



# Опыт реализации системы контентной фильтрации Интернет-ресурсов на базе технологий АРИОН



# Принципы контентной фильтрации

1. Эффективная контентная фильтрация обеспечивается глубоким анализом текстового содержимого Интернет-ресурсов
2. Релевантный поиск и проблемный анализ содержимого Интернет-ресурсов должен выполняться специальным программно-аппаратным комплексом
3. Задача комплекса – формирование и «веерное» распространение базы данных запрещенных ресурсов для их последующего блокирования
4. Ресурсы, относящиеся к запрещенной тематике, должны быть классифицированы в соответствии с их информационной направленностью
5. Для каждой категории пользователи может задаваться свой перечень запрещенных классов ресурсов

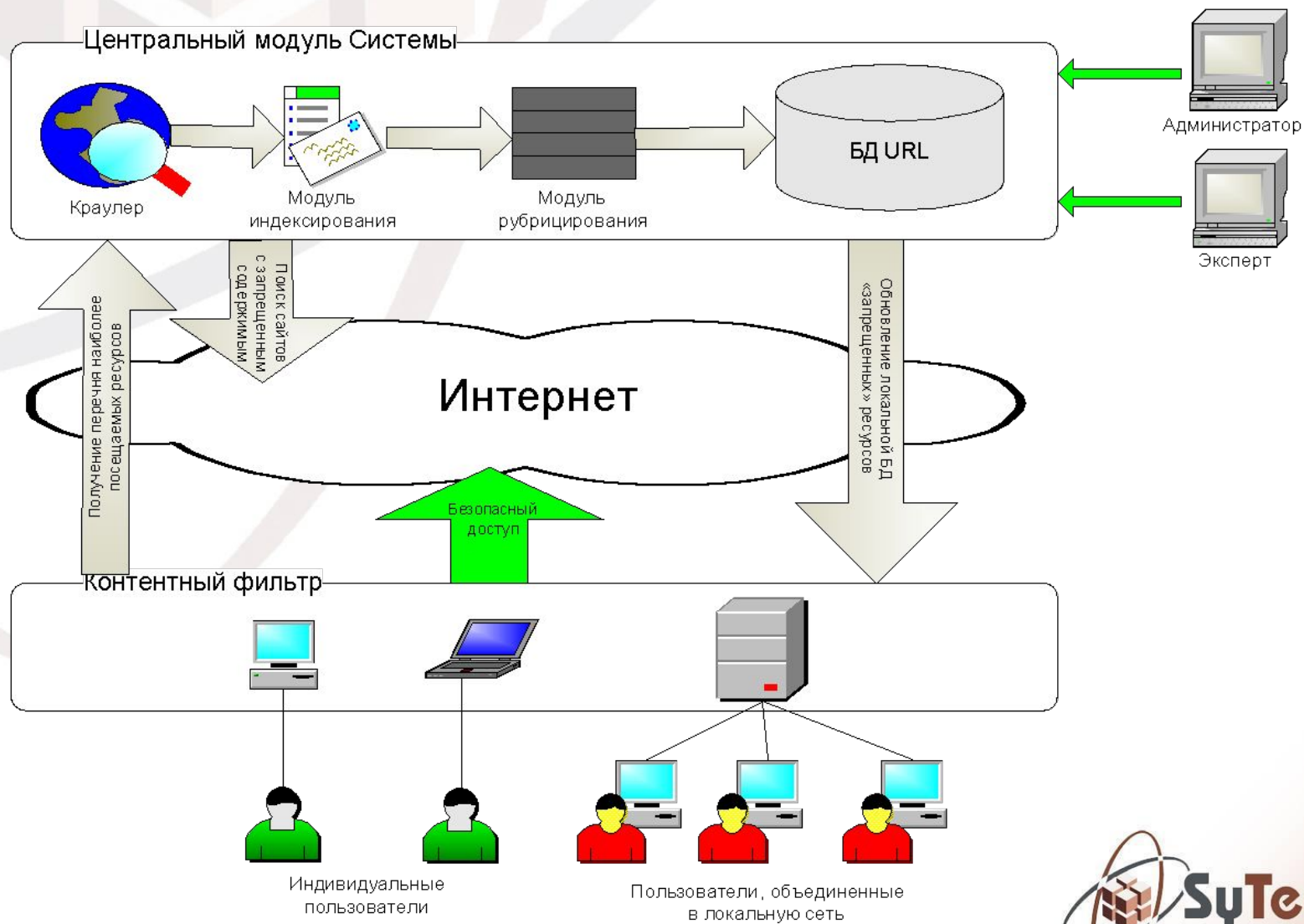
# Платформа – ИАС «АРИОН»



