год, сообщили "Абирегу" в пресс-службе компании.
Отметим, что ЧП произошло 25 января в около пяти часов утра. По сообщению представителя компании, в результате пожара уничтожено 1,5 тыс. кв. м. производственных площадей.
"Сейчас предприятие полностью уничтожено. Причины возгорания, а также ущерб устанавливаются. Однако мы надеемся, что в течение года мы сможем не только заново построить предприятие, но увеличить его мощности", - заключил собеседник "Абирега...

Рисунок 60. *Документ про пожар*

Кроме объектов в документах могут быть выделены (при помощи того же RCO FactExtractor или другим способом) связи данных и понятия, именуемые концептом. В качестве таковых обычно выделяются абстрактные понятия: финансы, здоровье, информация и т.д.

Например, по запросу

кнц(министр)

был найден документ, в котором содержатся упоминания о министрах

14. Давос играет в хоккей

... на подрастающее поколение, и заканчивая глобальными финансовыми прогнозами и головоломной задачей о том, как обеспечить устойчивое развитие всего населения нашей планеты.

Перед российской делегацией стоят свои задачи. В ее состав вошли вице-премьеры Аркадий Дворкович (руководитель делегации) и Дмитрий Козак, ·**министр** ·**экономического** ·**развития** Алексей Улюкаев, ·**министр** ·**связи** Николай Никифоров, первый заместитель председателя Центрального банка России Ксения Юдаева, а также ряд ·**заместителей** ·**министров** и руководителей российских регионов. Как всегда, на форуме будет представлено значительное число российских компаний.

"Наша программа сформирована исходя из приоритетных для нас направлений работы, - рассказал накануне Дворкович. - Прежде всего, что касается участия Дмитрия Николаевича Козака, будут организованы изначально запланированные...

Рисунок 61. *Документ с концептами «министр»*

Здесь также применим – для поиска всех концептов запрос кнц(*).

ГЛАВА 6. Методические приемы проведения поиска

В этой главе мы расскажем Вам о том, каким образом можно повысить эффективность поиска и отбора необходимой информации, используя соответствующие приемы. Вы научитесь создавать фильтры, сортировать документы по тематическим папкам, легко ориентироваться в незнакомой предметной области.

6.1. Кластеризация

Если Вы плохо ориентируетесь в предметной области, по которой осуществляете поиск информации, то после проведения основного запроса (перед его уточнением) можно воспользоваться функцией *кластеризации*, которая поможет в выборе объекта для уточнения запроса. Под ***кластером*** понимается группа близких, связанных между собой понятий.

Кластеризация объектов (слов и/или словосочетаний) *инфопортрета* производится таким образом, что в кластеры объединяются тематически близкие объекты. Близость объектов по теме определяется как *расстояние* между ними. *Расстояние* вычисляется на основе алгоритмов математической статистики.

Кластеризация может осуществляться как по отдельным спискам (словам или словосочетаниям), так и по объединенной статистике.

Примечание

Кластеризация работает только на одной базе.

Перед тем, как использовать функцию *кластеризации*, необходимо провести запрос. Для этого воспользуйтесь примером из предыдущих глав – запрос *иранская программа* (см. п. 3.2).

Откройте *инфопортрет* полученной выборки. Слева от каждого объекта (слова или словосочетания) находится значок «шарика» черного цвета . Выберите слово (словосочетание), которое определяет интересующую тематику и относительно которого будет производиться *кластеризация*, например, *«ядерное досье»*. Щелкните по «шарику» рядом с этим словом (словосочетанием) (*см.* Рисунок 62*)*. Это слово (словосочетание) называется *центром кластера* и обозначается значком .

Цвет «шариков» указывает на степень близости соответствующего объекта (слова или словосочетания) к центру кластера. Цвета «шариков», определяющие степень близости объекта к определенной теме (по убыванию): красный (наиболее близкий к теме), желтый, зеленый, синий, фиолетовый (наименее близкий к теме). Если «шарики» остаются окрашенными в черный цвет, то соответствующие им объекты никаким образом не относятся к рассматриваемой теме.

При наведении мышью на «шарик» возникают цифры процентов близости (величина расстояния) объекта к выбранному *центру кластеризации* (обозначенному ✖). Соответственно, чем ближе понятия между собой, тем меньше будет подсвечиваемая цифра.

Включение и отключение режима кластеризации производится на странице «Настройки» – «Внешний вид» – панель «Инфопортрет» галочкой «Использовать кластеризацию». Если необходимо отменить кластеризацию, следует воспользоваться кнопкой Сброс под или над таблицей инфопортрета.

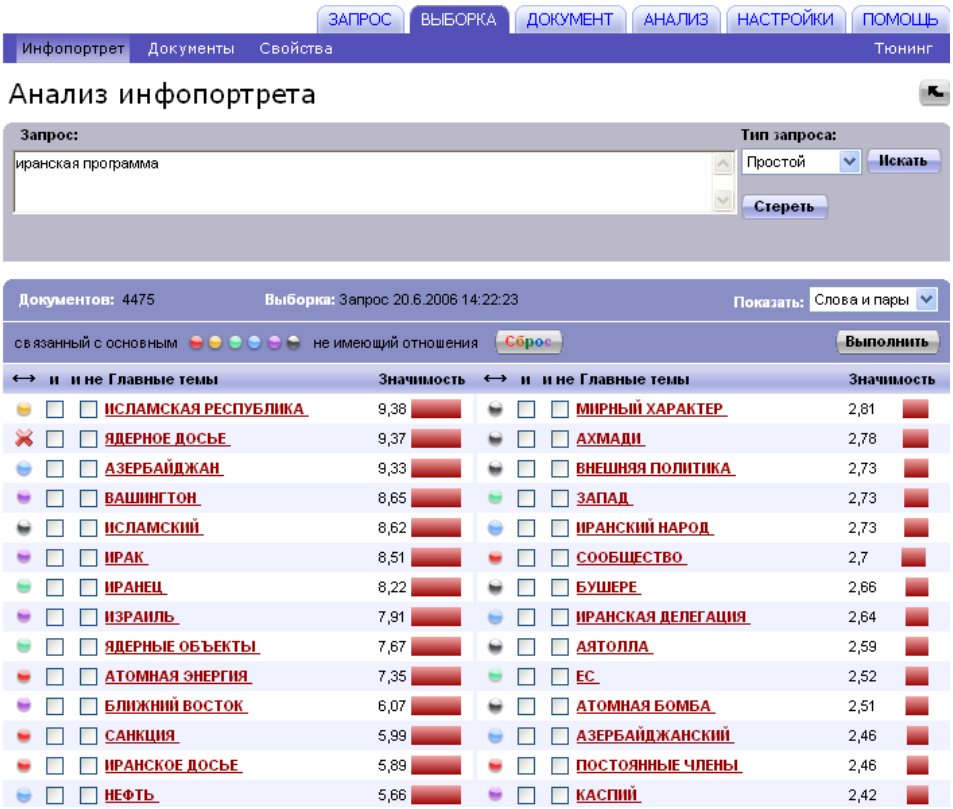


Рисунок 62. *Пример кластеризации, центр кластера – «ядерное досье»*

Примечание

В ряде случаев при «плохой» статистике (малый объём выборки, документы из разных областей, маленькие документы) *кластеризация* перестаёт работать. При этом выдаётся сообщение «Не получены кластерные данные», далее кластеризация не работает.

6.2. Папки. Страница «Запрос» – «Папка»

Занимаясь исследованиями нескольких предметных областей, удобно для каждой из них завести свой каталог, в котором будут храниться документы, относящиеся к одной тематике.

6.2.1. Назначение папок

В Системе реализован механизм создания виртуальных папок.

Папка – функциональная возможность Системы совместно хранить ряд документов.

Использование папок позволяет решить задачу распределения документов по темам.

6.2.2. Создание папок

Чтобы создать папку, перейдите в раздел *«Папка» (см.* Рисунок 63*)*. В поле *«Имя папки»* введите название папки, например *Папка1*, и нажмите кнопку *«Создать» (см.* Рисунок 64*)*. Итак, Вы создали собственную папку.

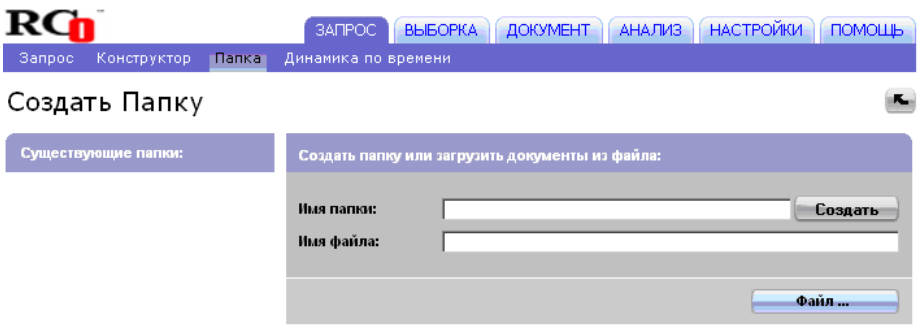


Рисунок 63. *Страница «Запрос» – раздел «Папка»*

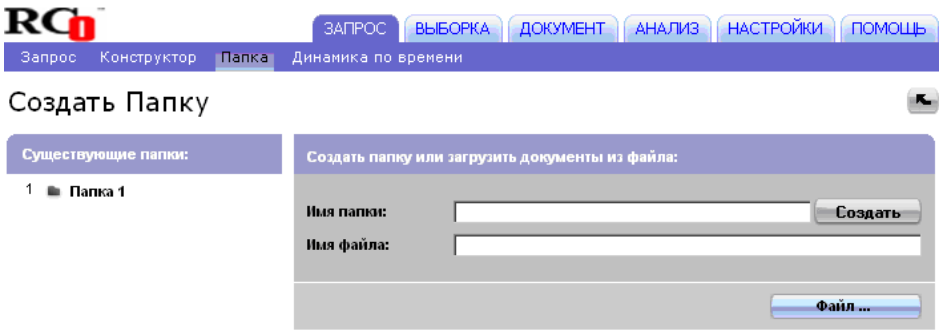


Рисунок 64. *Папка создана*

6.2.3. Работа с папками

Отправлять документы в папку можно двумя способами:

1. Вручную с помощью значка *(см.* Рисунок 65*)*. После проведения запроса на странице *«Выборка»* – раздел *«Документы»* ниже автореферата документа расположен значок . Нажмите на этот значок, выберите нужную папку, установив флажок в соответствующем поле, щелкните *«Добавить»*. Таким образом, документ оказывается в выбранной Вами папке. Кнопка *«Отказаться»* служит для того, чтобы закрыть окно без изменений состояния папок. Для того чтобы убрать документ из папки, нажмите значок и в

появившемся списке выберите папку, из которой он должен быть исключён. Затем щелкните *«Добавить»*.

В режиме просмотра папки для удаления документа поставьте флажок слева от заголовка документа и выберите в разделе *«Манипуляции»* пункт *«Удалить»* (Рисунок 66).

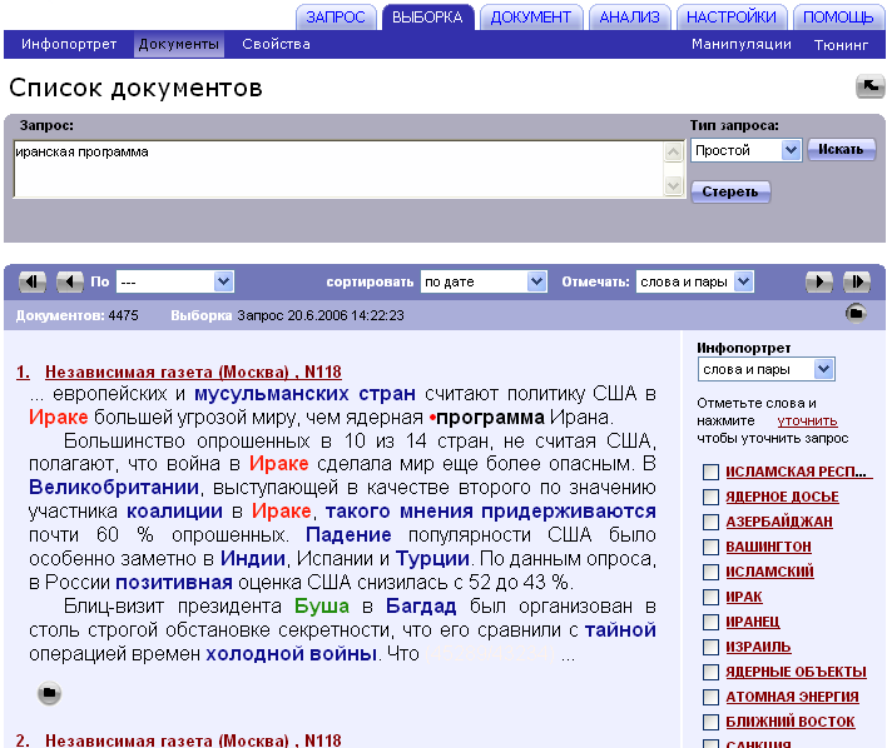


Рисунок 65. *Ручной режим переноса документов в папку помощью значка «папка»*

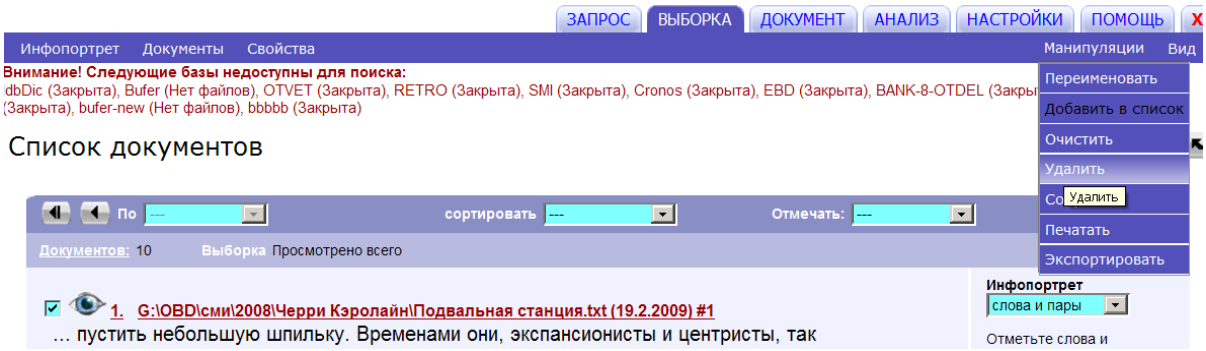


Рисунок 66. *Удаление документа из папки*

2. В автоматическом режиме на странице *«Выборка»* – раздел *«Свойства» (см. п. 4.4)*. На панели *«Перенос документов в папку»* выберите папку, в которую будет осуществлен перенос документов, а также количество переносимых документов – от 10 до 1000 первых документов. После установки этих параметров, нажмите кнопку *«Объединить»*. К сожалению, количество документов, помещаемых в папку, сдерживается ограничениями на размер папки.

Папку с документами можно сохранить на диске. Для этого после создания и заполнения папки перейдите на страницу *«Анализ»* - раздел *«Исследования» (см.* Рисунок 67*)* и щелкните по значку напротив названия папки, выберите каталог для сохранения и нажмите *«Открыть»*. Таким образом, вы сохраните папку в выбранном каталоге на диске.

Для того чтобы просмотреть содержимое папки, щелкните по значку слева от ее названия, Система автоматически перейдет на страницу *«Выборка»* и отобразится список документов, помещенных в папку.

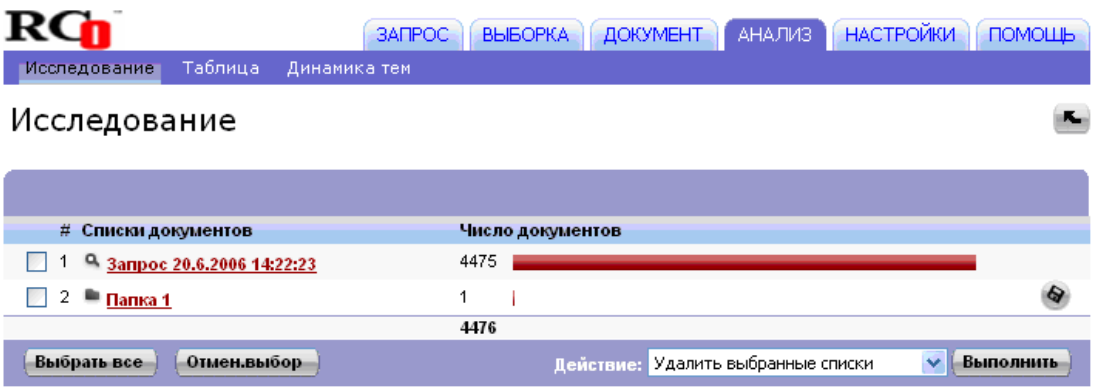


Рисунок 67. *Страница «Анализ» - «Исследования»*

6.2.4. Сортировка выборки по папке или запросу

С помощью папки можно проводить «гибкую» сортировку (ранжирование) документов.

На странице *«Выборка»* (раздел *«Свойства»*), на панели *«Отсортировать документы»* выберите в меню *«Сортировка» «по папкам»*. Нажмите на ссылку *«Выбрать папки»*. Система откроет таблицу *«Выборки и операции над ними»* с ранее проведенными Вами запросами. В соответствующих полях, слева от названий папок и запросов, отметьте галочкой ту папку или запрос, по которому хотите провести сортировку документов. Выберите *«Ранжировать»*, из списка *«Действие»*, расположенного ниже таблицы выборок. Нажмите *«Выполнить»*.

Способ сортировки по папке или запросу является гибким механизмом *уточнения* запроса. Т.к. при этом количество документов не изменяется, а происходит их сортировка таким образом, что список документов возглавляет документ, наиболее близкий к тематике папки (запроса), по которой происходит сортировка.

Пример: необходимо отобрать документы, касающиеся деятельности г-жи Кондолизы Райс относительно иранской ядерной программы. Для этого отсортируем запрос *«Кондолиза Райс»* по запросу *«иранская ядерная программа»*. Аналогично, можно провести ранжирование по папке *«иранская ядерная программа»*, предварительно создав ее и отправив туда документы по данной теме.

Примечание

Перед тем, как проводить ранжирование по запросу, необходимо сначала провести сам запрос.

На странице *«Настройки»* в поле *«Сохранять новый запрос»* установите флажок, щелкнув в соответствующем окошке. При этом каждый проведенный запрос будет сохраняться на странице *«Анализ»*. Таким образом, Вы сможете просмотреть любой проведенный Вами запрос, щелкнув по соответствующей ссылке в таблице запросов.

Проведите последовательно два запроса (со страницы «Запрос»):

иранская и ядерная и программа

кнт(кондолиза райс)

Затем перейдите на страницу *«Выборка» – «Свойства»* запроса *«Кондолиза Райс»*. В закладке *«Отсортировать документы»* щелкните по ссылке *«Выбрать папки»*. Затем в открывшейся таблице отметьте флажком выборку *«иранская ядерная программа» (см.* Рисунок 68*)*, по которой будет производиться сортировка, в поле *«Действие»* установите *«Ранжировать»* и, нажмите *«Выполнить»*.

Рисунок 68. *Страница «Исследование» выбор папки (запроса) для ранжирования*

В результате Вы получите выборку *«Кондолиза Райс»*, отсортированную по запросу *«иранская ядерная программа» (см.* Рисунок 69*)*.

Первыми в списке окажутся документы, речь в которых идет об иранской ядерной программе и связанной с ней деятельностью г-жи Кондолизы Райс.

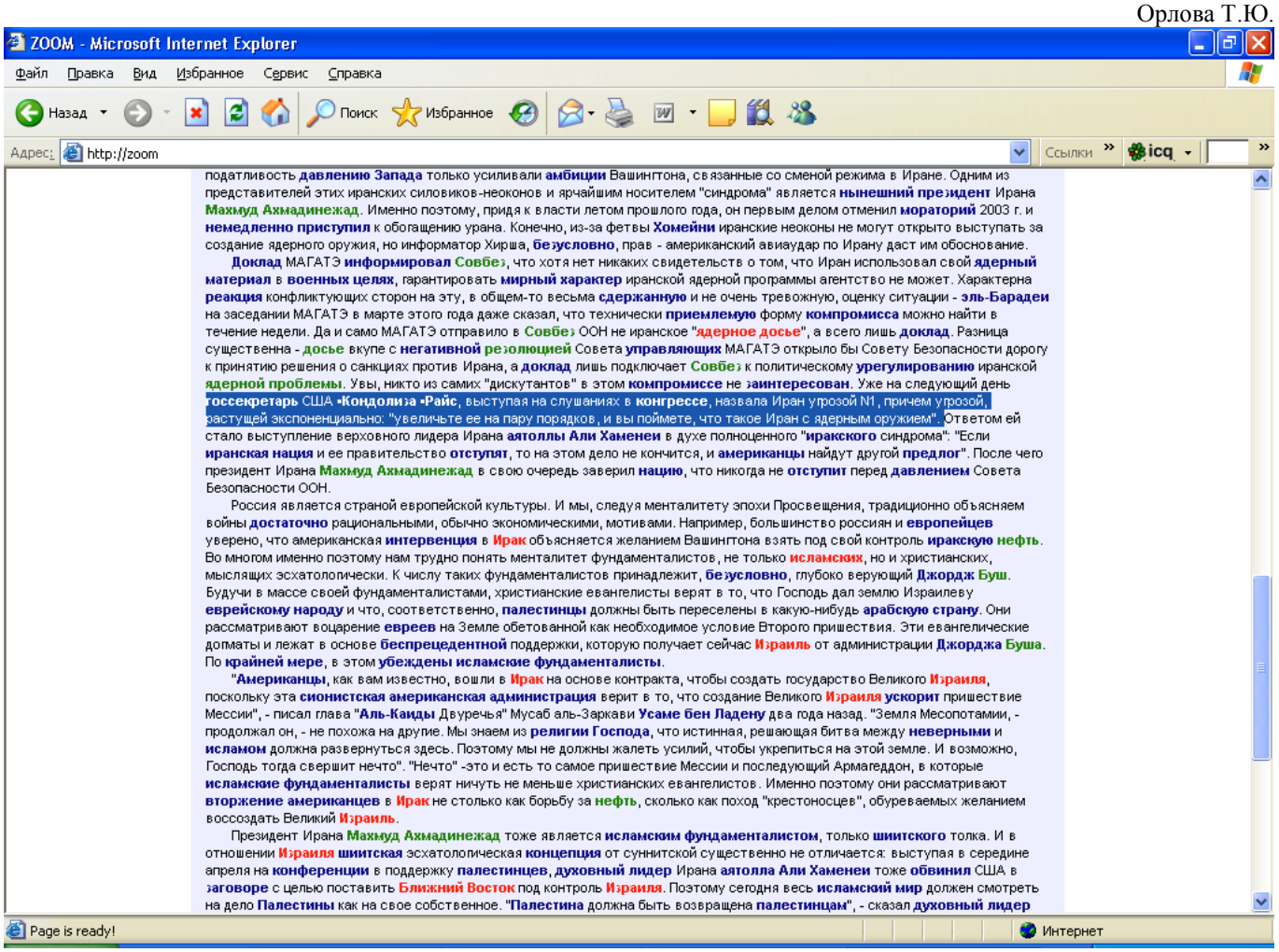


Рисунок 69. *Первый документ в выборке «Кондолиза Райс», отсортированной по запросу «иранская ядерная программа»*

ГЛАВА 7. Анализ

В предыдущих главах мы ознакомили Вас со всевозможными приемами проведения эффективного поиска, Вы научились составлять запросы по правилам языка запросов, создавать папки, использовать фильтры, сортировать документы различными способами для отбора нужной информации.

В этой главе мы хотим рассказать Вам о возможностях Системы, связанных с анализом информации, таких как: построение публикационного рейтинга, проведение сравнительного и динамического анализа.

7.1. Общие положения. Структура аналитического исследования

Любое аналитическое исследование предполагает следующие этапы:

1. Постановка задачи:
   - определение темы – объекта и предмета исследования;
   - определение цели исследования.
2. Сбор и поиск необходимой информации.
3. Анализ:
   - общее (общеизвестное, общепринятое);
   - частное (индивидуальное, свойственное, несвойственное, позитив/негатив, фактография);
   - связи – информационное окружение объекта (внутренние, внешние, явные, неявные);
   - прогноз;
   - выводы и заключения;
   - предложения.

В предыдущих главах речь шла о первых двух этапах аналитического исследования, особенно подробно мы рассмотрели темы, касающиеся поиска и отбора нужной информации.

В этой главе мы расскажем Вам о последнем этапе, а именно, о самом *анализе*.

Итак, после того, как предметная область четко определена, найдена вся необходимая информация, можно переходить к непосредственному исследованию объекта. Разработка гипотез, а также их дальнейшее подтверждение или опровержение является неотъемлемой частью аналитической работы.

Как правило, в ходе исследования объекта формируется несколько категорий – *«Общее», «Частное», «Связи»*. Информацию об этих категориях можно получить как из *информационного*

портрета объекта (см. п. 3.7), так и в ходе проведения *сравнительного анализа* нескольких объектов.

После проведения *сравнительного анализа*, для полноты исследования, необходимо сделать *прогноз*. Проследить тенденции развития главных тем во времени и сделать прогноз дальнейшего развития ситуации можно в результате проведения *динамического анализа главных тем*.

Часто бывает необходимо сравнить один и тот же объект на различных временных отрезках. Такое сравнение осуществляется в Системе при построении *публикационного рейтинга*.

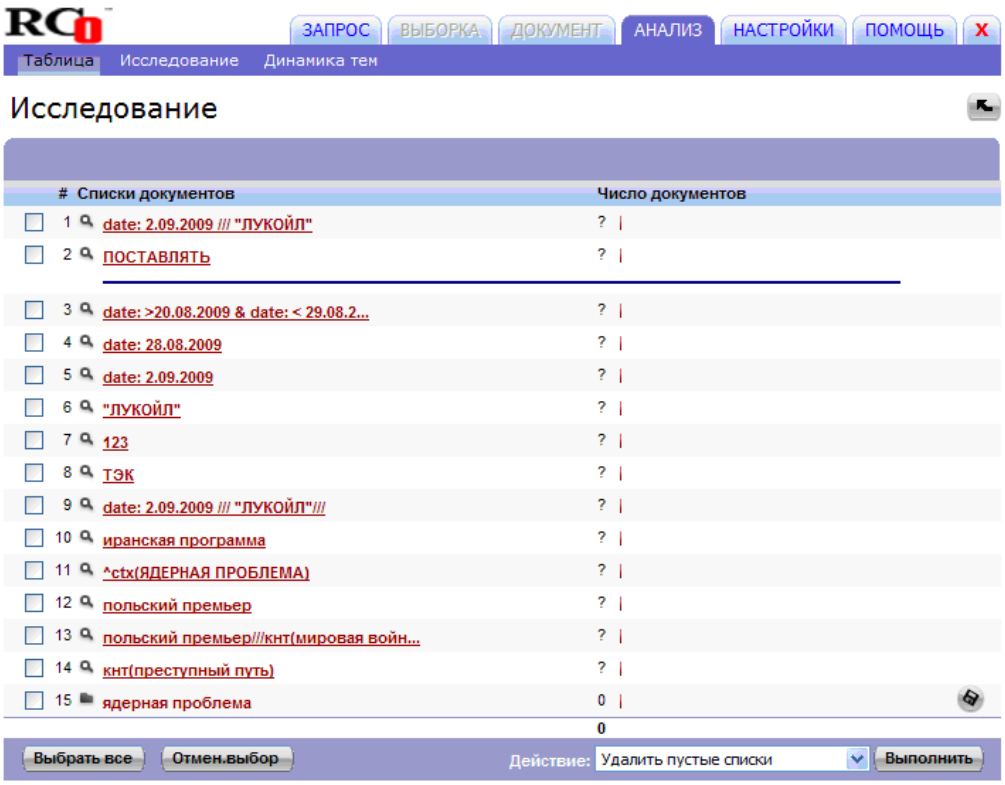


Рисунок 70. *Вид страницы «Анализ» – «Таблица» при входе в систему*

7.2. Страница «Анализ» – «Таблица»

На странице *«Анализ» – «Таблица»* мы можем просмотреть список сохраненных выборок. Таблица состоит из следующих столбцов:

- ✓ поля для отметки выборок для проведения над ними каких-либо действий;
- ✓ порядковый номер выборки;
- ✓ значок, означающий запрос (🔍) или папку (📁). При нажатии на него осуществляется переход к списку документов;
- ✓ название выборки. При нажатии на него осуществляется переход к инфопортрету;
- ✓ количество документов в выборке в числовом и графическом представлении соответственно;

- ✓ значок «Сохранить папку» ( ) для папок. Нажатие на него позволяет сохранить папку в виде файла на диске.

При входе в Систему количество документов в таблице не отображается. Вместо чисел Вы видите «?» для запросов и «0» для папок (см. Рисунок 70). Для отображения количества документов выделите интересующие вас выборки (флажками в первом столбце), а в поле *«Действие»* (справа внизу таблицы) выберите пункт *«Восстановить выбранные списки»* и нажмите кнопку *«Выполнить»*. (см. Рисунок 71)

Для выделения всех выборок служит кнопка *«Выбрать все»*, для удаления всех флажков служит кнопка *«Отмен.выбор»*.

