

Информационный центр "ЭЛВИСТИ"



**Система  
интеграции  
новостей из сети  
Интернет**

*Методическое пособие*

Киев-2003

Система интеграции новостей из сети Интернет /Методическое пособие/

Под общей редакцией к.т.н. Д.В.Ландэ

Редакторы: С.А.Бороденков, Т.Г.Селиверстова, О.А.Шевчук

Информационный центр "ЭЛВИСТИ"  
г. Киев, ул. Максима Кривоноса 2-А, "Internet-офис EIVisti"  
Телефон/факс: (380 44) 239 90 91, 247 39 40, 247 39 41  
E-mail: stream@visti.net

## Аннотация

Назначение данного пособия - ознакомить пользователей (как реальных, так и потенциальных) с системой интеграции новостей InfoStream™, научить их правилам составления запросов, использованию широких возможностей поиска информации с целью ее дальнейшей аналитической обработки.

Система InfoStream™ предназначена для нахождения в сети Интернет новостной информации по интересующим пользователя темам, оперативной доставки результатов поиска, и, таким образом, минимизации усилий пользователя на отсеивание дублирующейся информации, шума.

Руководство состоит из трех основных разделов. Первый – включает общее описание системы InfoStream™, решаемых ею задач и сервисных режимов.

Второй раздел посвящен описанию интерфейса пользователя, языка запросов, особенностей его применения.

В третьем разделе описаны дополнительные интеллектуальные возможности системы: средства уточнения запросов (так называемые “информационные портреты”), автоматического создания дайджестов, а также персонализации поискового интерфейса.

Для тех, кто еще не является пользователем системы, но желает ознакомиться с ее возможностями на практике, предоставляется бесплатный тестовый доступ к системе в течение недели.

## Оглавление

1. Система InfoStream™ .....	5
1.1. Спектр задач .....	5
1.2. Технологические аспекты .....	5
1.3. Виды сервисов .....	6
1.3.1. Он-лайн доступ к оперативным и ретроспективным базам данных .....	7
1.3.2. Варианты рассылки новостной информации .....	7
1.3.3. Поток новостей на сайт .....	8
1.3.4. InfoStream™ Port .....	8
2. Интерфейс пользователя системы InfoStream™ .....	9
2.1. Общие сведения о пользовательском интерфейсе .....	9
2.2. Язык запросов InfoReS .....	11
2.3. Особенности составления запросов .....	11
2.3.1. Термы .....	11
2.3.2. Словосочетания .....	11
2.3.3. Логические операторы и скобки .....	12
2.3.4. Источники, даты, рубрики .....	13
3. Дополнительные возможности .....	16
3.1. Информационные портреты .....	16
3.2. Дайджесты .....	17
3.3. Персонализация .....	18
3.3.1. Организация доступа к информации .....	18
3.3.2. Настройка подписки .....	19
Заключение .....	21

## 1. Система InfoStream™

На протяжении последних лет во всем мире и в Украине существенно развивается информационное наполнение сети Интернет, что обуславливает появление ряда серьезных проблем, в частности, проблему нахождения в Интернет актуальной новостной информации по необходимой пользователю тематике.

### 1.1. Спектр задач

Для решения задач автоматизированного сбора новостной информации из Интернет, ее обработки, систематизации, обобщения и обеспечения доступа к ней в Информационном центре «ЭЛВИСТИ» была разработана система InfoStream™ (Рис.1).

В настоящее время система InfoStream™ предоставляет пользователям широкий спектр сервисных возможностей по обеспечению доступа к новостным ресурсам Интернет. Ее использование позволяет:

- анализировать события, своевременно на них реагировать, оперативно получать необходимую информацию по мере ее появления в Интернет;
- формировать собственные информационные каналы, определяемые запросами на информационно-поисковом языке, формировать архивы для последующей обработки и ретроспективного анализа;
- проводить работы по анализу 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 PR-активности;
- оценивать возможные сферы влияния конфликтных или кризисных ситуаций, осуществлять информационный контроль вероятных источников рисков;
- находить потенциальных клиентов и партнеров.

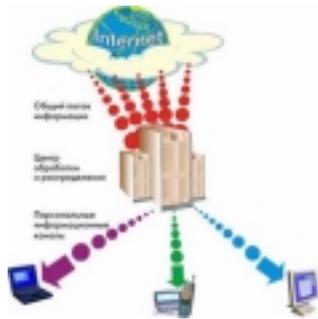


Рис.1.  
Общая схема функционирования системы InfoStream™

### 1.2. Технологические аспекты

Система 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.