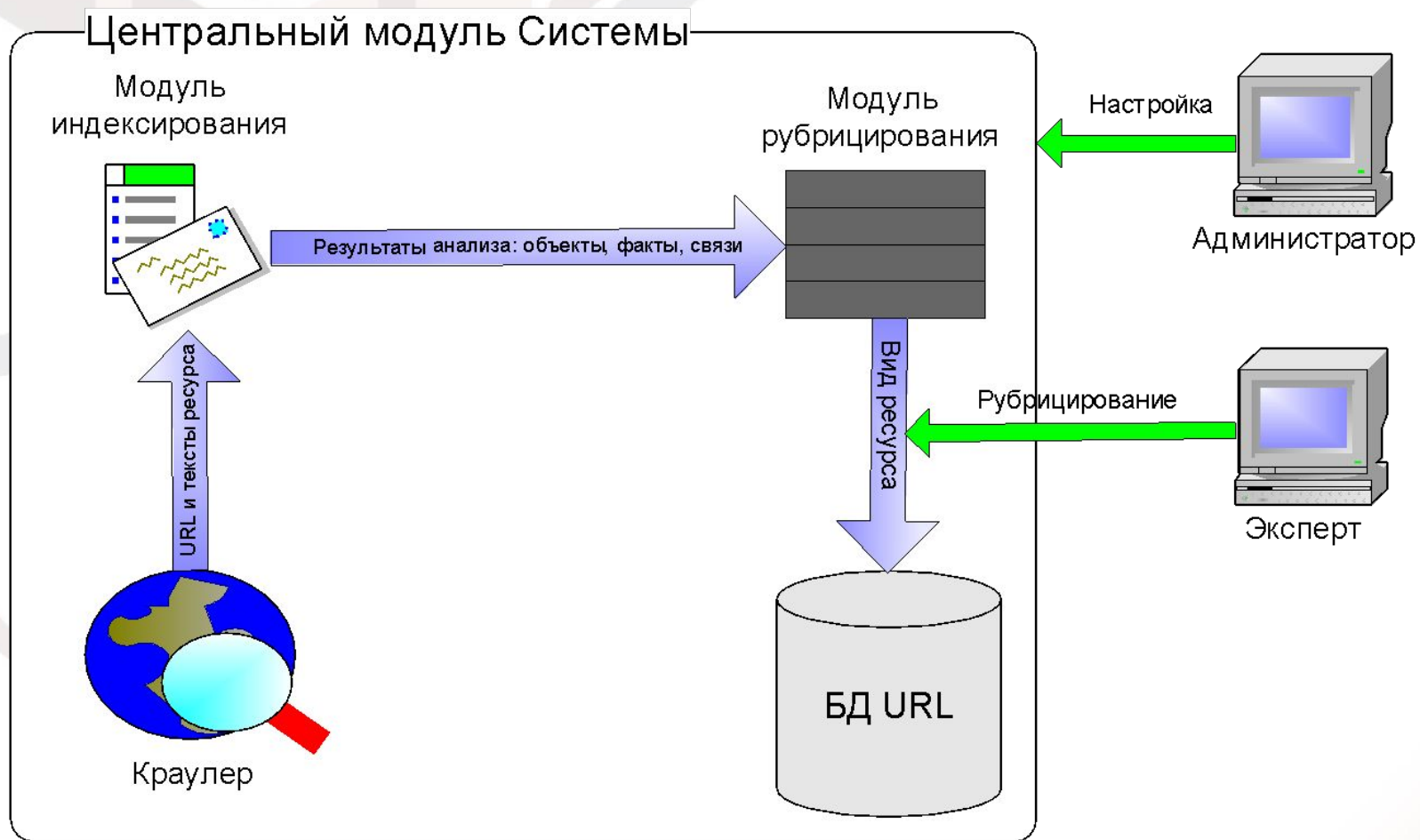
# Ключевые технологии

1. Автономный поиск и автоматическая загрузка текстового содержания ресурсов Интернета
2. Фактографический разбор и тональный анализ текстов на естественном языке
3. Анализ и классификация ресурсов с использованием полнотекстового и фактографического режимов

# Концептуальная схема



# Аналитическая часть



# Сбор информации - Краулер

## Назначение

- поиск в Интернете ресурсов по заданным критериям

## Основные возможности

- автономный переход по ссылкам
- использование поисковых машин
- извлечение текста из различных форматов: txt, html, rtf, doc, pdf
- первичный анализ текстового содержимого (с помощью фильтров)



# Сбор информации

Интернет - ресурсы

The screenshot shows a forum page on the website 'Психология' (Psychology). The forum is titled 'Психология' and has sub-sections for 'Психология', 'Секс', and 'Наркотики'. The main thread is titled 'Про легализацию марихуаны (rss)'. The thread has 158 messages and 130 users. The first post is by user 'Hunter' on 08/07/02, discussing the benefits of legalization. The second post is by user 'Герга' on 08/07/02, discussing the risks. The third post is by user 'дядюра' on 08/07/02, discussing the social aspects. The fourth post is by user 'Правило №9' on 24/07/02, discussing the legal aspects. The fifth post is by user 'Ульяна Громова' on 21/04/03, discussing the economic aspects. The sixth post is by user 'Рыбка с лицом' on 22/04/03, discussing the social aspects.





# Сбор информации

Менеджер загрузки

Менеджер загрузок

Управление задачами

Управление шаблонами

Свойства задачи

Свойства

Свойства

Параметры

Связи

Название	Тип	Значение	Проверена
Название конфигурации	Массив	Тестовый выходной адаптер для F ...	Да
Маска	Маска	Фильтр	Да

Редактор

Выражение

Учитывать

Инвертировано

(борьба OR закон OR меры OR силовые акции OR задержани	Нет	Нет
(возвращ* OR нормальная жизнь OR реабилитац* OR лечен* (	Нет	Нет
((Акция OR мероприятие) AND (против нарко* OR без наркот	Да	Нет
((хими* OR психотропн* OR психоактив*) AND способы AND (и	Да	Нет
(материал OR пособие OR история OR исследование OR терми	Нет	Нет

Загрузить из файла

Сохранить в файл

Добавить

Удалить

Сохранить

Отмена

# Модуль индексирования

## Назначение

- анализ текстового содержания найденных ресурсов

## Основные возможности

### Полнотекстовый отбор

- мощный язык запросов
- учет морфологии
- поддержка синонимов и опечаток

### Фактографический анализ

- графематический анализ
- морфологический анализ
- синтаксический анализ
- логико-семантический анализ

### Тональный анализ

# Индексирование

Список источников,  
загруженных в систему



» Запросы

» Каталоги

» Документы

» Отчеты

» Настройка

»» Протокол загрузки

Страница:  1 2 3 4 5 6 7 8 9 10 > >> 20 ▾

Установить фильтр

	630	14.12.2007-13:29:51	<a href="#">/болезни/Наркомания - серьезная болезнь.htm</a> - Импорт из XML - 14.12.2007-13:29:51		<a href="#">Исходный текст</a>   <a href="#">Опубликовать</a>   <a href="#">Редактировать</a>
	629	14.12.2007-13:29:39	<a href="#">/болезни/Наркомания - болезнь безответственности С Копонев.htm</a> - Импорт из XML - 14.12.2007-13:29:39		<a href="#">Исходный текст</a>   <a href="#">Опубликовать</a>   <a href="#">Редактировать</a>
	628	14.12.2007-13:28:34	<a href="#">/болезни/Наркомания - болезни! Физическая и психологическая зависимость при наркомании.htm</a> - Импорт из XML - 14.12.2007-13:28:34		<a href="#">Исходный текст</a>   <a href="#">Опубликовать</a>   <a href="#">Редактировать</a>
	627	14.12.2007-13:28:17	<a href="#">/болезни/MedicReferat - Наркомания болезнь века.htm</a> - Импорт из XML - 14.12.2007-13:28:17		<a href="#">Исходный текст</a>   <a href="#">Опубликовать</a>   <a href="#">Редактировать</a>
	626	14.12.2007-13:28:03	<a href="#">/болезни/Кафедра психиатрии с наркологией Воронежской государственной медицинской академии им Н Н Бурденко.htm</a> - Импорт из XML - 14.12.2007-13:28:03		<a href="#">Исходный текст</a>   <a href="#">Опубликовать</a>   <a href="#">Редактировать</a>
	625	14.12.2007-13:27:57	<a href="#">/болезни/Кафедра наркологии Харьковской медицинской академии последипломного образования.htm</a> - Импорт из XML - 14.12.2007-13:27:57		<a href="#">Исходный текст</a>   <a href="#">Опубликовать</a>   <a href="#">Редактировать</a>
	624	14.12.2007-13:27:46	<a href="#">/болезни/Контент сайта Танатология Edwin Ehrlich Конспект судмедэксперта Посмертное определение морфина Информационный сайт.htm</a> - Импорт из XML - 14.12.2007-13:27:46		<a href="#">Исходный текст</a>   <a href="#">Опубликовать</a>   <a href="#">Редактировать</a>
	537	13.12.2007-18:07:14	<a href="#">Про Наркотики 2 - Ввод документа - 13.12.2007-18:07:13</a>		<a href="#">Исходный текст</a>
	536	13.12.2007-18:06:24	<a href="#">Про Наркотики - Ввод документа - 13.12.2007-18:06:23</a>		<a href="#">Исходный текст</a>
	534	13.12.2007-18:05:30	<a href="#">Новый растительный наркотик из Африки завоевывает мир - Ввод документа - 13.12.2007-18:05:30</a>		<a href="#">Исходный текст</a>