Кроме восстановления можно удалить выбранные списки (*«Действие» – «Удалить выбранные списки»*) и удалить пустые списки (*«Действие» – «Удалить пустые списки»*).

Доступна возможность очистки всей истории просмотра ((*«Действие» – «Очистить историю просмотра»*). Это рекомендуется делать время от времени во избежание переполнения папки «Просмотрено всего», т.к. Система не может хранить в папках большие списки документов (примерное ограничение размера – 10000 документов. При переполнении папки она пропадет из сохраненной сессии). После очистки истории просмотра очистится и папка «Просмотрено в сессии».

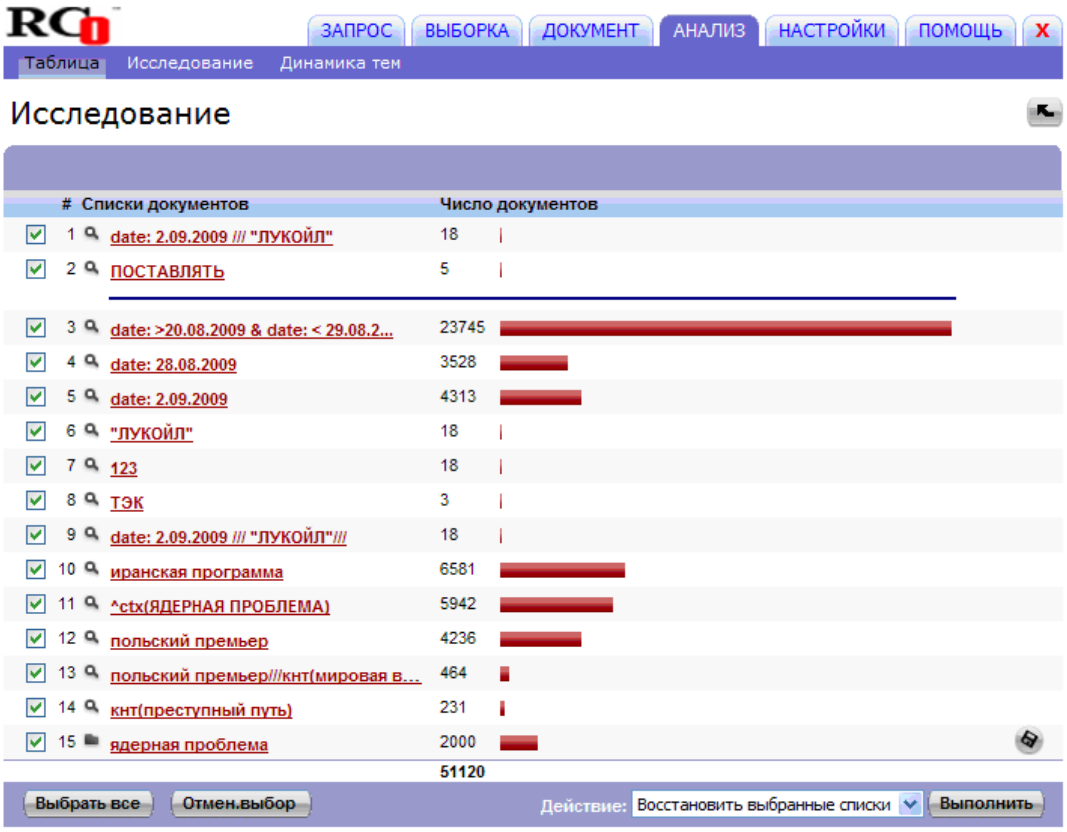


Рисунок 71. *Страница «Анализ» – «Таблица»*

7.3. Работа с «важными» выборками

Сохраненные выборки (см. страницу *«Анализ»*) можно разделить на две группы: «обычные» и «важные». В списке выборок эти две группы разделены чертой: сверху – «важные», снизу – «обычные». Группа «важных» выборок нужна для защиты от случайного стирания. Как это работает? Если установлены режимы «Сохранять новый запрос автоматически» и «Всегда добавлять в конец списка запроса» *(«Настройки»* – блок *«Анализ»*), и достигнут максимальный объем списка выборок (сейчас это 52 элемента), то для сохранения новых данных автоматически удаляются самые давние выборки из числа «обычных». «Важные» же выборки можно удалить из списка сохраненных только ручной маркировкой.

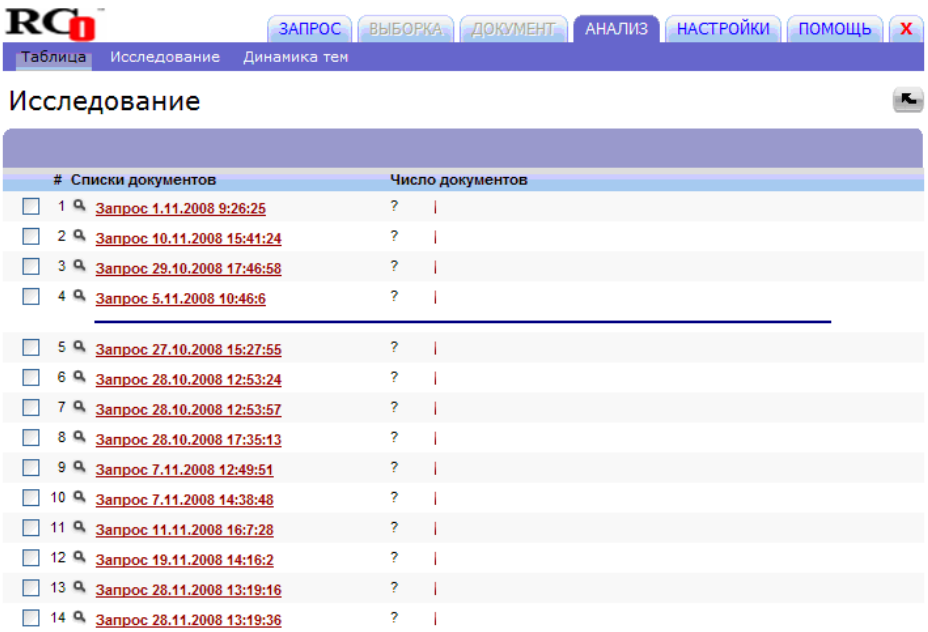


Рисунок 72. *«Важные» и «обычные» выборки*

Чтобы сделать выборку «важной», нужно выделить ее номер нажатием левой кнопки мыши и, не отпуская кнопку, перетащить вверх через разделительную черту. Все нажатия/отпускания кнопки мыши необходимо осуществлять в пределах колонки с номером сохраненной выборки.

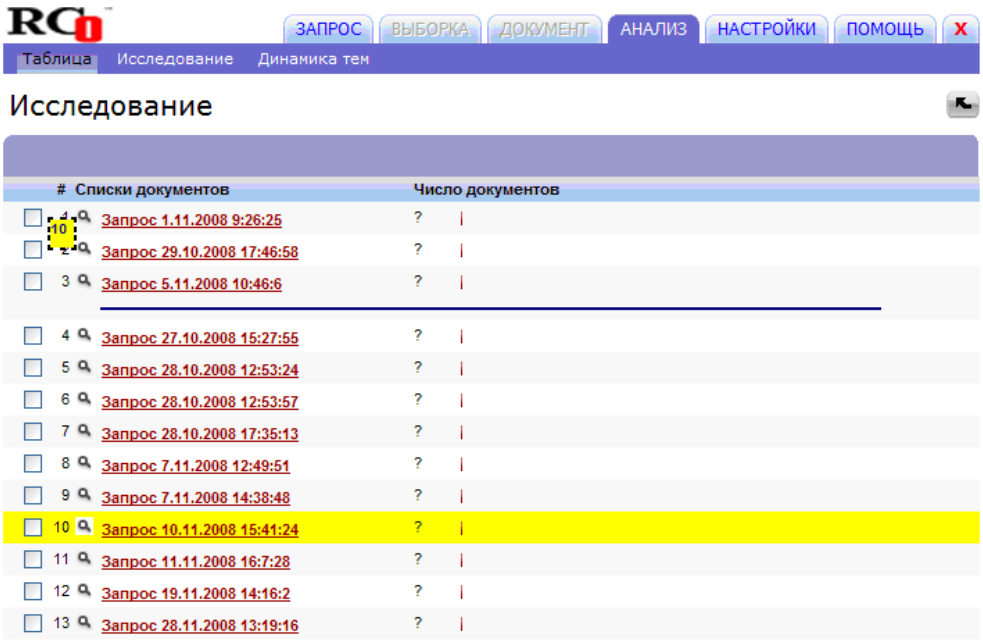


Рисунок 73. *Добавление «важной» выборки*

Группа «важных» выборок появляется при включении в нее первого элемента. Для этого нужно выполнить действия как при обычном добавлении, но образ выбранной выборки нужно установить над символом # (номер) в шапке таблицы сохраненных выборок.

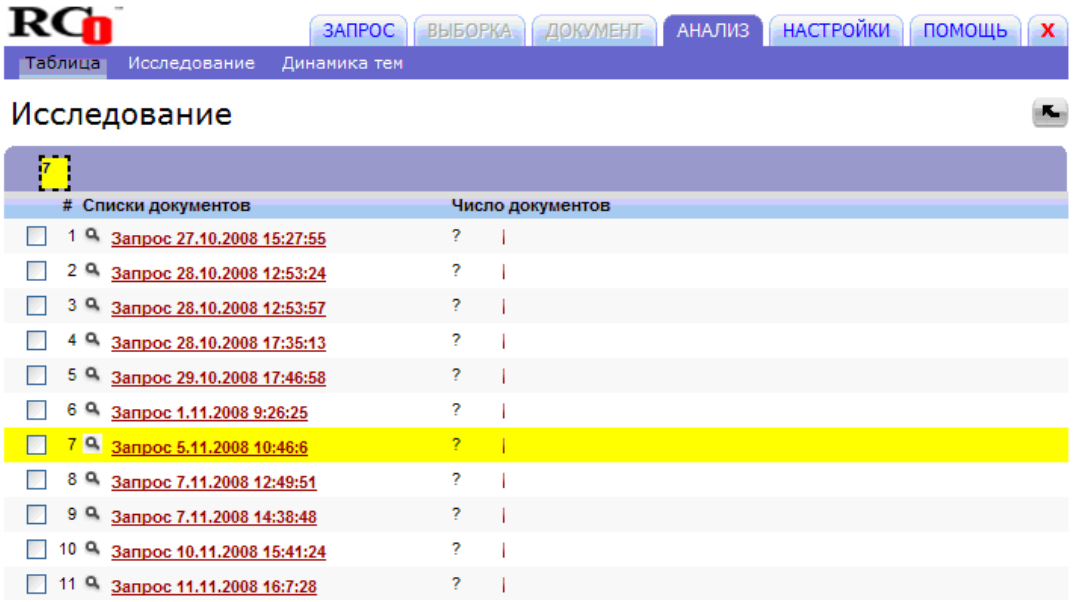


Рисунок 74. *Добавление первой выборки в группу «важных»*

Для исключения выборки из группы «важных» (без удаления) нужно выделить ее номер нажатием левой кнопки мыши и, не отпуская кнопку, перетащить вниз под разделительную черту. Если нет группы «обычных» выборок, то необходимо проделать то же, что и при обычном

исключении, но образ элемента установить над ограничительной чертой в конце списка сохраненных выборок.

7.4. Сравнительный анализ. Страница «Анализ» – «Исследование»

На странице *«Анализ» – «Исследование»* (Рисунок 75) мы можем видеть список сохраненных выборок, уже знакомый нам по странице *«Анализ» – «Таблица»*. Под списком справа расположена панель «Сравнить», на которой перечислены виды сравниваемых списков главных тем:

- Списки Главных тем отдельно – списки главных тем отдельно по каждому объекту;
- Уникальные Главные темы отмеченных выборок – темы, относящихся к каждому объекту исследования, которые не являются *общими*, а представляют собой особенности объектов;
- Общие Главные темы – темы, относящихся к каждому объекту исследования, являющиеся *общими* для исследуемых объектов;
- Объединённый список Главных тем – общий список главных тем.

Справа расположена панель «Для выборки показать» с выпадающим списком имеющихся выборок. Варианты показа:

- Все отсутствующие в ней **Главные темы** из отмеченных выборок – показываются главные темы, присутствующие в списке главных тем для отмеченных выборок, но отсутствующие в отдельной выборке;
- Общие отсутствующие в ней **Главные темы** из отмеченных выборок – показываются темы, общие для отмеченных выборок, но отсутствующие в отдельной выборке;
- Уникальные **Главные темы** – показываются темы, уникальные для данной выборки.

На панели, расположенной ниже, задаются способы сравнения главных тем:

- слова и словосочетания;
- слова;
- словосочетания.

Кроме того, указывается способ сортировки главных тем:

- по алфавиту;
- по общей значимости;
- максимальной значимости;
- минимальной значимости;
- градиенту.

Осуществив свой выбор, нужно нажать кнопку «Выполнить».

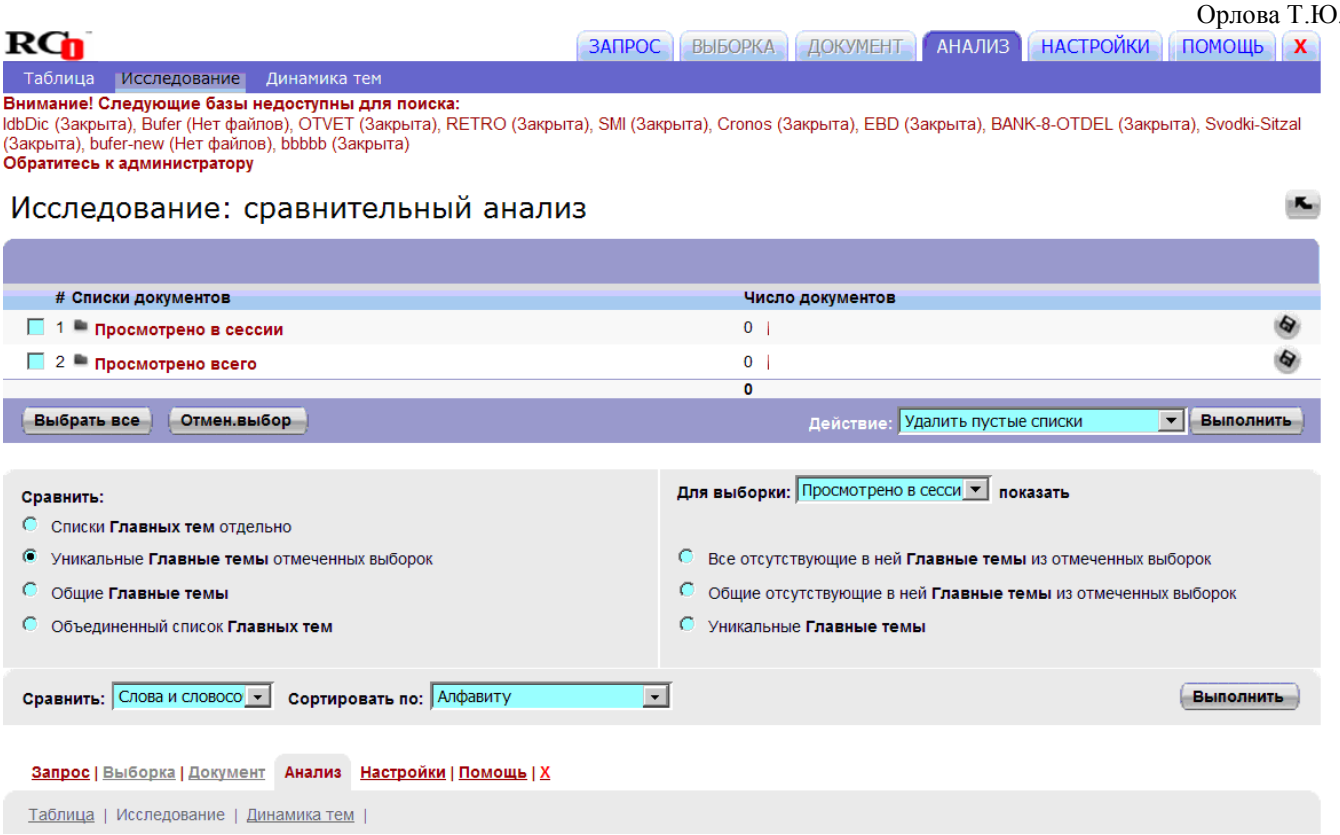


Рисунок 75. *Раздел «Исследование» страницы «Анализ»*

Сравнительный анализ проводится для выявления категорий *общего*, *частного* и *связей* (внутренних/внешних, явных/неявных, свойственных/несвойственных). В ходе проведения сравнительно анализа Система определяет, какие темы являются *главными* для объектов, какие *общими*, а какие *уникальными*. При этом необходимо помнить, что найденные слова, как правило, характеризуют связи между объектами, а словосочетания отражают свойства объектов.

Перед тем, как проводить *сравнительный анализ*, необходимо найти и отобрать то, что сравнивать, т.е. *объекты исследования*.

Примечание

Для того чтобы очистить исследование, либо удалить некоторые выборки, отметьте флажками в таблице выборок те выборки, которые необходимо удалить. Затем в поле *«Действие»* выберите *«Удалить отмеченные выборки»*, нажмите *«Выполнить»*.

Если Вы хотите восстановить ранее проведенные запросы, то в поле *«Действие»* выберите *«Восстановить отмеченные выборки»*. Нажмите кнопку *«Выполнить»*.

Пример: сравним между собой объекты – Герман Греф, Алексей Кудрин, Андрей Фурсенко и Михаил Зурабов.

Сравнительный анализ будем проводить по категориям *общее*, *частное*, *главные темы*.

На странице *«Настройки»* – раздел *«Внешний вид»* на панели *«Анализ»* в поле *«Сохранять новый запрос автоматически»* поставьте флажок, если он отсутствует. Кроме того, для достижения наибольшей точности на панели *«Инфопортрет»* (страница *«Настройки»* – раздел *«Внешний вид»*) в поле *«Число строк»* установите 500. Нажмите *«Установить»*.

Затем последовательно проведите контекстные запросы: *кнт(Герман Греф)*, *кнт(Алексей Кудрин)*, *кнт(Андрей Фурсенко)*, *кнт(Михаил Зурабов)* (см. п. 3.2, *«Тип статистики» – слова и пары*, *«Используемые базы» – SBase*, *«Сортировать» – по значимости*, *«Морфологический анализ»* вкл., *«Статистика»* вкл.). При этом на странице *«Выборка»* в разделе *«Свойства»* последовательно переименуйте запросы (например, им можно присвоить имена объектов) для удобства проведения исследования.

После проведения всех запросов перейдите на страницу *«Анализ»* – раздел *«Таблица»*, в таблице выборок должны быть отображены 4 запроса.

Перейдите в раздел *«Исследование» (см.* Рисунок 76*)*.

В таблице выборок, отметьте флажками все выборки, щелкая по полям напротив соответствующих запросов.

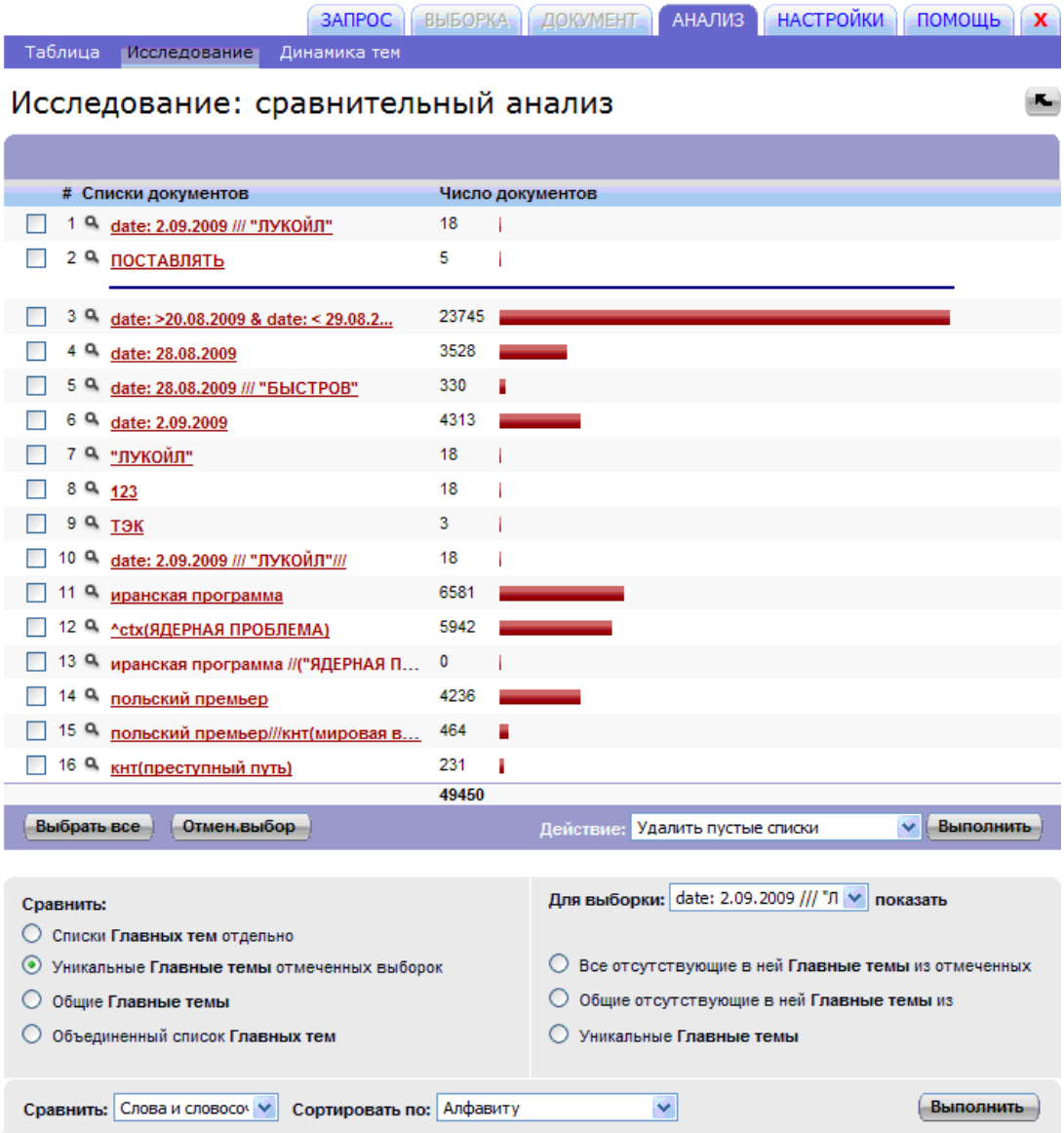


Рисунок 76. *Раздел «Исследование» страницы «Анализ»*

Затем на панели *«Сравнить»* отметьте *«Общие главные темы»*, в поле *«Сравнить»* установите слова *и словосочетания*, в поле *«Сортировать по»* установите по *Алфавиту* и нажмите *«Выполнить» (см.* Рисунок 77*)*.

Рисунок 77. *Список общих главных тем для объектов «Герман Греф», «Алексей Кудрин», «Андрей Фурсенко» и «Михаил Зурабов»*

Итак, мы получили список общих главных тем для выбранных объектов, отсортированный по *алфавиту*. Таким образом, мы выявили *связи* между объектами.

Помимо слов и словосочетаний в таблице для каждой из выборок указывается коэффициент значимости приведенных словоформ.

Можно также проводить сортировку главных тем по *общей значимости*, *градиенту* (*градиент* – скорость изменения значимости), *максимальной* или *минимальной значимости*. Сортировка имеет смысл только для общих и объединённых тем.

Теперь приступим к выявлению категории *частное* для объектов исследования. Для этого вернитесь в раздел *«Исследование»* и проделайте все действия, которые описывались выше. Отметьте *«Уникальные главные темы отмеченных выборок»* и нажмите *«Выполнить»*.

Таким образом, Вы получите список главных тем, относящихся к каждому объекту исследования, которые не являются *общими*, а представляют собой особенности объектов *(см.* Рисунок 78 *и* Рисунок 79*)*.

ЗАПРОС | ВЫБОРКА | ДОКУМЕНТ | АНАЛИЗ | НАСТРОЙКИ | ПОМОЩЬ

Исследование | Таблица | Динамика тем | Тюнинг

Сравнительный анализ - Уникальные Главные Темы

Сортировка по общей значимости

Герман Греф	Алексей Кудрин	Андрей Фурсенко	Михаил Зурабо
ТАМОЖЕННЫЙ	АМНИСТИЯ	ВЫСШЕЕ ОБРАЗОВАНИЕ	ЛЕКАРСТВО
ПОШЛИНА	ВАЛЮТНЫЙ	ОБРАЗОВАТЕЛЬНЫЙ	ЛЕКАРСТВЕННОЕ
КОНТРОЛЬНЫЙ ПАКЕТ	БЕДНЕЙШИЕ СТРАНЫ	УЧЕБНЫЕ ЗАВЕДЕНИЯ	ЛЬГОТНЫЕ ЛЕКА
ТАМОЖНЯ	ВАЛЮТА	ЕГЭ	ПЕНСИОННЫЙ Ф
РОСБР	БУДУЩИЕ ПОКОЛЕНИЯ	РЕКТОР	ПЕНСИОННЫЙ Е
ТАМОЖЕННАЯ СЛУЖБА	НАЛОГОВАЯ АМНИСТИЯ	МИНОБРНАУКИ	ЛЕКАРСТВЕННЫ
ФТС	ВОСЬМЕРКА	НАЦИОНАЛЬНЫЙ ПРОЕКТ	ОБЩАЯ ПРАКТИ
ЭКСПОРТНАЯ ПОШЛИНА	НАЛОГОВИК	БЮДЖЕТНЫЕ МЕСТА	АПТЕКА
ПАКЕТ	ДОСРОЧНАЯ ВЫПЛАТА	ВЫПУСКНИК	МЕДИЦИНСКАЯ
ТАМОЖЕННИК	КОПЕЙКИНА	УЧИТЕЛЬ	ДЕНЕЖНЫЕ ВЫП
ПРИВАТИЗАЦИЯ	РЕЗЕРВ	РОССИЙСКИЕ ВУЗЫ	РЕЦЕПТ
ЭКОНОМРАЗВИТИЯ	ПАРИЖСКИЙ	ОБУЧЕНИЕ	БЕСПЛАТНЫЕ ЛЕ
ИНОСТРАННЫЕ ИНВЕСТОРЫ	КОПЕЙКИН	ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАНИЕ	ПОЛИКЛИНИКА
ВВОЗ	ФИНАНСОВЫЕ РЫНКИ	СТИПЕНДИЯ	ЛЬГОТНЫЕ КАТЕ
КАЛИНИНГРАДСКАЯ ОБЛАСТЬ	ДОСРОЧНЫЙ	ВОЕННЫЕ КАФЕДРЫ	БАЗОВАЯ ЧАСТ
РОССИЙСКИЙ РЫНОК	ФИНАНСОВОЕ ВЕДОМСТВО	РОССИЙСКОЕ ОБРАЗОВАНИЕ	БЕСПЛАТНЫЙ ПІ
САВЕНКО	ШАТАЛОВ	ЗАВЕДЕНИЕ	ПРЕПАРАТ
СТРАТЕГИЧЕСКИЙ	ЗУРАБ	ЭКЗАМЕН	ИНВАЛИД
ГАЗОВАЯ МОНОПОЛИЯ	РОССИЙСКИЙ ДОЛГ	СПЕЦИАЛЬНОСТЬ	ПРОЕЗДНОЙ БИ
КВОТА	ФИНАНСИСТ	ПЕДАГОГ	МИНЗДРАВ
ВСТУПЛЕНИЕ	ПОСЛАНИЕ	РОССИЙСКАЯ АКАДЕМИЯ	МЕДИКАМЕНТ

Рисунок 78. *Список уникальных главных тем для объектов «Герман Греф», «Алексей Кудрин», «Андрей Фурсенко» и «Михаил Зурабов»*