В настоящее время система InfoStream™ охватывает мощнейший поток информации, превышающий 20 000 документов в сутки более чем с 500 Web-сайтов. Сервер системы InfoStream™ установлен на площадке ISP EIVisti, одного из ведущих провайдеров в Украине.

### 1.3. Виды сервисов

Существует несколько доступных пользователям вариантов использования системы InfoStream™ - сервисов. В рамках простейшего сервисного пакета пользователь может подписаться на ленту новостей по своей тематике, выраженной запросом, имеющим, например, такой вид: **банк&(защит~1/информаци)**. Данный запрос относится к защите информации в банковской сфере. Получение документов, соответствующих такому запросу, относится, скорее, к общему анализу отрасли, его субъектов и событий. Для анализа деятельности партнеров и конкурентов с помощью системы InfoStream™ можно подписаться на информационные ленты (определяемые запросами), сообщения которых включают названия соответствующих фирм, имена и фамилии др. Для работы пользователей, нуждающихся в постоянном изменении запросов, предусмотрен режим он-лайн, который идеально подходит для проведения постоянного контроля, например, за источниками рисков и конкурентами, 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 (информационными портретами) или технологией автоматического построения дайджестов, реализующих современный подход «глубинного анализа текстов» (Text Mining). Еще один очень важный аспект – это персонализация поискового интерфейса в режиме он-лайн – возможность сохранения запросов пользователями, организации подписки на них.

### 1.3.1. Он-лайн доступ к оперативным и ретроспективным базам данных

**InfoStream™ Online** – это сервисный пакет, обеспечивающий доступ к базам данных в режиме онлайн. Доступная пользователям база данных оперативной информации содержит документы, поступившие за последние 7 дней. Эта база данных обновляется в режиме реального времени. Кроме доступа к оперативной базе данных пользователи имеют возможность доступа к ретроспективной информации за последние три месяца и к базам данных, полученным из англоязычных источников. В рамках этого сервисного пакета предоставляется возможность поиска/просмотра записей, а также автоматического построения аналитических отчетов (дайджестов) по запросам пользователей.

### 1.3.2. Варианты рассылки новостной информации

Существуют три основных сервисных пакета, обеспечивающих подписку на результаты избирательно-го распределения информации по запросам пользователей:

- **MiniStream** - рассылка информации по одному запросу один или два раза в сутки. Пакет услуг MiniStream предусматривает доставку информации только по одному электронному адресу.
- **MidStream** - рассылка информации по одному запросу 1, 2 или 6 раз в сутки. Пакет услуг MidStream предусматривает доставку информации на один или два электронных адреса.
- **MainStream** - рассылка информации по одному запросу в режиме реального времени, круглосуточно, по мере появления информации в Интернет. По желанию пользователя, информация доставляется одновременно по трем электронным адресам или в виде заголовков и анонсов статей помещается непосредственно на сайт пользователя с помощью встраиваемого JavaScript-приложения.

### 1.3.3. Поток новостей на сайт

**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-ресурсов. Информация экспортируется с помощью встраиваемого JavaScript-приложения. Пакет услуг WebStream представляет собой идеальный инструмент для организации на Web-ресурсах пользователей колонок новостей профильной тематики, публикаций об отрасли, компаниях и т.п.

### 1.3.4. InfoStream™ Port



Рис. 2.  
InfoStream™ Port

InfoStream™ Port – это впервые созданное в Украине аппаратно-программное решение – реализация новостного поискового сервера, предназначенного для информационного обеспечения компаний разного уровня (Рис.2).

Информационное обеспечение InfoStream™ Port строится на основе использования информационного хранилища, формируемого на технической площадке 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.

Использование InfoStream™ Port обеспечивает:

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

## 2. Интерфейс пользователя системы InfoStream™

Пользовательские интерфейсы всех сервисов системы тесно связаны с ее поисковым ядром – полнотекстовой информационно-поисковой системой (ИПС) InfoReS.

