

## Системы интеграции Internet-контента

Internet представляет собой гигантское хранилище информации, объем которой удваивается каждый год. По экспертным оценкам, количество новостей только в украинском и российском сегментах Internet превышает 100 тыс. сообщений в сутки.

Очевидно, что такое разнообразие информации может быть полезным лишь при эффективном доступе к ней, что оказывается не просто осуществить на практике. Как уже упоминалось, по оценкам экспертов, около 79% журналистов обращаются к Internet в поисках новостей и лишь 20% из них находят необходимую информацию.

Попробуем проанализировать, почему традиционный поиск в Сети может оказаться неэффективным, и рассмотрим существующие уже решения, справляющиеся с подобной задачей гораздо лучше.

### 3.1. Статическая и динамическая составляющие Web-пространства

Как уже было сказано, все Internet-пространство можно с достаточной долей условности разделить на две составляющие — стабильную и динамическую. Стабильная составляющая Сети содержит информацию “долговременного” плана, в то время как динамическая включает постоянно обновляемые ресурсы. Некоторая часть динамической составляющей со временем вливается в стабильную, однако большая ее часть “исчезает” из Сети или попадает в сегмент “скрытого” Web-пространства, не доступного пользователям с помощью информационно-поисковых систем.

В традиционной сетевой поисковой системе информационное пространство, состоящее из стабильной и новостной частей и индексируемое этой ИПС, меняет свое содержимое через  $N$  дней: некоторые новостные документы уходят в стабильную часть в виде архивов, а остальные исчезают.

В этом случае пользователь при обращении к ИПС получает соответствующие запросу ответы из стабильной части, устаревшие ссылки из новостной части и ничего из обновленной новостной части.

Пользователь часто часами проводит время в Сети, посещая сотни сайтов с целью получения данных по определенной тематике. Ведь ни одна из традиционных поисковых систем в достаточном объеме не помогает в поиске актуальной новостной информации, находящейся в динамической части Сети.

Решение этой задачи требует создания своеобразного интеллектуального посредника между пользователем и Internet. Подобный посредник (или агент новостей) должен выполнять всю “черновую” работу по сбору и селекции информации и обеспечивать предпосылки для создания документальной базы данных.

Принцип индексирования, используемый посредником, несколько отличается от аналогичного принципа традиционных поисковых систем: индексируется не все пространство Internet, а только его новостная часть.

При этом, за счет относительно небольшого объема этих данных, частота индексирования выбирается достаточно малой — от нескольких минут до нескольких часов (в зависимости от источника).

В результате через  $N$  дней обрисовывается такая ситуация: пользователь получает необходимые ответы по новостной и “устаревшей” новостной части, подтвержденные документами из собственной архивной базы данных, но не получает полной выборки документов из стабильной части информационного наполнения Сети.

Таким образом, проблема получения полной информации из Сети в идеале может быть решена путем использования двух инструментов — традиционных ИПС и системы агентов новостей.

## 3.2. Недостатки традиционного поиска

Владельцы Web-сайтов уже давно осознали, что новостная информация привлекает посетителей, а потому количество источников такого рода информации в Сети постоянно возрастает, что само по себе усложняет задачу поиска необходимых данных.

Можно сказать, что извечная проблема поиска информации сегодня получила новое звучание: “поиск информации в неограниченной, неоднородной динамической информационной среде”, или, если перефразировать, — “поиск иголки в стог сена”.

Традиционные поисковые системы предлагают лишь частичное решение этой проблемы. Периоды индексации у них составляют от недель до нескольких месяцев. И несмотря на то, что практически все известные поисковые порталы (Google, Yahoo!, AltaVista, Lycos и другие) имеют новостные разделы, они, сами по себе, уже многих не устраивают. Традиционным подходам к организации поиска сетевой информации присущи такие недостатки, как низкая оперативность, зависимость от набора источников и ограниченность спектра этих источников, средние поисковые возможности, отсутствие средств уведомления о появлении новых данных.

Одна из проблем нахождения информации в Сети обусловлена основным форматом, в котором представлена эта информация, — HTML. Этот формат был разработан, в первую очередь, для решения задач отображения содержания на каждом конкретном Web-ресурсе, поэтому он не всегда удобен для автоматической обработки информации, в том числе и организации поиска. В результате информация в Internet оказалась ориентирована, прежде всего, на отдельные сайты и очень слабо приспособлена для автоматизированного обобщения, классификации и аналитической обработки.

При импортировании в Web-ресурс информации с другого сайта (включении новостных сообщений и т.п.) возникает вопрос однотипного представления их содержания (контента). Если этот вопрос не решается, то изменение HTML-оформления сайта-источника приводит к необходимости одновременной модификации программного обеспечения на всех сайтах, которые принимают от него информацию.

Итак, объективно назрела необходимость использования некоего унифицированного формата представления данных. Сегодня в качестве такового все чаще используется XML или его подмножество — RSS (об этих форматах речь пойдет ниже). XML представляет собой метаязык, т.е. язык, на базе которого можно определять новые языки. Он предназначен не только для организации обмена данными в Web, но и для распознавания их семантики. В отличие от HTML, XML предназначен для представления информации в “чистом” виде, предполагая структурную, а не оформительскую разметку данных.