ЗАПРОС | ВЫБОРКА | ДОКУМЕНТ | АНАЛИЗ | НАСТРОЙКИ | ПОМОЩЬ

Исследование | Таблица | Динамика тем | Тюнинг

Сравнительный анализ - Уникальные Главные Темы

Сортировка по общей значимости

РОССТАТ	ПОДОХОДНЫЙ НАЛОГ	МАГИСТРАТУРА	ФЕДЕРАЛЬНЫЙ
ЛУКОЙЛ	ТРЕХЛЕТНИЙ БЮДЖЕТ	МГУ	СЕМЕЙНЫЙ ВР
ТОРГОВАЯ ОРГАНИЗАЦИЯ	ПОВТОРНЫЕ ПРОВЕРКИ	БАКАЛАВРИАТ	ЛЕКАРСТВЕНН
НЕФТЕПРОДУКТ	ВАЛЮТНЫЙ ФОНД	СЕЛЬСКИЕ ШКОЛЫ	ЛЕКАРСТВЕНН
ТОННА	ТРИЛЛИОН	ГОСУДАРСТВЕННЫЕ ВУЗЫ	ПОТРЕБИТЕЛЬ
ТАМОЖЕННЫЕ ПОШЛИНЫ	УКРЕПЛЕНИЕ	УЧЕБНИК	ПЕРВИЧНОЕ ЗЕ
ИНВЕСТИЦИОННЫЕ ПРОЕКТЫ	НАЛОГОВАЯ НАГРУЗКА	БАКАЛАВР	БАТАН
ЕЭС	ДОПОЛНИТЕЛЬНЫЕ ДОХОДЫ	КЛАССНОЕ РУКОВОДСТВО	ЛЕЧЕНИЕ
СДЕЛКА	КОНВЕРТИРУЕМОСТЬ	ГОСУДАРСТВЕННЫЙ ЭКЗАМЕН	УЧАСТКОВЫЕ
ЭКОНОМИЧЕСКИЙ ЭКСТАЗ	БЕДНЕЙШИЙ	ОБРАЗОВАТЕЛЬНЫЕ КРЕДИТЫ	ЛЕЧЕБНЫЕ УЧР
ДВТУ	НАЛОГОВАЯ СЛУЖБА	САДОВНИЧИЙ	МЕДИЦИНА
РОССИЙСКАЯ ДЕЛЕГАЦИЯ	ПОГАШЕНИЕ	ЛУЧШИЕ УЧИТЕЛИ	ЛЬГОТНЫЕ РЕЦ
РЕЗИДЕНТ	ДОЛГОВОЙ	ШКОЛЬНЫЙ	УЧАСТКОВЫЕ
ФЕДЕРАЛЬНЫЕ ВЛАСТИ	ДОХОДНОСТЬ	НЕГОСУДАРСТВЕННЫЕ ВУЗЫ	ПЕДИАТР
ОСОБАЯ ЗОНА	БЕЗГРАМОТНОЕ УТВЕРЖДЕНИЕ	ПРЕПОДАВАТЕЛЬ	РЕГИСТР
КАЛИНИНГРАДСКИЙ	ДРУГИЕ МИНИСТРЫ	БЕСПЛАТНОЕ ОБРАЗОВАНИЕ	МИНСОЦЗДРА
КОНЦЕРН	ЭКОНОМИЧЕСКИЕ НОВОСТИ	ДВУХУРОВНЕВАЯ СИСТЕМА	ФАРМАЦЕВТИЧ
МЕДВЕДКОВ	РОССИЙСКАЯ ВАЛЮТА	НАУЧНЫЕ ОРГАНИЗАЦИИ	УЧАСТКОВЫЙ
ПРОМЫШЛЕННАЯ СБОРКА	ПОЛИТИЧЕСКИЕ ТЕХНОЛОГИИ	ОБРАЗОВАТЕЛЬНАЯ СИСТЕМА	СОЦИАЛЬНАЯ
СЛИЯНИЕ	ЭКСПОРТНЫЕ ПОШЛИНЫ	ОБРАЗОВАТЕЛЬНЫЕ УЧРЕЖДЕНИЯ	ДЕМОГРАФИЧЕ
ЭТОТ ПРОЕКТ	ЛЬГОТНАЯ СТАВКА	НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ	ПРОЖИТОЧНЫ
БИЗНЕСМЕН	ФЕДЕРАЛЬНОЕ СОБРАНИЕ	ВУЗОВСКИЙ	СОЦИАЛЬНЫЙ

Рисунок 79. *Список уникальных главных тем для объектов «Герман Греф», «Алексей Кудрин», «Андрей Фурсенко» и «Михаил Зурабов»*

ЗАПРОС ВЫБОРКА ДОКУМЕНТ АНАЛИЗ НАСТРОЙКИ ПОМОЩЬ

Исследование Таблица Динамика тем Тюнинг

Сравнительный анализ - Слова и словосочетания

Сортировка по общей значимости

Герман Греф	Алексей Кудрин	Андрей Фурсенко	Михаил Зурабо
ИНФЛЯЦИЯ	ИНФЛЯЦИЯ	ВЫСШЕЕ ОБРАЗОВАНИЕ	ЛЬГОТНИК
МЭРТ	ГРЕФ	ОБРАЗОВАТЕЛЬНЫЙ	ЛЕКАРСТВО
КУДРИН	СТАБИЛИЗАЦИОННЫЙ ФОНД	УЧЕБНЫЕ ЗАВЕДЕНИЯ	МОНЕТИЗАЦИ
ЭКОНОМИЧЕСКИЕ ЗОНЫ	СТАБФОНДА	ЕГЭ	ЛЬГОТНЫЙ
ФРАДКОВ	ФРАДКОВ	РЕКТОР	ПЕНСИЯ
ОЭЗ	ВНЕШНИЙ ДОЛГ	МИНОБРНАУКИ	ПЕНСИОННЫЙ
ВТО	НЕФТЬ	НАЦИОНАЛЬНЫЙ ПРОЕКТ	ЛЕКАРСТВЕНН
РОСНЕФТЬ	ФРАДКОВА	БЮДЖЕТНЫЕ МЕСТА	МИНЗДРАВСО
НЕФТЬ	СТАБФОНД	ВЫПУСКНИК	ЛЬГОТНЫЕ ЛЕ
ГАЗПРОМ	ПАРИЖСКИЙ КЛУБ	УЧИТЕЛЬ	ПЕНСИОННЫЙ
ЭКОНОМИЧЕСКИЙ РОСТ	ЭКОНОМИЧЕСКИЙ РОСТ	ВЫСШАЯ ШКОЛА	ПЕНСИОНЕР
МИНФИН	МИНЭКОНОМРАЗВИТИЯ	РОССИЙСКИЕ ВУЗЫ	ПЕНСИОННЫЙ
РОССИЙСКАЯ ЭКОНОМИКА	МИЛЛИАРД	ОБУЧЕНИЕ	ЛЬГОТНЫЕ ВЫ
СТАБИЛИЗАЦИОННЫЙ ФОНД	ГЕРМАН	ИННОВАЦИОННЫЙ	КУДРИН
ТЕМП	МЭРТ	ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАНИЕ	ЛЕКАРСТВЕНН
ВВП	ПРОГНОЗ	СТИПЕНДИЯ	ВЫПЛАТА
ПУТИН	НАЛОГОВЫЙ	ВОЕННЫЕ КАФЕДРЫ	ФРАДКОВ
МОНОПОЛИЯ	РОССИЙСКАЯ ЭКОНОМИКА	РОССИЙСКОЕ ОБРАЗОВАНИЕ	ФЕДЕРАЛЬНЫЕ
МЭРТА	НДС	ЗАВЕДЕНИЕ	ОБЩАЯ ПРАКТ
НЕФТЯНОЙ	ЭКОНОМИЧЕСКОЕ РАЗВИТИЕ	ЭКЗАМЕН	АПТЕКА
ПУТИНА	ЗУРАБОВ	СПЕЦИАЛЬНОСТЬ	МЕДИЦИНСКА

Рисунок 80. *Списки главных тем выборок «Герман Греф», «Алексей Кудрин», «Андрей Фурсенко» и «Михаил Зурабов»*

ЗАПРОС ВЫБОРКА ДОКУМЕНТ АНАЛИЗ НАСТРОЙКИ ПОМОЩЬ

Исследование Таблица Динамика тем Тюнинг

Сравнительный анализ - Слова и словосочетания

Сортировка по общей значимости

УДВОЕНИЕ	АЛРОС	ПЕДАГОГ	ДЕНЕЖНЫЕ ВЫГ
ТАМОЖЕННЫЙ	ВИЦЕ-ПРЕМЬЕР	РОССИЙСКАЯ АКАДЕМИЯ	РЕЦЕПТ
ПОШЛИНА	ПУТИНА	АБИТУРИЕНТ	БЕСПЛАТНЫЕ ЛЕ
ФРАДКОВА	ПУТИН	УСЫНОВЛЕНИЕ	ГРЕФ
ЕСТЕСТВЕННЫЕ МОНОПОЛИИ	СТАБИЛИЗАЦИОННЫЙ	АКАДЕМИЯ	ФЕДЕРАЛЬНЫЙ
НЕФТЯНЫЕ КОМПАНИИ	ЦБ	РАНА	ПОЛИКЛИНИКА
БАНК	НАЛОГ	МИНОБРАЗОВАНИЕ	ЕДИНАЯ РОССИ
ИНВЕСТОР	БЮДЖЕТНЫЙ	РОССИЙСКАЯ НАУКА	ФРАДКОВА
МИЛЛИАРД	ВВП	МАГИСТРАТУРА	НАЦИОНАЛЬНЫ
РАО	ФЕДЕРАЛЬНЫЙ БЮДЖЕТ	МГУ	ЛЬГОТНЫЕ КАТЕ
РОССИЙСКОЕ ПРАВИТЕЛЬСТВО	МОНЕТИЗАЦИЯ	БАКАЛАВРИАТ	БАЗОВАЯ ЧАСТ
ИНВЕСТИЦИОННЫЙ	ЗОЛОТОВАЛЮТНЫЕ РЕЗЕРВЫ	СЕЛЬСКИЕ ШКОЛЫ	БЕСПЛАТНЫЙ ПР
КАПИТАЛ	ЦЕНТРОБАНК	ГОСУДАРСТВЕННЫЕ ВУЗЫ	НАТУРАЛЬНЫЕ Л
ХРИСТЕНКО	ТЕМП	УЧЕБНИК	СОВЕЩАНИЕ
СИБНЕФТЬ	БАРРЕЛЬ	БАКАЛАВР	ПРЕПАРАТ
ЮКОС	ПРЕМЬЕР	КЛАССНОЕ РУКОВОДСТВО	ОТСТАВКА
ВИЦЕ-ПРЕМЬЕР	ЛЬГОТА	ГОСУДАРСТВЕННЫЙ ЭКЗАМЕН	ИНВАЛИД
КОНТРОЛЬНЫЙ ПАКЕТ	ДЕНЕЖНАЯ МАССА	ОБРАЗОВАТЕЛЬНЫЕ КРЕДИТЫ	ПРОЕЗДНОЙ
ЦБ	АМНИСТИЯ	БЮДЖЕТНЫЙ	ПЕНСИОННАЯ С
ТАМОЖНЯ	ДОСРОЧНОЕ ПОГАШЕНИЕ	САДОВНИЧИЙ	ПРОЕЗДНОЙ БИ
МИРОВЫЕ ЦЕНЫ	ТРЛН	ЛУЧШИЕ УЧИТЕЛИ	МИНЗДРАВ
АДМИНИСТРАТИВНАЯ РЕФОРМА	НЕФТЯНОЙ	ШКОЛЬНЫЙ	МЕДИКАМЕНТ

Рисунок 81. *Списки главных тем выборок «Герман Греф», «Алексей Кудрин», «Андрей Фурсенко» и «Михаил Зурабов»*

Для того чтобы получить списки главных тем отдельно по каждому объекту вернитесь в раздел *«Анализ» – «Исследование»*, отметьте *«Списки главных тем отдельно»*, выберите способ сортировки и нажмите *«Выполнить» (см.* Рисунок 80 *и* Рисунок 81*).*

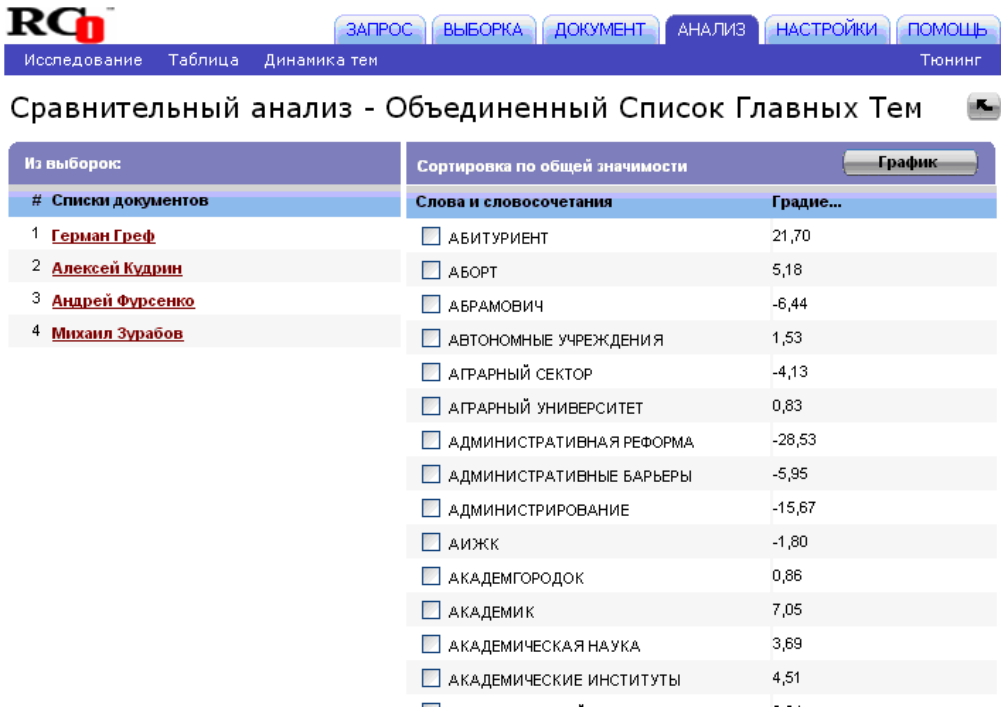


Рисунок 82. *Объединенный список главных тем выборок «Герман Греф», «Алексей Кудрин», «Андрей Фурсенко» и «Михаил Зурабов»*

Для получения объединенного списка главных тем по всем объектам перейдите в раздел *«Таблица»*, отметьте *«Объединенный список главных тем»*, выберите способ сортировки и нажмите *«Выполнить» (см.* Рисунок 82, Рисунок 83 *и* Рисунок 84*).*

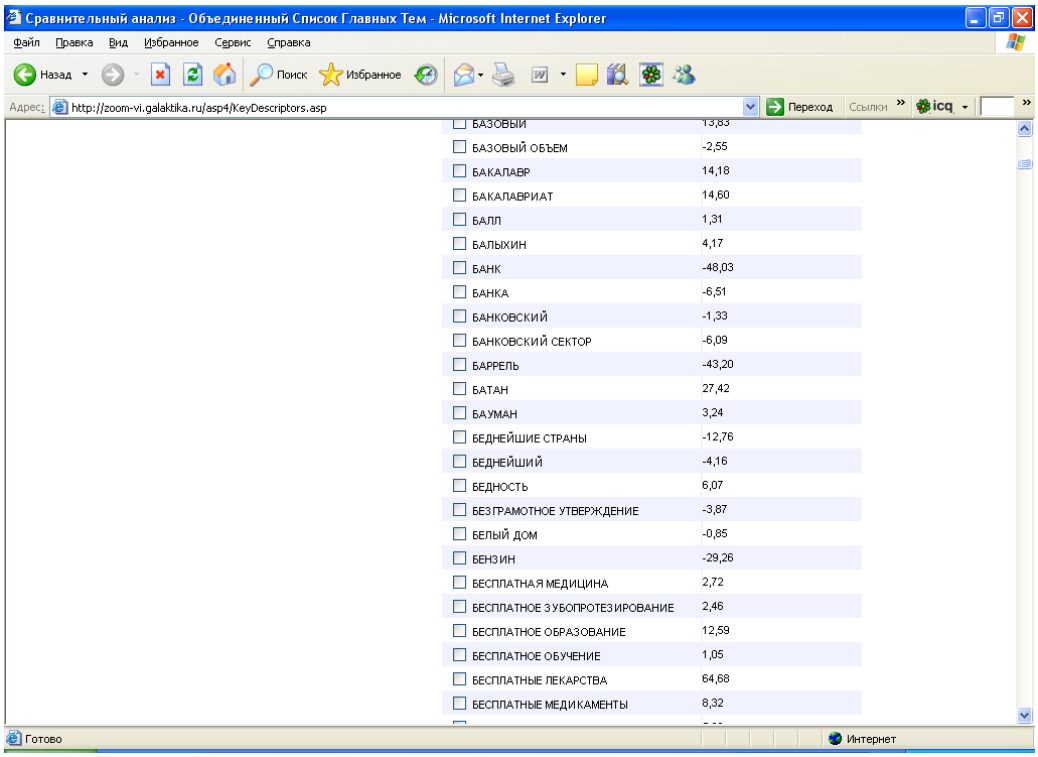


Рисунок 83. *Объединенный список главных тем выборок «Герман Греф», «Алексей Кудрин», «Андрей Фурсенко» и «Михаил Зурабов»*

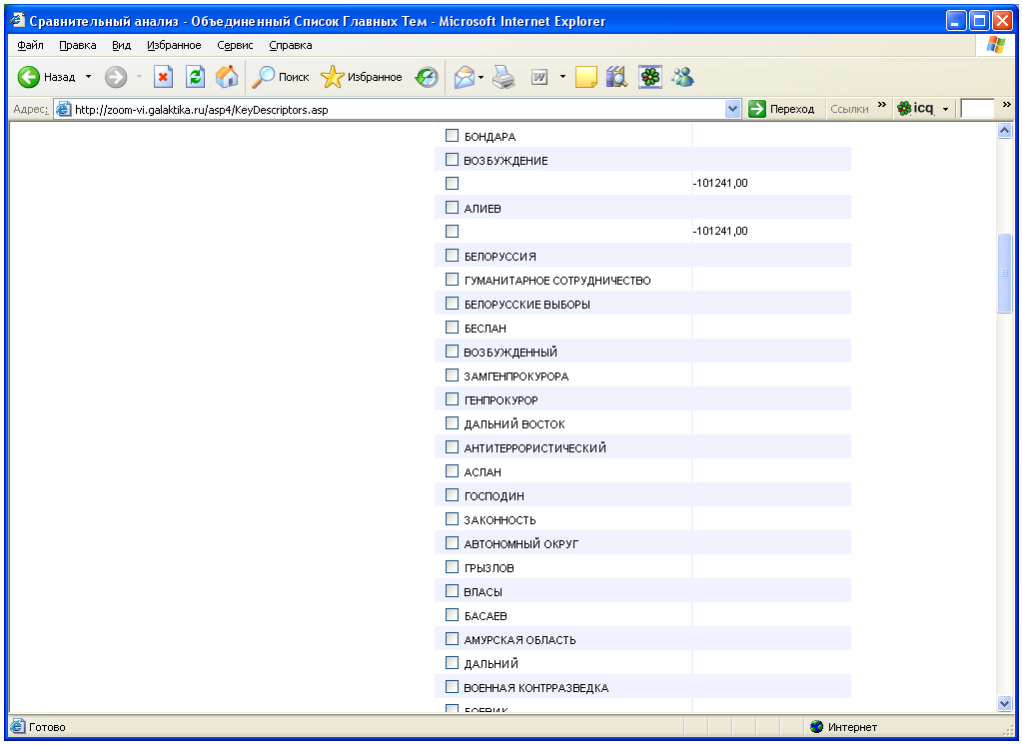


Рисунок 84. *Объединенный список главных тем выборок «Герман Греф», «Алексей Кудрин», «Андрей Фурсенко» и «Михаил Зурабов»*

Аналогично, можно выбрать в поле *«Для выборки показать»* конкретный объект (запрос) и сравнить его с отмеченными в таблице объектами (запросами).

В разделе *«Исследование»* отметьте запросы, например, *«Алексей Кудрин»*, *«Андрей Фурсенко»*, *«Михаил Зурабов»*, в поле *«Для выборки показать»* выберите запрос *«Герман Греф»*. Для того чтобы получить уникальные главные темы выбранного запроса *«Герман Греф»*, т.е. темы, которые содержатся в данном запросе, но отсутствуют в отмеченных выборках, отметьте *«Уникальные главные темы»* и нажмите *«Выполнить»* (*см.* Рисунок 85 *и* Рисунок 86).

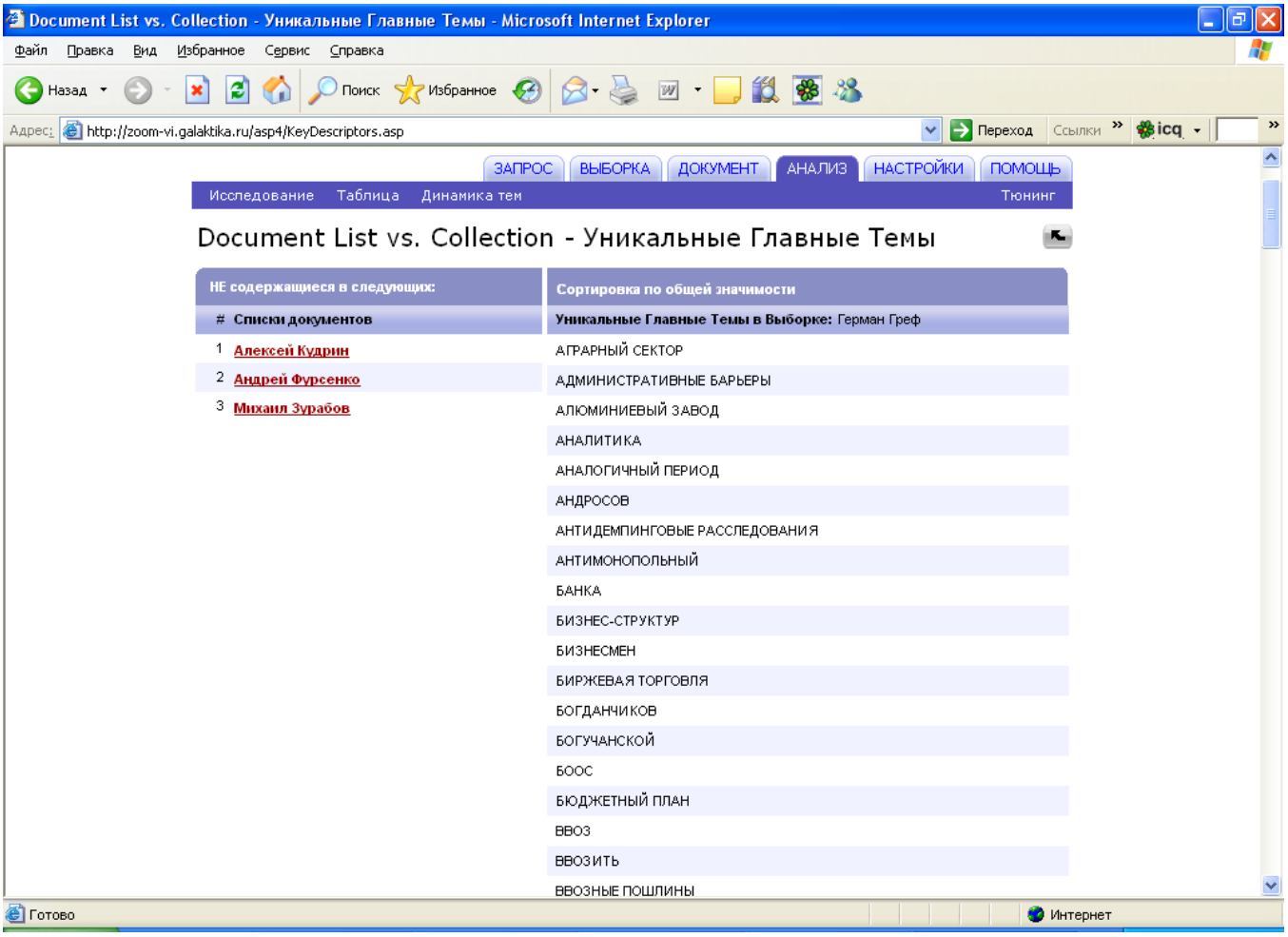


Рисунок 85. *Список уникальных главных тем выборки «Герман Греф»*

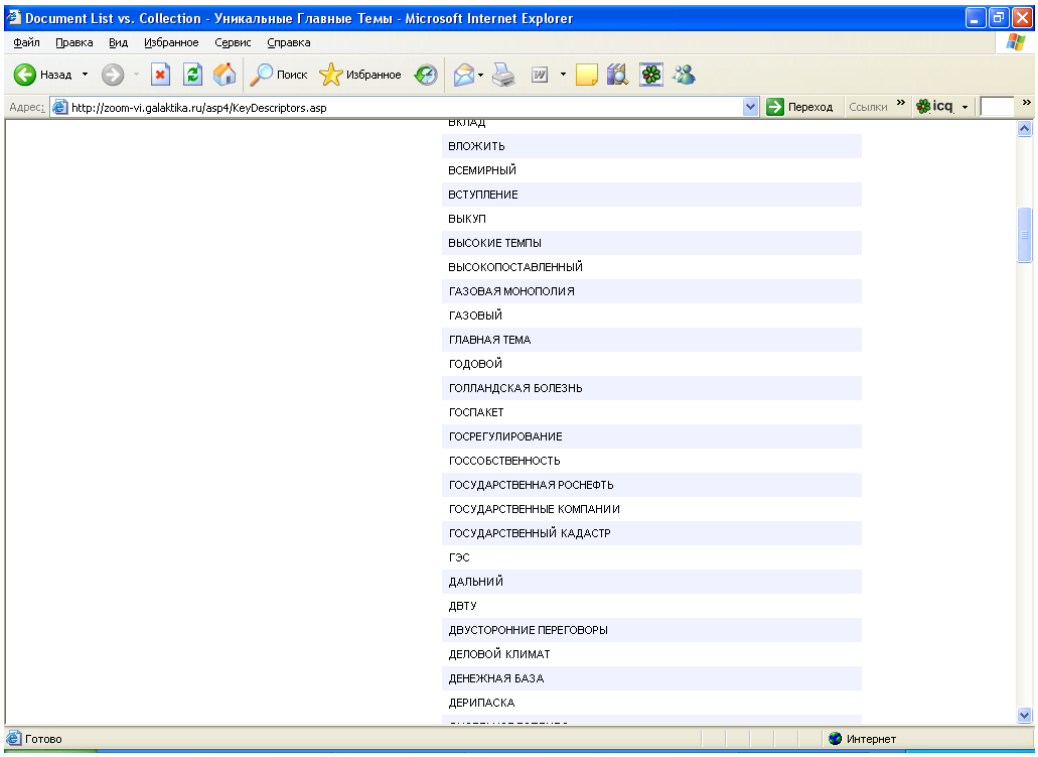


Рисунок 86. *Список уникальных главных тем выборки «Герман Греф»*

Для выявления главных тем отмеченных выборок, отсутствующих в выборке *«Герман Греф»*, проделайте все те же действия. Отметьте *«Все отсутствующие в ней главные темы из отмеченных выборок»* и нажмите кнопку *«Выполнить» (см.* Рисунок 87 *и* Рисунок 88*).*

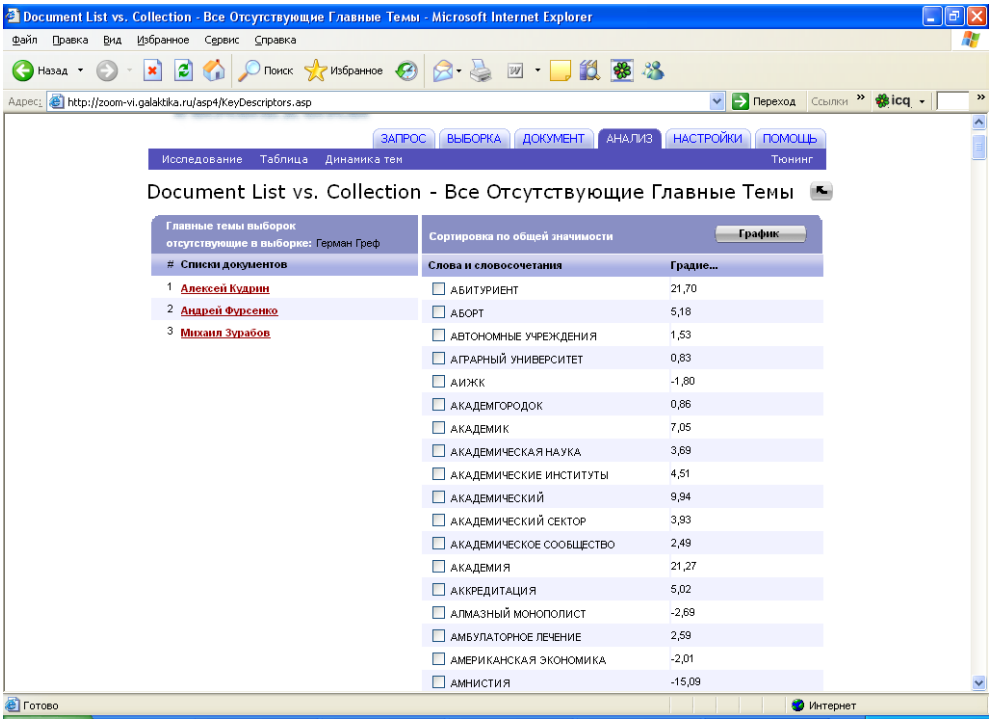


Рисунок 87. *Фрагмент списка главных тем выборок «Алексей Кудрин», «Андрей Фурсенко», «Михаил Зурабов», отсутствующих в выборке «Герман Греф»*

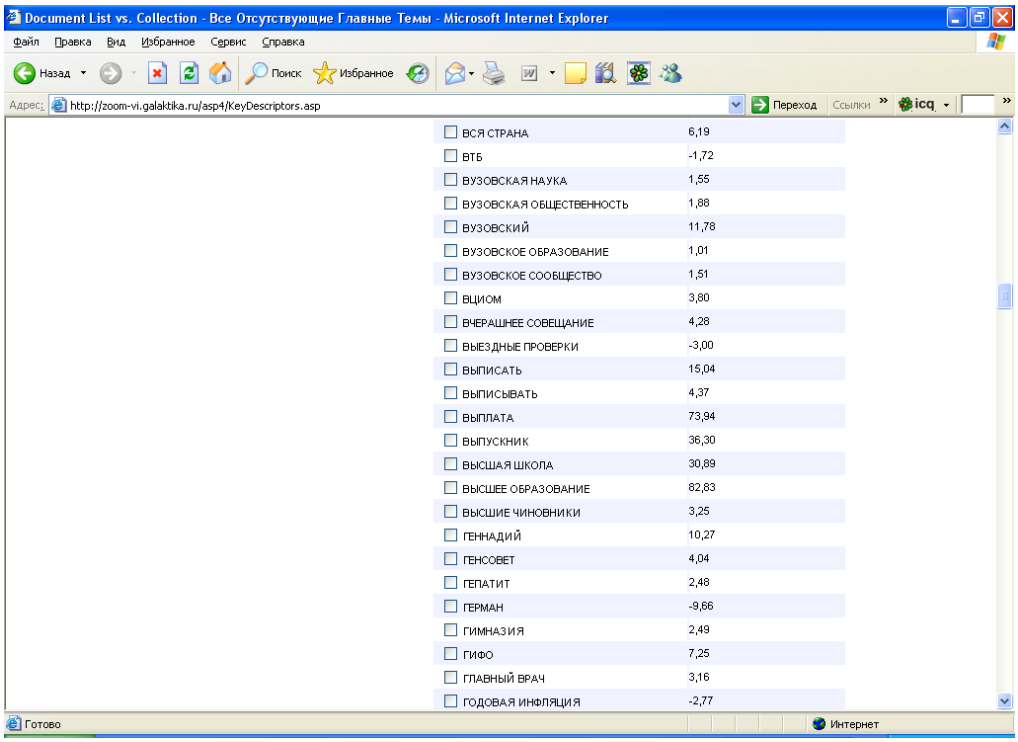


Рисунок 88. *Фрагмент списка главных тем выборок «Алексей Кудрин», «Андрей Фурсенко», «Михаил Зурабов», отсутствующих в выборке «Герман Греф»*

Для выявления общих тем отмеченных выборок *«Алексей Кудрин», «Андрей Фурсенко», «Михаил Зурабов»*, отсутствующих в выборке *«Герман Греф»*, отметьте *«Общие отсутствующие в ней главные темы из отмеченных выборок»* и нажмите *«Выполнить» (см.* Рисунок 89*)*.