### 2.1. Общие сведения о пользовательском интерфейсе

Интерфейс пользователя для работы с запросами системы рассмотрим на примере сервисного пакета InfoStream™ Online. Для его активизации необходимо при работе с браузером перейти по адресу <http://uaport.net/strm> и ввести логин и пароль. При этом отобразится главное окно системы – интерфейс работы с оперативной базой данных украинско- и русскоязычных источников, содержащей информацию, поступившую в систему за 7 дней (Рис.3).

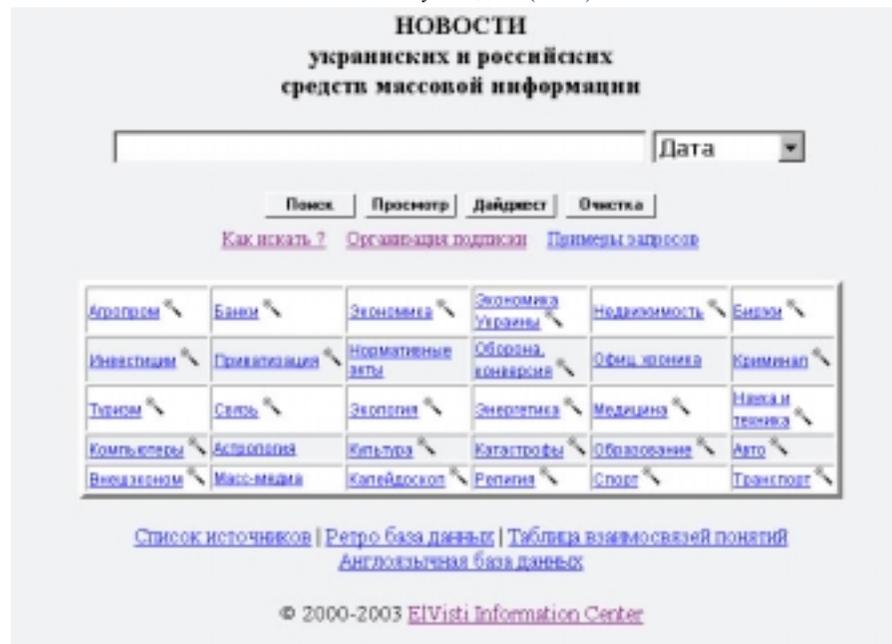


Рис.3. Главное окно сервисного пакета InfoStream™ Online

В этом окне представлено поле для ввода запросов на языке информационно-поисковой системы InfoReS. Рядом с полем ввода запросов находится окно для выбора диапазона дат. При отсутствии уточнения поиска по дате он производится по всему диапазону – за 7 дней. После ввода запроса (описание языка запросов приведено в разделе 2.3 этого пособия) для получения результатов поиска достаточно нажать на клавишу “Enter” на клавиатуре или активизировать кнопку «Поиск».

Результат поиска по запросу представляет собой список документов, распределенный по экранным страницам. Документы в списке результатов представлены заглавием, аннотацией (первым абзацем), датой и временем поступления, названием сайта-источника и гиперссылкой на документ-оригинал. Заглавие представляет собой гиперссылку на текстовый образ документа. Список результатов поиска отсортирован по дате и времени – на первых позициях представлены последние поступившие документы. На одной экранной странице представлено фиксированное количество ссылок – 15 (если общее число результатов поиска превышает это число). Если количество найденных документов превышает 10, то вывод списка документов сопровождается информационным портретом, представляющим собой таблицу с ключевыми словами, представленную в правой части экрана. Подробное описание работы с информационными портретами приведено в разделе 3.1.

При активизации гиперссылки, связанной с заглавием документа, отображается его текстовый образ, в котором цветом выделены слова, входящие в запрос. Текстовый образ документа также содержит название сайта-источника и гиперссылку на документ-оригинал.

Под окном ввода запроса находятся также кнопки «Просмотр», «Дайджест» и «Очистка». Нажатие кнопки «Просмотр» обеспечивает последовательный просмотр (в порядке, обратном поступлению) всего списка документов базы данных за 7 дней. Режим «Дайджест» обеспечивает автоматическое формирование дайджеста по введенному запросу. Подробное описание этого режима приведено в разделе 3.2. Активизация кнопки «Очистка» вызывает инициализацию формы ввода запросов.

В главном окне системы приведена таблица с темами – гиперссылками на информационные каналы, которые формируются в результате обработки запросов, приведенных в пункте 2.3.4.

В этом же окне содержатся гиперссылки, обеспечивающие переходы к режимам работы с

ретроспективной и англоязычной базами данных, а также к списку сайтов-источников информации.