Обновить

10

W

5

5

品

□

⌚ Задачи (1)

✉ Буфер обмена (0)

✉ Оповещения (0)

remote



# Индексирование

Арион

» Запросы    » Каталоги    » Документы    » Отчеты    » Настройка

» Протокол загрузки

http://ssa.ru:18080/?id=281 - Исходный текст - Windows Internet Explorer

Страница: 630

## » Исходный текст

Новый растительный **наркотик** из Африки завоевывает мир

01.07.08, Mignews. Не успело мировое сообщество более или менее разобраться с определениями основных видов **наркотиков** и способами **борьбы** с их **распространением**, **употреблением** и даже **легализацией**, как появилась еще одна опасная **"травка"**. Причем оттуда, откуда подобных вещей в принципе не ждали - из овощных лавочек и магазинчиков, торгующих зеленью. И цена, в общем, совсем невысока - всего 4 цента за пучок.

Название новой **"травки"** - кхат, а родиной ее является северное побережье Африки. Это растение, листья которого по химическому составу и **воздействию** на организм - напоминают **наркотики** группы **амфетаминов**, чрезвычайно популярно в среде выходцев из Сомали и Йемена, а в последнее время в больших количествах стало появляться и в других социальных группах.

По свидетельствам медиков, **употребление** (пережевывание) этого растения может **серьезно осложнить** вашу жизнь - от появления стойкой **депрессии** до развития **рака** ротовой полости.

Административные органы США и ряда европейских стран в срочном порядке приняли **постановления** о ее **запрете**, несмотря на протесты "национальных меньшинств".

Те, в свою очередь, уверены, что кхат не может являться причиной столь резких мер **противодействия** его появлению на широком **рынке**, поскольку сомалийцы и йеменцы знакомы с ним уже почти 700 лет и почти никогда не сталкивались с описанными медиками последствиями... Более того, если вам приходится работать дольше, чем вы привыкли, или у вас (это - для **женщин**) возникли **проблемы** с семейной жизнью, то кхат, по словам "компетентных" пользователей, попросту **незаменим**: он абсолютно натурален и не вызывает побочных эффектов.

-----

**Автор:** Зашибись 12 Января 2002

**Автор:** stanislav 13 Мая 2002  
Лично я **против наркотиков** в любом виде. Никотин, алкоголь и прочее от чего **человек** зависит. Я сам не раз сталкивался с **проблемной зависимостью**. Надеюсь, я от нее избавился.

**Автор:** Ебанокрып 14 Августа 2002

Обновить

Закреть

Установить фильтр

	Исходный текст   Опубликовать   Редактировать
	Исходный текст   Опубликовать   Редактировать
12.2007-13:28:34	Исходный текст   Опубликовать   Редактировать
	Исходный текст   Опубликовать   Редактировать
ko.htm - Импорт из XML -	Исходный текст   Опубликовать   Редактировать
L - 14.12.2007-13:27:57	Исходный текст   Опубликовать   Редактировать
формационный сайт.htm -	Исходный текст   Опубликовать   Редактировать
	Исходный текст
	Исходный текст
	Исходный текст

10 W 5 5

Задачи (1)    Буфер обмена (0)    Оповещения (0)

remote

Текст источника с выделенными объектами (желтый цвет) и тональными маркерами (серый цвет)

# Индексирование

Карточка  
объекта

Арион Аналитический

Запросы Карты

http://ssa.ru:18080/?id=14874&setId=534 - АМФЕТАМИН СТИМУЛИРУЮЩЕЕ - Карточка объекта - Windows In

Настройка

Маркеры АМФЕТАМИН СТИМУЛИРУЮЩЕЕ

Атрибуты Вложения Логические связи Ассоциативные связи Прототипы Маркеры

Тип	Значение	Вес
НАРКОТИК_ОПИСАНИЕ_ТОН	ПОПУЛЯРНЫЙ	-15.0
НАРКОТИК_ОПИСАНИЕ_ТОН	ВРЕД	-10.0

Установить фильтр

Объекты	Граф	Навигатор
Вид: Таблица	Страница: 1 2 3 4	
<input type="checkbox"/>		14870 ВОЗДЕЙСТВИЕ
<input type="checkbox"/>		14871 НАРКОТИК
<input type="checkbox"/>		14872 БОРЬБА
<input type="checkbox"/>		14873 ТРАВКА
<input type="checkbox"/>		14874 АМФЕТАМИН
<input type="checkbox"/>		14875 РАСПРОСТРАНЕНИЕ
<input type="checkbox"/>		14876 ТРАВКА
<input type="checkbox"/>		14877 УПОТРЕБЛЕНИЕ
<input type="checkbox"/>		14878 НАРКОТИК
<input type="checkbox"/>		14879 ЛЕГАЛИЗАЦИЯ
<input type="checkbox"/>		14880 НАРКОТИК
<input type="checkbox"/>		14881 УПОТРЕБЛЕНИЕ
<input type="checkbox"/>		14882 ЗАБОЛЕВАНИЕ

Закреть

Специальное мероприятие

Наркотическое средство

Употребление

РАК

remote



# Модуль рубрикации

## Назначение

- автоматическое рубрицирование ресурсов на основании выполненного анализа

## Способы рубрицирования

- анализ характерных слов и выражений (полнотекстовый поиск).
- анализ описываемых ситуаций (фактографический поиск)
- анализ отношения автора текста к предмету (анализ тональности высказываний)

# Запросы рубрикации

[» Запросы](#)[» Каталоги](#)[» Документы](#)[» Отчеты](#)[» Настройка](#)[» Каталоги](#)

Мои каталоги

- Запросы
  - Демонстрация
  - Информационные материалы и научные исследования по наркологии
    - Сайты о борьбе с незаконным оборотом наркотиков
    - Сайты, посвященные лечению наркозависимости
    - Сайты, посвященные пропаганде против наркотиков
    - Сайты, содержащие прямую или косвенную пропаганду наркотиков
      - Крайне негативные источники
      - Сайты обществ и движений, выступающих за легализацию наркотиков
      - Сайты, содержащие описание способов изготовления и употребления наркотиков