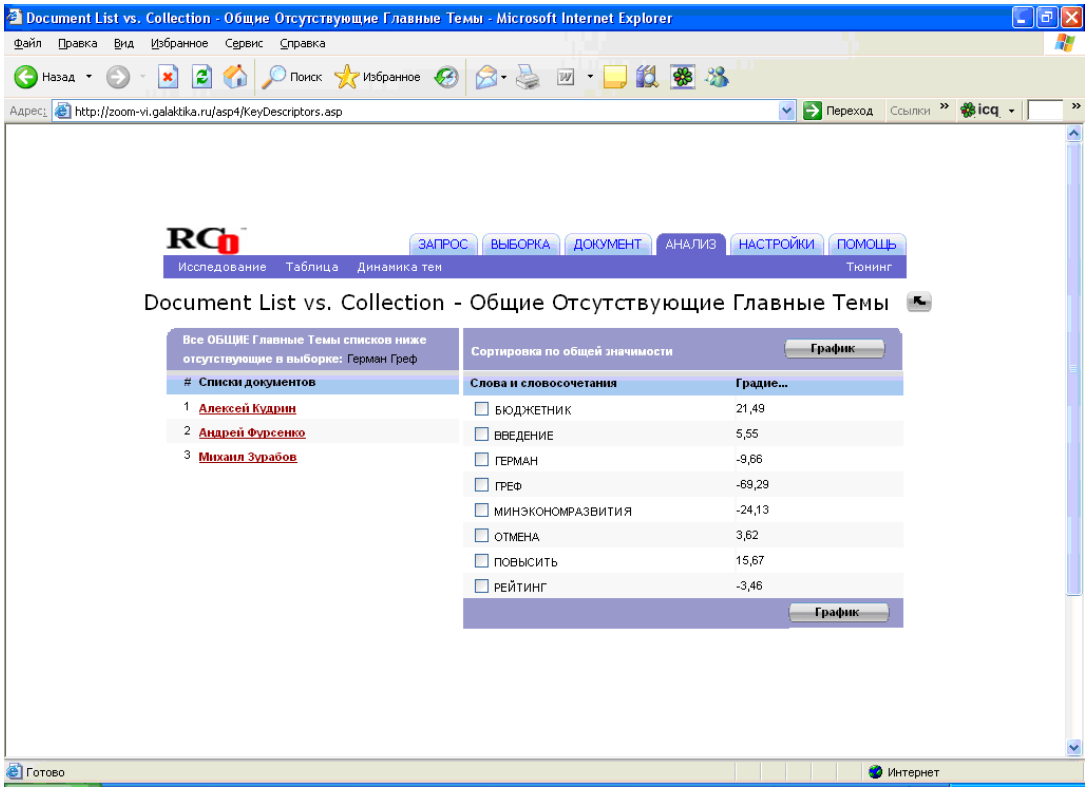


Рисунок 89. *Фрагмент списка общих главных тем выборок «Алексей Кудрин», «Андрей Фурсенко», «Михаил Зурабов», отсутствующих в выборке «Герман Греф»*

Кроме того, в полученных таблицах указывается коэффициент значимости приведенных словоформ, а в случае списка *общих тем*, либо *объединенного списка главных тем*, отображается *градиент* (*градиент* – показатель изменения значимости) Значения градиента могут принимать как положительные (повышение значимости темы), так и отрицательные значения (понижение значимости темы). Исследование градиента имеет смысл для выборок, связанных между собой по смыслу, например в динамике по времени. У уникальных тем значимость только в одной выборке – градиент считать бессмысленно.

Итак, для того чтобы определить *связи* между объектами можно воспользоваться одним из предлагаемых способов:

1. В *информационном портрете* (страница *«Выборка»*, раздел *«Инфопортрет»*) объекта использовать функцию кластеризации (см. п. 6.1). Следует заметить, что при использовании кластеризации «шарики» кластеров показываются у тех тем, которые относятся к одному из сравниваемых запросов и их набор меняется при переходе к другому запросу (запросы должны быть проведены по одной базе, иначе кластеризация не работает).
2. На странице *«Анализ»* (раздел *«Исследования»*) провести действия по выявлению общего между объектами *(см. выше)*. Если связи между объектами существуют, то пользователь получает вышеописанную таблицу. Отсутствие таблицы говорит об отсутствии связей между объектами.

Функцию кластеризации можно также использовать для определения связей между главными темами вышеописанных таблиц.

При включении кластеризации («Вид» – «Кластеризация» – «Применить») инфопортрет несколько изменится – появятся «шарики» и под инфопортретом появится раздел, позволяющий определять размер кластера (Рисунок 90). Выберите слово (словосочетание), которое определяет интересующую тематику и относительно которого будет производиться *кластеризация*, например, *«ЕГЭ»*. Щелкните по «шарику» рядом с этим словом. Это слово (словосочетание) станет *центром кластера* и обозначается значком . Большой размер кластера позволяет увидеть больше связей между понятиями в пределах выборки. Малый размер кластера позволяет детализировать картину распределения близких к центру понятий. Цвет «шарика» показывает степень близости слов/словосочетаний друг к другу. Красный – наиболее близкие, фиолетовый – далёкие друг от друга. При наведении указателя мыши на «шарик» подсвечивается число, характеризующее степень близости понятий (чем оно больше, тем дальше друг от друга понятия).

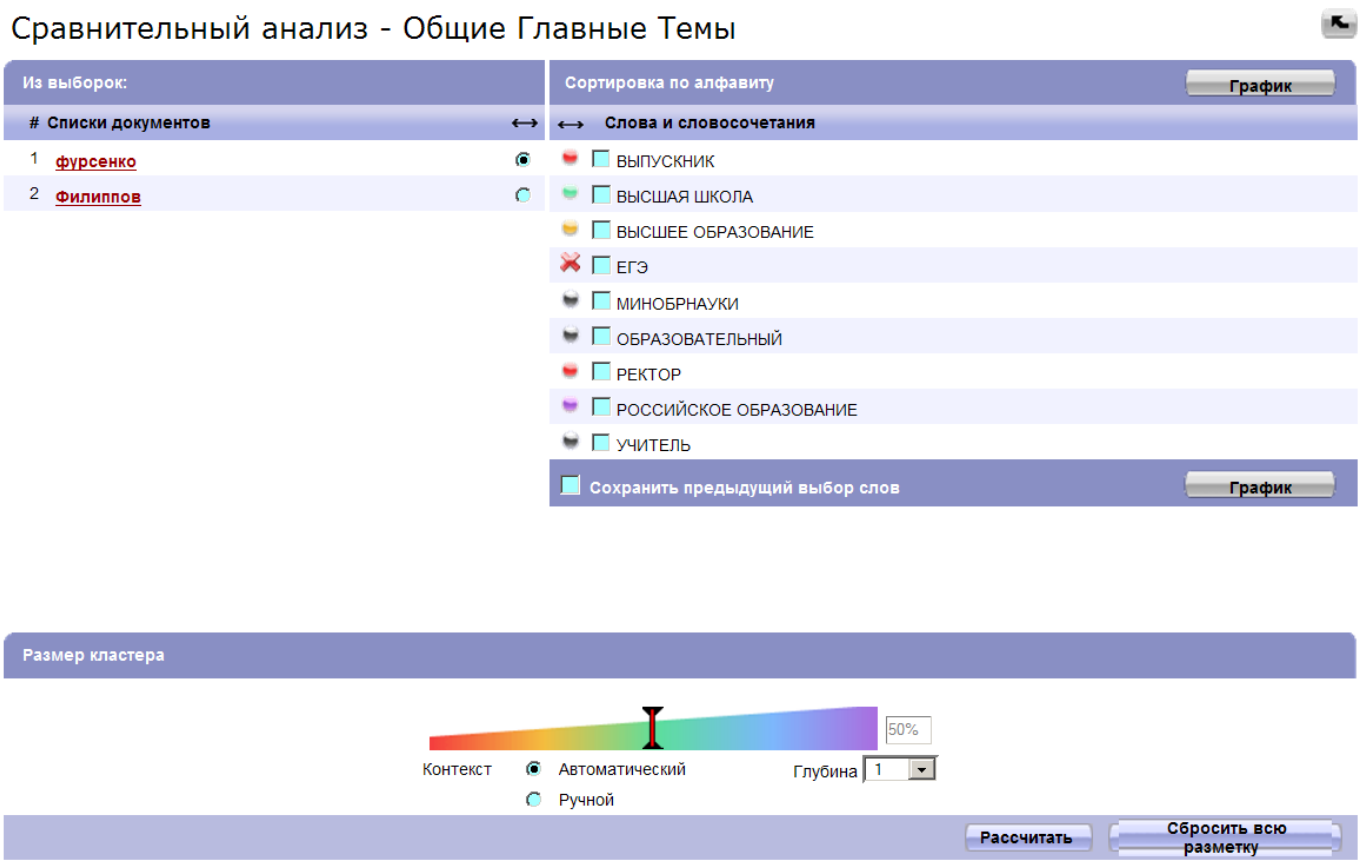


Рисунок 90. *Управление размером кластера*

Управление размером кластера осуществляется перемещением вертикального указателя. Вначале размер кластера устанавливается 50%.

Переключатель «Контекст» имеет два состояния:

- Автоматический – пересчёт значений (цвет «шариков», степени близости) осуществляется сразу после выбора центра кластера;
- Ручной – позволяет назначить в качестве единого центра несколько понятий (после назначения пересчёт осуществляется нажатием кнопки *«Рассчитать»*).

Для того чтобы проследить изменение значимости общих тем от объекта к объекту, а также наглядно представить результаты исследования, можно построить график, предварительно, отметив в таблице интересующие главные темы и нажав кнопку *«График»*, которая отображается выше/ниже списка общих главных тем выборок.

В результате, Вы получите следующий график: по оси X – исследуемые объекты (запросы), по оси Y – значимость главных тем.

Примечание:

Если у Вас уже были построены графики и Вы хотите добавить ещё несколько слов/словосочетаний, следует перед построением графика установить галочку «Сохранить

предыдущий выбор слов» (Рисунок 90). Если галочка не установлена, графики будут построены только по вновь выбранным словам.

Например, построим график по *общим главным темам* выборок *«Герман Греф», «Алексей Кудрин», «Андрей Фурсенко»* и *«Михаил Зурабов» (см.* Рисунок 91*).*

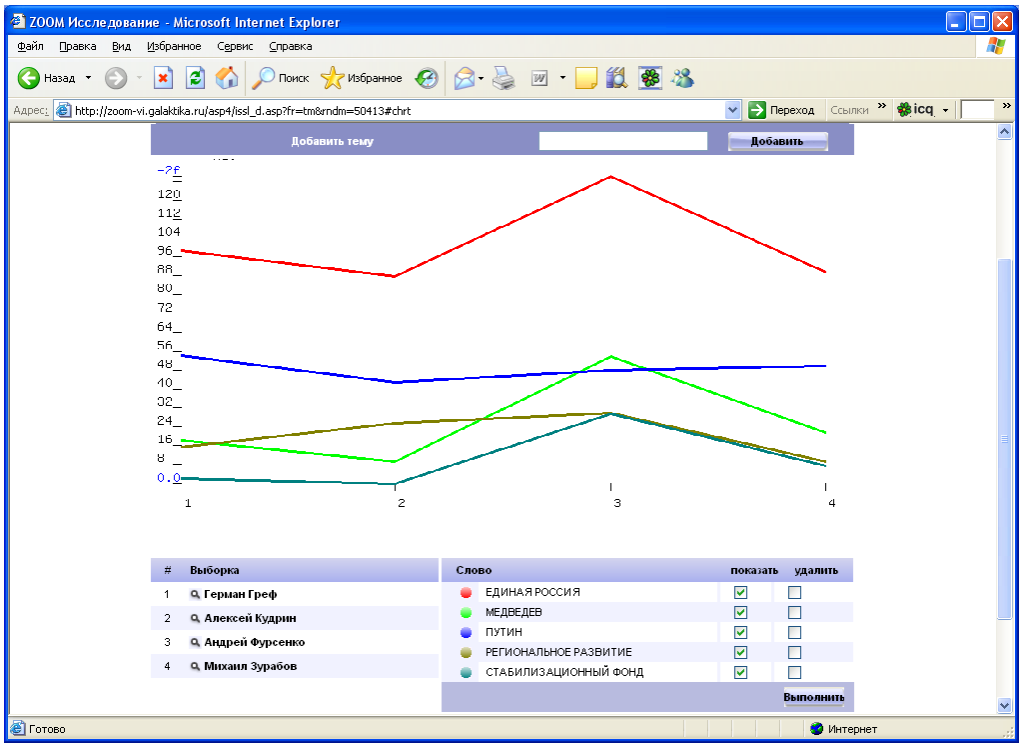


Рисунок 91. *График изменения значимости главных тем от объекта к объекту*

На графике можно скрыть или показать главные темы, для этого в поле *«Показать»* напротив темы необходимо поставить/убрать флажок, затем нажать *«Выполнить»*.

Чтобы удалить какую-либо тему, в поле *«Удалить»* напротив соответствующей темы поставьте флажок и нажмите кнопку *«Выполнить»*.

Для добавления слов или словосочетаний в пустом поле под таблицей главных тем введите слово или словосочетание, по которому хотите построить дополнительный тренд и нажмите кнопку *«Добавить слово»*. Может быть добавлено любое слово, но имеет смысл добавлять значимые слова – для слов, не имеющих значимости, график стоит бессмысленно.

После добавления или удаления главных тем обновите страницу, чтобы увидеть результат.

7.5. «Публикационный рейтинг». Страница «Запрос» («Динамика по времени»)

Помимо вышеперечисленных возможностей сравнения, Система позволяет сравнивать не только различные объекты, но и один и тот же объект на разных временных отрезках. Количество

документов – публикаций (в которых упоминается объект), соответствующее различным промежуткам времени, называется *«публикационным рейтингом»*.

«Публикационный рейтинг» в Системе представляет собой таблицу, в которой отображаются временные отрезки и соответствующее каждому из них количество документов (число).

Для того чтобы построить *публикационный рейтинг*, на странице *«Запрос»* перейдите в раздел *«Динамика по времени»*. Затем в окне запроса введите запрос, укажите его тип и выберите промежуток времени и единицу деления – год, квартал, месяц и т.д. Число делений не должно превышать 52. После того как все параметры указаны, нажмите кнопку *«Построить»*.

Результаты представляются на странице *«Анализ»* (раздел *«Исследования»*) в табличной форме, с указанием периода времени и количества документов, которое ему соответствует, а также с визуальным представлением самой выборки в виде столбчатой диаграммы. Указанные в таблице периоды времени являются ссылками, щелкнув по которым, Вы сможете открыть интересующую Вас выборку документов, соответствующую данному промежутку времени, и просмотреть документы. Таким образом, при построении *публикационного рейтинга*, исходная выборка документов разбивается на более мелкие выборки, соответствующие конкретным промежуткам времени.

После построения *публикационного рейтинга* можно проводить *сравнительный и динамический анализ* объектов (см. пред. и послед. пп.). В данном случае под объектом понимается выборка, соответствующая определенному временному промежутку.

Рассмотрим **пример**.

Построим *публикационный рейтинг* объекта – Дмитрий Медведев в период с 01.01.2006 по 21.06.2006.

В поле *«Запрос»* введите текст запроса – *кнт(Дмитрий Медведев)*. Затем укажите промежуток времени, на котором хотите проводить исследование – с 01.01.2006 по 21.06.2006. В поле *«Деление по»* установите значение – *месяцам*. Далее выберите требуемые параметры запроса (тип запроса, тип статистики, используемые базы, морфология и т.д.) *(см.* Рисунок 92*)*. Нажмите *«Искать»*. Затем Система выдаст сообщение о количестве интервалов разбиения. Если количество интервалов не превышает 52, то нажмите кнопку *«ОК»*. В случае если количество интервалов разбиения превышает значение 52, пересмотрите промежутки времени, либо интервалы деления. Система проведет деление и при количестве интервалов больше 52, но в этом случае будет сохранена лишь часть выборок, соответствующих начальным временным интервалам (определяется наличием свободных мест в полном списке выборок).

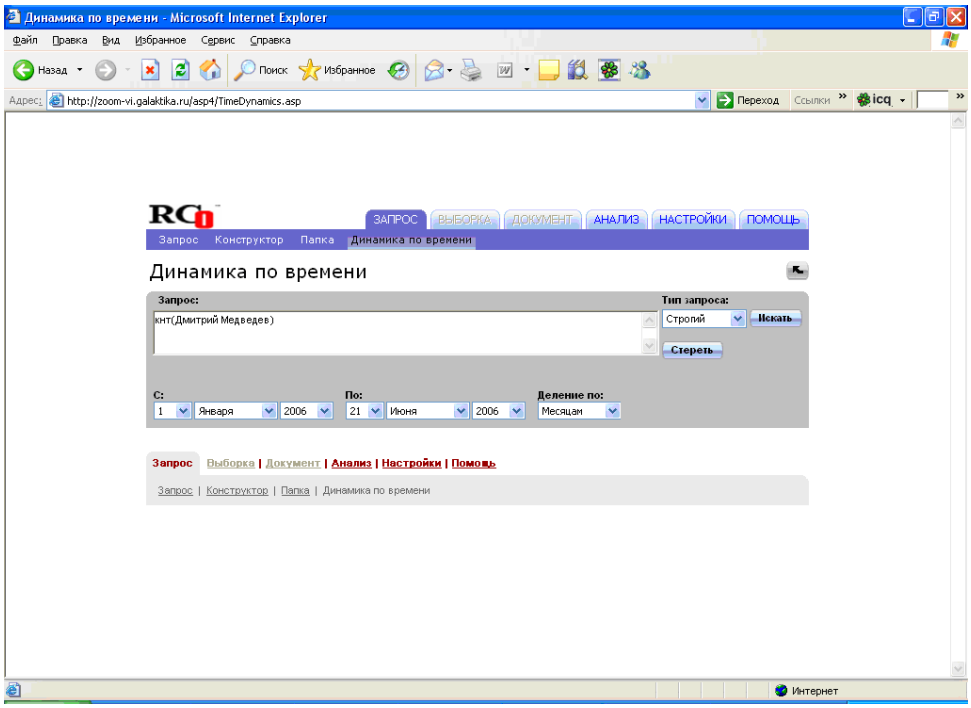


Рисунок 92. *Страница «Динамика по времени»*

Примечание

Если Система выдаст 0 документов на каждом промежутке, или вы сочтете получившийся результат некорректным, на странице *«Настройки»* проверьте все параметры запроса. При необходимости измените параметры. Подробно страница *«Настройки»* будет описана в следующей главе.

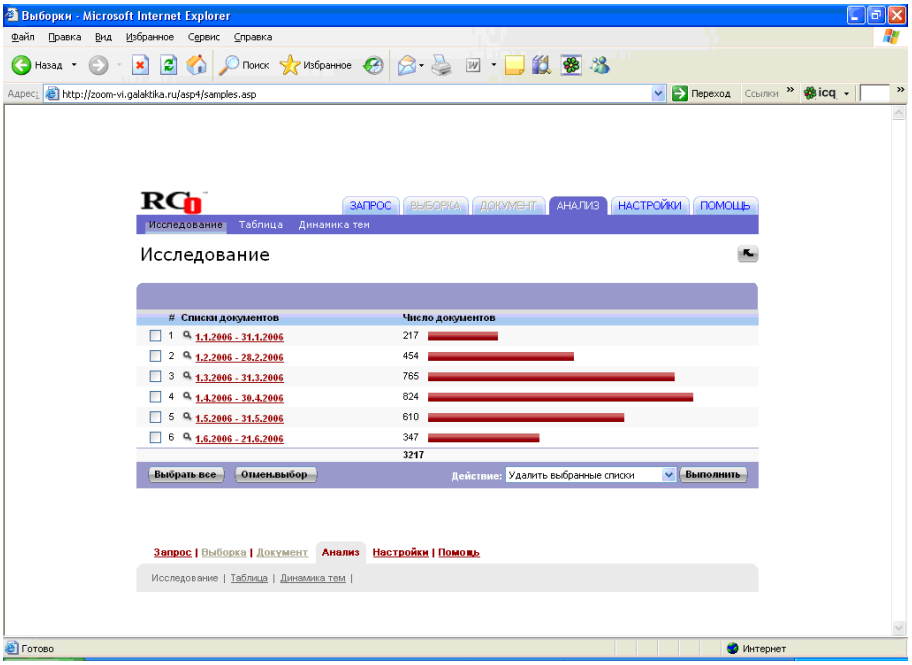


Рисунок 93. *Публикационный рейтинг Дмитрий Медведева*

(наибольшее количество публикаций приходится на апрель 2006 года)

Таким образом, Вы получите *публикационный рейтинг* Дмитрия Медведева за период с 01.01.2006 по 21.06.2006 с делением по месяцам *(см.* Рисунок 93*)*. По *публикационному рейтингу* обычно судят о PR-активности компаний, персон и других объектов в определенные периоды времени.

После построения публикационного рейтинга, Вы получаете возможность изучить объект во времени, т.е. проследить тенденции изменения главных тем с течением времени, сделать прогноз дальнейшего развития ситуации. Поэтому переходим к *динамическому анализу*.

7.6. Динамический анализ главных тем. Страница «Анализ» («Динамика тем»)

Система позволяет увидеть, как изменяется та или иная тема по отношению к изучаемому объекту во времени. По результатам анализа можно определить наиболее значимые периоды в развитии объекта, а также возможные взаимосвязи между объектами.

Перед тем, как приступать к *динамическому анализу* объекта, необходимо сформировать временные срезы или, так называемый, *публикационный рейтинг* объекта *(см. пред.п.)*.
Затем на странице *«Анализ»* перейдите в раздел *«Динамика тем»*. После этого Вам предоставляется возможность добавить собственные слова (словосочетания) в исследование.

Итак, главные темы можно добавлять в исследование вручную. Для этого, введите слово или словосочетание в поле *«Добавить тему»* и нажмите кнопку *«Добавить»*. После того, как добавлены все исследуемые темы, установите флажки в поле *«Показать»* напротив каждой темы, нажмите *«Выполнить»*.

Результаты исследования автоматически представляются в виде графика, с визуализацией легенды, в которой указаны главные темы. По оси X последовательно располагаются промежутки времени, по оси Y – значимость главных тем.

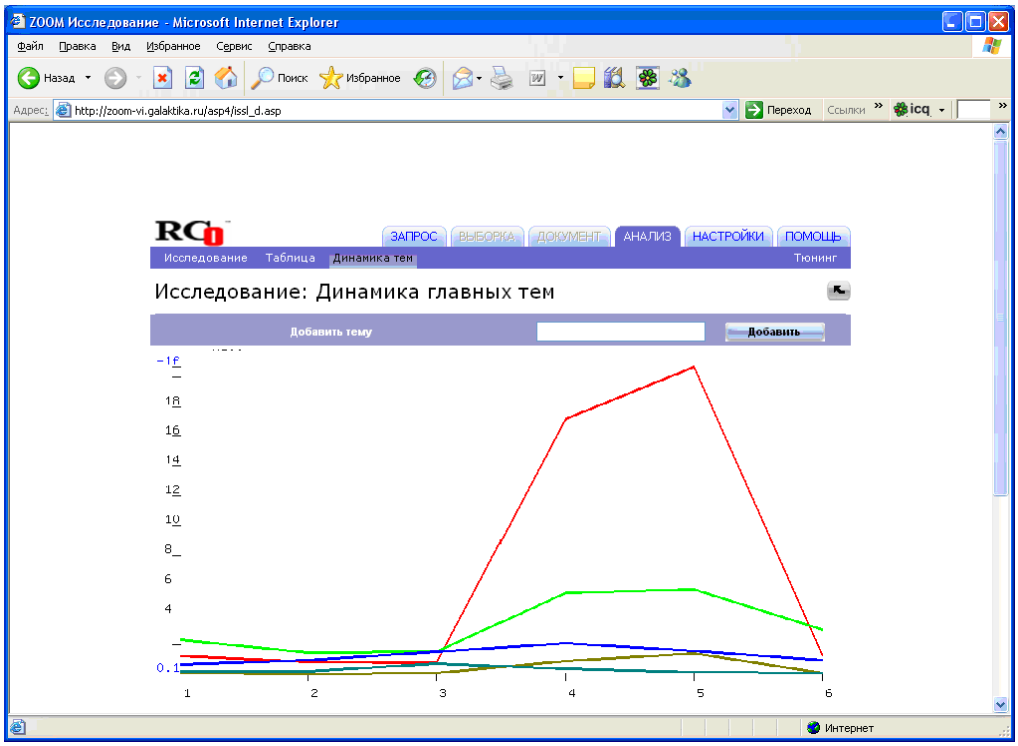


Рисунок 94. *Временные тренды главных тем объекта «Дмитрий Медведев»*

В ходе работы, можно отображать или не отображать главные темы, для этого в поле *«Показать»* напротив темы необходимо поставить/убрать флажок, затем нажать *«Выполнить»*. Чтобы удалить какую-либо тему, в поле *«Удалить»* напротив соответствующей темы поставьте флажок и нажмите *«Выполнить»*.

Для добавления слов или словосочетаний в поле *«Добавить тему»* под таблицей главных тем введите слово или словосочетание, по которому хотите построить дополнительный тренд и нажмите кнопку *«Добавить»*.

Рассмотрим предыдущий пример (пример построения *публикационного рейтинга*), проведем теперь *динамический анализ главных тем* объекта *Дмитрий Медведев*.

В поле *«Добавить тему»* введите *«газификация»*, нажмите *«Добавить»*. Аналогичным образом добавьте темы *«нацпроекты»*, *«газпром»*, *«голубое топливо»*, *«ипотечное кредитование»*. Затем установите флажки в поле *«Показать»* напротив всех тем, нажмите *«Выполнить»*.

Итак, мы построили график изменения главных тем объекта во времени, т.е. временные тренды. Под графиком располагается легенда.