### 3.3. Невизуальный Web

Сегодня Internet — это огромное хранилище информации, интегрированный доступ к динамической составляющей которого — новостным ресурсам — затруднен. Разнообразие информации, в том числе и новостных сообщений, в Сети не может быть полезным на практике при отсутствии эффективного доступа.

Поэтому, как было сказано выше, возникла необходимость в использовании унифицированного формата данных на сайтах, стандарта, обеспечивающего однотипный обмен данными в Internet. В качестве такого унифицированного формата все шире используется язык eXtensible Markup Language (XML) и его диалекты.

Одним из первых проектов унификации обмена данными в Internet стал “Семантический Web” (Semantic Web) [40, 59]. Основная идея проекта, задуманного в консорциуме W3C Тимом Бернерсом-Ли и его коллегами (рис. 3.1), за ключалась

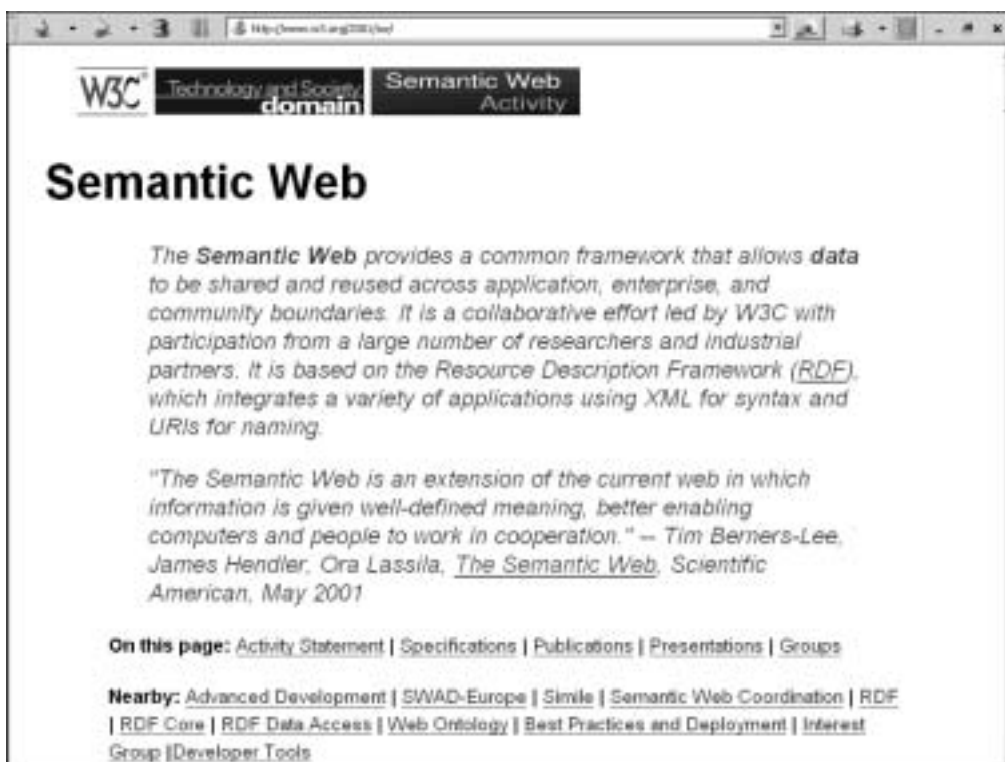


Рис. 3.1. Проект консорциума W3C “Семантический Web”

в такой организации данных: Web-серверы должны не только визуализировать, но и использовать их, чтобы программы разных производителей могли эффективно работать с Web-контентом. Именно для “Семантического Web” были разработаны спецификации XML, предусматривающие разделение средств визуализации и смыслового содержания.

В основу “Семантического Web” были положены три ключевых элемента:

- спецификация XML, позволяющая определить синтаксис и структуру документов;
- механизм описания ресурсов — Resource Definition Framework (RDF), обеспечивающий модель кодирования для значений, определенных в онтологии;
- система онтологий, позволяющая определять термины (или понятия) и отношения между ними.

“Семантический Web” также использует другие технологии и концепции, в частности универсальные идентификаторы ресурсов, цифровые подписи, системы логического вывода, обычные протоколы Internet и т.д.

XML представляет собой метаязык, т.е. язык, на базе которого можно определять новые языки. Но он предназначен не только для организации обмена данными в Web, но и для распознавания семантики этих данных. В отличие от HTML, XML обеспечивает представление информации в чистом виде, предполагая ее структурную, а не оформительскую разметку. При этом потребовались стандарты не только для синтаксической формы документов, но и для их семантического наполнения. В результате консорциумом W3C были разработаны стандарты языков XML и RDF, которые совместно позволяют поддерживать семантическую совместимость в Сети.

Вместе с тем, формально элементы разметки (теги) XML оторваны от определения их смыслового наполнения. Поэтому параллельно с XML была начата разработка стандарта для схемы описания источников RDF — языка формального описания содержимого Web-сайтов в рамках единого стандарта.

RDF является языком общего назначения для описания информации в Web. RDF-документы представляют собой совокупность RDF-предложений, состоящих из троек элементов: *ресурс — именованное свойство — значение свойства* (или *объект — атрибут — значение атрибута*). Ресурсом может выступать понятие, которому можно приписать некоторый URI (Uniform Resource Identifiers). Значение свойства или атрибута — это его контент, т.е. содержимое.

