



Новости проекта



**InfoStream  
Online**

## Новые возможности системы мониторинга новостного контента сети Интернет

За время своего существования система InfoStream® обрела 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 охватывает свыше 25000 документов в сутки с более чем 700 Web-сайтов), так и с необходимостью не только находить документы,

но и проводить эффективный анализ результатов поиска.

Для решения этих задач, наряду с развитием информационной базы и поисковых возможностей, был создан новый интерфейс системы InfoStream, включающий средства персонализации и содержательного анализа результатов поиска.

В новом интерфейсе сервиса InfoStream Online у пользователя появились возможности выявления смыслового дублирования, поиска подобных документов, уточнения запросов с помощью информационных портретов различного наполнения, эффективной работы со списком источников, персонализация.

Расширен также ряд опциональных возможностей системы, в состав которых входят, например, возможности подключения морфемной обработки слов, задания требуемых объемов сообщений, уровня насыщенности документов цифровой информацией, языков, стран и др.



Пользователи мобильных телефонов имеют доступ к оперативной информации системы InfoStream через WAP-сайт

<http://wap.uaport.net>

# Новый интерфейс системы InfoStream



## Информационные портреты

В системе InfoStream информационный портрет, соответствующий запросу, - это множество ключевых слов, рубрик, источников, языков и т.п., максимально отражающее информацию, получаемую в результате поиска.

С помощью информационного портрета запрос может детализироваться и уточняться. В частности, в новом интерфейсе системы информационный портрет существенно облегчает выбор источников информации, соответствующих заданному запросу.

Информационный портрет	
Уточнить запрос	
Рубрики (32)	
Языки (3)	
Размер (3)	
Цифровая насыщенность (3)	
Слова (80)	
AND	NOT
<input type="checkbox"/> EVisti	<input type="checkbox"/>
<input type="checkbox"/> InfoStream	<input type="checkbox"/>
<input type="checkbox"/> InfoReS	<input type="checkbox"/>
<input type="checkbox"/> Online	<input type="checkbox"/>
<input type="checkbox"/> Text Mining	<input type="checkbox"/>
<input type="checkbox"/> UAmedia	<input type="checkbox"/>
<input type="checkbox"/> UAport	<input type="checkbox"/>
<input type="checkbox"/> Информационный	<input type="checkbox"/>
<input type="checkbox"/> Интерфейс	<input type="checkbox"/>
<input type="checkbox"/> Провайдер	<input type="checkbox"/>
<input type="checkbox"/> Контент	<input type="checkbox"/>
<input type="checkbox"/> Мониторинг	<input type="checkbox"/>
<input type="checkbox"/> Сервис	<input type="checkbox"/>
Источники (50)	
Страны источников (15)	
Уточнить запрос	



## Сюжеты

Функция «Сюжеты» обеспечивает семантическое ранжирование результатов поиска и позволяет ответить на вопросы:

- что нового?
- о чем больше всего говорят?

При построении сюжетных цепочек система определяет лингво-статистические характеристики отобранных в результате поиска документов и автоматически выявляет наиболее значимые темы, освещаемые в информационных потоках. 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.

Для того чтобы воспользоваться новой возможностью, любому пользователю сервиса InfoStream Online достаточно ввести поисковый запрос и нажать на клавишу "Сюжеты". При этом составление запроса максимально упрощается - для получения точных результатов вполне достаточно указать одно-два слова, относящихся к необходимой тематике.



## Динамика встречаемости понятий

Форма представления динамики встречаемости понятий - это гистограмма, которая строится как результат информационного поиска по множеству запросов. Эти запросы представляют собой комбинацию ключевых слов, соответствующих понятию, и дат, которые определяют необходимый диапазон времени. Каждая дата в гистограмме является гиперссылкой, ведущей к результату поиска по указанному критерию.



## Кабинет пользователя

Наряду с изначально существующими в системе возможностями создания персональных информационных каналов на основе технологии RSS/XML, новый интерфейс системы *InfoStream Online* позволяет сохранять, модифицировать и постоянно обрабатывать персональные запросы пользователей непосредственно в рабочем разделе - "Кабинет пользователя".



## Оptionальные возможности

### Режим «Морфология»

Режим «Морфология» обеспечивает предварительную обработку слов, входящих в поисковый запрос. В каждом слове отбрасывается изменяемое окончание, что обеспечивает охват системой не только слов, но и их словоформ.

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

### Режим «Убрать дубли»

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

### Поиск подобных документов

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

В новом интерфейсе пользователю сервиса *InfoStream Online* стала 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.

### Учет размера 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.



## Новые сервисные пакеты

В системе *InfoStream*<sup>®</sup> недавно внедрено два новых пакета услуг, доступных по электронной почте - *InfoStream Topics* и *InfoStream Rating*. Первый из этих пакетов позволяет получать логически структурированные отчеты по запросам в виде сюжетных цепочек, каждая из которых представлена наиболее характерной для сюжета новостью и набором дополнительных ссылок. Это позволяет пользователю избавиться от получения формально релевантных, но не соответствующих по смыслу документов, что зачастую присуще рассылке сообщений по запросам. *InfoStream Rating* обеспечивает получение по электронной почте рейтингов - результатов анализа публикаций по заданным компаниям, брендам, персоналиям и т.п., например, рейтингов политиков, банков или пивоваров.



Помощь



Кабинет



Источники

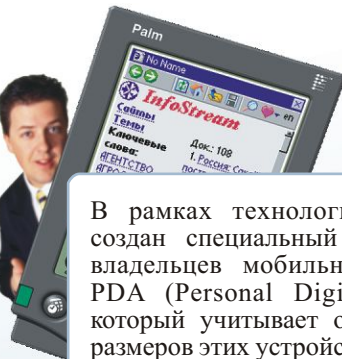


Статистика

## Присоединяйтесь!

С развитием информационных ресурсов Интернет вечная проблема поиска информации сегодня получила новое звучание: «поиск информации в неограниченной неоднородной динамической информационной среде». Или, другими словами, «поиск иголки в стог сена».

Способности людей генерировать, а также копировать информацию значительно превышают возможности ее анализировать. Это основная причина роста энтропии информационного пространства - информационного хаоса. Именно для решения этой проблемы, возможности нахождения важной информации, знаний, создаются современные системы мониторинга новостей, в частности, система InfoStream.



В рамках технологии InfoStream создан специальный веб-сайт для владельцев мобильных устройств PDA (Personal Digital Assistant), который учитывает ограниченность размеров этих устройств. С помощью PDA Palm или Pocket PC пользователи режима InfoStream Online получают простой, надежный и удобный доступ к новостям, сгруппированным по тематикам и источникам.

<http://pda.uaport.net>

Традиционные поисковые системы предлагают лишь частичное решение этой проблемы. Им присущи такие недостатки, как низкая оперативность, зависимость от спектра источников, слабые возможности ранжирования результатов поиска.

Система InfoStream решает проблему не только поиска необходимой информации, но и обобщения данных и их анализа.

Одним из самых перспективных направлений обработки информации в настоящее время является контент-мониторинг - непрерывный процесс анализа текстовых массивов. Именно непрерывная, конвейерная обработка информационных потоков является самой характерной чертой системы InfoStream, которая нашла широкое применение для поддержки принятия решений в таких областях, как государственное управление, анализ товарных рынков, реклама, маркетинг, поиск партнеров и клиентов, отслеживание деятельности конкурентов и др.