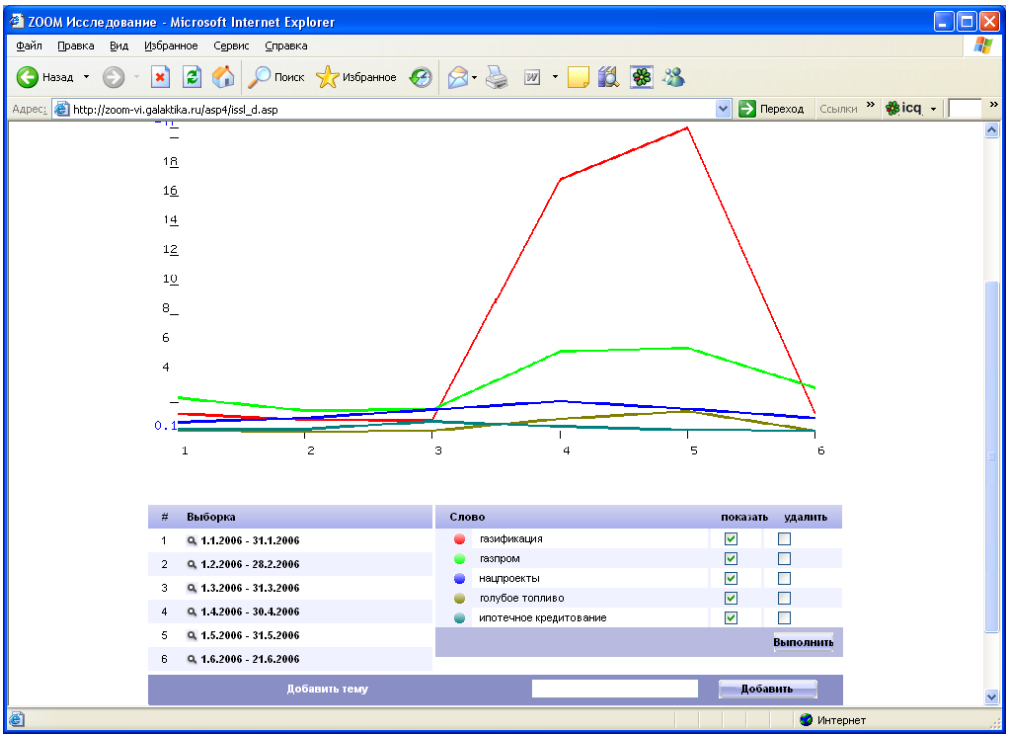


Рисунок 95. *Таблица значений*

При изучении графиков необходимо отмечать пики и спады значений параметров, выявляя, таким образом, причинно-следственные связи и основные тенденции, на основании чего можно построить прогноз.

В связи с этим *динамический анализ* применяется в совокупности с изучением информационного портрета объекта и соответствующих документов на каждом промежутке времени. Иногда бывает полезно провести *сравнительный анализ* главных тем объекта в различные периоды времени и выявить причины пиков и спадов.

7.7. Разработка отчетного документа

Отчетный документ составляется с учетом индивидуальных потребностей Руководителя.

Рекомендации по составлению **отчетного документа:**

- *Вступление* (цель, описание темы, объекта и предмета исследования; краткая справка с указанием периода и количества обработанных документов);
- *Основная часть* (описание *информационного образа* объекта; таблицы и графики – результаты сравнительного анализа и трендового исследования; примеры – значимые фрагменты из БД);
- *Выводы и предложения* (самостоятельная работа аналитика).

ГЛАВА 8. Редактирование документов

В Системе для документов некоторых баз (настроенных администратором) допускается редактирование отдельных полей документа.

Для баз, документы которых доступны для редактирования, несколько видоизменяется страница просмотра документа (Рисунок 96). В меню появляется пункт «Редактировать документ»

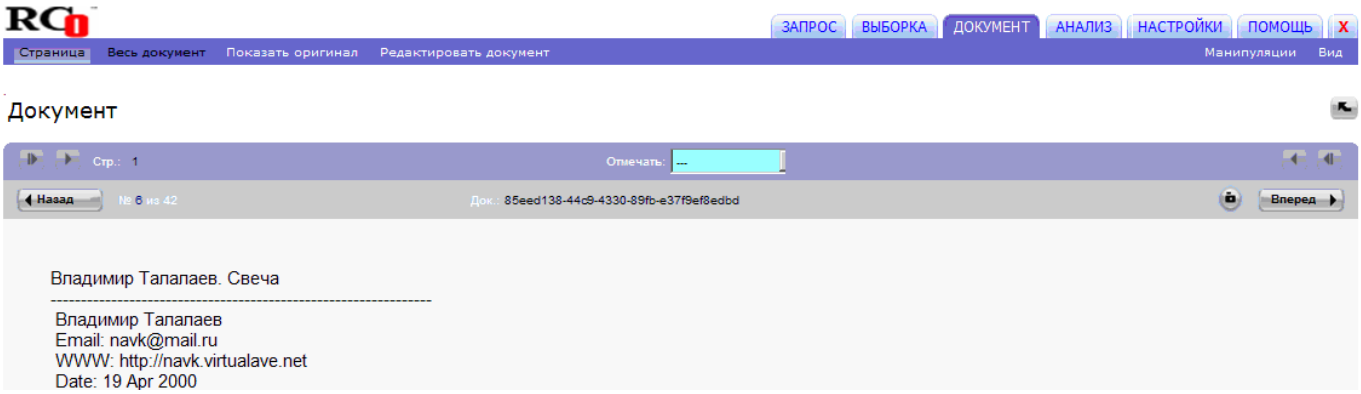


Рисунок 96. *Страница «Документ»*

При выборе пункта «Редактировать документ» открывается окно с перечнем доступных для редактирования полей (Рисунок 97). Для редактирования доступны только совпадающие для всех страниц документа поля. Не доступны для редактирования специальные поля и поле по умолчанию. После изменения значений полей можно сохранить сделанные изменения («Сохранить» Рисунок 98) или отказаться от них («Отменить» Рисунок 99).

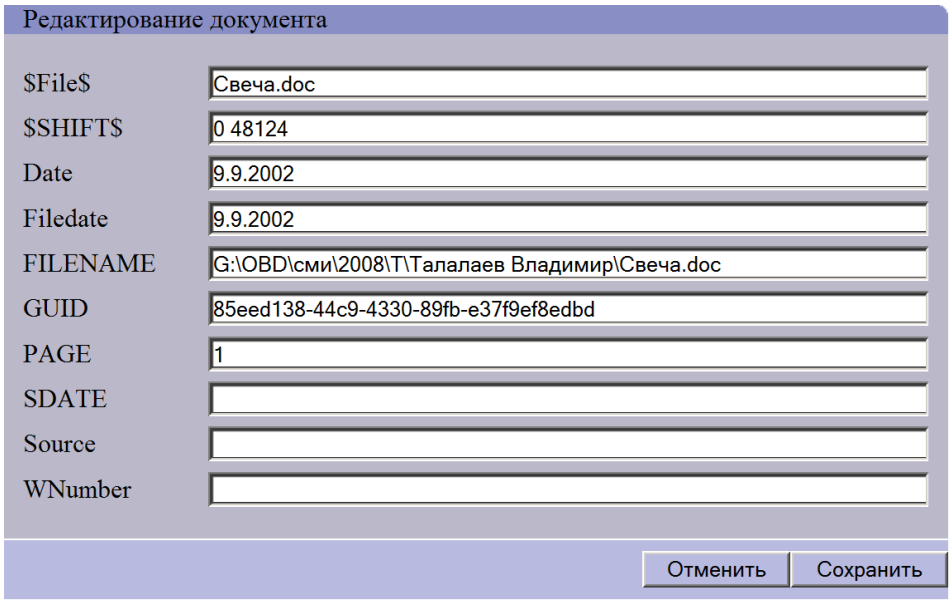


Рисунок 97. *Страница «Редактирование документа»*

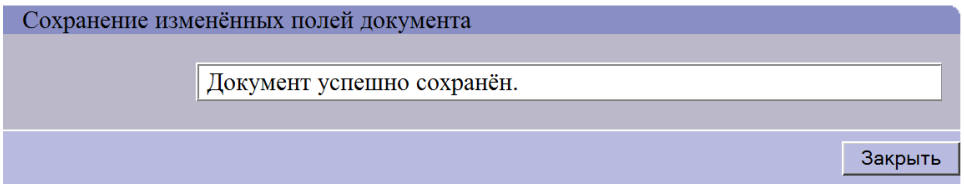


Рисунок 98. *Страница «Сохранение изменённых полей документа»*

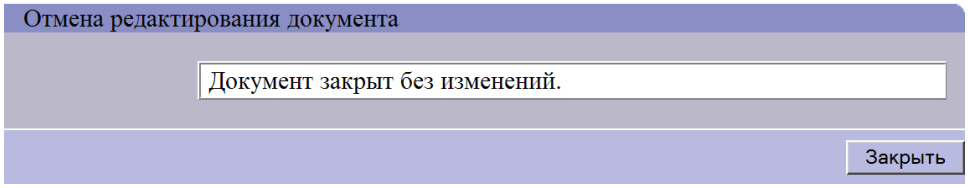


Рисунок 99. *Страница «Отмена редактирования документа»*

ГЛАВА 9. Настройки

Перед тем, как начинать работу с Системой, рекомендуем посетить страницу *«Настройки»*. Здесь Вы сможете устанавливать индивидуальные настройки Системы, например, настройки интерфейса, и при необходимости изменять их.

По умолчанию администратор устанавливает минимальный набор параметров, необходимых для обеспечения корректной работы Системы.

9.1. Основные настройки. Страница «Настройки» («Поиск»)

Перейдем на страницу *«Настройки»*. Главное меню страницы *«Настройки»* содержит разделы:

«Поиск» – для настройки параметров запроса;

«Внешний вид» – для настройки интерфейса пользователя;

«Статистика» – для настройки параметров статистики;

«Автоотсечение по рангам» – позволяет уменьшить и оптимизировать выборку;

«Смена пользователя» – раздел открывает регистрационную страницу для входа в систему.

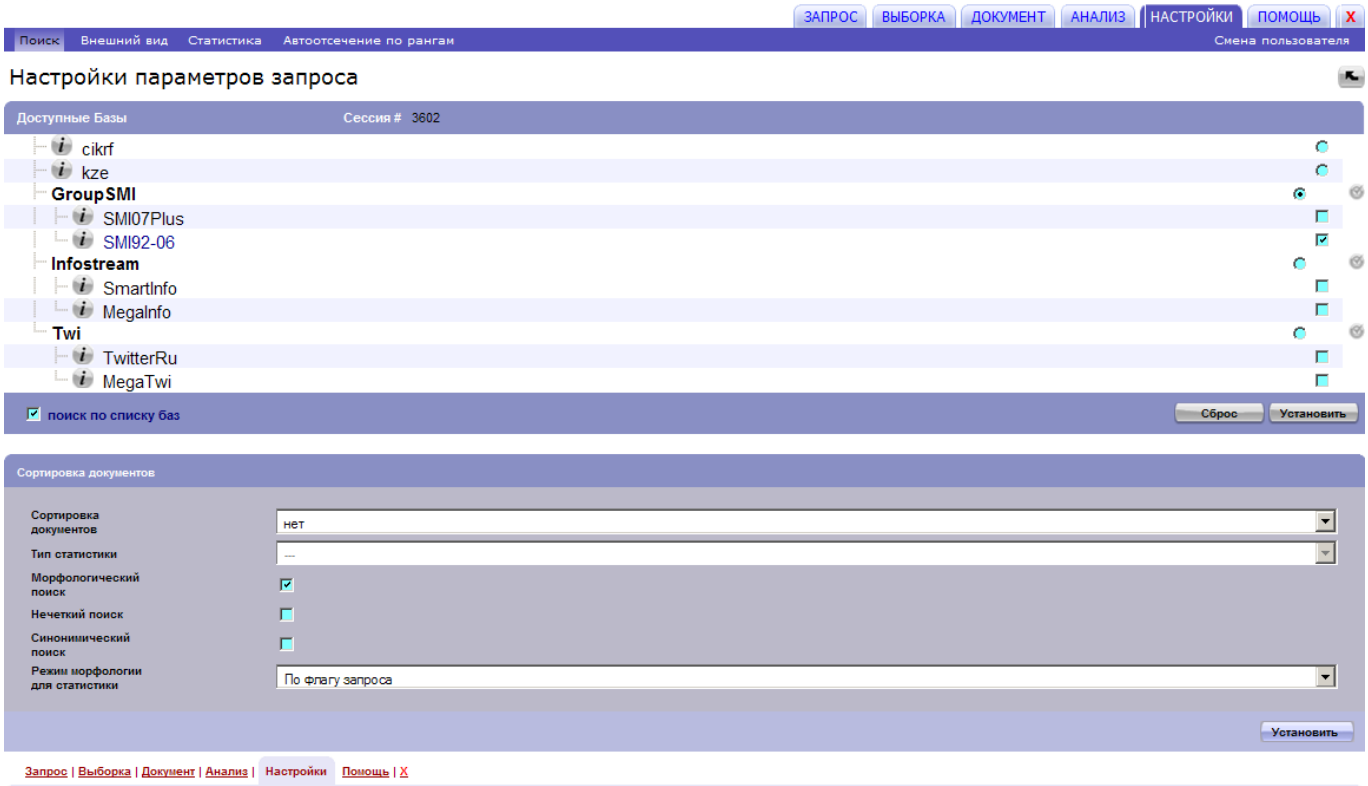


Рисунок 100. *Страница «Настройки параметров запроса»*

9.2. Изменение настроек системы. Страница «Настройки» («Поиск»)

Рассмотрим подробно раздел *«Поиск» (см.* Рисунок 100*)*. В данном разделе можно изменять установленные по умолчанию настройки Системы.

Поле «Доступные базы»:

Для работы в Системе нужно выбрать из предложенного списка одну базу, либо группу баз, установив флажки в соответствующих полях. Объединение баз в группы позволяет проводить запросы одновременно по нескольким базам. При этом для баз, объединённых в одну группу, статистика тоже объединена (значимость слов считается не в пределах одной базы, а для всей группы). Однако при поиске по группе не работает кластеризация.

Поле «Сортировка документов»:

- ✓ ***нет*** – функция отключена, документы в списке будут расположены в том порядке, в котором они находятся в базе
- ✓ ***по дате*** – в соответствии с датой документа, список найденных по запросу документов возглавит самый новый документ;
- ✓ ***по значимости*** – документы в списке будут расположены в порядке уменьшения значимости, список найденных по запросу документов возглавит документ, наиболее близкий к информационному портрету выборки;
- ✓ ***по имени файла*** – в соответствии с именем файла, из которого документ был загружен, список найденных по запросу документов возглавит документ, имя файла которого первое по алфавиту.

Поле «Тип статистики»:

- ✓ ***слова*** – в списке Главных тем будут отображены значимые слова;
- ✓ ***слова и пары*** – в списке Главных тем будут отображены значимые словосочетания.

Данное поле является доступным при установке *«Сортировка документов» «по значимости»*.

Поле «Морфологический поиск»:

- ✓ ***вкл. (стоит флажок)*** – морфология включена, т.е. поиск будет осуществляться по всем возможным формам слов, указанным в запросе. Например, по запросу «банк» найдутся все его словоформы, в том числе «банка», «банки», «банками» и пр. Установка флажка влияет на все поля текстового типа;
- ✓ ***выкл.*** – морфология выключена.

Поле «Нечеткий поиск»:

- ✓ ***вкл. (стоит флажок)*** – нечеткий поиск включён, т.е. Система отберет документы, релевантные тексту запроса, при этом учитывая возможность возникновения орфографической ошибки;
- ✓ ***выкл.*** – нечеткий поиск выключен.

Поле «Синонимический поиск»:

- ✓ ***вкл. (стоит флажок)*** – синонимический поиск включён, т.е. при указании в запросе одного из слов, имеющих синоним, автоматически выполняется поиск по всем синонимам (в том числе и многословным);
- ✓ ***выкл.*** – синонимический поиск выключен.

Поле «Режим морфологии для статистики»:

При подсчёте статистики по словам может учитываться или не учитываться морфология (при подсчёте статистики по словосочетаниям морфология учитывается всегда). Поле *«Режим морфологии для статистики»* влияет на то, будет ли при подсчёте статистики по словам учитываться морфология.

- ✓ ***по флагу запроса*** – при проведении запроса (для статистики!) учитывается состояние флажка «Морфологический поиск» в «Настройках» (или, что тоже самое, «Морфологический анализ» в конструкторе запроса);
- ✓ ***всегда использовать*** – морфология (для статистики!) используется вне зависимости от настроек поиска***;***
- ✓ ***никогда не использовать*** – морфология (для статистики!) не используется вне зависимости от настроек поиска.

Подробнее о режиме морфологии см. п. *2.2 Страница «Запрос»*

После того, как установлены необходимые параметры запроса, нажмите кнопку ***«Установить»***. С этого момента будут введены в действие новые настройки Системы.

9.3. Настройки параметров показа. Страница «Настройки» («Внешний вид»)

Для того чтобы установить настройки интерфейса Системы необходимо перейти в раздел *«Внешний вид» (см.* Рисунок 101 *и* Рисунок 102*).* Таким образом, Вы окажетесь на странице настроек интерфейса Системы.

Настройки внешнего вида

Инфопортрет:
Строгий поиск
Морфологический поиск
Отсекать мусор
Использовать кластеризацию
Показывать запрос в инфопортрете
Число строк 20

Список документов:
Сортировать документ по: Нет
Тип статистики Слова
Показывать рядом со списком: Объединенный спи
Кол-во документов на стр.: 10
Режим: добавить в папку: текущую страницу
Папка:
Фрагменты важный
Длина контекстного фрагмента: 2
Максимальная длина фрагмента 100

Анализ:
Сохранять новый запрос автоматически
Всегда добавлять в конец списка запросов
Имя запроса - это дата + время
При проведении нового запроса очищать исследование

Восстановленный документ:
Отмечать: Не показывать
Тип главных тем документа: Объединённый список
Кол-во главных тем документа: 5
Шаг многостр. просмотра: all
Навигация: + по найденным словам
Показ при открытии: найденное слово
Файл цитат: c:\Cite.txt
Ширина для оригинала: 75%
Файл документа:

По умолчанию Установить

Рисунок 101. *Страница «Настройки внешнего вида»*

Изменение пароля
Старый пароль
Новый пароль
Подвердите новый пароль
Изменить

Язык Russian

Установить специальный session timeout: 1230 минут Выполнить

Запрос | Выборка | Документ | Анализ | Настройки Помощь | X

Поиск | Внешний вид | Статистика | Автоотсечение по рангам Смена пользователя

Рисунок 102. *Страница «Настройки внешнего вида»*

Рассмотрим все поля, присутствующие на данной странице.

Панель «Инфопортрет»

Поле «Строгий поиск» – если в данном поле установлен флажок, то на странице *«Запрос»* будет установлен параметр тип запроса *строгий*, в противном случае – *простой*.

Поле «Морфологический поиск» – если в данном поле установлен флажок, главные темы в инфопортрете будут строиться с учетом морфологии. Подробнее о режиме морфологии см. п. *2.2 Страница «Запрос»*.

Поле «Отсекать мусор»: Рекомендуем всегда устанавливать флажок в этом поле для того чтобы получать выборку, наиболее соответствующую предметной области.

Поле «Использовать кластеризацию» – для включения в инфопортрет кластеризации. Подробнее о кластеризации см. п. 6.1 *Кластеризация*

Поле «Показывать запрос в инфопортрете» – для отображения окна запроса с текстом проведенного запроса.

Поле «Число строк» – для установки максимального количества главных тем, присутствующих в инфопортрете.

Панель «Анализ»

Поле «Сохранять новый запрос автоматически»:

- ✓ ***вкл.*** – если необходимо исследовать несколько выборок одновременно, то в поле необходимо установить флажок. Результаты нового запроса будут добавляться к уже имеющимся на странице исследований;
- ✓ ***выкл.*** – если предполагается работать с одной выборкой, то данное поле необходимо оставить пустым. При этом происходит полное обновление страницы исследования после проведения нового запроса.

Поле «Всегда добавлять в конец списка запросов»:

- ✓ ***вкл.*** – сохраненная выборка будет добавляться в конец списка обычных выборок. При нехватке места в исследовании для сохранения очередной выборки будет удалена последняя выборка в списке обычных выборок;
- ✓ ***выкл.*** – сохраненная выборка будет добавляться в начало списка обычных выборок.

Поле «Имя запроса»:

- ✓ ***дата + время*** – на странице исследований в заголовке запроса будет отображаться дата и время его проведения;
- ✓ ***текст запроса*** – на странице исследований в заголовке запроса будет отображаться его текст.

Поле «При проведении нового запроса очищать исследование»:

- ✓ ***вкл*** – при проведении нового запроса, включая динамику по времени, все выборки из исследования будут удалены;
- ✓ ***выкл.*** – все новые запросы будут включаться в конец списка выборок, пока их количество не превысит максимально допустимое количество.

Выбор того или иного параметра зависит от целей и задач пользователя.

Панель «Список документов»

Поле «Сортировать документ по»:

- ✓ ***нет*** – функция сортировки отключена, документы в списке будут расположены в том порядке, в котором они находятся в базе;
- ✓ ***дата*** – в соответствии с датой документа, список найденных по запросу документов возглавит самый новый документ;
- ✓ ***по значимости (слова)*** – документы в списке будут расположены в порядке уменьшения значимости слов, список найденных по запросу документов возглавит документ, наиболее близкий к информационному портрету выборки;
- ✓ ***по значимости (объединённый список)*** – документы в списке будут расположены в порядке уменьшения значимости слов и словосочетаний, список найденных по запросу документов возглавит документ, наиболее близкий к информационному портрету выборки.

Поле «Тип статистики»:

- ✓ ***слова*** – в списке Главных тем будут отображены значимые слова;
- ✓ ***словосочетания*** – в списке Главных тем будут отображены значимые словосочетания;
- ✓ ***объединенный список*** – в списке Главных тем будут отображены значимые слова и словосочетания.

Поле «Показывать рядом со списком»:

- ✓ ***не показывать*** – при выборе данной настройки не будет отображаться таблица – подсказка;
- ✓ ***слова*** – при выборе данной настройки в таблице – подсказке окажутся только слова;
- ✓ ***словосочетания*** – при выборе данной настройки в таблице – подсказке окажутся только словосочетания;
- ✓ ***объединенный список*** – при выборе данной настройки в таблице – подсказке окажутся слова и словосочетания.

Поле «Количество документов на странице» позволяет установить количество документов, отображаемое Системой на странице *«Выборка» («Документы»*) после проведения запроса.

Режим «Добавить в папку»:

- ✓ ***текущую страницу*** – позволяет добавить в папку текущую, просматриваемую страницу документа;
- ✓ ***найденные страницы*** – позволяет добавить в папку найденные страницы документа;
- ✓ ***все страницы*** – позволяет добавить в папку все страницы документа.

Поле «Папка» – позволяет выбрать папку на диске, в которую требуется поместить документ. Если появляется ошибка «Невозможно загрузить компонент», см. ГЛАВА 10. Аварийные ситуации

Поле «Фрагменты»

- ✓ --- – в списке документов отображаются только заголовки;
- ✓ ***важный*** – под заголовком будет отображаться наиболее значимый фрагмент документа;
- ✓ ***существенные*** – под заголовком будут отображаться несколько фрагментов, содержащих найденные слова (контекстные фрагменты).

Поле «Длина контекстного фрагмента» позволяет определить количество слов до и после найденного слова в контекстном фрагменте.

Поле «Максимальная длина фрагмента» позволяет определить общее количество слов во фрагменте. Для существенных фрагментов этот параметр определяет суммарную длину всех фрагментов. Если суммарная длина всех фрагментов документа превысит заданную в настройках, в конце документа появится значок «==>», означающий, что имеются другие, не показанные пользователю фрагменты.

<u>Панель «Восстановленный документ»</u>

Поле «Отмечать»:

- ✓ ***не показывать*** – при выборе данной настройки в восстановленном документе будет отсутствовать разметка значимых слов и словосочетаний;
- ✓ ***слова*** – при выборе данной настройки в восстановленном документе будут подсвечиваться значимые слова;
- ✓ ***словосочетания*** – при выборе данной настройки в восстановленном документе будут подсвечиваться значимые словосочетания;
- ✓ ***объединенный список*** – при выборе данной настройки в восстановленном документе будут подсвечиваться значимые слова и словосочетания.

Поле «Тип главных тем документа»:

- ***не показывать*** – главные темы не будут отображаться;
- ***слова*** – главными темами будут слова;
- ***словосочетания*** – главными темами будут словосочетания;
- ***объединенный список*** – главными темами будут слова и словосочетания;
- ***отдельные списки*** – отдельные списки главных тем из слов и главных тем из словосочетаний.

Поле «Количество главных тем документа» – позволяет установить максимальное количество главных тем, отображаемых в списке главных тем документа.

Поле «Шаг многостраничного просмотра» – позволяет установить количество страниц многостраничного документа, отображаемых при просмотре на странице *«Документ»*. При выборе режима ***all*** будут отображены все страницы документа. Например, если в настройках будет указано 5 страниц, то будут показаны найденная страница и 4 следующие страницы.

Поле «Навигация»:

- ***по умолчанию*** – навигация, настроенная администратором;
- ***стандартная*** – навигация по словам выключена;
- ***по найденным словам*** – навигация по словам включена.

Поле «Показ при открытии» определяет позиционирование окна браузера при переходе к просмотру документа.

- ***начало страницы*** – отображение документа с начала текста;
- ***найденное слово*** – отображение со строки, содержащей первое найденное слово;
- ***значимый фрагмент*** – отображение значимого фрагмента, показанного в списке документов.

Поле «Файл цитат» позволяет выбрать каталог для сохранения файла цитат при цитировании в текстовый файл (с расширением .txt).

Кнопка ***(документ Word)*** – для сохранения цитат в документ Word.

Поле «Ширина для оригинала» – позволяет установить масштаб страницы, в которой будет отображаться оригинальный текст документа.

Поле «Файл документа» – позволяет выбрать папку для сохранения документа.

Кнопка *«По умолчанию»:*

Вводит в действие настройки, установленные администратором по умолчанию.

Кнопка *«Установить»:*

Вводит в действие новые установленные параметры настройки.

Панель «Изменение пароля»

В поле *«Старый пароль»* введите текущий пароль. Если Вы впервые осуществили вход в Систему, то оставьте данное поле пустым. Поля *«Новый пароль»* и *«Подтвердите новый пароль»* предназначены для ввода нового пароля. Кнопка *«Изменить»* вводит в действие новый пароль.

В ***поле «Язык»*** можно выбрать русскоязычное или англоязычное представление.

Панель «Установить специальный session timeout:»

С помощью этой панели Вы можете изменять длительность пребывания на данной странице от 20 минут до 24 часов. Для подтверждения установки нажмите *«Выполнить»*.

9.4. Настройка статистики. Страница «Настройки» («Статистика»)

Настройки статистики влияют на то, какие главные темы будут фигурировать в инфопортрете запроса.

Для того чтобы установить настройки Статистики необходимо перейти в раздел «Настройки» *«Статистика»*. Появится окно настроек статистики (Рисунок 103).

Рисунок 103. *Раздел «Статистика» (стандартный вариант)*

Большинству пользователей достаточно возможности выбора статистических макросов. Для более тонкой настройки может быть предоставлен доступ к другим параметрам статистики. В этом

случае администратор Системы может настроить доступ к ним и вид страницы настроек статистики становится таким, как показано на рисунке (Рисунок 104).

Рассмотрим поля, присутствующие на странице настроек.

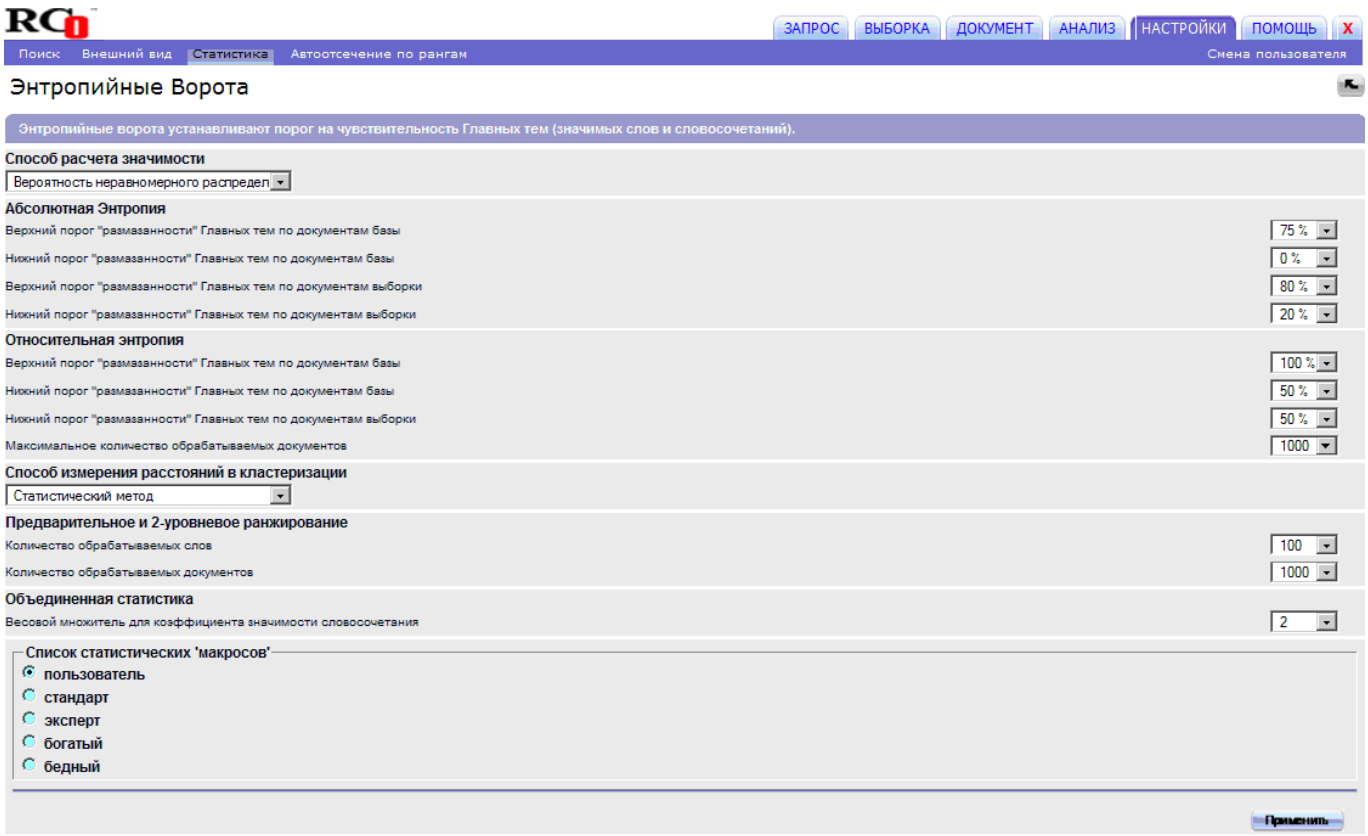


Рисунок 104. *Раздел «Статистика» (полный вариант)*

<u>Поле «Способ расчёта значимости»</u>

- ✓ ***Вероятность равномерного распределения*** – значимость рассчитывается исходя из равномерного распределения слов по тексту документов;
- ✓ ***Вероятность неравномерного распределения*** – значимость рассчитывается учитывая неравномерное распределения слов по документам.

