



# *InfoStream*

## **Мониторинг новостей из Интернет:**

Научно-методическое пособие

## **ТЕХНОЛОГИЯ, СИСТЕМА, СЕРВИС**

Киев - 2007



УДК 004.738.5  
ББК 32.973  
М77

*Рекомендовано к печати*  
*Ученым советом Государственного научно-исследовательского института*  
*автоматизированных систем в строительстве*  
*(протокол № 2 от 27 марта 2007 г.)*

**Рецензенты:**

А.А. Снарский - доктор физико-математических наук, профессор

А.Б. Качинский - доктор технических наук, старший научный сотрудник

**М77 Григорьев А.Н., Ландэ Д.В., Бороденков С.А., Мазуркевич Р.В., Пацьора В.Н.**  
InfoStream. Мониторинг новостей из Интернет: технология, система, сервис:  
научно-методическое пособие. - К., ООО "Старт-98", 2007. - 40 с.

В работе освещены теоретические и практические вопросы информационного, научно-методологического и программно-технологического обеспечения процессов поддержки принятия решений на основе мониторинга ресурсов сети Интернет.

Назначение данного пособия - ознакомить широкий круг пользователей с основными подходами к информационному поиску, научить их принципам и правилам составления запросов, использованию широких возможностей поиска информации и ее дальнейшей аналитической обработке с помощью системы InfoStream®.

При написании этого научно-методического пособия использовалась эксплуатационная документация системы InfoStream, разработчиком которой является ООО "Информационный центр "ЭЛВИСТИ".

ISBN 978-966-96800-0-6

Тираж 500 экз.

© Григорьев А.Н., Ландэ Д.В.,  
Бороденков С.А., Мазуркевич Р.В.,  
Пацьора В.Н., 2007

## Аннотация

Система InfoStream\*\* предназначена для нахождения в сети Интернет новостной информации по интересующим пользователя тематикам, оперативной доставки результатов поиска, предоставления единого интерфейса доступа к информации с тысяч web-сайтов, минимизации усилий на отсеивание дублирующейся информации, шума.

Пособие состоит из четырех основных разделов и трех приложений. Первый раздел включает общее описание технологии InfoStream, аспектов решаемых задач.

Второй раздел посвящен описанию возможностей системы InfoStream, среди которых: поисковые возможности, средства уточнения запросов, автоматического формирования сюжетных цепочек, дайджестов, построения гистограмм динамики понятий, таблиц взаимосвязей и т.д.

В третьем разделе описаны сервисы и решения, с помощью которых реализуются возможности системы для пользователей.

В четвертом разделе описана работа пользователя с системой на примере одного из ее режимов (онлайн-доступа к массивам оперативной и ретроспективной информации).

В приложениях к данному пособию содержатся дополнительные сведения: правила использования языка запросов ИПС InfoReS\*\* при работе с системой InfoStream, описание доступных пользователям возможностей технологии RSS, а также описание основных пакетов услуг.

Увеличение интенсивности и объемов информационных потоков (в настоящее время система InfoStream охватывает свыше 40 000 документов в сутки с более чем 2700 web-сайтов) обуславливает необходимость не только находить интересующие пользователя документы, но и проводить эффективный анализ и обработку результатов поиска. Для решения этих задач, наряду с развитием информационной базы и поисковых возможностей, постоянно расширяются возможности InfoStream для предоставления онлайн-доступа к базам данных, включая средства персонализации и содержательного анализа результатов поиска.

---

\*) InfoStream, InfoReS - зарегистрированные товарные знаки. Все права на использование данных товарных знаков принадлежат ООО "Информационный центр "ЭЛВИСТИ". Свидетельства на знак для товаров и услуг №37381, №37378 от 16.02.2004, выданные Государственным департаментом интеллектуальной собственности Украины.

**Оглавление**

<b>1. Технология.....</b>	<b>5</b>
1.1. Спектр задач.....	5
1.2. Технологические аспекты.....	5
<b>2. Система.....</b>	<b>7</b>
2.1. Поиск и отображение информации.....	7
2.2. Информационные портреты.....	7
2.3. Классификатор-навигатор.....	8
2.4. Дайджест.....	9
2.5. Основные сюжеты.....	9
2.6. Новые события.....	10
2.7. Динамика понятий.....	11
2.8. Взаимосвязь рубрик.....	11
<b>3. Сервис и решения на основе InfoStream.....</b>	<b>12</b>
3.1. Онлайн доступ к оперативной и ретроспективной информации.....	12
3.2. Рассылка новостной информации по e-mail.....	13
3.3. Лента новостей на web-сайт.....	14
3.4. Использование технологии RSS.....	14
3.5. Почтовый ящик для персональных рассылок.....	14
3.6. Сервер InfoStream Port.....	15
<b>4. Работа пользователя InfoStream Online.....</b>	<b>17</b>
4.1. Основной экран.....	17
4.2. "Кабинет пользователя".....	22
4.3. Каталог источников.....	24
4.4. Статистика поступления информации.....	25
4.5. Результаты поиска.....	25
4.6. Просмотр сообщения/документа.....	27
4.7. Дайджест электронной прессы.....	27
4.8. Обзор основных сюжетов.....	28
4.9. Новые события.....	28
4.10. Динамика понятий.....	28
4.11. Взаимосвязь рубрик.....	29
<b>Заключение.....</b>	<b>30</b>
<b>ПРИЛОЖЕНИЕ А</b>	
Использование языка запросов ИПС InfoReS.....	31
<b>ПРИЛОЖЕНИЕ Б</b>	
Технология RSS и система InfoStream.....	36
<b>ПРИЛОЖЕНИЕ В</b>	
Основные пакеты услуг InfoStream.....	38

## Введение

Сегодня ни у кого не вызывает сомнения необходимость использования в информационно-аналитической работе такого источника как Интернет. Информационное наполнение Сети интенсивно развивается во всем мире и в Украине, что обуславливает появление ряда проблем, в частности, проблему нахождения актуальной и полной информации по необходимой пользователю тематике. Современные средства поиска новостной информации в Интернет, находящиеся в бесплатном доступе, не настолько совершенны, чтобы в полной мере обеспечить решение этой задачи. Поэтому именно специализированные система мониторинга новостей могут предоставить необходимый пользователям сервис.

В Информационном центре "ЭЛВИСТИ" создана система InfoStream, с помощью которой охватываются новости из тысяч отечественных и зарубежных web-сайтов, осуществляется их обработка и обобщение. InfoStream обеспечивает:

- доступ к оперативной информации с единого интерфейса (по мере ее появления в Сети) в поисковом режиме с учетом возможного дублирования и семантической близости документов, языковых версий, размеров документов, их цифровой насыщенности и т. д.;
- доступ к уникальному ретроспективному фонду, превышающему 50 млн. записей;
- поддержку аналитической работы в режиме реального времени: построение сюжетных цепочек, дайджестов, диаграмм встречаемости и таблиц взаимосвязей понятий, медиа-рейтингов.

Главные преимущества InfoStream по сравнению с традиционными сетевыми информационно-поисковыми системами следующие:

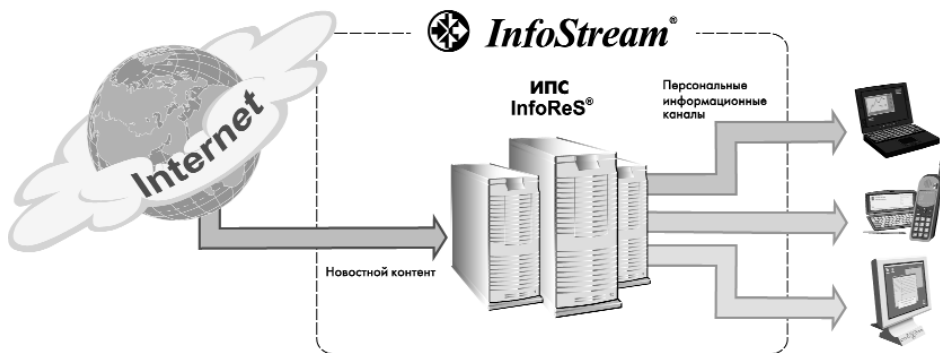
1. Оперативность - базы данных системы пополняются каждые 15 минут, источники сканируются в Сети по мере их обновления, в то время как период индексации традиционных информационно-поисковых систем может измеряться днями или даже неделями.
2. Доступность ретроспективного фонда - даже если информация удалена с web-сайта источника, она сохранена в информационном хранилище.
3. Наличие аналитического инструментария - пользователь может в режиме реального времени не только получать результаты поиска, но и формировать дайджесты, строить сюжетные цепочки, анализировать взаимосвязь рубрик, динамику понятий и т.д.
4. Возможность селекции дубликатов сообщений - система осуществляет автоматическую маркировку идентичной по содержанию новостной информации.
5. Наличие инструментария многоуровневого уточнения запроса.