Спецификации RDF обеспечивают поддержку тегов, позволяющих определять любые понятия (например, тегами PRICE и INVOICE можно пользоваться для обозначения цены и счета соответственно). Следует заметить, что данным в формате RDF присваиваются дескрипторы, которые могут определяться в отдельных файлах определения типов документов (Document Type Definitions — DTD). Сегодня практически в каждой отрасли знаний имеется свой, постоянно расширяющийся список DTD. На основе XML и RDF был создан формат RSS, специально предназначенный для организации информационных коммуникаций как между людьми, так и между серверами [57].

Предполагается, что третий элемент “Семантического Web” — онтологии — будет играть определяющую роль в обработке знаний в Сети, а также в их совместном использовании приложениями. При этом онтология определяется как

система, состоящая из набора понятий и набора утверждений этих понятиях, на основе которых можно строить классы, объекты и отношения. Онтология определяет семантику конкретной области и способствует установлению связей между значениями элементов предметной области.

В рамках “Семантического Web” предлагается в среде RDF описывать онтологии на языке RDF.

### **3.4. Синдикация новостной информации**

Оптимальное решение, способное помочь ориентироваться в новостной информации Internet, сегодня предоставляют информационные службы нового типа — системы синдикации новостей. Под синдикацией в данном случае понимается сбор информации в Internet и последующее распространение ее фрагментов в соответствии с потребностями пользователей [16, 25]. Кроме того, службы синдикации обеспечивают публикацию одних и тех же данных на различных сайтах (в том числе предназначенных для карманных компьютеров и мобильных телефонов).

Службы синдикации обеспечивают публикацию соответствующих потребностям пользователей данных на различных Web-страницах, сайтах и порталах (в том числе на Internet-ресурсах для карманных компьютеров и мобильных телефонов), а также доставку информации пользователям.

Технология синдикации Internet-новостей включает в себя “обучение” программ сбора структуре выбранных источников (Web-сайтов), непосредственное сканирование информации, приведение ее к общему формату (в последнее время — к XML), а также ее классификацию и доставку пользователям различными путями (e-mail, Web, WAP, SMS и т.д.).

### **3.5. От “поисковиков” — к “интеграторам”**

Необходимость сетевой интеграции новостей несколько лет назад осознали известные сетевые поисковые службы — Google, Excite, Lycos, AltaVista. На первых этапах они заключили соглашения с крупнейшими информационными агентствами, такими как Reuters, Associated Press, CNN и другие, и стали предоставлять возможности поиска и просмотра новостных сообщений. Таким образом у пользователя впервые появилась возможность бесплатно находить и просматривать новости реального (а не только “виртуального”) мира в Сети. Старейший навигационный портал Yahoo! также не обошел стороной идею интеграции новостей, создав службу Daily News (<http://dailynews.yahoo.com>), объединяющую информацию нескольких десятков информационных агентств и обеспечивающую графическое и мультимедийное представление отдельных тематических областей.

Практически одновременно с освоением традиционными СМИ виртуального пространства Internet и с настоящей экспансией он-лайнных изданий стали возникать службы, обеспечивающие обобщенное представление информации со страниц сетевых СМИ на своих сайтах, а также “проталкивание” (push-технология) информации, якобы интересующей пользователей, в рабочие области их браузеров. Получила развитие технология “персональных информационных агентов”, обеспечивающих клиентскую часть систем интеграции новостей.

Идея интеграции новостей породила несколько технологий, имеющих общие корни. Один из мировых лидеров на рынке интеграции Internet-контента,

компания MoreOver (<http://www.moreover.com>) опубликовала свою технологическую схему интеграции контента, которая в настоящее время “де-факто” признана типовой (рис. 3.2).