<u>Панель «Абсолютная Энтропия»</u>

<u>Поле «Верхний порог «размазанности» Главных тем по документам базы»</u>

Задаётся в процентах; отсеиваются наиболее общие темы для данной базы.

<u>Поле «Нижний порог «размазанности» Главных тем по документам базы»</u>

Задаётся в процентах; отсеиваются самые редкие для данной базы темы.

<u>Поле «Верхний порог «размазанности» Главных тем по документам выборки»</u>

Задаётся в процентах; отсеиваются наиболее общие темы для данной выборки.

<u>Поле «Нижний порог «размазанности» Главных тем по документам выборки»</u>

Задаётся в процентах; отсеиваются самые редкие для данной выборки темы.

Панель «Относительная Энтропия»

Поле «Верхний порог «размазанности» Главных тем по документам базы»

Задаётся в процентах; отсекает темы, наиболее равномерно распределённые по документам базы.

Поле «Нижний порог «размазанности» Главных тем по документам базы»

Задаётся в процентах; отсекает самые неравномерно распределённые темы.

Поле «Нижний порог «размазанности» Главных тем по документам выборки»

Задаётся в процентах; отсекает самые неравномерно распределённые темы для данной выборки.

Поле «Максимальное количество обрабатываемых документов»

Задаётся количество включаемых в обработку документов. Служит для управления скоростью обработки. При снижении количества повышается скорость, но снижает полноту информационного портрета.

Поле «Способ измерения расстояний в кластеризации»

- ✓ ***Статистический метод*** – ;
- ✓ ***Прямой корреляционный метод*** – ;
- ✓ ***Энтропийный метод (экспериментальный)*** – ;
- ✓ ***Метод пересечения выборок*** – .

Панель «Предварительное и 2-уровневое ранжирование»

Поле «Количество обрабатываемых слов»

Задаётся количество слов инфопортрета, используемых в ранжировании.

Поле «Количество обрабатываемых документов»

При сортировке выборки сперва проводится быстрая грубая оценка значимости документов (предварительное ранжирование). Затем из числа наиболее значимых отбирается указанное количество документов, обрабатываемых ранжированием 2-го уровня (более точный вид сортировки, но более медленный).

Панель «Объединённая статистика»

Поле «Весовой множитель для коэффициента значимости словосочетания»

Выбирается из диапазона от 2 (оптимально) до 10. Задаёт превышение значимости словосочетаний над отдельными словами.

Панель «Список статистических ‘макросов’»

Возможны следующие варианты (графическое представление режимов отображения главных тем см. Рисунок 105):

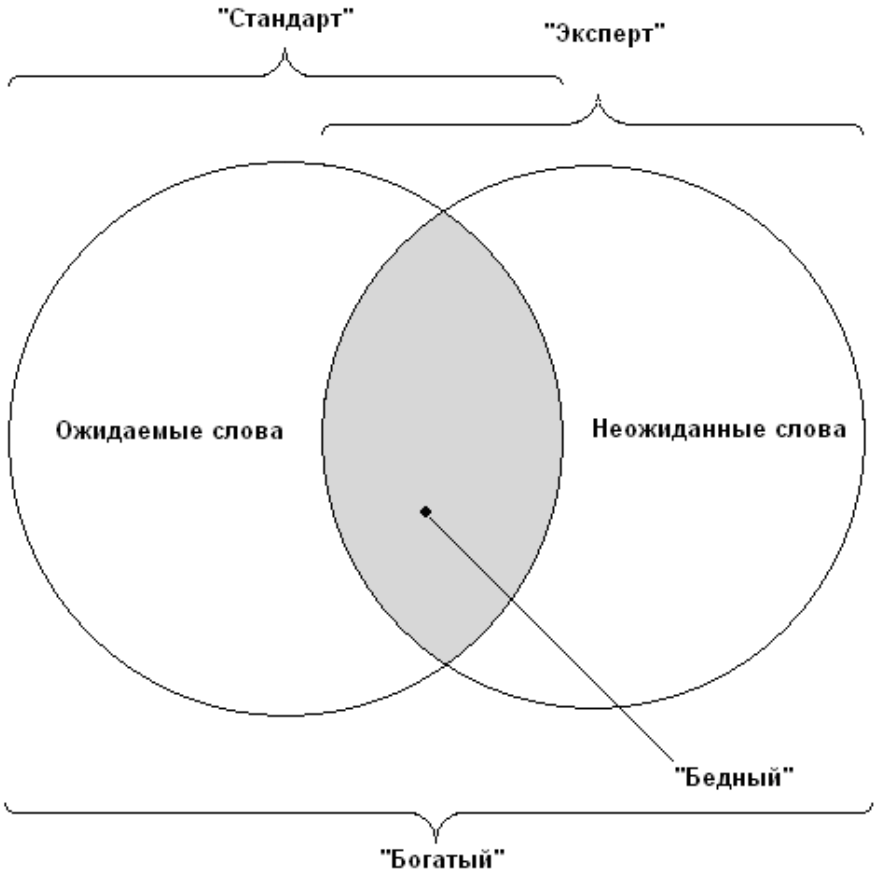


Рисунок 105. *Варианты отображения главных тем*

«Стандарт». В главных темах – обычные для запроса (скорее ожидаемые) слова;

«Эксперт». В главных темах появляются слова, не имеющие прямого отношения к запросу (скорее неожиданные);

«Богатый». В главных темах – и ожидаемые, и неожиданные слова;

«Бедный». В главных темах – слова, ожидаемые и неожиданные одновременно.

После выбора необходимых настроек, нажмите *«Применить»*.

9.5. Страница «Настройки» («Автоотсечение по рангам»)

Режим автоотсечения позволяет уменьшить и оптимизировать выборку путём удаления малозначимых документов.

Подробности:

Для этого выборка сперва сортируется в соответствие с режимом, заданным для автоотсечения. Для каждого документа вычисляется его ранг. Ранг это характеристика документа, связанная с

заданным режимом сортировки документа (по значимости, по папке, быстрое ранжирование). Чем больше ранг, тем более значим документ для данной выборки. Из отсортированной выборки отбрасываются (отсекаются) документы с малым рангом. Затем остаток выборки сортируется в соответствие с порядком, в котором пользователь хочет видеть документы (например, по дате).

Страница «Настройки» – «Автоотсечение» (Рисунок 106) содержит следующие строки:

Поле «Режим автоотсечения» – позволяет отключить автоотсечение или выбрать его тип:

- ✓ ***отключено*** – отключает режим автоотсечения;
- ✓ ***абсолютное значение*** – отбрасываются документы с рангом ниже указанного значения;
- ✓ ***относительное значение*** – отбрасываются документы с рангом ниже указанной доли от максимального ранга. Это удобнее, т.к. обычно максимальный ранг в выборке заранее не известен.

Поле «Минимальный ранг» – здесь указывается значение минимального ранга (при автоотсечении по абсолютному значению) или долю от максимального ранга (при автоотсечении по относительному значению). При значении, равном нулю, автоотсечения не происходит.

Поле «Сортировка» – выбирается режим сортировки, при котором будет проводиться отсечение:

- ✓ ***быстрое ранжирование (для простых запросов)*** – документы по простому запросу предъявляются в порядке их соответствия (релевантности) запросу;
- ✓ ***по значимости*** – документы показываются в порядке убывания значимости;
- ✓ ***по имени файла*** – документы показываются в алфавитном порядке имен файлов, из которых документ был загружен;
- ✓ ***по папкам*** – документы показываются в порядке их соответствия тематике папок (подробнее см.6.2.4 *Сортировка выборки по папке или запросу*).

Поле «Тип статистики» – это поле активно, если выбран вид сортировки по значимости или папкам. Можно установить один из трех вариантов: слова, словосочетания, объединенный список.

- ✓ ***Слова*** – в списке будут показаны ключевые слова;
- ✓ ***Словосочетания*** – список будет состоять из ключевых словосочетаний;
- ✓ ***Объединённый список*** – появятся и ключевые слова и словосочетания единым списком.

Если установлен режим сортировки документов (*«Настройки» – «Поиск» – «Сортировка документов»*), то оставшиеся после автоотсечения документы в выборке будут автоматически отсортированы по выбранному параметру (т.е. «по дате» или «по значимости»).

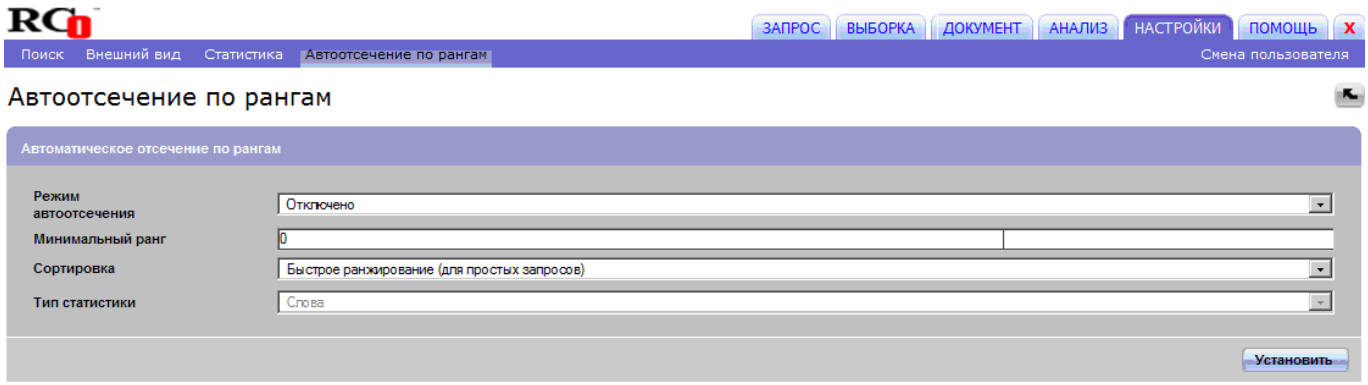
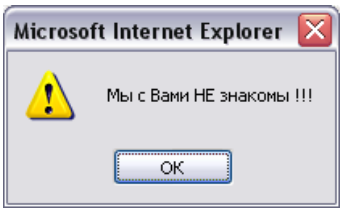
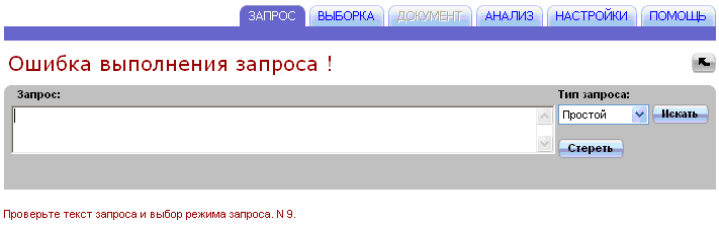
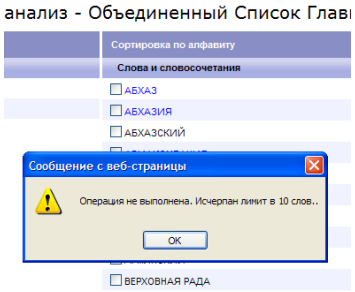
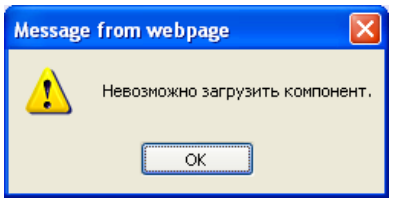


Рисунок 106. *Раздел «Автоотсечение по рангам»*

ГЛАВА 10. Аварийные ситуации

В процессе работы с Системой могут возникать различные аварийные ситуации, например при некорректном входе, либо при проведении запроса. Рассмотрим возможные аварийные ситуации и действия, которые необходимо предпринять для их преодоления.

Аварийная ситуация	Действия, предпринимаемые в случае возникновения аварийной ситуации
Сообщение об ошибке при входе в Систему  	1) Проверьте правильность ввода имени и пароля; 2) Повторите ввод имени и пароля. Если вы забыли пароль, обратитесь к администратору Системы, он обеспечит Вам доступ, затем Вы сможете завести новый пароль.
Ошибка выполнения запроса  	Данная ошибка возникает в том случае, когда Вы некорректно ввели текст запроса. 1) Нажмите кнопку 2) Повторите ввод запроса.
Сообщение об ошибке при занесении в динамику главных тем более 10-ти слов.  	Для наглядности представления на графике в динамике главных тем введено ограничение на количество трендов. 1) Нажмите «ОК»; 2) Сократите количество слов (словосочетаний) на графике до 10-ти.
Сообщение об ошибке при невозможности загрузки ActiveX элемента.	Для реализации полной функциональности системы используются ActiveX элементы.



ActiveX элементы используются для выполнения следующих операций:

- ✓ ***Цитирование документа;***
- ✓ ***Печать / сохранение документа;***
- ✓ ***Операции с папками;***
- ✓ ***Вход в систему при выключенной Windows-аутентификации.***

1) Включите ActiveX элементы в настройках браузера.

10.1. Включение ActiveX элементов

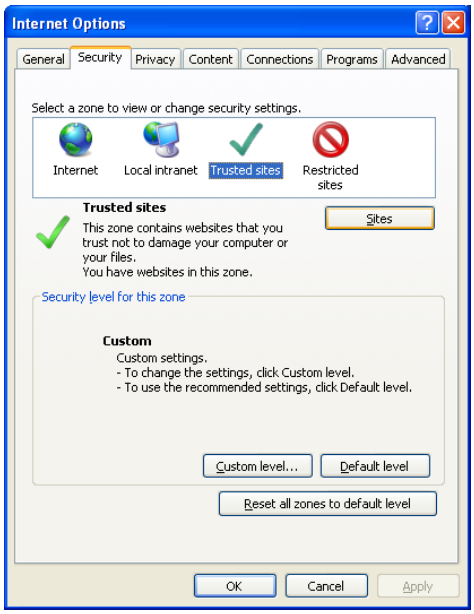
Для разрешения загрузки и запуска ActiveX элементов необходимо выполнить следующие действия:

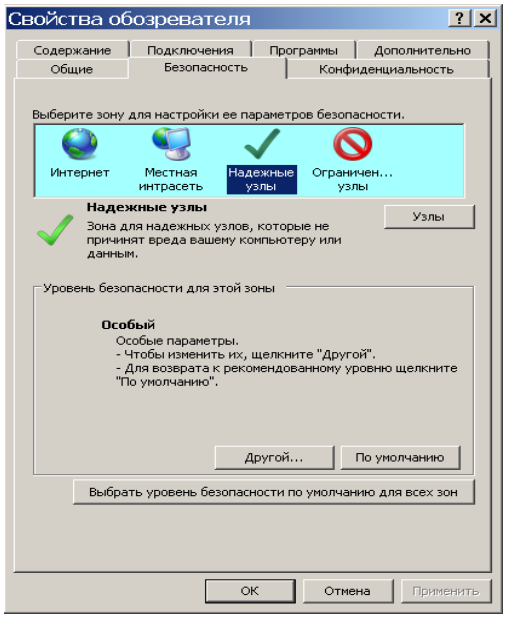
«Tools» (Сервис) – *«Internet Options» (Свойства обозревателя)* – вкладка *«Security» (Безопасность):*

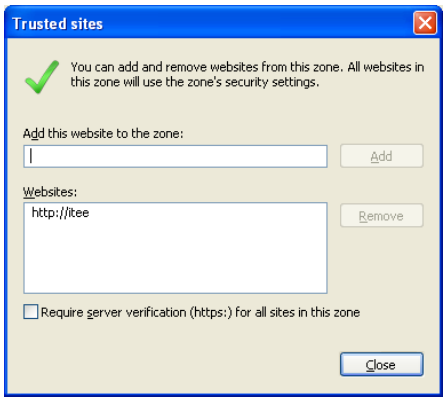
– установите значок *«Trusted sites» (Надёжные узлы)* в поле *«Select a zone to view or change security settings» (Выберите зону для настройки её параметров безопасности)* – нажмите кнопку *«Sites» (Узлы)* – появляется окно *«Trusted sites» (Надёжные узлы)* – наберите текст в строке *«Add this website to the zone» (Добавить в зону следующий узел)* – уберите флажок *«Require server vertification (…)» (Для всех узлов этой зоны требуется проверка серверов (https:))* – нажмите кнопку *«Add» (Добавить)*;

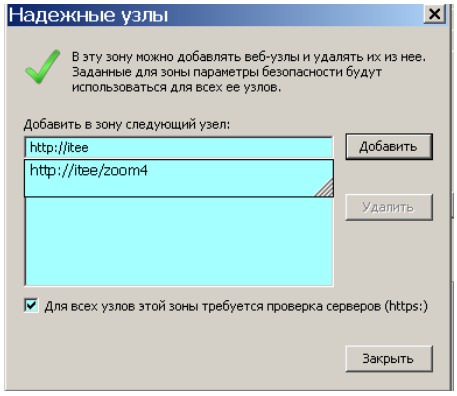
– *В поле «Security level for the zone» (Уровень безопасности для этой зоны)* нажмите кнопку *«Custom level…» (Другой…)* – появляется окно *«Security Settings – Trusted Sites Zones» (Параметры безопасности – зона надёжных…)* – разрешите все операции, связанные с загрузкой *ActiveX* (установите флажок *«Enable» (Включить)* в соответствующих пунктах) – нажмите кнопку *«OK»*;

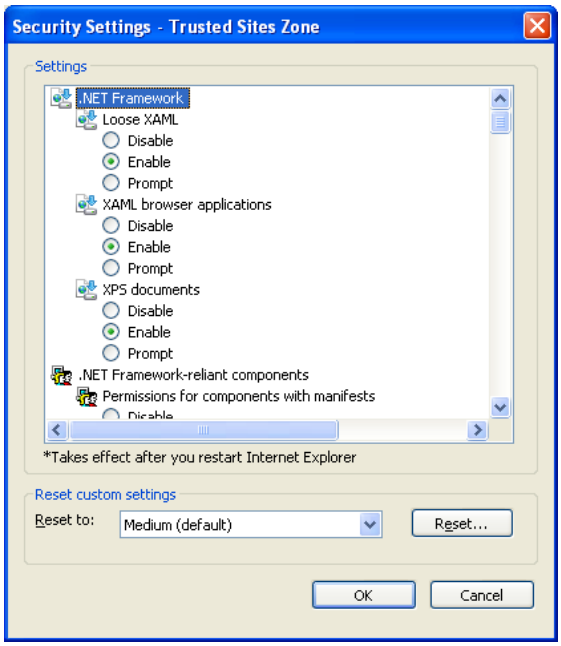
– нажмите кнопку *«OK»*.

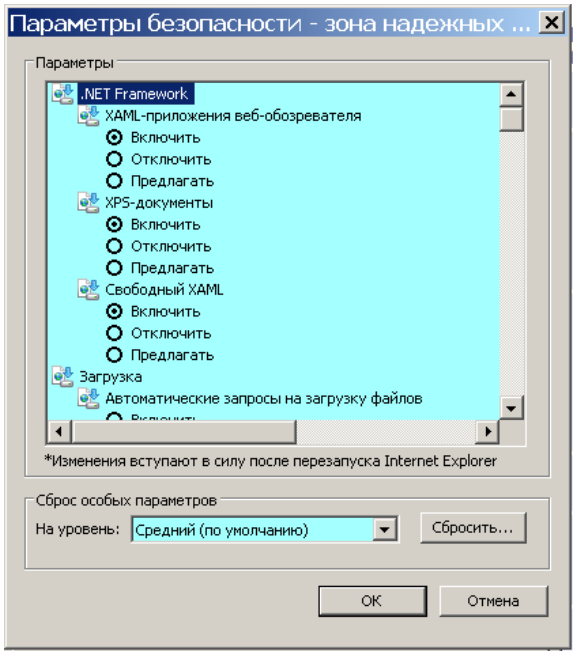

Internet Options
General
Security
Privacy
Content
Connections
Programs
Advanced
Select a zone to view or change security settings.
Internet
Local intranet
Trusted sites
Restricted sites
Trusted sites
This zone contains websites that you trust not to damage your computer or your files.
You have websites in this zone.
Sites
Security level for this zone
Custom
Custom settings.
- To change the settings, click Custom level.
- To use the recommended settings, click Default level.
Custom level...
Default level
Reset all zones to default level
OK
Cancel
Apply


Свойства обозревателя
Содержание
Подключения
Программы
Дополнительно
Общие
Безопасность
Конфиденциальность
Выберите зону для настройки ее параметров безопасности.
Интернет
Местная интрасеть
Надежные узлы
Ограничен... узлы
Надежные узлы
Зона для надежных узлов, которые не причинят вреда вашему компьютеру или данным.
Узлы
Уровень безопасности для этой зоны
Особый
Особые параметры.
- Чтобы изменить их, щелкните "Другой".
- Для возврата к рекомендованному уровню щелкните "По умолчанию".
Другой...
По умолчанию
Выбрать уровень безопасности по умолчанию для всех зон
OK
Отмена
Применить


Trusted sites
You can add and remove websites from this zone. All websites in this zone will use the zone's security settings.
Add this website to the zone:
Add
Websites:
http://itee
Remove
Require server verification (https:) for all sites in this zone
Close


Надежные узлы
В эту зону можно добавлять веб-узлы и удалять их из нее. Заданные для зоны параметры безопасности будут использоваться для всех ее узлов.
Добавить в зону следующий узел:
http://itee
http://itee/zoom4
Добавить
Удалить
Для всех узлов этой зоны требуется проверка серверов (https:)
Закрыть


Security Settings - Trusted Sites Zone
Settings
.NET Framework
Loose XAML
Disable
Enable
Prompt
XAML browser applications
Disable
Enable
Prompt
XPS documents
Disable
Enable
Prompt
.NET Framework-reliant components
Permissions for components with manifests
*Takes effect after you restart Internet Explorer
Reset custom settings
Reset to:
Medium (default)
Reset...
OK
Cancel


Параметры безопасности - зона надежных ...
Параметры
.NET Framework
XAML-приложения веб-обозревателя
Включить
Отключить
Предлагать
XPS-документы
Включить
Отключить
Предлагать
Свободный XAML
Включить
Отключить
Предлагать
Загрузка
Автоматические запросы на загрузку файлов
*Изменения вступают в силу после перезапуска Internet Explorer
Сброс особых параметров
На уровень:
Средний (по умолчанию)
Сбросить...
OK
Отмена

Рекомендации по освоению

Постановка задачи

Перед началом работы с Системой необходимо четко поставить цель и задачи.

Первичная оценка исходных данных,

определение объекта исследования и предметной области

После того, как определена цель исследования, необходимо проанализировать имеющиеся исходные данные и выбрать необходимые для работы источники информации (БД). Далее следует определить объект исследования и предметную область.

Сбор данных, поиск документов

Начать работу с Системой необходимо со страницы *«Запрос»* (раздел *«Запрос»*) (Рисунок 107). В окно запроса следует ввести исследуемый объект или составить запрос при помощи языка запросов *(см. Приложение 1)*, затем заполнить поля, формирующие условия поиска и нажать кнопку *«Искать»*.

Рисунок 107. *«Окно поиска документов»*

Итоги запроса могут быть представлены в двух режимах:

1. *Результаты поиска – выборка документов.*

   На экране отображается список документов.

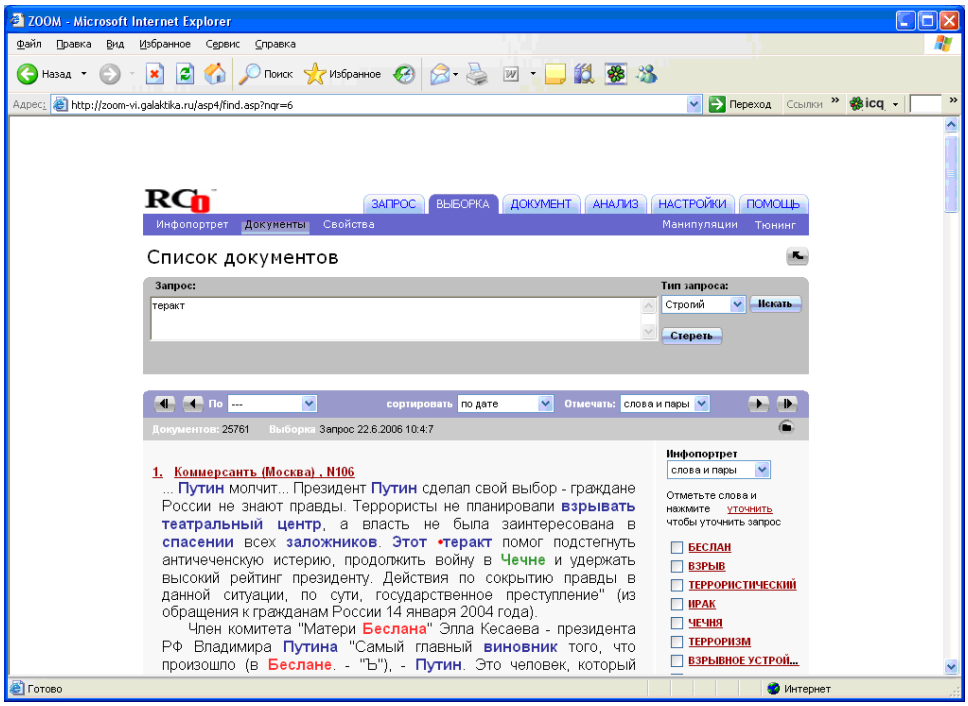


Рисунок 108. *«Результаты запроса»*

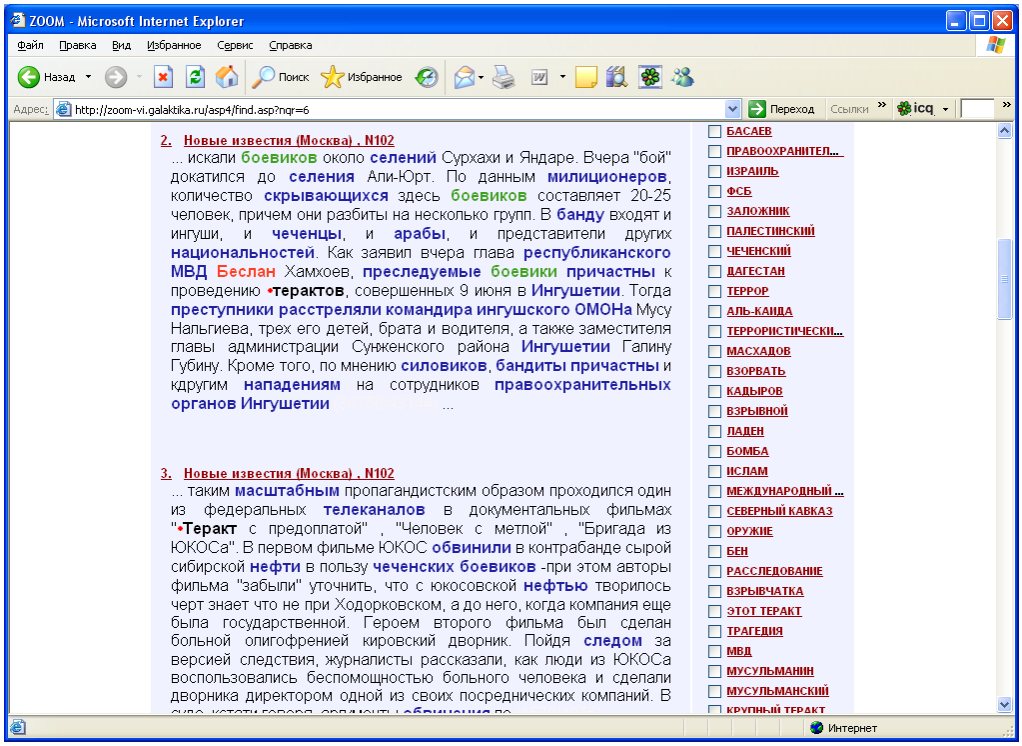


Рисунок 109. *«Результаты запроса»*

В левой части экрана, в соответствии со способом сортировки по убыванию размещены наиболее значимые фрагменты документов, релевантные запросу. В каждом из приведенных документов цветом выделены слова и словосочетания – в порядке возрастания значимости (синий, зеленый, желтый, красный).