Общие каталоги

По времени операций

По времени объектов

По типам объектов

По типам фактов

По объектам мониторинга

[/ / Мои каталоги | Запросы | Демонстрация](#)

Удалить

Копировать

В избранное

Каталог в избранное

Переместить

Фильтр

элементов: 6/6

<input type="checkbox"/>	Тип	Название	Дата создания
<input type="checkbox"/>		Рубрикация на основе фактографии - Лечение и реабилитация	13.12.2007 16:20:34
<input type="checkbox"/>		Рубрикация на основе фактографии - Маршруты транспортировки	13.12.2007 15:57:34
<input type="checkbox"/>		Определение направленности ресурсов	13.12.2007 00:15:05
<input type="checkbox"/>		Список запрещенных ресурсов	12.12.2007 22:42:17
<input type="checkbox"/>		Рубрикация на основе фактографии - Акции против наркотиков	12.12.2007 20:27:51
<input type="checkbox"/>		Запрос рубрикации	10.12.2007 17:22:24

# Рубрикация на основе полнотекстовового поиска

The screenshot displays the 'Арион' search interface. At the top, navigation links include 'Запросы', 'Каталоги', 'Документы', 'Отчеты', and 'Настройка'. The main header shows 'Запросы' and a callout box with the text 'Текст запроса рубрикации'. The left sidebar contains sections for 'Области', 'Фильтры', 'Режимы', 'Сервисы', and 'Множества'. The main content area is divided into 'Операции' and 'Параметры' tabs. Under 'Параметры', the 'Название' field contains 'Запрос рубрикации'. The 'полнотекстовый поиск' section shows a search text: '(борьба OR закон OR меры OR силовые акции OR задержание OR пресечение OR транспорт\* OR распространен\* OR продажи OR перевозки) AND нарко\*'. Other parameters include 'Режим поиска: Все слова', 'Искать в: названии документа, теле документа', and 'Искать с учетом: синонимов, опечаток'. The 'Фильтр по типам объектов' section shows 'Типы объектов: Источник'. The 'Перемещение в каталоги' section lists 'Сайты о борьбе с незаконным оборотом наркотиков'. The 'Начальное множество' section includes 'Ограничение на количество объектов: 10000', 'Очистить текущее', 'Итерация: отключена', and 'Выборка: вся база'. The 'Полнотекстовый поиск' section shows a search text: '(возвращ\* OR нормальная жизнь OR реабилитаци\* OR лечен\* OR перестали употреблять OR избавление OR консультация ) AND (наркозависим\* OR токсикомания OR наркоман\*)'. At the bottom, there are buttons for 'Найти', 'Сохранить', 'Редактировать', 'Очистить', and 'Импорт'.

Арион

» Запросы » Каталоги » Документы » Отчеты » Настройка

» Запросы

Текст запроса рубрикации

Инструменты Библиотека История

Операции Параметры

Название:

полнотекстовый поиск

Текст: (борьба OR закон OR меры OR силовые акции OR задержание OR пресечение OR транспорт\* OR распространен\* OR продажи OR перевозки) AND нарко\*

Режим поиска: Все слова

Искать в: названии документа, теле документа

Искать с учетом: синонимов, опечаток

Фильтр по типам объектов

Типы объектов: Источник

Перемещение в каталоги

Каталоги:

Сайты о борьбе с незаконным оборотом наркотиков

Начальное множество

Ограничение на количество объектов: 10000

Очистить текущее

Итерация: отключена

Выборка: вся база

Полнотекстовый поиск

Текст: (возвращ\* OR нормальная жизнь OR реабилитаци\* OR лечен\* OR перестали употреблять OR избавление OR консультация ) AND (наркозависим\* OR токсикомания OR наркоман\*)

Найти Сохранить Редактировать Очистить Импорт



# Рубрикация на основе полнотекстовового поиска

Арион

» Запросы » Каталоги » Документы » Отчеты » Настройка

» Кат

Дерево рубрик

Список источников рубрики «Сайты обществ и движений, выступающих за легализацию наркотиков»

Мои каталоги

- Запросы
- Информационные материалы и научные исследования по наркологии
  - Сайты о борьбе с незаконным оборотом наркотиков
    - Сайты, посвященные лечению наркозависимости
    - Сайты, посвященные пропаганде против наркотиков
    - Сайты, содержащие прямую или косвенную пропаганду наркотиков
      - Крайне негативные источники
      - Сайты обществ и движений, выступающих за легализацию наркотиков
      - Сайты, содержащие описание способов изготовления и употребления наркотиков
      - Умеренно негативные

Общие каталоги

- По времени операций
- По времени объектов
- По типам объектов
- По типам фактов
- По объектам мониторинга

Удалить Копировать В избранное Каталог

Фильтр

элементов: 132/132

<< < Страница 1 из 14 > >>

<input type="checkbox"/>	Тип	Название	Дата создания
<input type="checkbox"/>		/html/Трип репорты - narkotrip_ru.htm - Импорт из XML - 12.12.2007-18:38:02	12.12.2007 18:38:02
<input type="checkbox"/>		/html/Осознанное Изменение Сознания.htm - Импорт из XML - 12.12.2007-18:37:33	12.12.2007 18:37:33
<input type="checkbox"/>		/html/Школа выживания.htm	11.12.2007 19:07:00
<input type="checkbox"/>		/html/Тропой Шамана_ Психопрофилактические экстрим программы_ Расширение сознания.htm	11.12.2007 19:06:26
<input type="checkbox"/>		/html/Рок-группа Глас вопиющего.htm	11.12.2007 19:04:18
<input type="checkbox"/>		/html/Общественное движение Матери против наркотиков.htm	11.12.2007 19:04:13
<input type="checkbox"/>		/html/Новый растительный наркотик из Африки завоевывает мир.htm	11.12.2007 19:02:46
<input type="checkbox"/>		/html/Новочеркасск-город_RU - Новочеркасский городской информационный портал.htm	11.12.2007 19:01:34
<input type="checkbox"/>		/html/НарКом - СМИ и наркомания.htm	11.12.2007 19:00:56
<input type="checkbox"/>		/html/НарКом - Каждый получает свою долю беды_ Здесь нет героев - одни только жертвы.htm	11.12.2007 18:57:57

# Запросы рубрикации

[» Запросы](#)[» Каталоги](#)[» Документы](#)[» Отчеты](#)[» Настройка](#)[» Каталоги](#)

Мои каталоги

- Запросы
  - Демонстрация
  - Информационные материалы и научные исследования по наркологии
    - Сайты о борьбе с незаконным оборотом наркотиков
    - Сайты, посвященные лечению наркозависимости
    - Сайты, посвященные пропаганде против наркотиков
    - Сайты, содержащие прямую или косвенную пропаганду наркотиков
      - Крайне негативные источники
      - Сайты обществ и движений, выступающих за легализацию наркотиков
      - Сайты, содержащие описание способов изготовления и употребления наркотиков