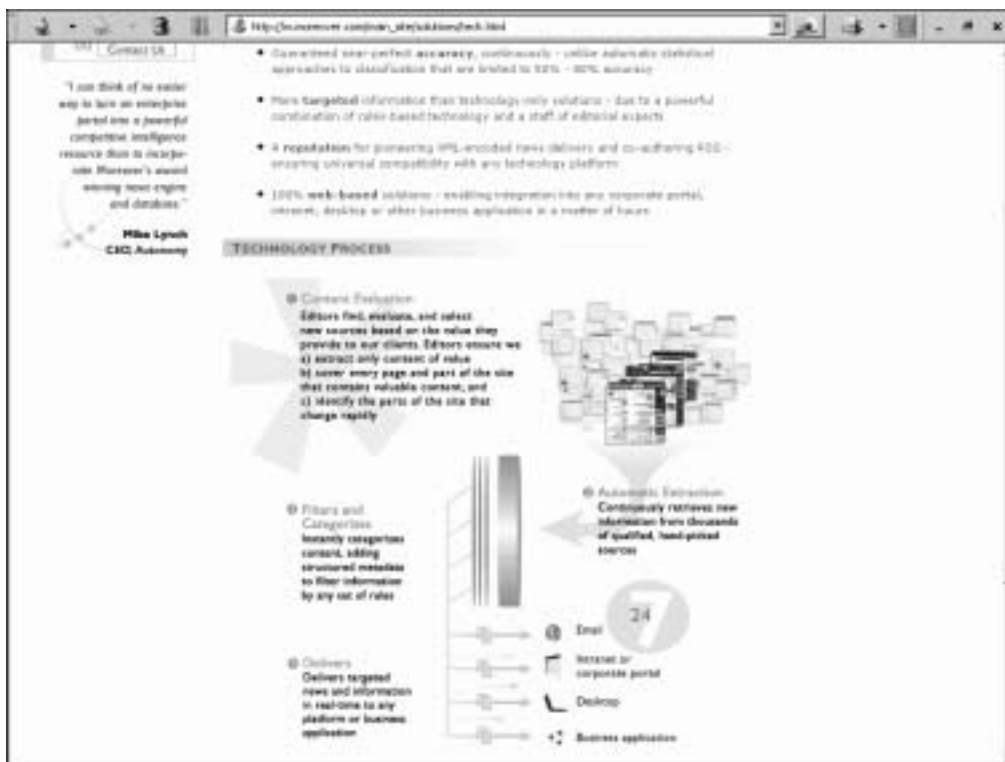


Рис. 3.2. Технология интеграции новостей компании MoreOver

На первом уровне типовой системы интеграции контента выполняется оценка источников информации: специалисты-эксперты находят необходимые сайты-источники, анализируют их содержание, форматы представления данных, частоту обновления результатов. Конечный результат их работы — шаблоны (или макроописания) информационных ресурсов.

На втором уровне осуществляется автоматизированный сбор обновлений с целевых Web-сайтов — непрерывное сканирование материалов из тысяч источников с помощью программ-роботов.

На третьем уровне обрабатывается контент, что выполняется в несколько этапов.

1. Приведение собранной информации к выбранному внутрисистемному формату (как правило, это подмножество XML).
2. Систематизация информации, добавление к ней метаданных для обеспечения последующей фильтрации.
3. Категоризация.
4. На последнем этапе осуществляется доставка релевантных информационных материалов пользователям путем:

- рассылки результатов избирательного распределения сообщений по электронной почте;
- публикации на определенных страницах Web-сайтов (как открытых, так и закрытых, корпоративных);
- загрузки в intranet-сети (информационные хранилища или корпоративные порталы);
- подачи информации на вход разнообразным бизнес-приложениям.

### 3.6. Форматы синдикации новостей

Для решения задачи синдикации новостей было создано несколько форматов описания данных на основе XML. Самый распространенный формат получил название RSS [29, 30], что означает Really Simple Syndication, Rich Site Summary, хотя изначально он назывался RDF Site Summary. Смысл всех этих аббревиатур заключается в простом способе обобщения и распределения информационного наполнения Web-сайтов, т.е. синдикации контента.

Изначально формат RSS создавался компанией Netscape для портала Netcenter как одно из первых XML-приложений, но затем он стал использоваться на многих других сайтах. Сегодня практически все ведущие новостные сайты, — “Живые журналы”, работающие в Internet, — используют RSS в качестве инструмента оперативного представления своих обновлений. Например, сегодня экспорт в RSS осуществляют крупнейшие порталы, включая CNN, BBC News, Amazon, CNet News, MSNBC, The Register, Wired и т.д.

RSS действительно обеспечивает согласованный способ резюмировать содержимое Web-сайтов. Кроме того, его применение позволило администраторам новостных сайтов, он-лайн-дневников-блогов, форумов и других часто обновляемых Web-ресурсов представить информацию в унифицированном виде.

По признанию многих экспертов, 2004 год стал “Годом RSS”, т.е. в этом году началось широкое внедрение этого формата. При этом только в начале 2004 года Internet-пользователи по-настоящему открыли для себя все прелести технологии RSS. Сегодня для работы с новостями в формате RSS разрабатываются все новые программы, сайты и поисковые системы, которые все более востребованы, в частности, пользователями карманных компьютеров.

Итак, RSS — это формат данных и технический стандарт, который обеспечивает интегрированный доступ к новостной информации, представленной на Web-сайтах, и специально создан для обмена их контентом.

Развитие RSS началось с версии 0.90, разработанной компанией Netscape, но его посчитали слишком сложным, и Netscape разработала упрощенную версию — 0.91, которую, после бума порталных технологий, передала компании UserLand Software. Это самый простой и доступный стандарт, который применяется сегодня, когда требуется несложный экспорт заголовков. Одновременно еще одна организация — RSS-DEV Working Group — создала свою версию RSS (1.0), близкую к исходной версии RSS 0.90 и максимально приближенную к стандарту RDF; эта версия предоставляет больше возможностей, чем все 0.9x, — например, допускает расширение при помощи модулей. Компания же UserLand решила развить ветвь 0.9x и создала версии 0.92, 0.93, 0.94, которые позволяют представлять метаданные, и, наконец, 2.0. При этом RSS 2.0 — это не новая версия RSS 1.0, а логическое продолжение ветви 0.9x. В ней также добавлена поддержка модулей. В настоящее

время существует 7 независимых версий RSS — 0.90, 0.91, 0.92, 0.93, 0.94, 1.0, 2.0. Эти версии отличаются друг от друга, хотя все они ориентированы на один тип информации и содержат одинаковые базовые поля. При этом многие считают все версии, кроме 2.0, устаревшими и “отмененными”, но это далеко не так. Пока еще самой популярной является RSS 0.91. Что же касается версии 0.94, то ее спецификаций не сохранилось даже на авторском сайте UserLand. Так, по адресу <http://backend.userland.com/rss094> находится спецификация версии RSS 2.0.