По сравнению с обычными новостными web-сайтами InfoStream обеспечивает такие преимущества:

1. Охват источников - пользователь имеет доступ к новостям по интересующей его тематике одновременно с большого числа web-сайтов, включая и те избранные, которые он привык просматривать ежедневно.
2. Принцип "одного окна" - пользователь системы мониторинга имеет доступ к информации с многих web-сайтов из одного интуитивного интерфейса.
3. Поисковые возможности - новостные web-сайты, в отличие от системы InfoStream, не всегда обладают развитыми поисковыми возможностями.
4. Доступность ретроспективного фонда.

В последнее время в Интернет появились в свободном доступе и, так называемые, службы синдикации новостей. По сравнению с ними InfoStream обладает такими преимуществами:

1. Непрерывно развивающийся список источников - ежегодно количество сканируемых источников увеличивается на 300-500 сайтов.
2. Учет местных ресурсов - даже крупнейшие платные зарубежные интеграторы новостей лишь частично охватывают, например, украинские web-ресурсы. В систему InfoStream возможно включение новых источников по заявкам пользователей сервиса.
3. Расширенный доступ к информации - пользователям системы InfoStream доступны не только заголовки или анонсы, но и полные тексты новостных сообщений, ссылки на подобные документы и др.
4. Наличие аналитического инструментария.
5. Развитая служба поддержки абонентов - пользователи системы при необходимости могут обращаться с вопросами в службу поддержки и получать исчерпывающие ответы и консультации.



## 1. Технология

### 1.1. Спектр задач

InfoStream предоставляет пользователям широкий спектр сервисных возможностей по обеспечению доступа к новостным ресурсам Интернет. Ее применение позволяет:

- оперативно получать необходимую информацию по мере ее появления в Интернет, анализировать события, своевременно на них реагировать;
- формировать персональные информационные каналы, определяемые запросами на информационно-поисковом языке, формировать архивы для последующей обработки и ретроспективного анализа;
- анализировать поток поступающей информации в режиме реального времени;
- своевременно выявлять тенденции развития и состояния рынков товаров или услуг;
- отслеживать в Интернет информацию о деятельности конкурентов и партнеров, их медиа-присутствие;
- оценивать возможные сферы влияния конфликтных или кризисных ситуаций, осуществлять информационный контроль вероятных источников рисков;
- находить потенциальных клиентов и партнеров.

### 1.2. Технологические аспекты

InfoStream обеспечивает интеграцию сетевых информационных ресурсов на базе эффективных средств сбора, обработки, хранения данных и организации эффективного доступа к ним. С помощью InfoStream выполняется автоматизированный сбор информации с web-сайтов в режиме реального времени, ее структурирование, группировка по семантическим признакам, а также эффективное тематическое избирательное распределение и предоставление доступа к информационным базам данных в поисковых режимах.

Программно-технологическим ядром InfoStream является полнотекстовая информационно-поисковая система InfoReS. Система InfoStream обеспечивает обработку информации в трех основных режимах, и,

соответственно, состоит из трех комплексов:

- сбора и избирательного распространения информации (ИРИ);
- интерактивного доступа к полнотекстовым базам данных;
- контент-мониторинга.

Комплекс избирательного распространения информации позволяет:

- выполнять автоматическое сканирование доступных информационных ресурсов сети Интернет;
- приводить информацию к единому формату;
- автоматически классифицировать информацию, выполнять ее избирательное распространение.

Комплекс интерактивного доступа к базам данных обеспечивает:

- автоматическое создание оперативных и ретроспективных баз данных;
- интерактивный доступ пользователей к базам данных;
- санкционированный доступ пользователей к базам данных.

Комплекс контент-мониторинга, базирующийся на технологии Text Mining, обеспечивает:

- формирование информационных портретов и дайджестов;
- выявление основных сюжетных цепочек и новых событий;
- построение диаграмм динамики понятий и таблиц их взаимосвязей.

В настоящее время InfoStream охватывает мощнейший поток информации, превышающий 40 000 документов в сутки более чем с 2700 web-сайтов. Сервер системы InfoStream установлен на площадке ISP EIVisti, одного из ведущих провайдеров в Украине.



## 2. Система

### 2.1. Поиск и отображение информации

Поиск и отображение найденной информации - это основные задачи InfoStream, которые реализуются средствами информационно-поисковой системы InfoReS.

Для выполнения поиска в онлайн-режиме пользователю необходимо ввести запрос в соответствующее поле ввода. Запросы вводятся с использованием определенных правил, называемых **информационно-поисковым языком** или **языком запросов** системы InfoReS.

Правила и возможности использования языка запросов при работе с системой подробно описаны в Приложении А.

Следует отметить, что язык запросов системы InfoReS используется и в режиме избирательного распространения информации.

Вместе с тем, язык запросов может рассматриваться скорее как внутренний язык системы. В настоящее время пользователь может использовать расширенные аналитические возможности информационно-поисковой системы для ввода и уточнения запросов. InfoStream включает и такие семантические средства, как выявление содержательных дублей, поиск подобных документов, "информационные портреты" и др.

При выводе найденных по запросу пользователя (релевантных) документов обеспечивается их удобное и максимально информативное отображение.

### 2.2. Информационные портреты

Построение информационных портретов выполняется на основе эмпирических и частотно-статистических методов, основу которых составляет определение весов отдельных терминов в информационном массиве (результате выполнения поискового запроса).

С помощью информационного портрета визуально можно детализировать и уточнять поисковый запрос. В частности, информационный портрет существенно облегчает выбор источников информации, релевантных

первичному запросу.

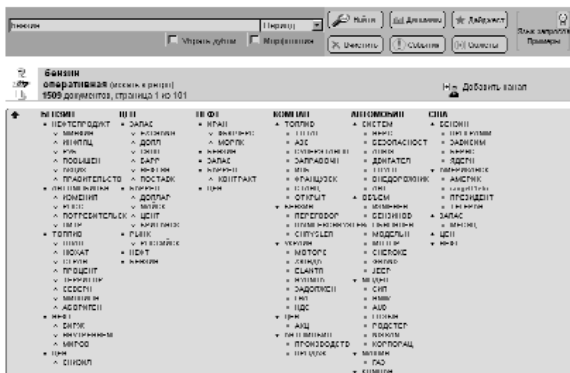
В информационном портрете отображаются такие характеристики массива документов, соответствующих запросу, как:

- рубрики базы данных;
- языки;
- размер сообщений (малый, средний, большой);
- цифровая насыщенность (малая, средняя, большая);
- страны источников;
- названия источников;
- тональность документов;
- географические названия из текстов документов;
- персоналии из текстов документов;
- наиболее характерные для данной выборки документов слова (термины).

Все приведенные характеристики ранжируются с учетом их "веса" в информационном массиве.

### 2.3. Классификатор-навигатор

Онлайн-интерфейс InfoStream включает реализацию многоуровневого классификатора-навигатора по откликам информационно-поисковой системы. На основе определения близости слов построен интерфейс уточнения запросов, реализующий принцип "поисковых папок пользователя".



Благодаря возможностям этого адаптивного инструмента, кластеризации результатов первичного

поиска, InfoStream позволят решать задачу динамической визуализации, предоставив пользователям интерфейс для последующего уточнения запроса и аналитического обобщения данных.

## 2.4. Дайджест

Дайджест строится на основе алгоритмов автоматического реферирования массивов документов - результатов поиска. Автоматическое реферирование, как и построение информационных портретов, выполняется на основе частотно-статистического метода. При этом его основу составляет определение весов как отдельных терминов, предложений и абзацев, так и целых документов.