Общие каталоги

По времени операций

По времени объектов

По типам объектов

По типам фактов

По объектам мониторинга

[/ / Мои каталоги | Запросы | Демонстрация](#)

Удалить

Копировать

В избранное

Каталог в избранное

Переместить

Фильтр

элементов: 6/6

<input type="checkbox"/>	Тип	Название	Дата создания
<input type="checkbox"/>		Рубрикация на основе фактографии - Лечение и реабилитация	13.12.2007 16:20:34
<input type="checkbox"/>		Рубрикация на основе фактографии - Маршруты транспортировки	13.12.2007 15:57:34
<input type="checkbox"/>		Определение направленности ресурсов	13.12.2007 00:15:05
<input type="checkbox"/>		Список запрещенных ресурсов	12.12.2007 22:42:17
<input type="checkbox"/>		Рубрикация на основе фактографии - Акции против наркотиков	12.12.2007 20:27:51
<input type="checkbox"/>		Запрос рубрикации	10.12.2007 17:22:24



# Рубрикация на основе фактографического поиска



» Запросы

» Каталоги

» Документы

» Отчеты

» Настройка

» Запросы

Инструменты Библиотека История

Операции

Название: Рубрикация на основе фактографии - Акции против

Поиск по шаблону ситуации

Группы

Мероприятие

Атрибут "Тип" содержит "Акция"

Мероприятие

Атрибут "Наименование" содержит "Акция"

Наркотические вещества

Документ

Связи

Мероприятие, Мероприятие (1) и Документ (3) (тип: Логическая, окраска: Любой, количество связей: 1)

Наркотические вещества (2) и Документ (3) (тип: Логическая, окраска: Любой, количество связей: 1)

Контекстный анализ

Глубина: 1

Объекты

Источник

Перемещение в каталоги

Каталоги: Акции против наркотиков

Контекстный анализ

Глубина: 1

Объекты

Документ

Перемещение в каталоги

Каталоги: Акции против наркотиков (документы)

Найти

Сохранить

Редактировать

Очистить

Импорт

Критерии отбора источников в рубрику «Акции против наркотиков» и текстовых документов в рубрику «Акции против наркотиков» (документы)

# Рубрикация на основе фактографического поиска

The screenshot shows the 'Арион' search interface. The top navigation bar includes links for 'Запросы', 'Каталоги', 'Документы', 'Отчеты', and 'Настройка'. The main area is titled 'Запросы' and has tabs for 'Инструменты', 'Библиотека', 'История', 'Операции', and 'Результаты'. The search results are displayed in a table with columns 'Название' and 'Дата'. The results list two entries: 'Найденные источники - Рубрикация на основе фактографии - Акции против наркотиков - Поиск - 25.01.2008-18:50:13 - 9 объектов' and 'Рубрикация на основе фактографии - Акции против наркотиков - Поиск - 25.01.2008-18:50:13 - 9 объектов'. A callout box highlights the first result with the text 'Результаты запроса - источники и текстовые документы'. The interface also features a left sidebar with a tree view of operations, a search input field, and a bottom toolbar with buttons for 'Найти', 'Сохранить', 'Редактировать', 'Очистить', and 'Импорт'.

Арион

» Запросы » Каталоги » Документы » Отчеты » Настройка

» Запросы

Инструменты Библиотека История

Операции Результаты

Название: Рубрикация на основе фактографии - Акции против

Результаты запроса [открыть в виде  таблицы  графа  отчета  статистики ]

- Найденные источники - Рубрикация на основе фактографии - Акции против наркотиков - Поиск - 25.01.2008-18:50:13 - 9 объектов
- Рубрикация на основе фактографии - Акции против наркотиков - Поиск - 25.01.2008-18:50:13 - 9 объектов

Выбрать все Открыть В буфер

**Результаты запроса - источники и текстовые документы**

Название Дата

Найти Сохранить Редактировать Очистить Импорт

# Рубрикация на основе фактографического поиска

[» Запросы](#)[» Каталоги](#)[» Документы](#)[» Отчеты](#)[» Настройка](#)

» Аналитический документ

Рубрикация на основе фактографии - Акции против н... (Объектов: 9)

Объекты | Граф | Навигатор

Список найденных документов

Вид: Таблица | Страница: 1 | 20

Найти

Установить фильтр

■	Тип	ID	Атрибуты	Цвет	Тональность
<input type="checkbox"/>		6202	/html/Радуга над Невою против наркотиков.htm - Импорт из XML - 12.12.2007-22:24:47	Wed Dec 12 22:24:47 MSK 2007	remote
<input type="checkbox"/>		6336	/html/Афиша Концерты_г Владимир_Владимир Онлайн.htm - Импорт из XML - 12.12.2007-22:25:50	Wed Dec 12 22:25:50 MSK 2007	remote
<input type="checkbox"/>		6965	/html/Новочеркасск-город_RU - Новочеркасский городской информационный портал.htm - Импорт из XML - 12.12.2007-22:36:51	Wed Dec 12 22:36:51 MSK 2007	remote
<input type="checkbox"/>		7178	/html/Общественная организация Родители против наркомании.htm - Импорт из XML - 12.12.2007-22:38:46	Wed Dec 12 22:38:47 MSK 2007	remote
<input type="checkbox"/>		7239	/html/Grand-TV Спортсмены против наркотиков.htm - Импорт из XML - 12.12.2007-22:43:08	Wed Dec 12 22:43:08 MSK 2007	remote
<input type="checkbox"/>		7288	/html/Рок-группа Глас вопиющего.htm - Импорт из XML - 12.12.2007-22:43:35	Wed Dec 12 22:43:35 MSK 2007	remote
<input type="checkbox"/>		7925	/html/Вечерний челябинск - Молодые сапиенсы выбирают___ праздник.htm - Импорт из XML - 12.12.2007-22:51:42	Wed Dec 12 22:51:42 MSK 2007	remote
<input type="checkbox"/>		10272	/html/Воспитанники детских домов - против наркотиков.htm - Импорт из XML - 12.12.2007-23:35:16	Wed Dec 12 23:35:16 MSK 2007	remote
<input type="checkbox"/>		10720	/html/Всероссийская государственная телевизионная и радиовещательная компания дирекция Санкт-Петербург - Масштабная акция Город против наркотиков стартовала утром у ворот Смольного.htm - Импорт из XML - 12.12.2007-23:44:09	Wed Dec 12 23:44:09 MSK 2007	remote

# Рубрикация на основе фактографического поиска

[» Запросы](#)[» Каталоги](#)[» Документы](#)[» Отчеты](#)[» Настройка](#)