## 2.2. Язык запросов InfoReS

Формирование запросов – это искусство, но искусство, доступное каждому. Запросы вводятся с использованием определенных правил, называемых в совокупности “языком запросов”.

В он-лайн режиме запрос вводится в область ввода текста 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.

## 2.3. Особенности составления запросов

### 2.3.1. Термы

Термы – это слова естественного языка или их правые усечения, состоящие как минимум из 2-х 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 не различает прописных и строчных букв, поэтому для поисковой процедуры запросы **завод]** и **Завод]** равнозначны.

### 2.3.2. Словосочетания

Словосочетания – это термы, состоящие из нескольких слов. Для реализации возможности поиска по словосочетаниям используется специальный оператор контекстной близости **ADJ** (возможно сокращенное написание - “~”).

Оператор контекстной близости обеспечивает отбор документов, в которые входят слова, связанные 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 расстояния между словами в запросе: /0/ - соседние слова (по умолчанию), /1/ - не более 1 слова в тексте документа между словами из запроса; /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.

### 2.3.3. 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** равносильно “!” или “^”;
- **AND** равносильно пробелу или “&” или “+”;
- **OR** равносильно “|” или “,” или “;”.

Например, запрос **банк&кредит&украин** равносильно 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 термы-словосочетания.

*Пример запроса:*

**((кофе | какао) & (молот | раствор)) ! зерн**

Такому запросу удовлетворяют документы, в которых присутствует термин **кофе** или **какао** (или оба вместе), а также либо термин **молот** (*молотый, молоть* и др.), либо термин **раствор** (*растворимый, растворимое* и др.), но при этом отсутствует термин **зерн** (*зерно, зерновое* и др.).

### 2.3.4. Источники, даты, рубрики

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

При поиске по источникам в качестве термов можно использовать правые части соответствующих доменных имен, например, [www.elvisti](http://www.elvisti.com), [www.lenta](http://www.lenta.com) или [4vlada.net](http://4vlada.net). Можно использовать и коды источников, перечень которых приведен в соответствующем разделе сайта системы.

Для поиска по датам в базах данных, доступных в режиме он-лайн как термы для поиска можно задавать даты в формате ГГГГ.ММ.ДД, например, **2003.06.12**. Допускаются также правые усечения дат, например, если указать в запросе **2003.06.0**, то будут выданы документы с 1 по 9 июня 2003 года.

В системе применяется двубуквенное 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 “&”.

В запросе также как обычные термы можно использовать коды рубрик. Например, запрос **rubr02&(нбу | (наибанк~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, соответствующие таким тематикам:

Код	Тематика	Запрос
Rubr01	Агропром	(агропром   сельск   сільськ   АПК   аграрн)& (коров   зерн   животнов   тварин   врожа   урожа   растений   рослиннов   птицевод   птахівн   рыб   риб)
Rubr02	Банки	((банковск~сфер)   (банківськ~сфер)   (банковск~сист)   (банківськ~сист)   (банковск~сект)   (банківськ~сект)   (банковск~струк)   (банківськ~струк)   (банковск~круг)   (банківськ~кол))& (банкрот   кредит   банк   банки   банке   банках   банков)
Rubr03	Мировая экономика	((макрэкон   минизкон   эконоом)& (стран   государств))& (ввл   (экспорт&импорт)   (индекс~/2/цен))
Rubr04	Экономика Украины	(украи   украї)& (економі   економи   макрэкон   макроекономі)& (бюджет   тариф   торгов   бизнес   кредит   инвест   інвест   банко   банкі)