В программе автоматического формирования дайджестов определяется заданное количество наиболее весомых по статистическим критериям документов, которые берутся в качестве его основы. При формировании дайджеста всегда используются заголовки выбранных документов. Кроме того, в дайджест включаются абзацы, которые имеют наивысшие весовые показатели, и не включаются дублирующиеся фрагменты. Для каждого фрагмента дайджеста указывается дата и ссылка на первоисточник.

Дайджест представляет собой самостоятельный документ, который можно при необходимости распечатать или сохранить в виде файла. Вместе с тем, электронный дайджест можно также рассматривать как аннотированный источник гиперссылок на документы, лежащие в его основе.

## 2.5. Основные сюжеты

Функция "Сюжеты" позволяет ответить на вопросы:

- что нового?
- о чем больше всего пишут?

путем семантического ранжирования результатов поиска.

При построении сюжетных цепочек система автоматически выявляет наиболее значимые темы, освещаемые в информационных массивах. Все весомые сообщения группируются по принадлежности автоматически определяемым сюжетам. В качестве

названий сюжетных цепочек используются заголовки сообщений, наиболее точно отражающих их суть. Порядок отображения сюжетов определяется количеством сообщений в сюжетной цепочке, что отражает общий интерес к данной теме, и временем публикации сообщений.

Процесс составления запроса для получения сложных цепочек максимально упрощен - для получения точных результатов вполне достаточно указать одно-два слова, относящихся к необходимой тематике.

## 2.6. Новые события

Для нахождения сообщений, соответствующих запросам и описывающим новые события, в отличие от полной выборки релевантных документов, реализован специальный режим "События".

Задача выявления новых событий из потока новостей является задачей прогноза того, о чем будут сообщать многие последующие документы. Эта задача актуальна как с практической, так и с научной точки зрения.

Реализованный в рамках InfoStream подход базируется на таких предположениях:

- а) минимальное время, прошедшее с момента публикации документа;
- б) близость лексического состава документа к лексическому составу массива документов за локальный промежуток времени;
- в) существенное различие лексического состава документа от лексического состава массива документов за значительный период времени;
- г) наличие в документе терминов, входящих в плюс-словарь (включающий важные для содержания новостей слова типа "теракт", "конфликт", "сенсация" и т.п.);
- д) высокий ранг "авторитетности" источника, а также допустимости лексики заглавий новостей;
- е) отсутствие дублирования информации.

## 2.7. Динамика понятий

Форма представления динамики встречаемости понятий - это гистограмма, которая строится как результат информационного поиска по множеству запросов за указанный период времени. Каждая дата в гистограмме является гиперссылкой, ведущей к результату поиска по указанному критерию.

В системе InfoStream реализован вариант гистограммы динамики понятий с указанием соотношений сообщений с различной тональностью.

## 2.8. Взаимосвязь рубрик

Таблица взаимосвязей рубрик строится как статистический отчет, отражающий близость (совместную встречаемость в новостных сообщениях) отдельных понятий. Элементы этой таблицы - коэффициенты взаимосвязей тематических рубрик, пропорциональные количеству документов, которые соответствуют обоим рубрикам.

С целью выявления блоков - множеств наиболее взаимосвязанных рубрик - применяется алгоритм кластерного анализа.

### 3. Сервис и решения на основе InfoStream

Существует целый ряд вариантов использования InfoStream (см. Приложение В). В рамках простейшего сервисного пакета пользователь может подписаться на получение по e-mail новостей по интересующей тематике, выраженной запросом, имеющим, например, такой вид: **(защита информации) & банк**. Данный запрос относится к защите информации в банковской сфере. Получение документов, соответствующих такому запросу, относится, к общему анализу отрасли, его субъектов и событий. Для анализа деятельности партнеров и конкурентов с помощью системы InfoStream можно подписаться на определяемые запросами информационные ленты, сообщения которых включают названия соответствующих фирм, имена и фамилии, бренды и т.п. Для работы пользователей, нуждающихся в постоянном варьировании запросов, предусмотрен режим онлайн, который идеально подходит для проведения постоянного контроля, например, источников рисков и конкурентов, оценки состояния рынков и т.д. Для того чтобы избежать информационной "перегрузки", получить наиболее существенные документы необходимой широкой тематики, можно воспользоваться средствами обобщения и уточнения запросов, реализующих современный подход "глубинного анализа текстов" (Text Mining). Еще один очень важный аспект - это персонализация поискового интерфейса в режиме онлайн - возможность сохранения запросов пользователями.

#### 3.1. Онлайн доступ к оперативной и ретроспективной информации

Для пользователей, информационные потребности которых не являются постоянными, предусмотрен режим InfoStream Online, сервисный пакет, обеспечивающий доступ в режиме онлайн к следующим базам данных:

- **оперативной** - содержащей документы, поступившие за последние несколько дней. Эта база данных обновляется в режиме реального времени и содержит документы, поступившие за последние 7 дней;

- **ретроспективной** - содержащей документы, поступившие за несколько последних месяцев;
- **табличной** - содержащей документы со структурированной информацией, например, прайс-листы, котировки, цены.

В рамках пакета услуг InfoStream Online предоставляются следующие возможности:

- поиск/просмотр документов и построения их информационных портретов;
- построение/просмотр дайджестов;
- построение/просмотр цепочек сюжетов;
- поиск/просмотр документов, содержащих описание новых событий;
- построение/просмотр диаграмм распределения и динамики встречаемости понятий;
- построение/просмотр таблиц взаимосвязей рубрик.

### 3.2. Рассылка новостной информации по e-mail

Пользователь сервиса InfoStream может подписаться на получение **по электронной почте** новостей по интересующей тематике. При этом запросы пользователей сохраняются в базе данных, а поиск и доставка релевантных документов происходят в режиме избирательного распространения информации. Гибкая система настройки и фильтрации по источникам, странам, предметным рубрикам, датам, персонам позволяет пользователю задав запрос, периодически получать по электронной почте подборки релевантных сообщений.

В настоящее время существуют несколько вариантов подписки на **получение по электронной почте** результатов поиска - от простейшего пакета MiniStream - рассылки информации по одному запросу один раз в сутки по одному электронному адресу - до MainStream - рассылки в режиме реального времени по нескольким адресам (см. Приложение В). В режиме рассылки по электронной почте пользователи также могут получать информацию в виде сюжетов, каждый из которых представлен наиболее характерной для него новостью и пятью заголовками близких по тематике публикаций со ссылками на первоисточники, а также медиа-рейтинги по заданным предметным областям.

### 3.3. Лента новостей на web-сайт

**InfoStream web (WebStream)** - это специальный сервисный пакет, предназначенный для информационной поддержки web-ресурсов, который обеспечивает экспорт данных, определяемых запросами пользователей, на страницы их web-сайтов и порталов. В рамках этого режима информация, соответствующая запросам, в виде заголовков и анонсов статей помещается на страницы web-сайтов пользователей. Информация экспортируется с помощью встраиваемого JavaScript-приложения. Предоставляемая пользователю (администратору web-ресурсов) возможность управления составом новостей (модерации) делает WebStream идеальным инструментом для организации на web-сайтах пользователей колонок профильной тематики, публикаций об отрасли, компаниях и т.п.

### 3.4. Использование технологии RSS

Персонализация интерфейса пользователей, работающих в режиме онлайн, может быть реализована на основе современной технологии RSS (Really Simple Syndication), формата данных и технического стандарта, который обеспечивает интегрированный доступ к данным. Технологию RSS можно рассматривать как один из видов избирательного распространения информации.

Использование технологии RSS позволяет пользователям, которые имеют доступ к InfoStream в онлайн-режиме, сохранять свои постоянные запросы и организовывать подписку на них с помощью разнообразных программ, которые имеют название "RSS-агрегаторы". Сегодня функции RSS-агрегаторов включены в некоторые браузеры, например в программу Opera.

Подробная инструкция по работе с RSS в рамках InfoStream с использованием агрегатора FeedReader приведена в Приложении Б.

### 3.5. Почтовый ящик для персональных рассылок

У пользователей режима избирательного распространения информации, которые получают



подборки новостей по электронной почте, имеется возможность использования бесплатного электронного почтового ящика, предоставляемого провайдером EIVisti.

Доступ пользователя InfoStream Client к этому почтовому ящику может осуществляться либо через web-интерфейс с помощью любого из существующих браузеров, установленных на его компьютере, либо путем добавления новой учетной записи в почтовой программе.