» Аналитический документ

Найденные источники - Рубрикация на основе фактог... (Объектов: 9)

Объекты | Граф | **Навигатор**

**Список найденных источников**

Вид: Таблица | Страница: 1 | 20

<input type="checkbox"/>	Тип	ID	Атрибуты	Цвет	Тональность
<input type="checkbox"/>		6167	/html/Радуга над Невой против наркотиков.htm	...	
<input type="checkbox"/>		6319	/html/Афиша Концерты_г Владимир_ Владимир Онлайн.htm	...	
<input type="checkbox"/>		6780	/html/Новочеркасск-город_RU - Новочеркасский городской информационный портал.htm	...	
<input type="checkbox"/>		7042	/html/Общественная организация Родители против наркомании.htm	...	
<input type="checkbox"/>		7185	/html/Grand-TV Спортсмены против наркотиков.htm	...	
<input type="checkbox"/>		7240	/html/Рок-группа Глас вопиющего.htm	...	
<input type="checkbox"/>		7877	/html/Вечерний челябинск - Молодые сапиенсы выбирают___ праздник.htm	...	
<input type="checkbox"/>		10220	/html/Воспитанники детских домов - против наркотиков.htm	...	
<input type="checkbox"/>		10695	/html/Всероссийская государственная телевизионная и радиовещательная компания дирекция Санкт-Петербург - Масштабная акция Город против наркотиков стартовала утром у ворот Смольного.htm	...	



# Рубрикация на основе фактографического поиска



» Запросы

» Каталоги

» Документы

» Отчеты

» Настройка

» Каталоги

Мои каталоги

- Запросы
  - Демонстрация
  - Информационные материалы и научные исследования по наркологии
    - Сайты о борьбе с незаконным оборотом наркотиков
      - Сайты, посвященные лечению наркозависимости
      - Сайты, посвященные пропаганде против наркотиков
        - Акции против наркотиков**
        - Акции против наркотиков (документы)
      - Сайты, содержащие прямую или косвенную пропаганду наркотиков
        - Крайне негативные источники
        - Сайты обществ и движений, выступающих за легализацию наркотиков

Общие каталоги

По времени операций

По времени объектов

По типам объектов

По типам фактов

По объектам мониторинга

/ / Мои каталоги | Сайты, посвященные пропаганде против наркотиков | Акции против наркотиков

Удалить

Копировать

В избранное

Каталог

Список источников, добавленных в рубрику «Акции против наркотиков»

Фильтр

элементов: 9/18

<input type="checkbox"/>	Тип	Название	Дата создания
<input type="checkbox"/>		/html/Всероссийская государственная телевизионная и радиовещательная компания дирекция Санкт-Петербург - Масштабная акция Город против наркотиков стартовала утром у ворот Смольного.htm	12.12.2007 23:43:58
<input type="checkbox"/>		/html/Воспитанники детских домов - против наркотиков.htm	12.12.2007 23:34:40
<input type="checkbox"/>		/html/Вечерний челябинск - Молодые сапиенсы выбирают ___ праздник.htm	12.12.2007 22:51:09
<input type="checkbox"/>		/html/Рок-группа Глас вопиющего.htm	12.12.2007 22:43:11
<input type="checkbox"/>		/html/Grand-TV Спортсмены против наркотиков.htm	12.12.2007 22:42:24
<input type="checkbox"/>		/html/Общественная организация Родители против наркомании.htm	12.12.2007 22:37:18
<input type="checkbox"/>		/html/Новочеркасск-город_RU - Новочеркасский городской информационный портал.htm	12.12.2007 22:35:15
<input type="checkbox"/>		/html/Афиша Концерты_г Владимир_Владимир Онлайн.htm	12.12.2007 22:25:42
<input type="checkbox"/>		/html/Радуга над Невой против наркотиков.htm	12.12.2007 22:24:21



# Запросы рубрикации

Арион

» Запросы    » Каталоги    » Документы    » Отчеты    » Настройка

» Каталоги

Мои каталоги

- Запросы
  - Демонстрация
  - Информационные материалы и научные исследования по наркологии
    - Сайты о борьбе с незаконным оборотом наркотиков
    - Сайты, посвященные лечению наркозависимости
    - Сайты, посвященные пропаганде против наркотиков
    - Сайты, содержащие прямую или косвенную пропаганду наркотиков
      - Крайне негативные источники
      - Сайты обществ и движений, выступающих за легализацию наркотиков
      - Сайты, содержащие описание способов изготовления и употребления наркотиков

Общие каталоги

- По времени операций
- По времени объектов
- По типам объектов
- По типам фактов
- По объектам мониторинга

/ / Мои каталоги | Запросы | Демонстрация

Удалить    Копировать    В избранное    Каталог в избранное    Переместить    Фильтр

элементов: 6/6

<input type="checkbox"/>	Тип	Название	Дата создания
<input type="checkbox"/>		Рубрикация на основе фактографии - Лечение и реабилитация	13.12.2007 16:20:34
<input type="checkbox"/>		Рубрикация на основе фактографии - Маршруты транспортировки	13.12.2007 15:57:34
<input type="checkbox"/>		Определение направленности ресурсов	13.12.2007 00:15:05
<input type="checkbox"/>		Список запрещенных ресурсов	12.12.2007 22:42:17
<input type="checkbox"/>		Рубрикация на основе фактографии - Акции против наркотиков	12.12.2007 20:27:51
<input type="checkbox"/>		Запрос рубрикации	10.12.2007 17:22:24

# Рубрикация на основе тональности высказываний

Арион

» Запросы    » Каталоги    » Документы    » Отчеты    » Настройка

» Запросы

Инструменты    Библиотека    История

Области

- » Начальное множество
- » Область запроса

» Фильтры

Режимы

- » Контекстный анализ
- » Поиск цепочек фактов
- » Поиск по шаблону ситуации
- » Полнотекстовый поиск
- » Поиск по шаблону ситуаций

» Сервисы

» Множества

Остальное

- » Добавление документов
- » Построение отчетов
- » Раскраска объектов
- » Сохранение результатов
- » Установка тональностей объектов
- » Выделение тональностей фактов и документов

Операции

Название:

Текущая папка с шаблонами:    Шаблоны   

Начальное множество

Ограничение:     Итерация:

Тональности

Установка тональностей объектов

Пакеты правил тональности

Фильтр по тональностям объектов

Значение

Режим редактирования запроса

Выбор пакетов правил тональностей

# Рубрикация на основе тональности высказываний

Арион

» Запросы    » Каталоги    » Документы    » Отчеты    » Настройка

» Запросы

Инструменты    Библиотека    История

Операции

Области

- » Начальное множество
- » Область запроса

Фильтры

- » Фильтр по дате
- » Фильтр по дате актуальности
- » Фильтр по каталогам
- » Фильтр по разделам
- » Фильтр по типам объектов
- » Фильтр по тональностям объектов