Спецификации отдельных версий формата RSS приведены на таких Web-страницах:

- RSS 0.90: <http://www.purplepages.ie/RSS/netscape/rss0.90.html>
- RSS 0.91: <http://my.netscape.com/publish/formats/rss-spec-0.91.html>
- RSS 0.92: <http://backend.userland.com/rss092>
- RSS 0.93: <http://backend.userland.com/rss093>
- RSS 1.0: <http://web.resource.org/rss/1.0>
- RSS 2.0: <http://backend.userland.com/rss>

Во всех версиях RSS есть некоторые особенности, но, как уже говорилось, объединяет их ориентация на один тип информации, вследствие чего они содержат общие базовые поля: основной блок данных (channel), который содержит такие атрибуты, как “заглавие канала” (title), “ссылки” (link), “данные о языке сообщений” (language) и “логотип” (image), после них идет список самих сообщений, где в каждом пункте (item) указывается заголовок (title), краткое описание (description) и ссылка на новость (link). Кроме того, каждый RSS-файл начинается обязательными элементами xml и rss. Первый из этих элементов содержит атрибуты version (версия) и encoding (кодировка).

Среди множества необязательных элементов RSS можно назвать самые распространенные — язык (language), авторское право (copyright), категория информации (category), дата и время публикации сообщения (pubDate), программа, которая использовалась для создания файла (generator), картинка, которую следует показывать наряду с текстовой информацией (image).

Кроме заголовка блока данных, в формате RSS предусмотрено описание отдельных информационных элементов (item). Каждый элемент — это отдельная статья или краткая аннотация и ссылка на полную версию статьи. Канал (channel) может содержать любое число элементов, содержащих только два обязательных вложенных элемента — название (title) и описание (description). Кроме того, часто используются такие вложенные элементы: ссылка на первоисточник (link), категория (category), комментарий (comments) и автор (author).

В качестве примера новостного канала формата RSS 0.91 можно привести динамический файл, формируемый по адресу <http://uaport.net/cgi-bin/infostream.rss> (обзор основных событий дня “Электронні Вісті”). Он имеет такой вид:

```
<?xml version="1.0" encoding="windows-1251" ?>
<!DOCTYPE rss PUBLIC "-//Netscape Communications//DTD RSS 0.91//EN"
"http://my.netscape.com/publish/formats/rss-0.91.dtd">
<rss version="0.91">
<channel>
```



```

<title>Електронні Вісті</title>
<language>ru</language><image>
<title>Електронні Вісті</title>
<url>http://www.elvisti.com/images/export/elvisticom3_88x31.gif</url>
<link>http://www.elvisti.com</link>
<width>88</width>
<height>31</height>
</image>

<item><title>РАДАР СЛЕДИТ ЗА КОСМИЧЕСКИМ МУСОРОМ</title>
<description>В японской префектуре Окаяма с 6 апреля начал работать радар
с дистанционным управлением, основная функция которого состоит в
отслеживании перемещения космического мусора.<cription>
<link>http://elvisti.com/2004/04/06/sci-tech.shtml#3</link>
</item>

<item><title>В ИВАНО-ФРАНКОВСКОЙ ОБЛАСТИ КУРИЦА СНЕСЛА ЯЙЦО ВЕСОМ 143 Г</title>
<description>В селе Делиев Галицкого района Ивано-Франковской области
курица снесла яйцо весом 143 г. </description>
<link>http://elvisti.com/2004/04/06/misc.shtml</link>
</item>

<item><title>В США БОЛЕЕ 60% КОРПОРАЦИЙ В 1990-Е ГОДЫ НЕ ПЛАТИЛИ НАЛОГИ</title>
<description>Более 60% американских корпораций в период бума
американской экономики с 1996 по 2000 годы не платили налоги
в государственную казну, сообщило Главное бюджетно-контрольное
управление США.</description>
<link>http://elvisti.com/2004/04/06/biz.shtml#2</link>
</item>

<item><title>СЕДЬМОЕ АПРЕЛЯ - ВСЕМИРНЫЙ ДЕНЬ ЗДОРОВЬЯ</title>
<description>В нынешнем году по рекомендации ВОЗ этот день пройдет под
лозунгом "Безопасность на дорогах зависит от каждого из нас".<cription>
<link>http://elvisti.com/2004/04/06/health.shtml#2</link>
</item>

</channel>
</rss>

```

Помимо формата RSS, недавно появился формат Atom (<http://www.mnot.net/drafts/draft-nottingham-atom-format-02.html>), пока окончательно не утвержденный, но используемый на крупнейшем поисковом портале Google, что предопределяет его популярность. Открытый стандарт Atom (актуальная версия — 3.0) совершенствуется командой программистов из IBM, Google и других компаний.

По информации журнала New Scientist, сегодня компания Google снова рассматривает возможность использования в некоторых своих сервисах формата RSS, хотя еще в феврале 2004 года сообщила, что для генерации и доставки срочных сообщений подписчикам он-лайнового журнала Blogger Google будет использовать формат Atom, разработанный в 2003 году компанией IBM.