### 3.6. Сервер InfoStream Port

InfoStream Port - это впервые созданное в Украине аппаратно-программное решение - реализация корпоративного новостного поискового сервера, предназначенного для информационного обеспечения компаний разного уровня.

Программно-технологическое и информационное обеспечение комплекса InfoStream Port устанавливается на аппаратном сервере, подключенном к корпоративной сети заказчика. Комплекс InfoStream Port по своему назначению работает в интересах пользователей этой сети.

Информационное обеспечение InfoStream Port строится на основе использования информационного хранилища, формируемого на технической площадке ISP EIVisti в результате последовательности технологических операций:

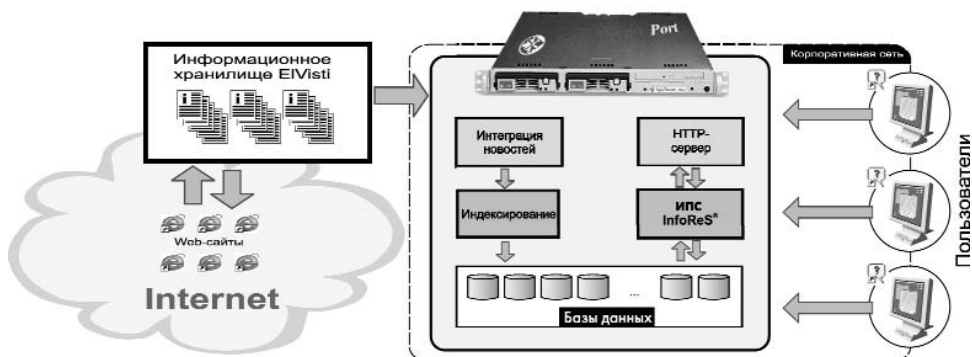
- сбор информации из сети Интернет;
- нормализация информации, приведение ее к единому формату;
- автоматическая классификация информации;
- помещение данных в информационное хранилище;
- предоставление санкционированного доступа к информационному хранилищу.

Программно-технологическое обеспечение InfoStream Port включает как компоненты утилиты обмена данными с хранилищем данных информационного провайдера и информационно-поисковую систему InfoReS. Комплекс InfoStream Port предоставляет доступ к базам данных оперативной, ретроспективной и табличной информации, которые

формируются непосредственно на нем.

Использование InfoStream Port обеспечивает:

- существенную экономию интернет-трафика;
- формирование и хранение ретроспективных баз данных практически неограниченного объема;
- интерактивный доступ корпоративных пользователей к базам данных;
- комфортную работу пользователей с неограниченного количества рабочих мест;
- экономию затрат на администрирование;
- высокий уровень защиты данных.



Таким образом, InfoStream Port решает задачи автоматизированного ведения хранилища новостей, которые собираются в сети Интернет, и мониторинга этих новостей со стороны пользователей корпоративной сети с помощью развитых средств поиска и обеспечения доступа к информации.

## 4. Работа пользователя InfoStream Online

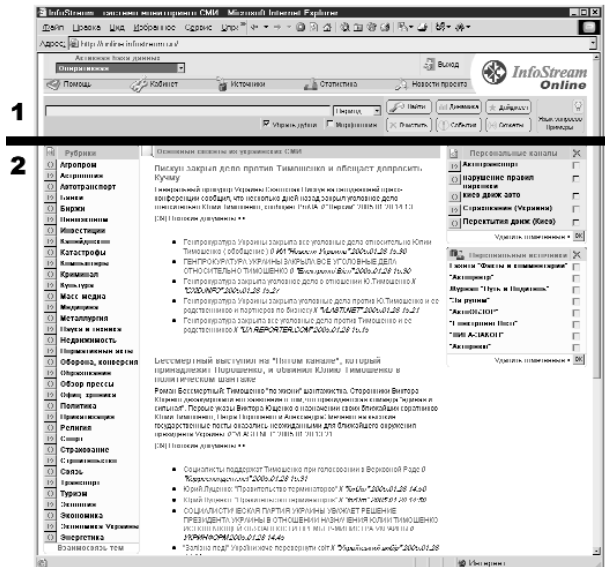
Для работы с системой InfoStream в режиме онлайн пользователь должен обратиться к серверу системы в сети Интернет по адресу <http://online.infostream.ua> и ввести свои регистрационные данные (логин и пароль). После авторизации пользователь получает в распоряжение удобный, интуитивный интерфейс с возможностью персонализации.

Ниже приведено описание основных экранов интерфейса и предоставляемых пользователю инструментов.

### 4.1. Основной экран

Первое, что видит пользователь после входа в систему - это "Основной экран". На всех экранах интерфейса системы можно выделить две области:

- 1) верхняя (основное "Меню системы");
- 2) рабочая (отображение результатов работы системы).



#### 4.1.1. Меню онлайн-доступа к системе InfoStream



Меню системы (1) постоянно и располагается на всех экранах интерфейса. Это меню предоставляет пользователю следующие возможности:

##### *Выбор базы данных*

Пользователь одновременно может работать (просматривать и производить поиск сообщений) только с одной из баз данных, подключенных к системе.

К системе подключены следующие базы данных:

- оперативная;
- ретроспективная;
- англоязычная;
- англоязычная/ретроспективная;
- табличная.

Выбор активной базы данных производится при помощи селектора, расположенного в левом верхнем углу экранной формы. Каждая база данных, подключенная к системе, предполагает свой вид "Основного экрана" работы с ней.

Так, на основном экране "**Оперативной**" базы данных располагаются:

- 1) меню системы;
- 2) форма ввода поискового запроса;
- 3) перечень рубрик;
- 4) перечень основных сюжетов из украинских СМИ;
- 5) перечень персональных информационных каналов пользователя системы;
- 6) перечень источников, выделенных пользователем как персональные.

На основном экране "**Ретроспективной**" базы данных располагаются:

- 1) меню системы;
- 2) форма ввода поискового запроса;
- 3) календарь выбора интервала дат для поиска документов.

На основном экране "**Англоязычной**" базы данных располагаются:

- 1) меню системы;
- 2) форма ввода поискового запроса;
- 3) перечень рубрик;
- 4) перечень основных сюжетов из зарубежных СМИ об Украине.

На основном экране "**Ретроспективной/Англо-язычной**" **базы данных** располагаются:

- 1) меню системы;
- 2) форма ввода поискового запроса;
- 3) календарь выбора интервала дат для поиска документов.

На основном экране "**Табличной**" **базы данных** располагаются:

- 1) меню системы;
- 2) форма ввода поискового запроса;
- 3) рубрики базы данных.

#### *Помощь*

Здесь содержится описание системы, интерфейса пользователя, методическое пособие по использованию технологии InfoStream, описание языка запросов, примеры сложных запросов.

#### *Кабинет пользователя*

Здесь пользователь имеет возможность:

- изменить настройки системы,
- настроить пользовательский интерфейс,
- просмотреть персональную статистику работы в системе.

#### *Источники*

Каталог источников, подключенных в систему. Пользователь имеет возможность объединить отдельные источники в группу "Персональные". Каталог позволяет также отслеживать активность источников.

#### *Статистика поступлений*

Представляет статистику поступлений сообщений в базу данных системы.

#### *Выход из системы*

Команда выхода из системы.

### **4.1.2. Рабочая область экрана (отображение результатов работы системы)**

На рабочей области экрана располагаются:

*Форма поиска* (2) позволяет вводить, редактировать, сохранять поисковые запросы, таким образом



конфигурируя информационный канал, организуемый посредством системы. Специальными возможностями системы являются автоматическое формирование "Дайджестов электронной прессы" и "Сюжетов", а так же выделение "Событий".

В качестве информационно-поисковой системы в InfoStream Online используется система InfoReS. Язык запросов описан в Приложении А.

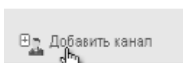
Поисковая форма предоставляет возможность указания критериев запроса (учитывать / не учитывать).

**Критерии запроса:**

- поисковые термины;
- логические операторы;
- морфология естественного языка (учитывать / не учитывать);
- дублирующиеся сообщения (показывать / не показывать);
- диапазон дат сообщения.

Кроме этого, к критериям запроса относятся термины и операнды из "Информационного портрета".