		финанс   фінанс)& country.ua
Rubr05	Недвижимость	(недвижимост   нерухоміст)& (украин   украї   росс   СНГ)
Rubr06	Биржи	(бирж   бірж)&(межбанк   міжбанк   котиров)&(валют   акц)
Rubr07	Инвестиции	(инвестици   інвестиці)& (эконом   эконоом   бизнес)&(инвестор   инвестор)
Rubr08	Приватизация	(приватизац   приватизир   приватизов)& (инвест   інвест   экономик   эконоом   держав   государств)& (имущест   майн   сертификат   сертифікат)& (украи   росс   украї   киев   одес   львов   харьков)
Rubr09	Законодательство	((закон]~україні)&(цей~закон))   www.dina   www.nau.   www.rada.   search.liga   (постанова~кабінету~міністрів)   (постанова~верховн~ради)   (указ]~президента~україн)   (указ]~президента~від)
Rubr10	Оборона	(Миноборон   міноборон   (военн~учени)   (військ~навчанн)   (военн~действ)   (боев~действ))& (полигон   навчання   полігон   учения   учений   маневр   стратег   тактич)
Rubr11	Официальная хроника	(новини~дня)   (міжнародні~новини)   (новости~минувшего~дня)   ((новини~/2/україн)&www.dwell)   ((міжнародні~новини)& www.dwell)   (новини~плюс)   (коротко~україн)   (коротко~украина)   (коротко~планет)   (www.facts&панорама)   ((только~факты)&(рабочая~газет))   ((мир~сегодня)& (рабочая~газет))   ((без~лишних~слов)& сегодня)
Rubr12	Криминал	(росс   росі   киев   одес   украи   украї)& (криминальн   убийств   вбивств   краж   крадіж   контрабан   мафі   мафи   преступ   злочин)& (милиц   полиц   поліц   міліц   прокур   расслед   розслід)
Rubr13	Обзоры прессы	(огляд~преси)   (огляд~укр~преси)   (обзор~прессы)   (обзор~росс~прессы)   (обзор~зарубеж~прессы)   (обзор~мировой~печати)   (обзор~немецк~печати)   (пресс~дайджест)   (обзор~еженед~прессы)
Rubr14	Связь	(телекомм   телекому   укртелеком   госкомсвяз   держкомзв   (отрасл~связ)   (галуз~зв'яз))& (мобильн   (междунар~/1/канал)   интернет   провайд   интернет   internet   траф   тариф   домен   спутников   супутников   междугород)
Rubr15	Экология	(эколог   еколог   довкілля   (окружающ~/1/сред))& (защита   захист   гринп   green   охрана   отход   загрязнен   забруднен)
Rubr16	Энергетика	(энергети   енергети   електроенерг   електроенерг   (атомн~энергетик)   (ядерн~энергетик)   (теплов~электро))& (гідроел   гидроэл   (нетрадиц~энергет)   електростанц   електростанц   (топлив~комплекс)   (палив~комплек)   вугіл   уголь   нефт   нефт)
Rubr17	Здоровье	(медицин   здравоохранени   (охран&здоров)   (охорон&здоров)   мінздрав   минздрав   медицин)& (лікар   врач   лікувальний   лечебн   санитар   санітар   медицинский   медициний)& (пандем   болезн   заболев   захвор   эпидем   епідем)
Rubr18	Наука, техника	(научн   исслед)&

		(космос   вчені   експеримент   експеримент   відкрит   учены   разработк   открыти   монограф   диссертаци   науков   дослід)
Rubr19	Компьютеры	(монитор]   ноутб   компьют   комп'ютер)& ((информацион~технологи)   (информацій~технолог)   celeron   ibm)   amd]   shareware   Intel   Sun]   Pentium)
Rubr20	Астрология	(астролог   гороскоп   эзотер   эзотер)& (зодиак   зодіак   козерог   водолей   стрелец   скорпион   скорпіон)
Rubr21	Культура	(муз   кино   кіно   театр   картин)& (концерт   писат   письмен   режис   акт   митец   митц   худож)& (культур   искусств   мистец   фестивал)
Rubr22	Катастрофы	((крупн~авари)   (велик~аварі)   катастроф   (стихийн~бедств)   (стихий~лих)   ураган   паводок   засух   повнь   инцидент   крушен   пожар   пожерж)& (жертв   спасател   рятувальн   погилб   загину)
Rubr22	Катастрофы	гаи~уполномоч~заявит
Rubr23	Образование	((образов&школ)   (вуз&студент))&учеб
Rubr23	Образование	((освіт&школ)   (вуз&студент))&навча
Rubr24	Ukraine (eng)	Ukrain&(the   each   have)
Rubr25	Международные отношения	(Україн   Украин)& (МИД]   МЗС]   (иностр~дел)   (закордон~справ)   президент   премьер   прем'єр)& (визит   візит   (пров~переговор))
Rubr26	СМИ	(медиа   медіа   массмедиа   масмедіа   СМИ]   ЗМІ])& (газет   журнал   радио   радіо   телев   телеб)& редак
Rubr27	Калейдоскоп	(феномен   шарлотан   парадокс   сенсаци   апапона)& (екстрасен   комета   экстрасенс   НЛО]   внеземн   диноз   приключен   нетрадици   иллюз   мистич)
Rubr28	Религия	(религ   relig)& (христианс   православ   католиц   протестант   бапти   адвенти   евангел   евангел   будизм   будізм   іудаї   иудаи   мусульманс   сект)
Rubr29	Спорт	(олимпиад   олімпіад   тренировк   турнир   соревнован   змаган   тренуван)& (чемпіон   чемпион   спорт   футбол   бокс   естафет   эстафет   медал   шах   вело)
Rubr30	Туризм	((тур~агент)   путевк   путівк   экскурс   путешеств   экскурс)&(туризм   турист)
Rubr31	Транспорт	(морск   железнодорожн   авиа   морськ   залізнич   авіа)& (транспорт   авиатранспорт   авиатранспорт)& (перевозк   перевез   движени   рух   минтранс   мінтранс)
Rubr32	Автотранспорт	(автотранспорт   автозавод   автосалон   автопроизводств   (автомоб~транспорт))& (автоваз   Fiat   BMW   тойота   джип   опель   фольксва   мерседес   Jeep   Mitsubis   внедорожн   заз]   маз]   Chrysler   КАМАЗ   Toyota   VW   General   Rover   Жигули)