И Atom, и RSS имеют свои преимущества. Atom обеспечивает подписчикам большую гибкость в выборе профиля, поскольку поддерживает больше метаданных, и является открытым стандартом, развитие которого далеко от завершения.

Он позволяет рассылать не только сами сообщения, но и комментарии читателей, что формату RSS пока не под силу.

Как и RSS, Atom является подмножеством XML. Приведем пример файла в этом формате, чтобы подчеркнуть его близость с RSS:

```
<?xml version="1.0" encoding="utf-8"?>
<feed version="0.3" xmlns="http://purl.org/atom/ns#">
  <title>Наименьший возможный фид в формате Atom 3.0</title>
  <link rel="alternate" type="text/html"
href="http://diveintomark.org/" />
  <modified>2004-04-09T18:30:02Z</modified>
  <author>
    <name>Иванов Петр</name>
  </author>
  <entry>
    <title>Atom 0.3 пример</title>
    <link rel="alternate" type="text/html"
href="http://uaport.ua/2004/04/09/atom03" />
    <id>tag:uaport.ua,2004:4.2397</id>
    <issued>2004-04-09T08:29:29-04:00</issued>
    <modified>2004-04-09T18:30:02Z</modified>
  </entry>
</feed>
```

Дэйв Уинер (Dave Winer), один из главных разработчиков RSS, недавно призвал всех разработчиков объединить свои усилия и разработать единый формат, совместимый как с RSS, так и с Atom, чтобы слить конкурентные стандарты в единое целое. “Новый формат можно назвать RSS/Atom, — заявил Уинер. — Он бы имел всю функциональность, которую разработчики Atom обещают внедрить. Максимально авторитетный формат получил бы наиболее полную поддержку от всех разработчиков.” Уинер предлагает, чтобы в RSS/Atom было как можно меньше отличий от RSS 2.0.

### 3.7. OPML — формат для хранения списка RSS-фидов

Еще один диалект XML — язык OPML (Outline Processor Markup Language) — используется для описания совокупности RSS-фидов. Его спецификация размещена по адресу <http://opml.scripting.com/spec>. С помощью OPML обеспечивается эффективный унифицированный обмен списками RSS-фидов.

Для того чтобы обеспечить удобную подписку с помощью RSS-агрегаторов сразу на несколько фидов, разработан специальный механизм, также ориентированный на формат XML. Список RSS-фидов заносится в файлы специальных форматов, имя и путь (адрес в сети Internet) к которым указывается в программах-агрегаторах при подписке. Применение этих файлов возможно в большинстве современных агрегаторов и не зависит от операционных систем, в которых эти агрегаторы работают.

В настоящее время для создания списка RSS-фидов применяется два основных, базирующихся на XML, открытых формата: OCS (Open Content Syndication) и OPML (Outline Processor Markup Language). Наиболее распространенным является формат OPML, спецификация которого приведена по адресу <http://www.opml.org/spec>. По этому же адресу можно подписаться на рассылку бюллетеня, в котором публикуются новости, относящиеся к спецификации и применениям OPML.

Рассмотрим, для примера, файл, описывающий RSS-каналы системы InfoStream® на портале UAport (<http://uaport.net/feeds.opml>).

```
<?xml version="1.0" encoding="windows-1251" ?>
<opml version="1.0">
<head>
  <title>InfoStream News</title>
  <ownerName>InfoStream.ua</ownerName>
  <ownerEmail>stream@visti.net</ownerEmail>
</head>

<body>
<outline text="АПК"
  htmlUrl="http://uaport.net/UAnews?rub=01"
  language="ru" title="http://uaport.net/UAnews?rub=01"
  type="rss"
  xmlUrl="http://uaport.net/cgi-bin/infostream.rss?rubr01" />
<outline text="Банковская сфера"
  htmlUrl="http://uaport.net/UAnews?rub=02"
  language="ru" title="http://uaport.net/UAnews?rub=02"
  type="rss"
  xmlUrl="http://uaport.net/cgi-bin/infostream.rss?rubr02" />
<outline text="Экономика"
  htmlUrl="http://uaport.net/UAnews?rub=03"
  language="ru" title="http://uaport.net/UAnews?rub=03"
  type="rss"
  xmlUrl="http://uaport.net/cgi-bin/infostream.rss?rubr03" />
<outline text="Недвижимость"
  htmlUrl="http://uaport.net/UAnews?rub=05"
  language="ru" title="http://uaport.net/UAnews?rub=05"
  type="rss"
  xmlUrl="http://uaport.net/cgi-bin/infostream.rss?rubr05" />

</body>
</opml>
```

Первая строка OPML-файла несет информацию о том, что этот файл размечен на подмножестве формата XML версии 1.0, а также о том, что в нем используется кодировка Windows 1251. Вторая строка содержит обязательный атрибут `version`. Данный OPML-файл состоит из двух основных частей: заголовка (`head`) и тела (`body`). Заголовок может содержать следующие основные теги, описывающие данный OPML-документ:

- `<title>` — заголовок документа, представляющего собой список RSS-фидов;
- `<dateCreated>` — дата создания документа в формате RFC 822 (с указанием дня недели и часового пояса);
- `<dateModified>` — дата модификации документа;
- `<ownerName>` — владелец документа;
- `<ownerEmail>` — электронный почтовый адрес владельца.