» Режимы

» Сервисы

» Множества

Остальное

- » Добавление документов
- » Построение отчетов
- » Раскраска объектов
- » Сохранение результатов
- » Установка тональностей объектов
- » Выделение тональностей фактов и документов

Название:

Установка тональностей объектов

Пакеты правил тональности: Тональности - мероприятие, Тональности - наркотическое вещество

Фильтр по тональностям объектов

Значение больше -90

Выделение тональностей фактов и документов

Режим работы: обработка документов

Фильтр по типам объектов

Типы объектов: Документ

Сохранение результатов

Название: Умеренно негативные источники  
Добавлять в результаты

Фильтр по тональностям объектов

Значение меньше -90

Сохранение результатов

Название: Крайне негативные источники  
Добавлять в результаты

Найти    Сохранить    Редактировать    Очистить    Импортировать

10    W    3    4       Задачи (0)    Буфер обмена (0)    Оповещения (0)    remote

Критерии запроса для отбора  
«Умеренно негативных источников»  
и «Крайне негативных источников»



# Рубрикация на основе тональности высказываний

The screenshot shows the 'Арион' software interface. The top navigation bar includes 'Запросы', 'Каталоги', 'Документы', 'Отчеты', and 'Настройка'. The main area is titled 'Запросы' and contains a search bar with the text 'Определение направленности ресурсов'. Below the search bar, there are tabs for 'Инструменты', 'Библиотека', and 'История'. The 'История' tab is active, showing a list of operations with a tree view on the left and a table of results on the right. The tree view shows folders for 'Сегодня', 'Эта неделя', 'Прошлая неделя', 'Этот месяц', 'Прошлый месяц', 'Этот год', and 'Прочее'. The table of results has columns 'Название' and 'Дата'. The search results are displayed as a list of links with checkboxes, and there are buttons for 'Выбрать все', 'Открыть', and 'В буфер'. A yellow callout box highlights the search results.

Название:

Результаты запроса [открыть в виде  таблицы  графа  отчета  статистики]

[Негативные публикации - Определение направленности ресурсов - Поиск - 25.01.2008-19:14:11 - 6 объектов](#)

[Крайне негативные публикации - Определение направленности ресурсов - Поиск - 25.01.2008-19:14:11 - 3 объекта](#)

**Результаты запроса - негативные и крайне негативные источники**

	Название	Дата
	Определение направленности ресурсов	25.01.2008 19:14:06
	Рубрикация на основе фактографии - Акции против наркотиков	25.01.2008 18:50:08
	Рубрикация на основе фактографии - Лечение и реабилитация	25.01.2008 18:57:40
	Рубрикация на основе фактографии - Маршруты транспортировки	25.01.2008 18:47:02

# Рубрикация на основе тональности высказываний

Арион

» Запросы    » Каталоги    » Документы    » Отчеты    » Настройка

» Каталоги

Мои каталоги

- Сайты, посвященные лечению наркозависимости
- Сайты, посвященные пропаганде против наркотиков
- Сайты, содержащие прямую или косвенную пропаганду наркотиков
  - Крайне негативные источники
  - Сайты обществ и движений, выступающих за легализацию наркотиков
  - Сайты, содержащие описание способов изготовления и употребления наркотиков
  - Умеренно негативные источники**
- Избранное
- Буфер обмена
- Шаблоны
- Результаты

Общие каталоги

- По времени операций
- По времени объектов
- По типам объектов
- По типам фактов
- По объектам мониторинга

Удалить    Копировать    В избранное    Каталог

**Список источников, добавленных в рубрику «Умеренно негативные источники»**

Фильтр

элементов: 6/6

<input type="checkbox"/>	Тип	Название	Дата создания
<input type="checkbox"/>		/html/Про Наркотики2.htm - Импорт из XML - 12.12.2007-18:41:34	12.12.2007 18:41:34
<input type="checkbox"/>		/html/Про Наркотики.htm - Импорт из XML - 12.12.2007-18:41:15	12.12.2007 18:41:15
<input type="checkbox"/>		/html/Марихуана-Гашиш - описание, последствия.htm - Импорт из XML - 12.12.2007-18:39:52	12.12.2007 18:39:52
<input type="checkbox"/>		/html/Лечение алкоголизма и наркомании Наркотическая зависимость.htm - Импорт из XML - 12.12.2007-18:39:23	12.12.2007 18:39:23
<input type="checkbox"/>		/html/Plan.htm - Импорт из XML - 12.12.2007-18:38:42	12.12.2007 18:38:42
<input type="checkbox"/>		/html/Новый растительный наркотик из Африки завоевывает мир.htm - Импорт из XML - 12.12.2007-18:37:02	12.12.2007 18:37:02



Шаблон

Статистика мониторинга 1

Статистика Гистограмма

Обновить Назад Печать Excel

Шаблон «Статистика мониторинга»