### 3. Дополнительные возможности

Дополнительные возможности включают средства интеллектуализации и персонализации поиска.

#### 3.1. Информационные портреты

Портрет можно рассматривать как модель реального объекта (или субъекта), выраженную его наиболее узнаваемыми чертами. В случае использования системы InfoStream™, информационный портрет, соответствующий запросу, – это множество ключевых слов, наиболее точно (по статистическим и смысловым алгоритмам) отражающее информацию, получаемую в результате поиска. Построение информационных портретов выполняется на основе эмпирических и частотно-статистических методов, основу которых составляет определение весов отдельных терминов в информационном канале.

С помощью информационного портрета (Рис.4.) запрос может детализироваться, уточняться. Информационный портрет реализован в виде таблицы в правой части экрана с результатами поиска системы. В случае, если в результате поиска найдено менее 10 документов, информационный портрет не формируется.

Для уточнения запроса по одному слову из информационного портрета достаточно активизировать гиперссылку, соответствующую этому слову, подведя к нему курсор и нажав на правую клавишу мыши. При этом уточнение будет производиться с учетом морфологии русского и украинского языков (из слова, входящего в из информационный портрет будет автоматически удалено окончание). Для уточнения запроса сразу несколькими словами из информационного портрета достаточно активизировать окошки для выбора (checkbox), находящиеся рядом с этими словами, а затем нажать на клавишу “Уточнить запрос”. При уточнении в этом режиме также будет учитываться морфология.

Рис.4.  
Фрагмент страницы с  
результатами поиска;  
Таблица справа -  
информационный портрет



### 3.2. Дайджесты

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

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

Электронный дайджест представляет собой документ, состоящий из заданного заранее количества фрагментов, который можно при необходимости распечатать. Вместе с тем, электронный дайджест можно также рассматривать как аннотированный источник гиперссылок на документы, лежащие в его основе.

### 3.3. Персонализация

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

#### 3.3.1. Организация доступа к информации

Пользователи могут получить доступ к данным в формате RSS с помощью специальных программ, называемых RSS-агрегаторами (Рис.5).

Для получения тематической ленты (фида) RSS от системы InfoStream™ в соответствующее поле RSS-агрегатора следует ввести адрес в формате:

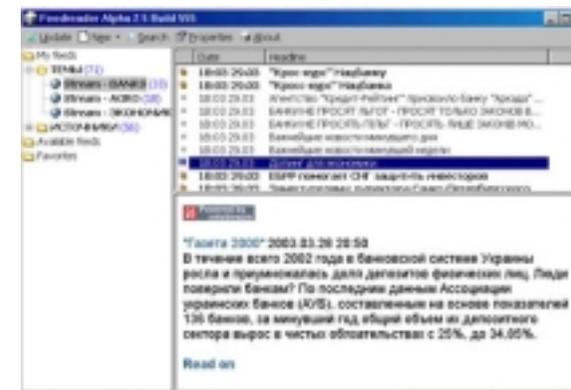
**http://uaport.net/cgi-bin/infostream.rss[?<ЗАПРОС>]**

где в качестве ЗАПРОСА можно ввести слово или словосочетание на языке запросов InfoReS.