Остальные элементы, упомянутые в спецификации, имеют второстепенное значение.

Тело OPML-документа теоретически состоит из неограниченного количества тегов. Также теоретически допускается использование тегов других типов, однако они чаще всего не поддерживаются популярными программами-агрегаторами. Каждый тег может содержать следующие основные атрибуты:

- `type` — тип элемента; имеет значение `rss` для фидов;
- `title` — название RSS-фида, которое соответствует тегу `<title>` элемента `<channel>` в формате RSS;
- `description` — краткое описание фида, которое соответствует тегу `<description>` элемента `<channel>` в формате RSS;
- `language` — язык документа;
- `xmlUrl` — гиперссылка на фид в виде RSS;
- `htmlUrl` — гиперссылка на HTML-страницу данного фида, которая соответствует тегу `<link>` в `<channel>` для RSS.

### 3.8. Источники новостного контента

Основным применением RSS в настоящее время являются новостные фиды (feed). Фид — это файл в формате RSS, в который записывается новостной контент Web-ресурса. Если есть необходимость оперативно отслеживать изменения на содержащем фид сайте, то можно делать это с помощью программы-агрегатора, не посещая самого сайта с помощью стандартных программ-браузеров.

Ниже приведены адреса самых популярных в Internet RSS-фидов:

```
http://w.moreover.com/categories/ocs/ocsdirectory.rdf
http://10.am/extra/ocsdirectory.php
http://www.newsisfree.com/ocs/directory.xml
http://blogspace.com/rss/feeds/converted.ocs
http://www.groksoup.com/ocs/ocsdirectory.xml
http://theweb.startshere.net/channels.phtml?format=OCS
http://myrss.com/catalog/ocs04.rdf
http://www.syndic8.com/xml.php
```

В настоящее время в русскоязычной части Internet представлены тысячи RSS-фидов, наиболее популярные из которых такие:

- NEWSru.com — <http://www.newsru.com/plain/rss/all.xml>
- Газета.ru - Все новости (RSS) — [www.gazeta.ru/export/gazeta\\_rss.xml](http://www.gazeta.ru/export/gazeta_rss.xml)
- Lenty.RU — <http://www.lenty.ru/export/bestnews.rss>
- Подробности — <http://www.podrobnosti.com.ua/export>
- Lenta.ru — <http://lenta.ru/l/r/EX/import.rss>
- Полит.РУ — <http://www.polit.ru/rss/index.xml>
- Портал “Юридическая Россия” — <http://law.edu.ru/rss/news.rss>
- Водка он-лайн — <http://vodka.com.ua/export/rss.xml>
- Портал “ПлейМобайл” — <http://playmobile.ru/news/rss>
- 3Dnews — <http://www.3dnews.ru/expnews/rss/newsrss.xml>

Обширный список RSS-фидов русскоязычного сегмента Internet находится по адресу <http://my.yandex.ru/rss.opml>; приведем лишь некоторые, наиболее интересные новостные фиды:

- Аргументы и Факты — <http://www.aif.ru/info/rss.php?magazine=aif>
- АвтоОБЗОР — <http://auto.obzor.ru/news/autonews.xml>
- АвиаПорт.Ру — [http://www.aviaport.ru/news/yandex\\_export.xml](http://www.aviaport.ru/news/yandex_export.xml)
- Деловая Хроника — <http://www.chronicle.ru/l/r/EX/rsschannel.xml>
- К2Kapital — <http://ad.k2kapital.com/cbp/mynetscape/mynews.news>
- Linux.org.ru — <http://images.linux.org.ru/getrss.php3>
- PalmQ Online — <http://www.palmq.net/backend.php>
- СПОРТ сегодня — [http://www.sports.ru/sports\\_docs.xml](http://www.sports.ru/sports_docs.xml)
- TRAVEL.RU. Все о путешествиях — <http://www.travel.ru/inc/side/yandex.rdf>
- АПК-Информ — <http://www.apk-inform.com/yandexr.php>
- ФОНТАНКА.РУ —  
[http://www.fontanka.ru/\\_transmission\\_for\\_yandex.shtml](http://www.fontanka.ru/_transmission_for_yandex.shtml)
- IMA Press. Тема дня —  
<http://www.ima-press.ru/rss.php?newsblock=theme&limit=1>
- Журнал “Итоги” — <http://www.itogi.ru/WebExport.nsf/Anons/itogi.xml>
- Остров. Новости Донбасса — <http://www.ostro.org/yandex.php>
- ПОЛИТ.РУ — [http://www.polit.ru/rss/index.xml?yandex\\_mode=1](http://www.polit.ru/rss/index.xml?yandex_mode=1)
- PRAVDA.Ru — <http://export.pravda.ru/yandex.txt>
- PR NEWS (все пресс-релизы компаний) — <http://www.prnews.ru/yandex/business.asp>
- Энциклопедия поисковых систем — <http://www.searchengines.ru/news/news.rdf>
- Сетевой журнал — <http://www.setevoi.ru/weekly/export1.txt>

На сегодня существует уже множество служб синдикации новостей, которые предоставляют тематические фиды, построенные на основе использования многочисленных источников. Такой фид, к примеру, доступен на портале UAport (<http://uaport.net>) и позволяет получить интегрированный доступ к потоку украинских и российских новостных сообщений, собираемому системой InfoStream. С помощью RSS-шлюза системой InfoStream предоставляется унифицированный доступ к информации более чем с 600 Web-сайтов, сгруппированной по тематикам, языкам, странам, источникам. Объем этой информации сегодня превышает 20 000 сообщений в сутки. RSS-каналы UAport могут генерироваться системой по собственным запросам пользователей к поисковой системе.