В правой части экрана расположена таблица-подсказка, с помощью которой можно уточнить запрос. Дальнейшая работа происходит в рамках одной выборки, и документы, отражаемые после уточнения запроса, приводятся в порядке возрастания выбранного способа сортировки.

2. *Результаты поиска – информационный портрет объекта*.

На экране отображается Информационный портрет объекта (Рисунок 110).

Информационный портрет объекта представляет множество слов и словосочетаний, которые содержаться в найденных по запросу документах. Информационный портрет позволяет выявлять ключевые объекты (термины), которые всесторонне описывают объект. В результате, Вы можете сконцентрироваться на связях (словах) и свойствах (словосочетаниях) объекта.

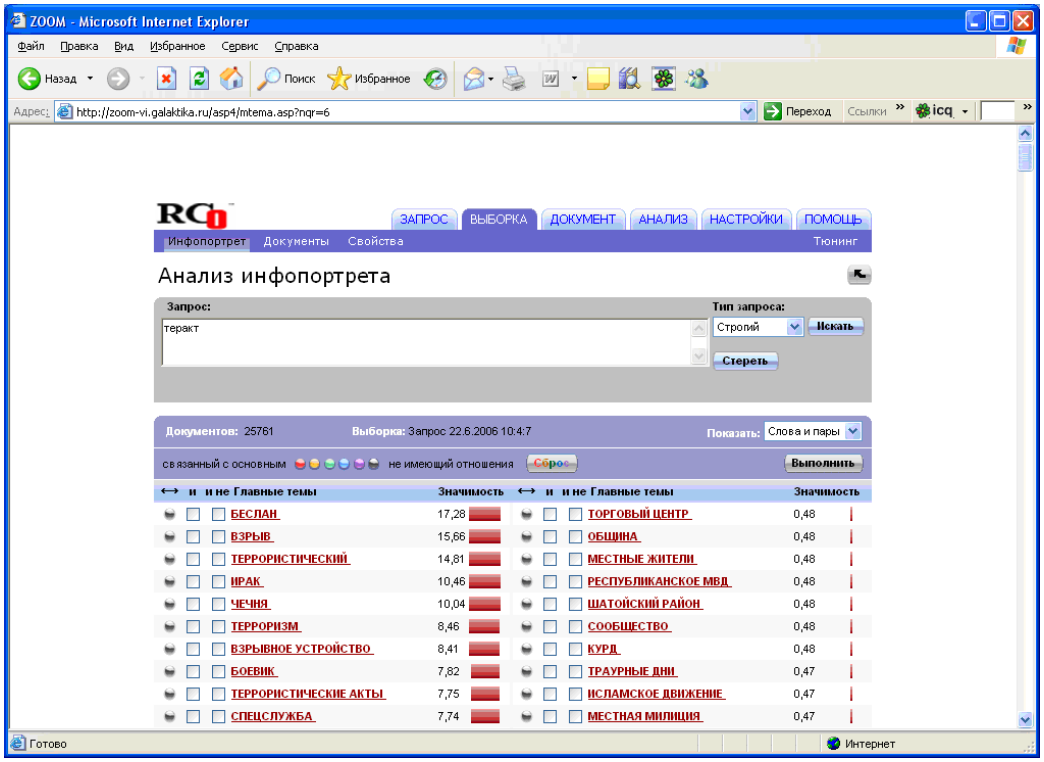


Рисунок 110. *«Информационный портрет объекта»*

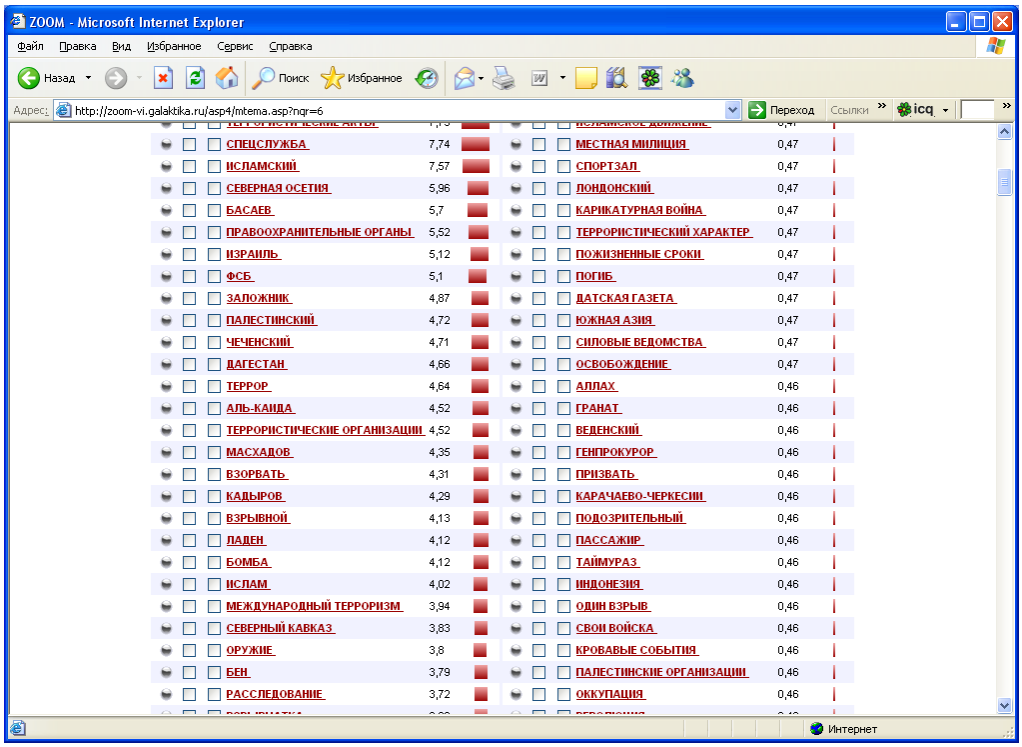


Рисунок 111. *«Информационный портрет объекта»*

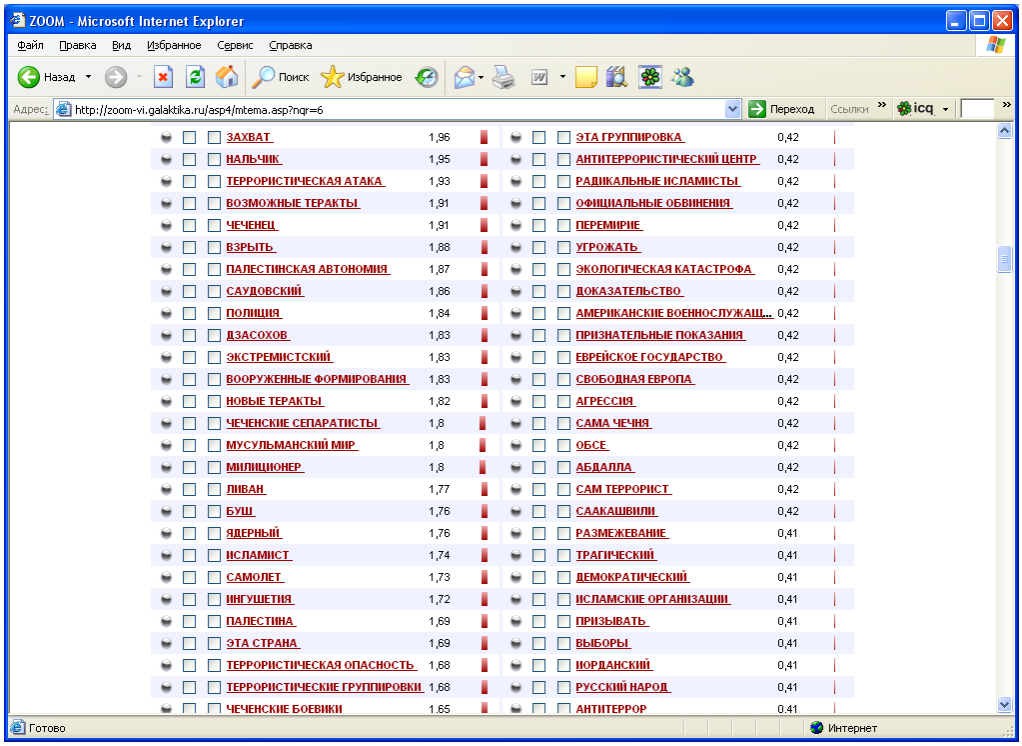


Рисунок 112. *«Информационный портрет объекта»*

Анализ информационного образа объекта

Слова и словосочетания, имеющие наибольшую степень значимости, как правило, отражают **очевидные, общеизвестные вещи, характерные для объекта**. Например (Рисунок 110), к этой категории относятся такие слова, как *«Беслан»*, *«терроризм»*, *«взрыв»*, *«Чечня»* и др.

Слова и словосочетания со средней степенью значимости показывают **некоторые особенности объекта и частные явления, которые с ним происходят**. Например, к этой категории относятся слова *«Аль-Каида»*, *«ислам»*, *«ФСБ»* и др.

Слова и словосочетания с наименьшей степенью значимости показывают **нечто редкое, малоизвестное, не типичное для объекта**. Например, к этой категории относятся слова *«лондонский»*, *«генпрокурор»*, *«оккупация»* и др.

При необходимости можно использовать функцию кластеризация[2], которая позволяет определить связь в тексте между объектами внутри Информационного портрета. Кластеры расположены слева от слов и словосочетаний и изначально окрашены в черный цвет (Рисунок 110– Рисунок 112). Для того чтобы выявить связь между объектами необходимо нажать на интересующее слово или словосочетание. После этого связанные объекты отмечаются различными цветами – в порядке возрастания значимости синий, голубой, зеленый, желтый и красный. При наведении мышью на «шарик» возникают цифры процентов близости объекта к выбранному *центру кластеризации* (обозначенному).

Уточнение предметной области объекта

Если Вы находитесь в режиме поиска документов, то уточнить предметную область можно с помощью таблицы-подсказки (Рисунок 108), в которой необходимо отметить интересующие слова и нажать кнопку *«Уточнить»*. При этом запрос уточняется в рамках одной выборки документов. После того, как проведены уточнения запроса предлагается перейти к изучению информационного портрета (страница *«Выборка»* – раздел *«Инфопортрет»*).

Если Вы находитесь в Информационном портрете, то уточнить предметную область объекта можно тремя способами:

1. Использовать функцию «И». Слова и словосочетания, которые Вас заинтересовали необходимо отметить оператором «И» и нажать на кнопку *«Уточнить»*. В результате происходит уточнение темы в рамках одной выборки. При этом количество документов уменьшается.
2. Использовать функцию «И НЕ». Слова и словосочетания, которые не соответствуют исследуемой Вами предметной области, необходимо отметить оператором «И НЕ» и нажать на кнопку *«Уточнить»*. Таким образом, Вы «удаляете» ненужные слова и словосочетания, и тем самым уменьшается количество документов в выборке.

[2] Функция *Кластеризация* применяется в Информационном портрете, длина которого не более 200 слов.

После того, как предметная область четко определена, можно перейти на страницу *«Анализ»* (раздел *«Исследования»*) и продолжить изучение объекта (Рисунок 113).

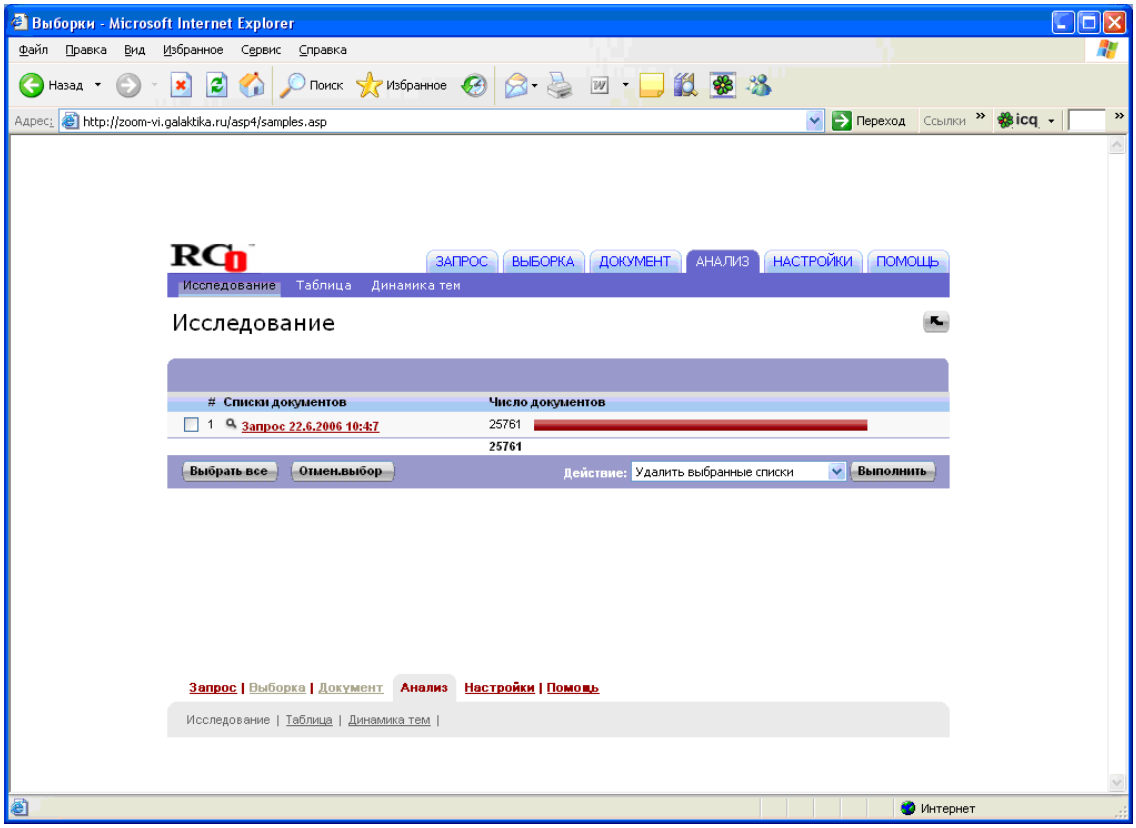


Рисунок 113. *Страница «Анализ»*

Разработка гипотез, подтверждение (опровержение)

Разработка гипотез, а также их дальнейшее подтверждение или опровержение является неотъемлемой частью аналитической работы.

Определение основных категорий

Как правило, в ходе исследования объекта формируется несколько категорий – *«Общее»*, *«Частное»*, *«Связи»*. Информацию для этих категорий можно получить как в Информационном портрете (Рисунок 110), так и в ходе проведения сравнительного анализа нескольких объектов.

Работа с несколькими выборками, объектами

Работа с несколькими выборками проводится на странице *«Анализ»* и обозначает проведение сравнительного (раздел *«Таблица»* – Рисунок 114) или динамического (раздел *«Динамика тем»*) анализа.

Сравнительный анализ применяется для выявления общего, частного, а также связей (внутренних, внешних, неявных, несвойственных). При этом необходимо помнить, что найденные слова, как правило, показывают связь между объектами, а словосочетания отражают свойства объекта.

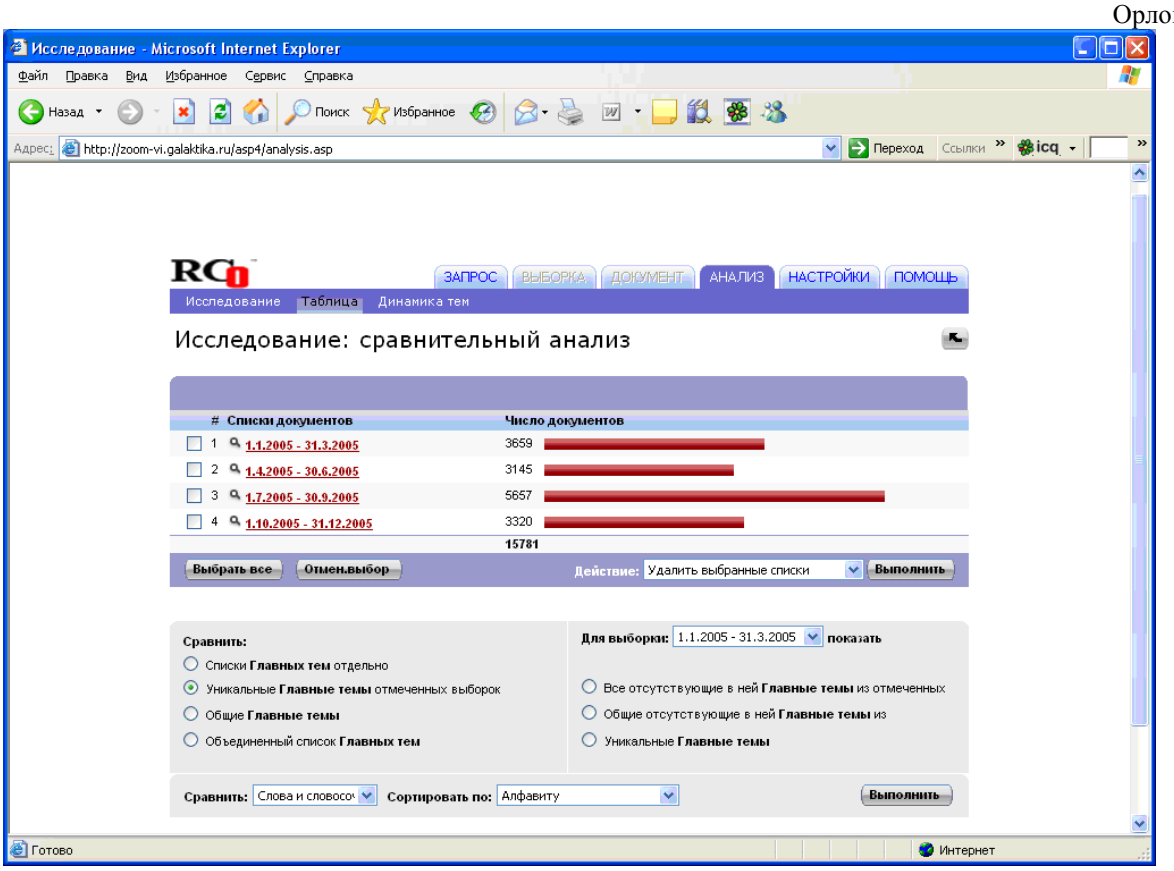


Рисунок 114. *Страница «анализ» – раздел «Таблица»*

Сравнительный анализ

(определение содержания пересечений выборок с документами)

Для нахождения общего между исследуемыми объектами необходимо перейти в раздел *«Таблица»* страницы *«Анализ»*. Затем отметить необходимые выборки, на аналитической панели *«Сравнить»* указать *«Общие главные темы»* и выбрать способ сортировки (например, по *«общей значимости»* или *«алфавиту»*), тип сравниваемых объектов (слов, словосочетаний, слов и словосочетаний). Далее необходимо нажать кнопку *«Выполнить»*. Результат представляется в виде таблицы, которая содержит общие для изучаемых выборок слова или словосочетания. Кроме того, в полученной таблице указывается коэффициент значимости приведенных словоформ для каждой из выборок.

Таким образом, по итогам проведенного исследования Вы получаете общие главные темы, которые можно отнести к категории *«Общее»*.

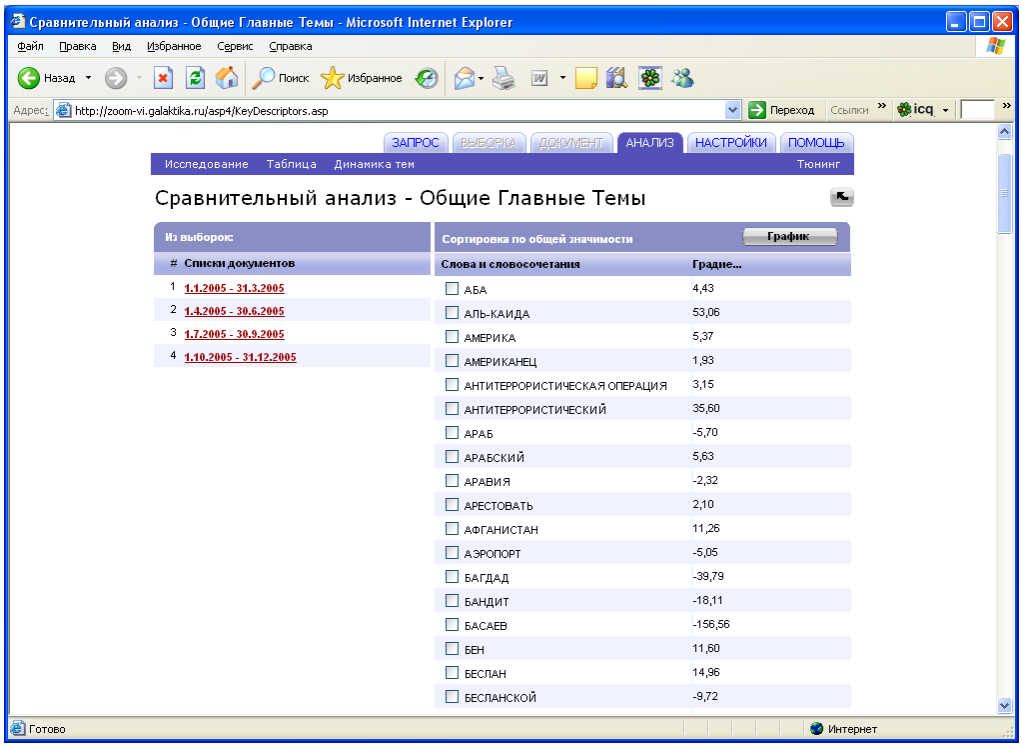


Рисунок 115. *Результаты сравнительного анализа (категория общее)*

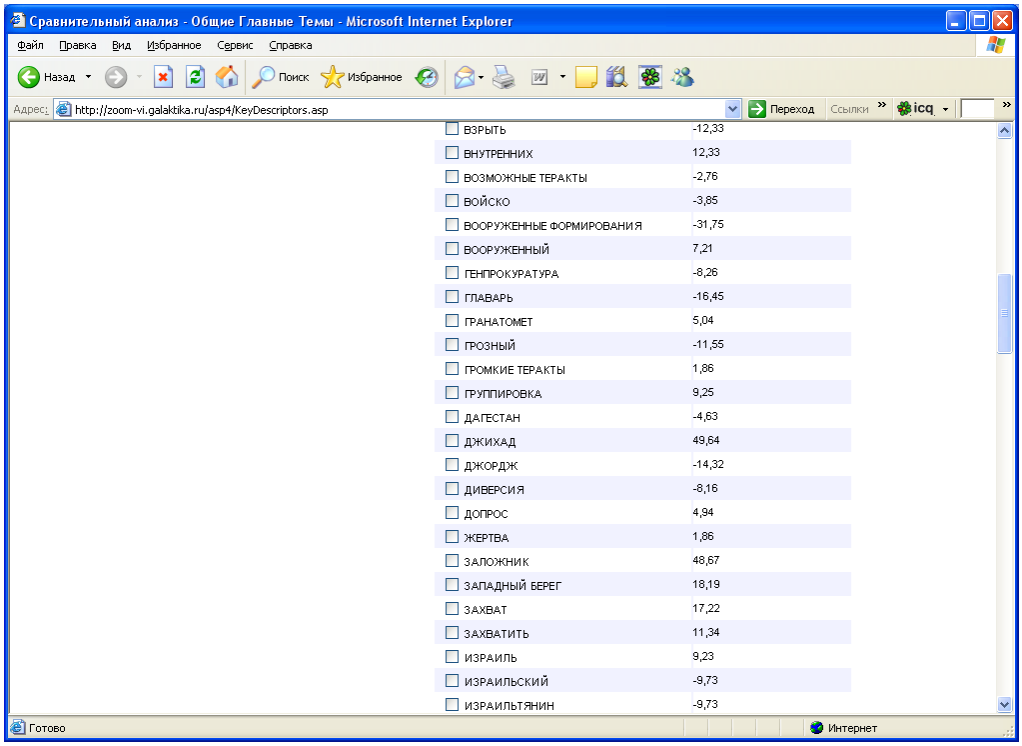


Рисунок 116. *Результаты сравнительного анализа (категория общее)*

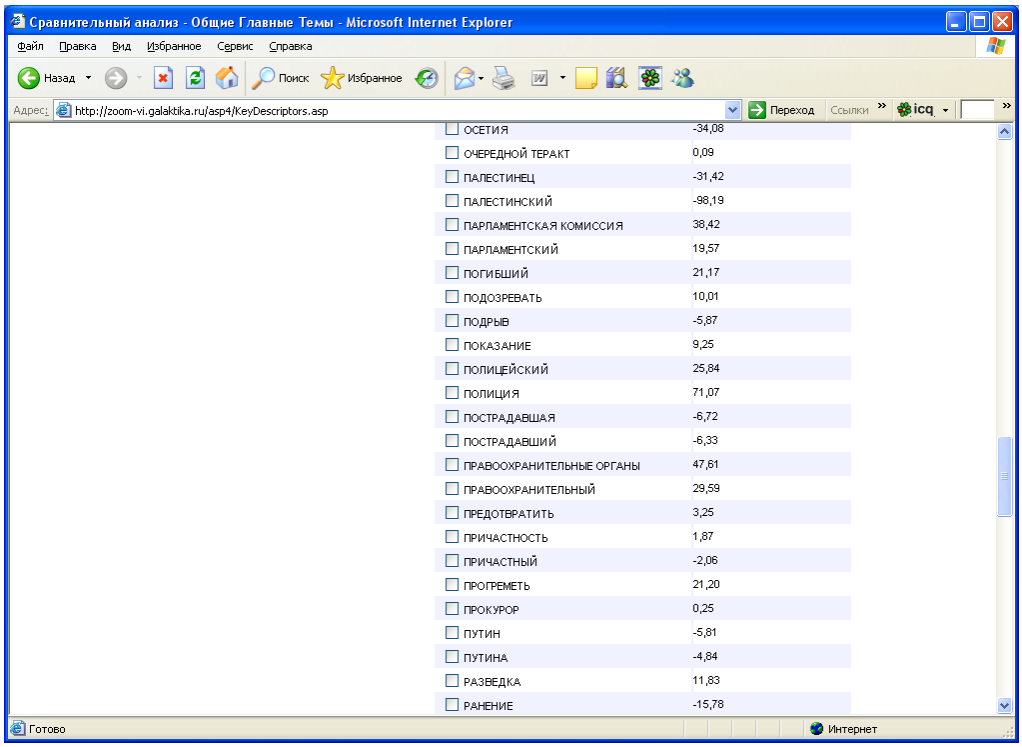


Рисунок 117. *Результаты сравнительного анализа (категория общее)*

Для нахождения частного, а точнее уникальных характеристик объекта, необходимо на странице Исследования отметить выборки по интересующим объектам, отметить *«Уникальные Главные темы»* и выбрать способ сортировки. Затем необходимо нажать кнопку *«Выполнить»*.

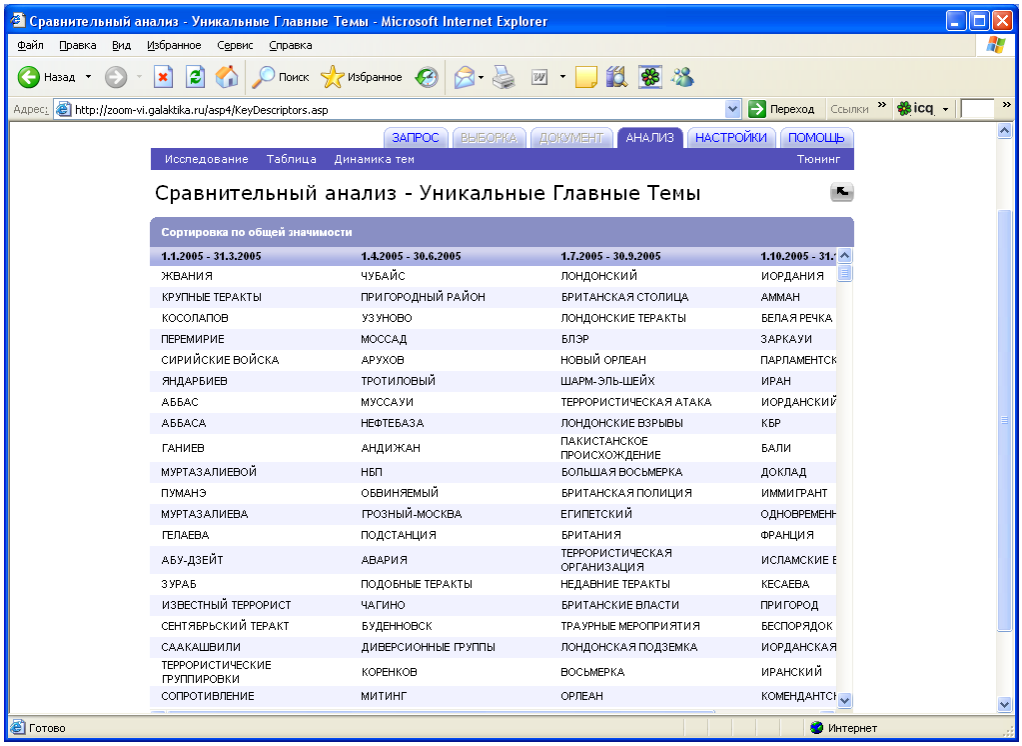


Рисунок 118. *Результаты сравнительного анализа – категория частное*

Результат работы Системы – списки слов и словосочетаний, которые можно отнести к категории *«Частное»*.