В качестве RSS-агрегатора настоятельно рекомендуется использовать **FeedReader** версии **2.4**.

(дистрибутив приведен по адресу:  
**http://infostream.com.ua/prg/feedreader24.exe**).

Рис.5.  
Интерфейс RSS-агрегатора

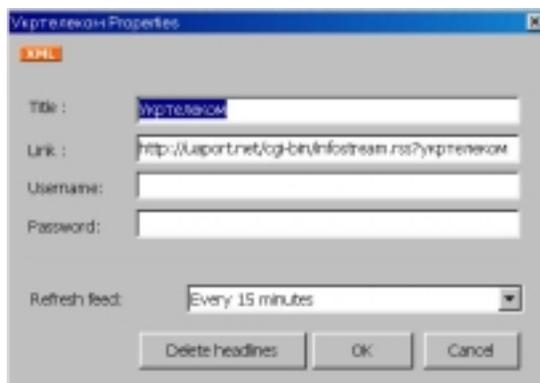


### 3.3.2. Настройка подписки

Для настройки подписки на тематическую ленту (фид) следует в основном окне системы InfoStream™ отладить запрос, после чего запустить программу FeedReader, активизировать опцию New (Рис.6) и ввести следующую информацию:

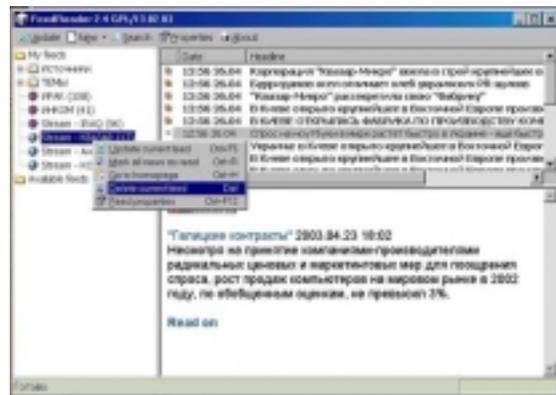
- адрес RSS-фида, включая запрос на информационно-поисковом языке системы InfoReS в формате, приведенном выше;
- название фида, которое может быть определено пользователем;
- периодичность обновления.

Рис.6.  
Окно подписки



Имеется возможность изменения кодировки, размеров шрифтов, помещения фида в отдельную папку, группировки фидов и т.д. (см. общий интерфейс программы).

Рис.7.  
Опции FeedReader



Для управления подпиской в этом режиме существуют дополнительные опции (Рис.7), активизируемые нажатием правой клавиши мыши при установке курсора на конкретном фиде:

- 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, заголовок и аннотация которого вызвали интерес, следует:

- произвести двойное нажатие левой клавиши мыши на заголовке или
- нажать на ссылку "Read on" в поле аннотации или
- нажать на соответствующую кнопку, стоящую перед заголовком или
- нажать правую клавишу мыши, находясь курсором на заголовке, при этом можно открыть текст сообщения в новом окне браузера или
- активизировать ссылку первоисточника и выйти через сеть Интернет на первоисточник.

## Заключение

InfoStream™ представляет собой систему интеграции новостных ресурсов сети Интернет, охватывающую в настоящее время практически все основные информационные Web-сайты Украины и России. Профессиональное использование возможностей системы InfoStream™, доступное пользователям ее сервисов, обеспечивает качественно новые возможности для информационно-аналитической работы в самых различных областях деятельности – от политики, макроэкономики, банковской деятельности – до управления персоналом или индустрии развлечений.

Сегодня системой InfoStream™ охватывается ежедневно свыше 20000 документов из более чем 500 информационных источников, перечень которых постоянно изменяется. Количество этих источников постоянно растет. Сведения о новых информационных источниках поступают как непосредственно от разработчика, так и от пользователей сервисов InfoStream™, в результате чего реализуется эффективный механизм обратной связи между службой сопровождения системы и пользователями.

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, использование математической логики, поэтапное уточнение критериев поиска и многие другие подходы и приемы будут способствовать повышению эффективности информационно-аналитической деятельности в любой области.

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  
E-mail: [stream@visti.net](mailto:stream@visti.net)

ДЛЯ ЗАМЕТОК