В момент, когда результаты обработки поискового запроса системой удовлетворят требованиям пользователя, запрос может быть сохранен для дальнейшего неоднократного обращения к нему. Таким образом осуществляется добавление в систему нового (или сохраняются новые настройки уже существующего) персонального информационного канала.



*Перечень рубрик (3)*

В качестве тематических рубрик в системе используются предустановленные поисковые запросы. Примеры запросов, соответствующих рубрикам, приведены в Приложении 2. Кликнув на названии рубрики, можно просмотреть полный перечень соответствующих ей документов. "Обзор основных сюжетов" по рубрике - кнопка "(·)".



*Перечень основных сюжетов из украинских СМИ (4)*

На основном экране системы представлен перечень наиболее популярных сюжетов из украинских СМИ на момент подключения пользователя к системе. Кликнув по ссылке "Похожие документы", можно просмотреть все сообщения, хранящиеся в базе данных и относящиеся к данному сюжету.

Просмотреть сообщение можно, кликнув по его заголовку.



*Перечень персональных информационных каналов пользователя системы (5)*

Персональные информационные каналы, организованные пользователем, приводятся в "Перечне каналов", в котором представлены:

- "Обзор основных сюжетов" для персонального канала - кнопка "(-)";
- название канала.

Управление организованными каналами:

- просмотр персонального канала;
- просмотр "Обзора основных сюжетов";
- удаление персонального канала производится непосредственно из "Перечня каналов".



*Перечень источников, выделенных пользователем как персональные (6)*

В "Перечне персональных источников" приводятся:

- название источника;
- кнопка удаления источника из перечня.

Управление "персональными источниками":

- просмотр сообщений из "персонального" источника;
- удаление источника из перечня "персональных" производится непосредственно из "Перечня персональных источников";
- добавление источников в "персональные" происходит из каталога-перечня источников системы;
- уточнение запроса путем поиска только в персональных источниках.



*Календарь выбора интервала дат для поиска документов (7)*

Интервал дат для поиска документов в ретроспективной базе данных можно указать, отметив поля выбора соответствующих месяцев года.

Интервал за весь год можно выбрать, кликнув по номеру года. Также можно использовать максимальный интервал для поиска (выбрав "Весь диапазон дат").

## 4.2. "Кабинет пользователя"



Здесь, помимо "Меню системы", пользователь имеет доступ (1) к:

- настройкам персональных информационных каналов;
- настройкам персональных источников;
- статистике работы пользователя в системе.

### *Настройка персональных информационных каналов (2)*

В момент, когда результаты обработки системой поискового запроса удовлетворяют требованиям пользователя, запрос может быть сохранен для дальнейшего многократного обращения к нему.

Таким образом осуществляется добавление в систему нового (или сохраняются новые настройки уже существующего) персонального информационного канала. Сохраняющиеся параметры канала можно видеть в строке ввода в "форме ввода поискового запроса", а именно, это критерии запроса:

- поисковые термины;
- логические операторы;
- морфология языка (учитывать / не учитывать);
- дублирующиеся сообщения ( показывать / не показывать).

Кроме этого, к критериям запроса относятся термины и операнды, выбранные из "Информационного портрета запроса":

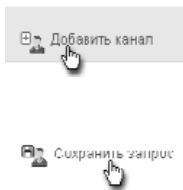
- рубрики;
- языки;
- размер сообщений;
- цифровая насыщенность сообщений;
- страны источников;
- источники;
- тональность;
- география;
- персоны;
- слова.

Персональные информационные каналы хранятся и отображаются в виде перечня.

Создание нового канала происходит следующим образом:

- анализируется результат отработки поискового запроса;
- если необходимо, корректируется поисковый





запрос (либо на языке запросов, либо при помощи "Информационного портрета") до получения приемлемых результатов;

- параметры запроса сохраняются в системе под собственным именем ("Добавить канал").

Конфигурирование существующего канала производится при его просмотре. Новые настройки можно сохранить для этого же канала ("Сохранить настройки") или добавить новый канал ("Добавить канал") с этими параметрами. В этом случае параметры исходного канала останутся без изменений.

Удаление персонального информационного канала происходит путем выбора чекбокса соответствующего канала в перечне и нажатия кнопки "Удалить отмеченные".

### *Настройка персональных источников (3)*

Любой источник, подключенный к системе, может быть выделен в перечень "Персональные источники". Персональные источники хранятся и отображаются в виде перечня.



Добавление источника в перечень персональных происходит в "Каталоге-перечне источников". В строке с названием источника необходимо отметить чекбокс и нажать кнопку "Добавить отмеченные в Персональные источники".



Удаление источника из перечня персональных происходит путем выбора чекбокса соответствующего источника в перечне и нажатия кнопки "Удалить отмеченные".

### *Статистика работы пользователя в системе (4)*

Статистика отображает:

- количество страниц, просмотренных пользователем;
- количество документов, просмотренных пользователем;
- количество уникальных документов, просмотренных пользователем;
- дату первого события;
- дату последнего события за отчетный период.



### 4.3. Каталог источников

Открытые источники сети Интернет, сканируемые системой, называются "Каталог источников". Все источники для удобства отображения сгруппированы по категориям:

- информационные агентства;
- теле-, радиоканалы;
- газеты;
- еженедельники и журналы;
- официальные источники;
- ассоциации, компании;
- интернет-издания;
- другие;
- архивы.

Источник описывается следующими полями:

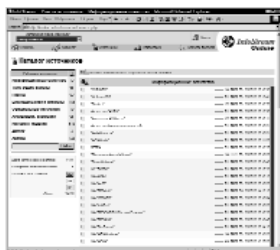
- категория (рубрика в каталоге);
- название источника;
- URL web-сайта источника;
- страна принадлежности источника;
- язык источника;
- периодичность обновления источником своего содержания;
- дата поступления последнего сообщения данного источника в систему;
- индикатор активности источника.

Каталог является функциональным, т.е.:

- имеется возможность поиска источника по названию и/или адресу web-сайта;
- нажав на название источника, можно просмотреть все его документы, хранящиеся в базе данных системы;
- нажав на надпись www, можно перейти непосредственно на web-сайт источника в сети Интернет;
- по индикатору активности источника можно судить о частоте обновления источником своего содержания и, соответственно, поступлении документов из него в базу данных системы;
- подсветка строк перечня говорит о принадлежности источника к Украине (желтый тон) и России (синий тон).

Любой источник, подключенный к системе, может быть добавлен в перечень персональных источников пользователя системы.

Источники к системе подключаются



администраторами. Пользователь может обратиться к администратору системы с соответствующей заявкой, если не найдет в Каталоге необходимого для себя источника.

#### 4.4. Статистика поступления информации

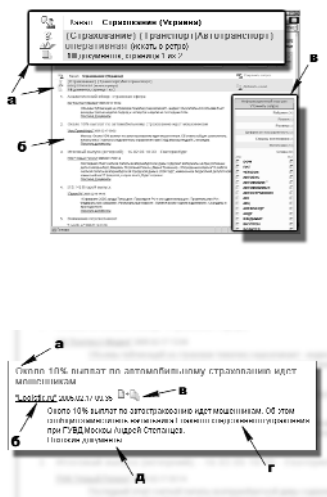


База данных (БД) системы наполняется документами из открытых источников сети Интернет, подключенных к системе. Функция анализа поступлений в БД представляет гистограмму распределения количества документов, которые поступили с указанием количественных параметров распределения по дням.

Нажав на дату, можно просмотреть детальную статистику по поступлениям документов в БД за выбранный день из подключенных к системе источников.

Следующий уровень анализа - детальный просмотр документов из источника. Представление документов аналогично представлению результатов поиска или просмотра персонального информационного канала .

#### 4.5. Результаты поиска



Результаты поиска отображаются в виде перечня документов, которые отвечают критериям поискового запроса и включают:

- а) заголовок перечня, содержащий:
  - название просматриваемой рубрики (если был выбран просмотр рубрики) или поисковый запрос;
  - выбранные базы данных и ссылку на ее смену;
  - количество найденных в БД документов и страниц выдачи;

- б) собственно перечень документов;
- в) "Информационный портрет" запроса/рубрики.