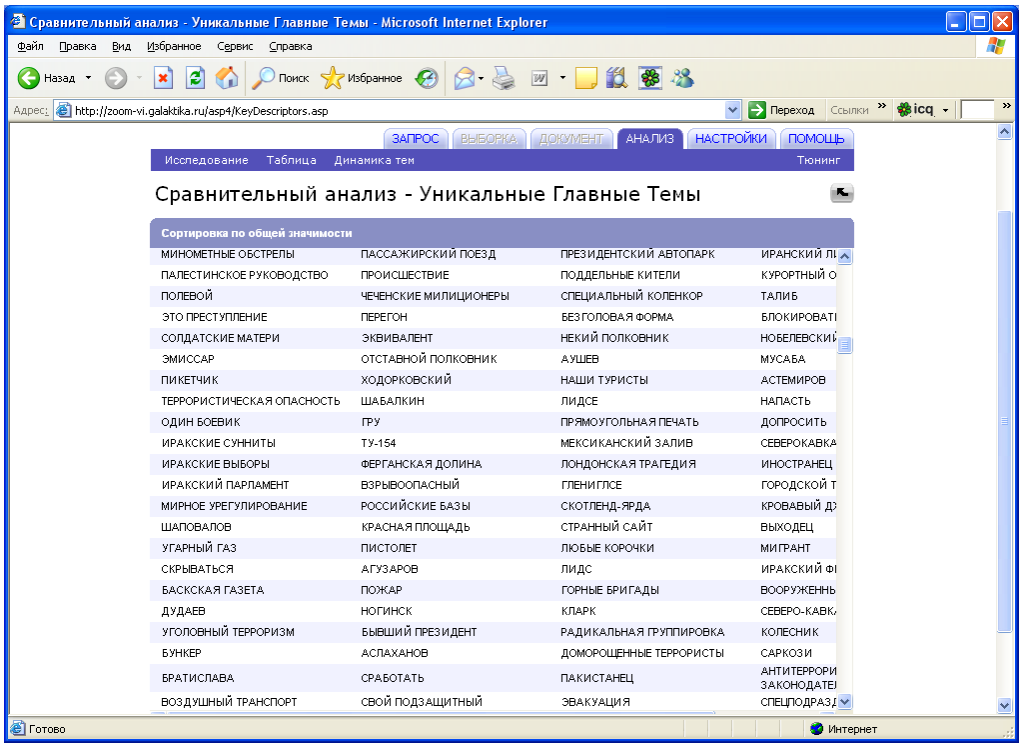


Рисунок 119. *Результаты сравнительного анализа – категория частное*

Для того чтобы определить связи между объектами можно воспользоваться одним из предлагаемых способов:

1. В Информационном портрете объекта применить функцию Кластеризация.
2. На странице *«Анализ»* (раздел *«Таблица»*) провести действия по выявлению общего между объектами (Рисунок 115). Если связи между объектами существуют, то пользователь получает вышеописанную таблицу. Отсутствие таблицы – говорит об отсутствии связей между объектами.

Динамический анализ

(формирование временных срезов, определение параметров для временных срезов, переход в графическое исследование, выявление поведенческих тенденций)

Для проведения динамического анализа необходимо перейти в раздел *«Динамика по времени»* страницы *«Запрос»*. Затем в окно запроса ввести исследуемый объект, указать тип запроса и выбрать единицу деления – год, квартал, месяц, неделя, день и т.д. Число делений не должно превышать 52.

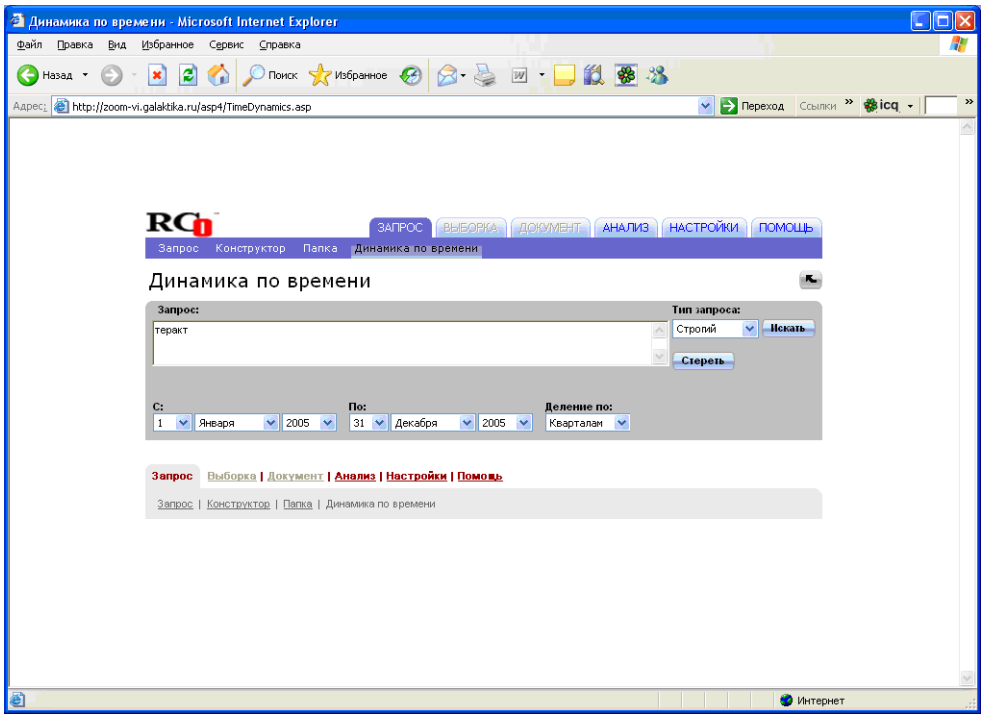


Рисунок 120. *Страница «Запрос» – раздел «Динамика по времени»*

После того как все параметры указаны, необходимо нажать на кнопку *«Искать»*. Результаты представляются в табличной форме, с указанием периода времени, количества документов, которое ему соответствует и визуального представления самой выборки. Таким образом, Система предлагает Вам *публикационный рейтинг* объекта (Рисунок 121).

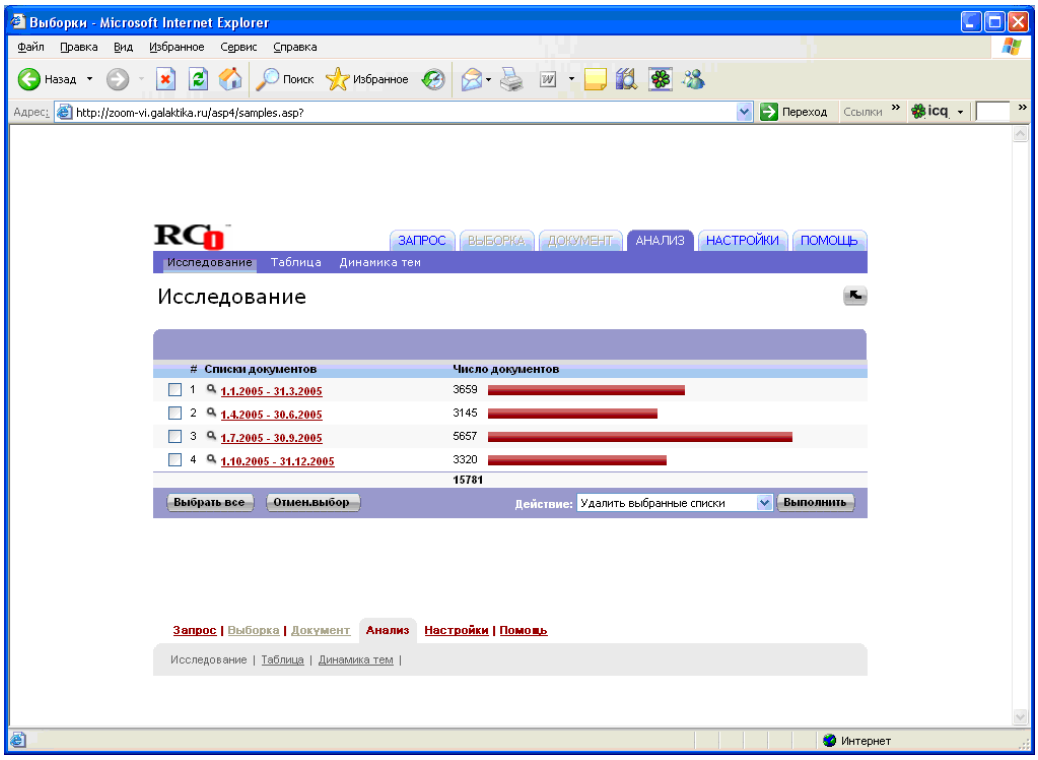


Рисунок 121. *«Окно исследование – построение публикационного рейтинга».*

После построение *публикационного рейтинга*, Вы получаете возможность изучить развитие объекта во времени. Для этого на странице *«Анализ»* в разделе *«Динамика тем»* необходимо добавить интересующие темы (в поле *«Добавить тему»* ввести слово/словосочетание, нажать *«Добавить»*). Затем в поле *«Показать»* установить флажки напротив соответствующих тем, нажать *«Выполнить»*.

Результаты исследования автоматически представляются в графическом режиме, с визуализацией легенды (Рисунок 122).

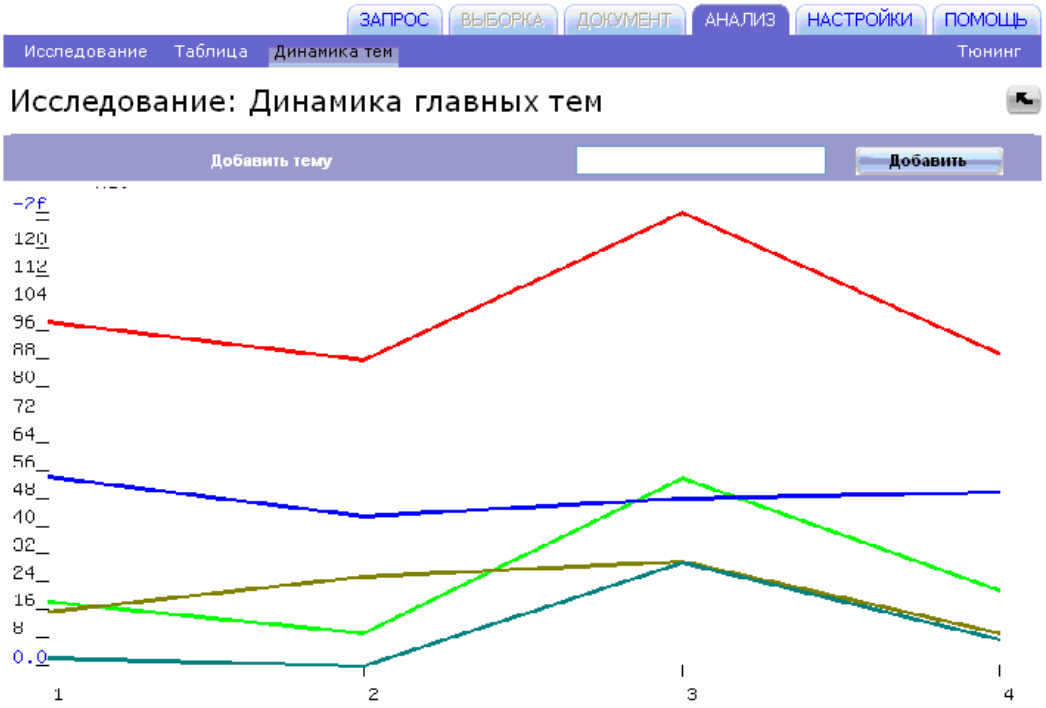


Рисунок 122. *«Результаты трендового исследования»*

По оси X расположены промежутки времени, а по оси Y значимость параметров.

При изучении графиков необходимо отмечать пики и спады параметров, выявлять прослеживается ли тенденция при изменении значимости параметров.

В ходе работы параметры можно как добавлять, так и удалять.

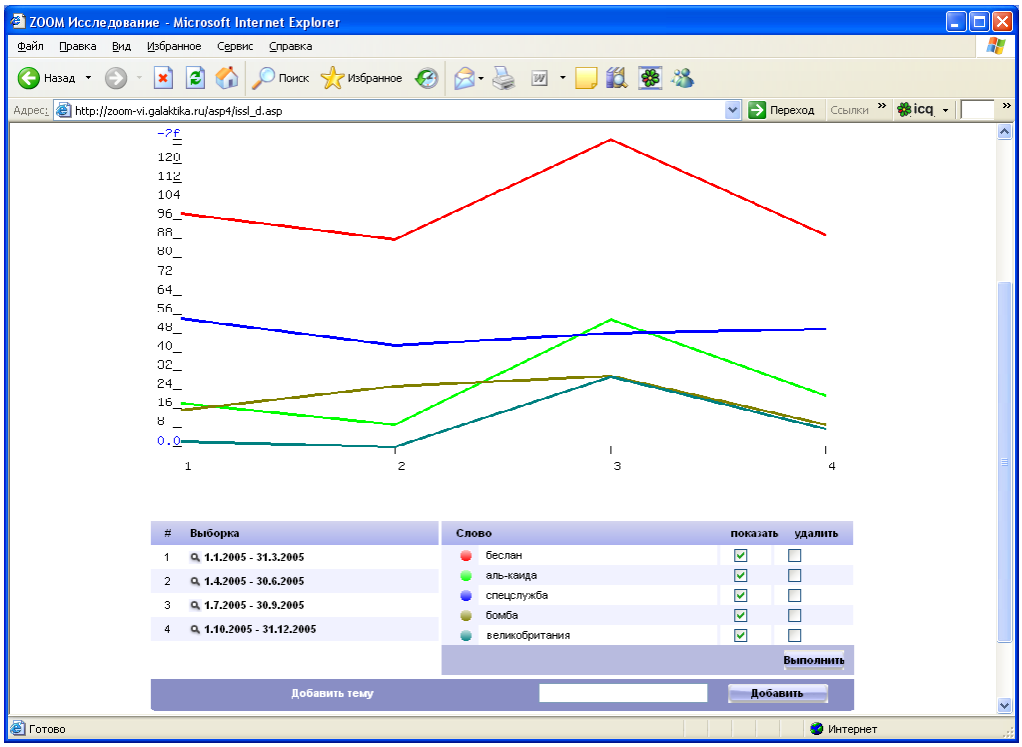


Рисунок 123. *«Результаты трендового исследования»*

Разработка отчетного документа

Отчетный документ составляется с учетом индивидуальных потребностей Руководителя.

Рекомендации по составлению отчетного документа

- *Вступление* (цель, описание темы, объекта и предмета исследования; краткая справка, с указанием периода и количества обработанных документов);
- *Основная часть* (описание Информационного образа объекта; таблицы и графики – результаты сравнительного анализа и трендового исследования; примеры – значимые фрагменты документов из БД);
- *Выводы и предложения* (самостоятельная работа аналитика).

Примечание

Для оптимизации работы с Системой настраиваются индивидуальные параметры. Перед началом работы Вы можете установить:

- морфологический поиск,
- способ сортировки документов
- тип запроса,
- БД,
- отсечение мусора.

Если Вы предполагаете работать с одним объектом, то поле *«Сохранять новый запрос»* (страница *«Настройки» – «Внешний вид»*) необходимо оставить пустым. Если предполагается изучать несколько объектов одновременно (т.е. проводить анализ), то в вышеназванном поле необходимо установить флажок. Затем нажать *«Установить»*.

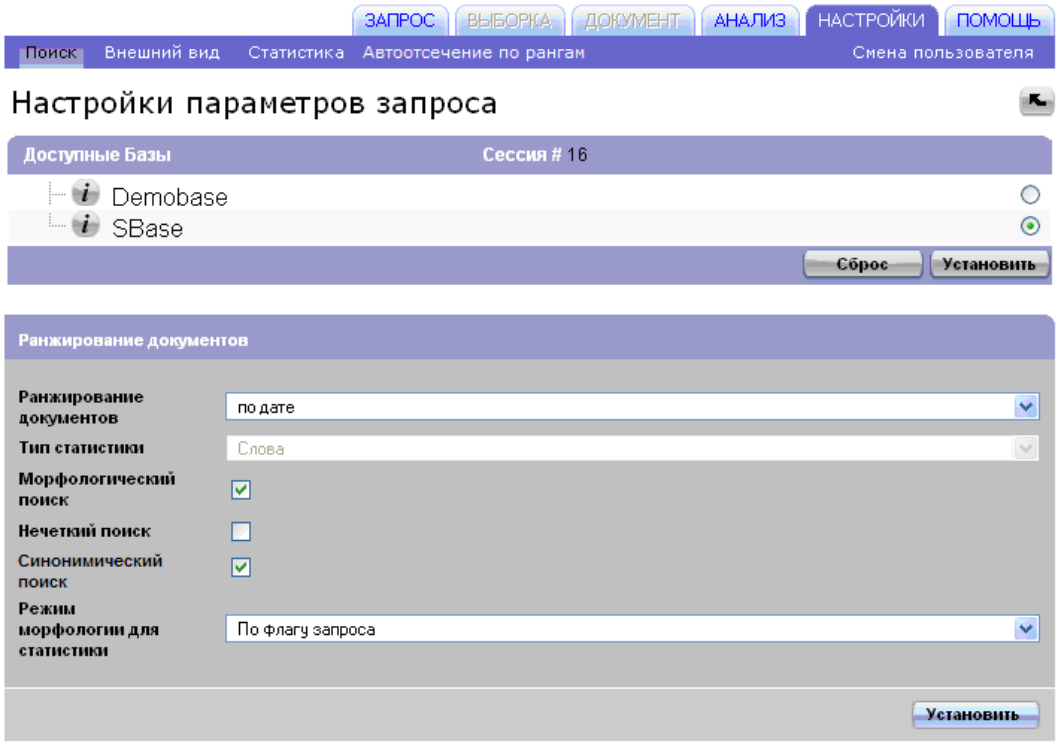


Рисунок 124. *Страница «Настройки»*

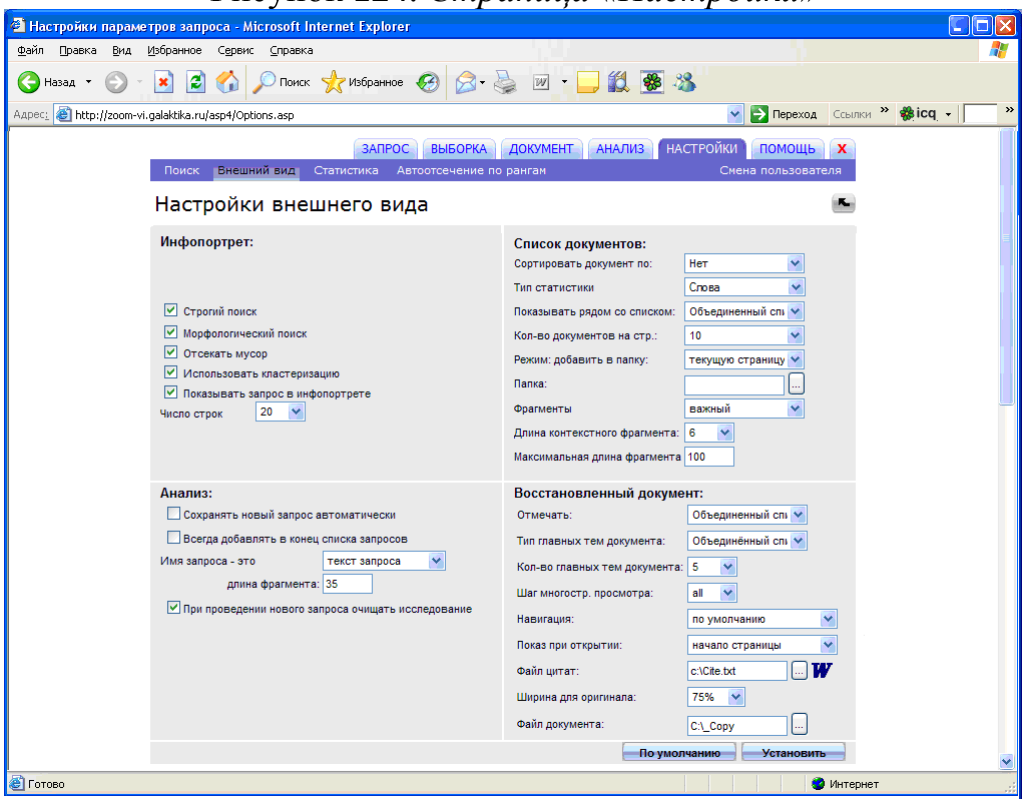


Рисунок 125. *Страница «Настройки»*

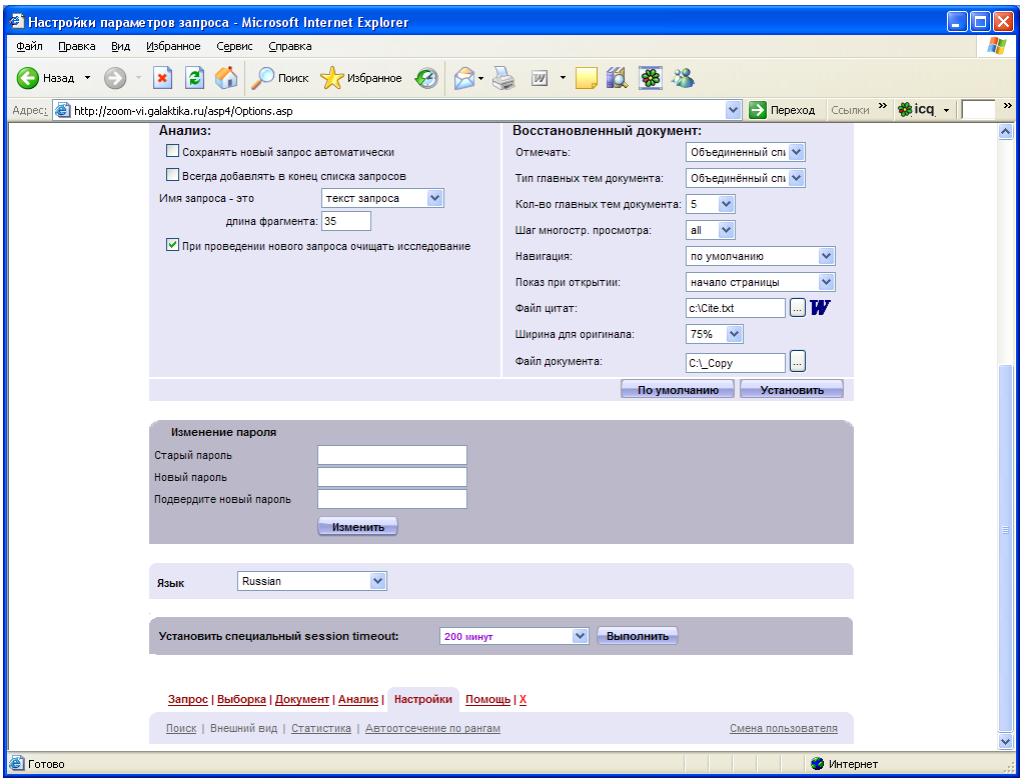


Рисунок 126. *Страница «Настройки»*

Приложение

Таблица операторов языка запроса Системы

Примечание: при проведении запроса регистр (строчные и прописные буквы) значения не имеет

Название оператора	Написание в Системе	Пример использования	Описание результата
Общие операторы			
Контекстный поиск	**кнт(...)** вариант: ctx(...)	кнт(выборы президента)	Будут найдены все имеющиеся в базе документы, где упоминаются "выборы президента"
Дата	**date: дд.мм.гггг** вариант: date:д.м.гг	date:01.07.2004 вариант: date:1.7.04	Будут найдены все имеющиеся в базе документы за указанную дату
Источник	**source: '...'**	source:'известия'	Будут найдены все имеющиеся в базе документы из газеты "Известия"
		source:'москва'	Будут найдены все имеющиеся в базе документы, с источником г. Москва
№ документа в базе	**#...**	#534	Будет найден документ с указанным номером
Поиск по любому текстовому полю	**@(…)**	@(выборы)	Будут найдены все имеющиеся в базе документы, в текстовых полях у которых есть слово "выборы" (в названии, в источнике, в тексте документа и любом другом)
Логические операторы			
и	**и** варианты: and, &	кнт(выборы президента) **и** date:01.07.2004	Все имеющиеся в базе документы, где упоминаются выборы президента за указанную дату
или	**или** вариант: or	кнт(выборы президента) **или** кнт(выборы государственной думы)	Все имеющиеся в базе документы, где упоминаются или выборы президента или выборы государственной думы
не	**и не** варианты: ^,	кнт(выборы президента) и **не** source:'известия'	Все имеющиеся в базе документы, где упоминаются выборы президента, но кроме документов из источника газета "Известия"

	and not		

Операторы сравнения

Меньше	<	#<534	Все имеющиеся в базе документы с номером в базе меньшим 534
Меньше или равно	<=	source:<= 'известия'	Все имеющиеся в базе документы из источников находящихся до газеты "Известия" включительно (перечень источников выстраивается по алфавиту)
Больше	>	date:>01.07.2004	Все имеющиеся в базе документы с датой позднее указанной
Больше или равно	>=	date:>= 01.07.2004	Все имеющиеся в базе документы, начиная с указанной даты и позднее
Диапазон	**…**	date:01.07.2004…15.07.2004	Все имеющиеся в базе документы в указанном диапазоне дат
		#534…700	Все имеющиеся в базе документы с указанным диапазоном номеров

Морфология

При включенной функции Морфология		человек	Будут найдены все имеющиеся в базе документы, в тексте которых есть слово человек, а также где есть упоминания человека, человеку, люди, людей и т.п.
Маска (шаблон)	*****	страх* 'Московский Комсомолец*'	Звездочка (*) заменяет любую последовательность символов в начале или в конце слова (символы могут отсутствовать). Звездочка не может стоять в середине слова. В примере *страх* будут найдены документы: страх, перестрахование, страховка, страховой и т.д. Шаблоны, содержащие пробелы и прочие разделители нужно заключать в апострофы.
Расширенный шаблон	**сл(…)** варианты: word(…), w(…)	сл(банк* и не(банкрот* или банкир*))	Расширенный шаблон используется, когда поиск со звездочкой (*) дает слишком много информационного шума.
Любой символ (**необязательно** должен присутствовать)	**!**	мо!хамед	Знак **!** заменяет любой символ. Запрос мо!хамед будет означать поиск следующих словоконструкций: мо**хх**амед или мо**х**амед
Любой символ (**обязательно** должен присутствовать)	**?**	ма?аме??	Запрос ма?аме?? = махамед**дд**, махаме**тт**, магаме**тт**, магаме**тт**... (махаме**т** или магаме**д** не найдется).
		ал?!джаз!ир!	Знаки ! и ? используются как **в начале** и **в конце** слов, так и **в середине**, при отсутствии четкого варианта написания слова.
Включение морфологии	**%+**	банк%+	Будут найдены документы, где упоминаются все морфологические конструкции слова (банкир, банковский, банкетка…).
Отключение	**%-**	банк%-	Будут найдены документы, где слово банк упоминается только в такой форме.

морфологии			

Позиционные запросы

Модификатор - это два числа, определяющие границы поиска относительно слова, после которого этот модификатор стоит.	**[N,N]**	кнт(в в [-2,1] путин)	кнт (**X** [-2,1] **Y**) Цифра с минусом означает расположение **Y** слева от **X** на количество позиций, равных цифре; цифра без указания знака, иными словами (+), означает расположние **Y** справа от **X** **на** количество позиций равных цифре. В данном случае будут найдены тексты, в которых есть контекст В.В. Путин или Путин В.В.
		кнт (транспортировка [3,5] газ)	Будут найдены документы, в которых упоминаются слово **транспортировка** и через **три** или **пять** позиций слово **газ.**
Вариант модификатора	**рдм (слово [n] слово)** вариант:near (слово [n] слово)	рдм (газпром [8] 1998)	Будут найдены документы, в тексте которых **1998 год** располагается через **восемь** символов либо **до,** либо **после** названия "**Газпром".**

Запросы по части документа

В пределах одного абзаца	**пргф(…)** варианты: prgf(…), P(…)	пргф(путин и (чубайс или березовский))	Будут найдены документы, где Путин упоминаются либо с Чубайсом, либо с Березовским в пределах одного **абзаца.**
В пределах одного предложения	**пред(…)** варианты: sent (...), S(…)	пред(путин и (чубайс или березовский))	Будут найдены документы, где Путин упоминаются либо с Чубайсом, либо с Березовским в пределах одного **предложения.**

Примечание: внутри операторов **пргф и пред** (в скобках) возможно использование только **слов** (словосочетания не принимаются). Для возможности поиска словосочетаний, необходимо соединять слова оператором "и". Например: пред(ельцин и российская и федерация)

Сложные запросы

Любой контекстный запрос может выступать как часть логического выражения. Для группировки логических выражений используются круглые скобки - (…)