### Статистика мониторинга тональной окраски ресурсов

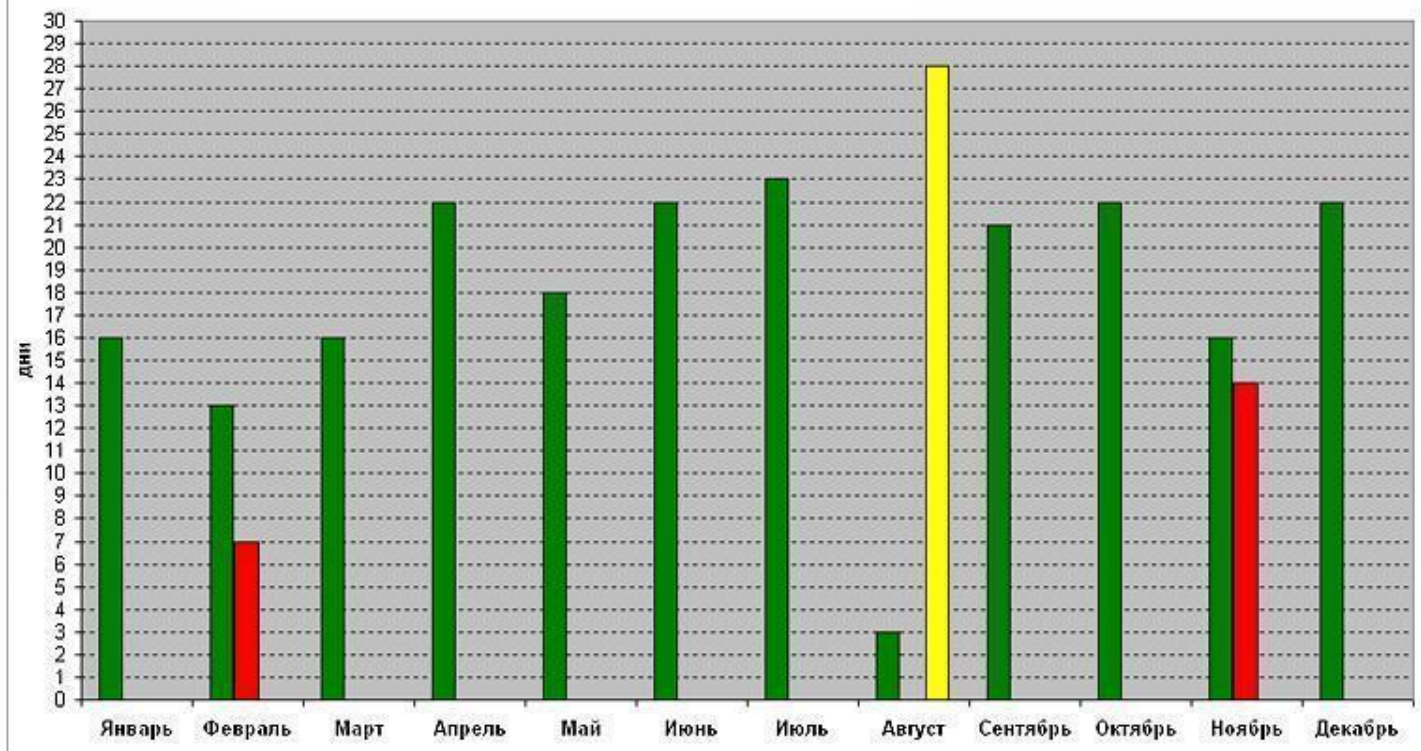
Категория  
Все

Категории выборки

Период

В интервале  
 Без ограничений  
 За год  
 За месяц  
 За день  
 В интервале:  
С    
По

Период выборки



■ Позитивные ■ Негативные ■ Нейтральные

# Список запрещенных ресурсов

- Формирование и выгрузка списка запрещенных ресурсов

# Формирование списка запрещенных ресурсов



» Запросы

» Каталоги

» Документы

» Отчеты

» Настройка

» Запросы

Критерии запроса для формирования списка запрещенных ресурсов

Инструменты Библиотека История

Операции

Области

- » Начальное множество
- » Область запроса

» Фильтры

Режимы

- » Контекстный анализ
- » Поиск цепочек фактов
- » Поиск по шаблону ситуации
- » Полнотекстовый поиск
- » Поиск по шаблонам ситуаций

» Сервисы

» Множества

Остальное

- » Добавление документов
- » Построение отчетов
- » Раскраска объектов
- » Сохранение результатов
- » Установка тональностей объектов
- » Выделение тональностей фактов и документов

Название: Список запрещенных ресурсов

Мероприятие

Атрибут "Любой атрибут" содержит "наркобизнес"

Наркотические вещества

Атрибут "Наименование" содержит "ГЕРОИН"

Контекстный анализ

Ограничение на количество объектов: 1000

Глубина: 2

Использовать при поиске: только логические связи

Объекты

Источник

Начальное множество

Ограничение на количество объектов: 1000

Итерация: отключена

Выборка:

Сайты обществ и движений, выступающих за легализацию наркотиков

Фильтр по типам объектов

Типы объектов: Источник

Построение отчетов

Отчет: sources 2.doc

Найти

Сохранить

Редактировать

Очистить

Импорт

# Формирование списка запрещенных ресурсов

The screenshot shows the 'Арион' web interface. The top navigation bar includes links for 'Запросы', 'Каталоги', 'Документы', 'Отчеты', and 'Настройка'. The main header displays 'Запросы' and a sub-header with 'Инструменты', 'Библиотека', and 'История'. The 'Операции' tab is active, showing a form with the name 'Список запрещенных ресурсов' and a 'Сформировать' button. Below this, the 'Отчеты:' section contains a link to 'Список запрещенных ресурсов'. On the left, a sidebar titled 'По времени операций' lists folders for 'Сегодня', 'Эта неделя', 'Прошлая неделя', 'Этот месяц', 'Прошлый месяц', 'Этот год', and 'Прочее'. At the bottom, a table with columns 'Название' and 'Дата' is visible. A footer bar contains buttons for 'Найти', 'Сохранить', 'Просмотреть', 'Очистить', and 'Импорт'.

Арион

» Запросы » Каталоги » Документы » Отчеты » Настройка

» Запросы

Инструменты Библиотека История

Операции Результаты

Название:

Отчеты:

[Список запрещенных ресурсов](#)

По времени операций

- Сегодня
- Эта неделя
- Прошлая неделя
- Этот месяц
- Прошлый месяц
- Этот год
- Прочее

Название	Дата
----------	------



# Формирование списка запрещенных ресурсов

Программой комплекта АРИОН Сформировано: 25.01.2008 19:18

**Список запрещенных ресурсов - Поиск - 25.01.2008-19:16:53**

Дата формирования отчета: 25.01.2008 19:18

**Список источников**

№	URL
1	/html/MEDMA.htm
2	/html/Хослов И. И. Производство опийных наркотиков (героина) в Афганистане. Международный наркобизнес - финансовая инфраструктура и транспортные потоки наркобизнеса.htm
3	/html/Сказ и наркотикам НЕТ - Наркотики без призра: .htm
4	/html/Million_iv.htm
5	/html/Трип репортаж - narkotip_ru.htm
6	/html/Европейская государственная телевизионная и радиовещательная компания дирекция Санкт-Петербург - Массовый акция Герои против наркотиков стартовала утром у ворот Смольного.htm
7	/html/Трип репортаж - narkotip_ru.htm
8	/html/Как наркоманы употребляют героин.htm
9	/html/Наркоконтроль в г. Сочи - Есть такая Служба в г. Сочи.htm
10	/html/Осознаюсе Изменение Сознания.htm
11	/html/Школа выживания.htm
12	/html/Про Наркотики.htm
13	/html/Драгс__ - Сайт Якутской Молодежи.htm

Печатная версия списка запрещенных ресурсов

Сохранить    Просмотреть    Очистить    Импортировать

# Заключение

Технологии ИАС «АРИОН» в рамках системы контентной фильтрации Интернет-ресурсов обеспечили решение следующих задач:

- Сбор и загрузка текстового содержания искомых ресурсов Интернета
- Анализ и классификация ресурсов с использованием полнотекстового и фактографического режимов
- Тональный анализ текстов на естественном языке
- Формирование списка запрещенных ресурсов

# Контактная информация

- **Адрес:** 115230, Россия, Москва, Электролитный проезд, дом 5Б, строение 8
- **Телефон:** +7 (495) 589-3929
- **Факс:** +7 (499) 317-4663
- **Электронная почта:** [info@sytech.ru](mailto:info@sytech.ru)
- **Web:** [www.sytech.ru](http://www.sytech.ru)