Документ в перечне представляется так:

- а) название сообщения (является переходом к полному тексту документа);
- б) источник сообщения (название источника является переходом на просмотр всех сообщений из данного источника в активной базе данных),

- дата поступления документа в БД;
- в) признак того, что это сообщение дублирует другое сообщение, поступившее в БД ранее;
- г) аннотация сообщения;
- д) переход к перечню похожих сообщений в активной базе данных ("Похожие документы").

"Информационный портрет" запроса/рубрики представляет собой множество ключевых слов и других параметров, наиболее точно отображающих информацию, полученную в результате поиска.

В информационном портрете отображаются такие характеристики массива документов, соответствующего критериям запроса:

- рубрики;
- языки;
- размер сообщений (малый, средний, большой);
- цифровая насыщенность (малая, средняя, большая);
- страны источников;
- источники;
- географические названия;
- персоналии;
- тональность;
- слова.

Все характеристики ранжируются с учетом их "веса" в портрете. Наиболее весомые (характеризующие большинство документов в результате обработки поискового запроса) имеют обозначение "\*" (звездочка).

При помощи информационного портрета можно легко детализировать и уточнять поисковый запрос, либо конфигурировать персональный информационный канал.

Для уточнения запроса по информационному портрету достаточно кликнуть на одной из характеристик. При этом уточнение будет производиться с учетом морфологии русского и украинского языков (от слова будет автоматически отделено окончание). Для уточнения запроса сразу по нескольким характеристикам из информационного портрета, можно отметить несколько чекбоксов "AND" - логическое "И" (слева от слова) или "NOT" - логическое "НЕТ" (справа от слова), а затем кликнуть на кнопке "Уточнить запрос". При уточнении в этом режиме также будет учитываться морфология.

**Информационный портрет**  
Уточнить запрос

**Рубрики (3)**

AND NOT

Строительство \*\*\*

Образование

Автотранспорт

**Языки (1)**

**Размер (2)**

AND NOT

средний \*\*\*\*

малый

**Цифровая насыщенность (2)**

**Страны источников (1)**

AND NOT

Украина \*\*\*\*

**Источники (9)**

**Тональность (1)**

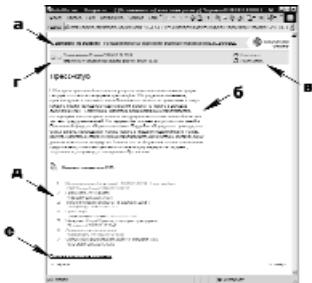
**География (8)**

**Персоны (10)**

**Слова (80)**

Уточнить запрос

#### 4.6. Просмотр сообщения/документа



Документ открывается в отдельном окне программы просмотра web-страниц. Его представление включает:

- а) название рубрики или запроса, или канала, которому соответствует данный документ;
- б) название и полный текст сообщения (в случае непосредственного ввода запроса в "форму ввода поискового запроса" найденные ключевые слова в тексте документа выделяются цветом);
- в) команд "Распечатать" (печать на принтере) и "Сохранить" (сохранение текста документа в отдельный файл на компьютере пользователя);
- г) название источника, дата поступления документа в БД системы, признак того, что это сообщение дублирует другое, которое поступило в БД раньше;
- д) перечень названий документов в активной базе данных, похожих на данный по содержанию;
- е) переход к полному перечню похожих документов в активной БД системы.

#### 4.7. Дайджест электронной прессы



Дайджест формируется на основе алгоритмов реферирования массивов документов - результатов поиска по запросу (а). При формировании дайджеста используются заголовки (б). Кроме того, в дайджест включаются абзацы, которые имеют наивысшие весовые показатели (в). Для каждого фрагмента дайджеста указывается дата его публикации и гиперссылка на первоисточник.

Дайджест электронной прессы представляет документ, состоящий из заранее заданного количества фрагментов, который, при необходимости, можно распечатать (г).

Вместе с тем дайджест электронной прессы можно также рассматривать как аннотированный источник гиперссылок на документы, которые лежат в его основе (в).

## 4.8. Обзор основных сюжетов



Функция "Обзор основных сюжетов" обеспечивает семантическое ранжирование результатов поиска. Все сообщения, соответствующие запросу/рубрике/каналу (а), группируются по принадлежности автоматически построенным сюжетам (б). В качестве названия сюжета используется заголовок сообщения, которое наиболее точно отражает суть данного сюжета (в). Порядок отображения сюжетов определяется количеством сообщений в сюжете (длиной сюжетной цепочки), которое отражает общий интерес к данной теме, и временем публикаций входящих в сюжет сообщений.

Представление сообщений в сюжетах практически аналогично представлению сообщений, выдаваемых в результате поиска, за исключением аннотаций.

Обзор основных сюжетов представляет собой документ, который, при необходимости, можно распечатать (г).

## 4.9. Новые события



Для выявления сообщений, соответствующих поисковым запросам и описывающим новые события, в отличие от полной выборки релевантных документов, реализован специальный режим "События".

Реализованный в рамках InfoStream подход базируется на предположении минимального времени, прошедшего с момента публикации документа, близости слов документа к словам массива документов за локальный промежуток времени и существенного различия лексического состава документа от словаря массива документов за значительный период времени и т. д.

## 4.10. Динамика понятий



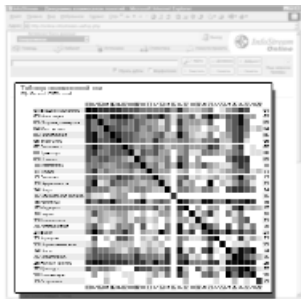
Для динамического анализа встречаемости понятий строится гистограмма распределения сообщений в базе данных системы по датам поступления.

Гистограмма динамики понятий позволяет проанализировать, как изменяются упоминания того или иного понятия, изучить его развитие во времени.

Непосредственно из гистограммы можно перейти к перечню документов, в которых данные понятия встретились в течение выбранного периода.

В системе InfoStream реализовано также отображение гистограммы динамики понятий с указанием соотношения тональностей сообщений.

#### 4.10. Взаимосвязь рубрик



Для наглядности отображения таблицы взаимосвязей рубрик отдельные ее элементы окрашиваются в различные оттенки серого цвета (в зависимости от значений коэффициентов взаимосвязи).

При наведении курсора на ячейку таблицы взаимосвязей рубрик отображается нормированный вес, а также количество документов, одновременно соответствующих рубрикам столбца и строки данной ячейки.

Каждая ячейка таблицы взаимосвязей рубрик представляет собой гиперссылку, которая указывает на список документов, одновременно соответствующих рубрикам столбца и строки данной ячейки.

## **Заключение**

InfoStream представляет собой систему интеграции новостных ресурсов сети Интернет, охватывающую в настоящее время практически все основные информационные web-сайты Украины и России.

За время своего существования технология, система и сервис InfoStream обрели широкую популярность и надежную клиентскую базу на украинском рынке. Вместе с тем требования, которые предъявляют пользователи к системе мониторинга новостных ресурсов Интернет, продолжают расти. Это связано как с увеличением информационных потоков (в настоящее время система InfoStream сканирует свыше 40 000 уникальных сообщений в сутки), так и с необходимостью не только поиска, но и проведения эффективного анализа его результатов.

InfoStream является одним из необходимых звеньев для принятия управленческих решений. Профессиональное использование возможностей системы InfoStream, доступное ее пользователям, обеспечивает качественно новые возможности для информационно-аналитической работы в самых различных областях деятельности - от политики, макроэкономики, прессы, банковской деятельности - до управления персоналом.

Навыки, получаемые пользователями в процессе работы с InfoStream, могут быть использованы при поиске и обобщении информации многих типов (не только новостной) с помощью самых разнообразных поисковых систем. Формализация поисковых предписаний, поиск значимых ключевых слов, учет формальных логических и лингвистических особенностей, поэтапное уточнение критериев поиска и многие другие подходы и приемы будут способствовать повышению эффективности информационно-аналитической деятельности в любой области.



## ПРИЛОЖЕНИЕ А

### Использование языка запросов ИПС InfoReS

Формирование запросов - это искусство, доступное каждому. Запросы составляются с использованием определенных правил, называемых в совокупности информационно-поисковым языком или "языком запросов".

Запрос вводится в область ввода текста поисковой формы и передается поисковой системе при нажатии на кнопку "Поиск". В режимах подписки запросы сохраняются администратором системы в базе данных.

Запросы состоят из термов (слов или их правых усечений) и операторов.

#### Особенности составления запросов

##### Термы

Термы - это слова естественного языка или их правые усечения, состоящие как минимум из двух букв. По умолчанию каждое введенное слово воспринимается как основа для поиска, т.е. введя, например, запрос **завод**, можно найти документы, содержащие словоформы: "завода", "заводить", "заводы" и др. При необходимости нахождения точного вхождения слова, при вводе запроса следует добавить к слову символ "]"", например: **завод]**.

Система InfoReS не различает прописных и строчных букв, поэтому для поисковой процедуры запросы **завод]** и **Завод]** равнозначны.

##### Словосочетания

Словосочетания - это термы, состоящие из нескольких слов. Для реализации возможности поиска по словосочетаниям используется специальный оператор контекстной близости ADJ (возможно равнозначное написание - "~").

Оператор контекстной близости обеспечивает отбор документов, в которые входят слова, связанные с этим оператором. Эти слова должны находиться в документах в указанной последовательности рядом друг с другом.

По умолчанию предполагается, что это соседние слова в документе (между ними отсутствуют какие-либо другие слова). Существует возможность задания в запросе расстояния между словами: /0/ - соседние слова (по умолчанию), /1/ - не более 1 слова в тексте документа между словами; /2/ - не более 2-х слов и т. д. Например, запрос **транспорт~1/нефти** обеспечивает нахождение документов, в состав которых входят словосочетания **"транспортировка нефти"** и **"транспорт иранской нефти"**, в то время, как запрос **транспорт~нефти** позволит выбрать только документ с первым словосочетанием.

## Логические операторы и скобки

В системе используется следующий набор логических операторов:

- NOT - логическое НЕТ, понимаемое как И-НЕТ;
- AND - логическое И;
- OR - логическое ИЛИ.

При употреблении операторов допускается также их сокращенное написание:

- NOT равнозначно "!" или "^";
- AND равнозначно пробелу или "&" или "+";
- OR равнозначно "|" или "," или ";".

Например, запрос **банк&кредит&украин** равнозначен запросам **банк кредит украин**, **банк+кредит+украин** и обеспечивает отбор документов, в которые входят все три термина - **"банк"**, **"кредит"**, **"украин"**.

Запрос может быть многоуровневым. Различные уровни определяются с помощью круглых скобок. С помощью скобок также рекомендуется выделять термины-словосочетания.

## Опциональные возможности

Язык запросов позволяет использовать в качестве термов определенные сочетания символов, которые могут трактоваться как рубрики, коды источников, даты и т. п.

### *"Источники"*

При поиске по источникам в качестве термов можно использовать правые части соответствующих доменных имен, например, **www.elvisti.com**, **www.lenta** или **4vlada.net**.

### *"Даты"*

Для поиска по датам в базах данных, доступных в режиме онлайн, как термы для поиска можно задавать даты в формате ГГГГ.ММ.ДД, например, **2006.06.12**. Допускаются также правые усечения дат, например, если указать в запросе **2006.06.0**, то будут выданы документы с 1 по 9 июня 2006 года.

### *"Страны"*

В системе применяется двухбуквенное кодирование стран, к которым относятся сайты - источники информации. Например, для поиска по сайтам, относящимся к Украине, достаточно уточнить запрос термом **country.ua**, соединив его с остальной частью запроса оператором "&".

### *"Рубрики"*

В запросе, так же как обычные термы, можно использовать коды рубрик. Например, запрос **rubr02&(нбу | (нацбанк~укра)|(нац~банк~укра))** обеспечивает отбор документов по банковской тематике, в которых есть информация о Национальном банке Украины. В качестве тематических рубрик в системе используются предустановленные запросы.

### *"Морфология"*

Режим "Морфология" обеспечивает предварительную обработку слов, входящих в поисковый запрос. В каждом слове отбрасывается изменяемое окончание, что приводит к охвату системой не только слов, но и их словоформ.

Важно, что пользователь всегда имеет возможность как активизировать этот режим, так и отменить его.

### *"Убрать дубли"*

Эта возможность позволяет исключить из результатов поиска сообщения, дублирующиеся не только целиком, но и по смыслу. Выявление дублей на основе лингво-статистических алгоритмов происходит на этапе формирования базы данных системы.

### *"Подобные документы"*

При выводе результатов поиска каждое сообщение дополнено ссылкой "Подобные документы", которая обеспечивает переход к списку содержательно близких сообщений. Содержательная близость, как и смысловое дублирование, выявляется на основе лингвостатистического анализа сообщений, но отличается более мягкими критериями.

### *"Насыщенность цифровой информацией"*

Пользователю сервиса InfoStream Online доступна возможность указания уровня насыщенности документов цифровой информацией. Эта возможность полезна, например, при поиске аналитических документов, ценовых таблиц, результатов рейтингов и т.п. В системе выделено три уровня насыщенности документов цифровой информацией: высокая (numb.large) - свыше 10%, средняя (numb.medium) - свыше 3% и низкая (numb.small) - до 3%. Обращение к этой возможности, в частности, предусмотрено в информационном портрете.

### *"Размер сообщений"*

Пользователю сервиса InfoStream Online также доступна возможность задания характеристик размеров искомых документов. Эта возможность может быть использована, например, как при поиске объемных аналитических материалов, обзоров, законодательных актов, так и при поиске кратких, насыщенных цифрами таблиц котировок, курсов валют или сводок погоды. В системе предусмотрено три уровня размера сообщений: высокий (leng.large) - свыше 10000 символов, средний (leng.medium) - свыше 1000 символов и низкий (leng.small) - до 1000 символов. Как и в предыдущем случае, эта возможность также отражена в информационном портрете.

### *"Тональность"*

Эмоциональная окраска (тональность) сообщений в системе InfoStream определяется по вероятностному

принципу на основании учета лексики из специальных словарей. Тональность присваивается не каждому сообщению. В случае достаточно надежного определения эмоциональной окраски сообщения, ему приписывается параметр "Позитивно" или "Негативно". Значение тональности используется для уточнения поискового запроса в информационном портрете, а также отображается в специальной диаграмме встречаемости понятий.

### *"Персоны"*

В системе InfoStream автоматически выявляются фамилии людей, которые затем отображаются в информационном портрете. Эти фамилии (папка "Персоны") можно использовать как уточняющие запрос параметры.

### *"География"*

В качестве параметров, уточняющих первоначальный поиск в информационном портрете системы InfoStream предусмотрены географические названия, которые автоматически определяются из текстов документов.

## ПРИЛОЖЕНИЕ Б

### Технология RSS и система InfoStream

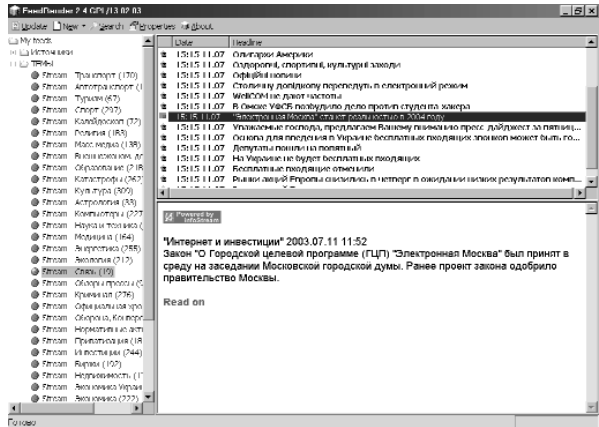
#### Персонализация на основе технологии RSS

Персонализация интерфейса пользователей, работающих в режиме онлайн, т.е. сохранение их постоянных запросов и организация подписки на них, может быть реализована на основе современной технологии RSS (Really Simple Syndication), формата данных и технического стандарта, который обеспечивает интегрированный доступ к новостной информации на web-сайтах.

#### Организация доступа к информации

Пользователи могут получить доступ к данным в формате RSS с помощью специальных программ, называемых RSS-агрегаторами.

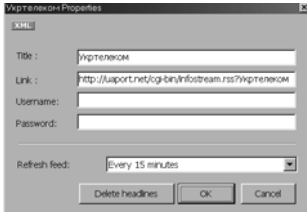
Интерфейс RSS-агрегатора



В качестве RSS-агрегатора рекомендуется использовать FeedReader версии 2.4. (дистрибутив приведен, например, по адресу: <http://infostream.ua/prg/feedreader24.exe>)

Для получения тематической ленты (RSS-фида) от системы InfoStream в соответствующее поле RSS-агрегатора следует ввести адрес в формате: **[http://online.infostream.ua/rss.php\[?<ЗАПРОС>\]](http://online.infostream.ua/rss.php[?<ЗАПРОС>])** где в качестве ЗАПРОСА можно ввести слово или словосочетание на языке запросов InfoReS.

## Настройка подписки



Окно подписки

Для настройки подписки на тематическую ленту (RSS-фид) следует в режиме InfoStream Online отладить запрос, после чего запустить программу FeedReader, активизировать опцию New и ввести следующую информацию:

- адрес RSS-фида, включая запрос на информационно-поисковом языке системы InfoReS в формате, приведенном выше;
- название фида, которое может быть определено пользователем;
- периодичность обновления.

Имеется возможность изменения кодировки, размеров шрифтов, помещения фида в отдельную папку, группировки фидов и т.д.

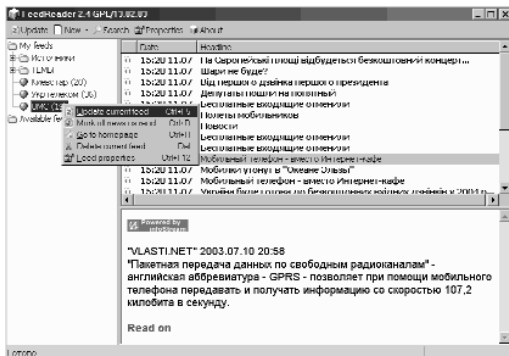
Для управления подпиской в этом режиме существуют дополнительные опции, активируемые для каждого RSS-фида отдельно:

- обновление фида (списка активных сообщений);
- отметка всех сообщений как уже прочитанных;
- удаление списка сообщений;
- изменение свойств подписки, включая тему, периодичность и др.

Для получения полного текста сообщения, заголовков и аннотация которого вызвали интерес, следует:

- произвести двойное нажатие левой клавиши мыши на заголовке,
- нажать на ссылку "Read on" в поле аннотации или
- нажать на соответствующую кнопку, стоящую перед заглавием, или
- нажать правую клавишу мыши, находясь курсором на заглавии, при этом можно открыть текст сообщения в новом окне web-браузера.

Опции FeedReader



## ПРИЛОЖЕНИЕ В

### Основные пакеты услуг InfoStream

Основные пакеты услуг сервиса InfoStream Client

- 1. InfoStream Email** - пакет услуг, предусматривающий получение информации на один или несколько электронных адресов пользователя по одному тематическому запросу с периодичностью от одного раза в сутки до режима реального времени (по мере появления сообщений). Стоимость данного пакета услуг определяется объемом рассылки (средним количеством документов за период), периодичностью отправления электронных отчетов и количеством электронных адресов для получения рассылок:
  - **MiniStream** - получение информации пользователем по одному запросу один раз в сутки. Пакет услуг MiniStream предусматривает доставку информации только по одному электронному адресу.
  - **MidStream** - получение информации пользователем по одному запросу 1 или 2 раза в сутки. Пакет услуг MidStream предусматривает доставку информации на один или два электронных адреса.
  - **BizStream** - получение информации пользователем по одному запросу один, два или шесть раз в сутки. Пакет услуг BizStream предусматривает рассылку информации на один или два электронных адреса.
  - **MainStream** - получение информации пользователем по одному запросу в режиме реального времени, круглосуточно, по мере появления информации в Интернет.
- 2. InfoStream Ukrainian Day** - ежедневная рассылка (утром или вечером) наиболее важных событий дня, произошедших в Украине или касающихся Украины. Рассылка осуществляется в виде сюжетов.
- 3. InfoStream Topics** - ежедневная рассылка логически структурированной информации по запросу в виде последовательных автоматически формируемых сюжетов, каждый из которых представлен наиболее "характерной" для сюжета новостью (приводится полностью) и пяти последних заголовков подобных публикаций (ссылки на первоисточники). Количество сюжетов определяется объемом документов, относящихся к запросу пользователя (до 20). Пакет предоставляет идеальный инструмент для оперативного ознакомления с новостями рынков, сфер деятельности, событий, явлений, происшествий и т.д.
- 4. InfoStream Rating** - аналитический отчет (еженедельный или ежемесячный), содержащий результаты анализа медиа-присутствия заданных объектов (компаний,



брендов, персон и др). Медиа-рейтинг содержит информацию о количестве упоминаний объектов, объектах, имеющих наилучшую абсолютную и относительную динамики медиа-присутствия, общую оценку медиа-активности анализируемых объектов, сравнительную диаграмму медиа-присутствия самых упоминаемых объектов, список интернет-СМИ, уделяющих наибольшее внимание освещению вопросов рынка, участниками которого являются анализируемые объекты.

5. **InfoStream Web (WebStream)** - специальный пакет для интернет-проектов, с помощью которого на сайте пользователя организуется автоматически обновляемая лента новостей по тематике, интересующей пользователя (в виде заголовков и анонсов публикаций). При этом пользователю предоставляется возможность управления содержимым ленты новостей с помощью простого и удобного web-интерфейса.
6. **InfoStream Online** - пакет услуг, в рамках которого предоставляется доступ к электронным массивам новостной информации в режиме online. Кроме оперативной информации, содержащей новости за последние 7 дней, пользователю доступна и ретроспективная информация за последние три месяца. В рамках пакета услуг пользователям предоставляется возможность поиска и просмотра новостей, а также автоматического построения дайджестов, сюжетов и временной диаграммы встречаемости понятия, заданного запросом пользователя. Для одного пакета услуг InfoStream Online установлено максимальное количество новостей, к которым пользователь может получить доступ - 1000 (одна тысяча) документов за месяц (количество запросов, просмотров результатов поиска и автоматической обработки этих результатов не ограничено).
7. **InfoStream Month/Year** - отчет об упоминаемости компании, организации, персоны, бренда в электронных СМИ за месяц/год. Отчет содержит количество упоминаний компании, организации, персоны, бренда за весь месяц/год по дням, неделям и месяцам, а также обоснование наиболее значительных изменений (увеличения или уменьшения) упоминаемости. Кроме этого, в отчете приводится сводная таблица интернет-источников, опубликовавших информацию о компании, организации, персоне, бренде с указанием количества публикаций.

С подробной информацией, а также с примерами рассылок и отчетов можно ознакомиться на сайте InfoStream (<http://infostream.ua>)

Пакет услуг/ Продукт	Количество документов в месяц	Способ получения	Максимальная периодичность получения
InfoStream Email	до 600 (MiniStream)	e-mail: 1 адрес	1 раз в сутки
	до 1500 (MidStream)	e-mail: до 2 адресов	1 или 2 раза в сутки
	до 3000 (BizStream)	e-mail: до 2 адресов	1, 2 или 6 раз в сутки
	до 4500 (MainStream)	e-mail: до 3 адресов	по мере появления
	более 4500	e-mail: до 3 адресов	по мере появления
InfoStream Online	до 1000	онлайн	по мере появления
InfoStream Web	до 1000	JavaScript	по мере появления
InfoStream Topics	-	e-mail: 1 адрес	1 или 2 раза в сутки
Ukrainian Day	-	e-mail: 1 адрес	1 или 2 раза в сутки
InfoStream Rating	-	e-mail: 1 адрес	1 или 4 раза в месяц
InfoStream Month	-	e-mail: 1 адрес	1 раз в месяц
InfoStream Year	-	e-mail: 1 адрес	1 раз в год

Для ознакомления с сервисом InfoStream возможно предоставление бесплатной тестовой рассылки новостей на адрес электронной почты или бесплатный тестовый доступ к электронным массивам оперативной и ретроспективной информации в режиме онлайн, что позволит сравнить всем интересующимся уже имеющиеся и используемые методы сбора и обработки информации с предлагаемыми возможностями системы автоматического мониторинга и поможет аргументированно определиться с решением о целесообразности сотрудничества.



Информационный центр "ЭЛВИСТИ"  
г. Киев, ул. Максима Кривоноса, 2-А, Internet-офис ElVisti  
Телефон/факс: (380 44) 239-90-91, 247-39-40, 247-39-41  
<http://infostream.ua>  
e-mail: [stream@visti.net](mailto:stream@visti.net)