Рассмотрим функциональность отдельных служб синдикации новостей, предоставляющих информацию в формате RSS.

## Moreover

Для интеграции соответствующего запросам пользователей контента в корпоративные сети или порталы служба Moreover (<http://www.moreover.com>) использует собственное решение — Connected Intelligence. Прием информации в систему от 6500 источников в режиме реального времени происходит каждые 15 минут, сообщения классифицируются и группируются по темам.

На сайте Moreover содержатся сведения о технологических подходах к интеграции новостей, которые были созданы в этой службе и де-факто стали стандартами в системах мониторинга. Определена следующая технологическая цепочка: сначала выполняется оценка информационного содержания Web-ресурса и построение конфигурационных профилей, описывающих данный ресурс. Затем Web-ресурсы автоматически сканируются в соответствии с профилями и происходит преобразование информации в формат XML с добавлением мета-тегов. После этого осуществляется классификация информации и ее распределение по запросам пользователей. На последнем этапе происходит вывод и доставка информации клиентам.

В июле 2003 года технология Moreover была интегрирована в новостной портал Yahoo!, с сайта которого (<http://news.yahoo.com>) возможен доступ к информации из 3500 источников (рис. 3.3).



Рис. 3.3. RSS-фиды новостной службы Yahoo

## Google

В 2002 году популярная поисковая система Google запустила свой новостной сервис — Google News (<http://news.google.com>), который охватывает информацию с 4500 различных сайтов за последние 30 дней. Данные на сайте системы отсортированы по нескольким категориям, таким как международные новости, деловой мир, шоу-бизнес, технологии и спорт.

Новости в системе отбираются в зависимости от времени их публикации, популярности источника информации и количества статей, появившихся в Internet, на данную тему. Компания Google — популяризатор и один из разработчиков формата Atom, применяемого, в основном, в блогах.

Вместе с тем, компания Google с подозрением относится к широким возможностям RSS-синдикации, углядев в этой технологии возможности для нарушений авторских прав. Так, недавно Google запретила британскому Web-мастеру использовать результаты поиска в системе Google News на другом сайте в виде RSS-фида. Британский программист Джулиан Бонд создал сценарий на языке PHP, который берет введенный пользователем запрос, направляет его на Google News, а результат выдает в формате RSS. Полученный результат можно использовать в любом RSS-агрегаторе. Сам скрипт под названием `gnews2rss` можно найти на сайте <http://www.voidstar.com/gnews2rss.php>. По словам Бонда, основной протест со стороны Google вызвал не сам скрипт, а использование его для формирования новостной ленты на постороннем сайте. Сам скрипт по-прежнему доступен в Internet, и его можно использовать в программах-агрегаторах. Тем не менее в письме Бонду в Google указывали на то, что предпочтительным вариантом является применение службы Google News Alerts.

## NewsIsFree

Одна из самых перспективных в Сети служб синдикации новостей NewsIsFree (<http://www.newsisfree.com>) охватывает свыше 9000 источников (в том числе российских и украинских). Сообщения обновляются каждые 15 минут и группируются по 15 основным категориям (<http://www.newsisfree.com/sources/browse>). Примечательно, что режим поиска в RSS-ресурсах обеспечивается поисковым механизмом компании Google. Основная особенность службы NewsIsFree — это полная интеграция с XML, в частности с RSS 0.91. Большинство разделов сайта службы содержит ссылки Syndicate, активизация которых приводит к отображению кода разделов в формате XML.

Несмотря на то что основой информационных ресурсов, охватываемых службой, являются англоязычные источники, NewsIsFree сегодня крупнейший интегратор и русскоязычных RSS-фидов, каталог которых доступен по адресу <http://newsisfree.com/sources/bylang/?lang=ru>.

## MSDN

Учитывая существующие в мире тенденции, служба MSDN (<http://msdn.microsoft.com>) также приступила к публикации своих новостей в формате RSS, выбрав версию 2.0 (рис. 3.4). Ниже приведен список некоторых тем и адресов новостных фидов MSDN:

- .NET Framework — <http://msdn.microsoft.com/netframework/rss.xml>
- ASP.NET — <http://msdn.microsoft.com/asp.net/rss.xml>

- Longhorn — <http://msdn.microsoft.com/longhorn/rss.xml>
- Mobile and Embedded — <http://msdn.microsoft.com/mobility/rss.xml>
- MSDN Subscriber Download — <http://msdn.microsoft.com/subscriptions/rss.xml>
- Office — <http://msdn.microsoft.com/office/rss.xml>
- Security — <http://msdn.microsoft.com/security/rss.xml>
- Visual Basic — <http://msdn.microsoft.com/vbasic/rss.xml>
- Visual C# — <http://msdn.microsoft.com/vcsharp/rss.xml>
- Visual C++ — <http://msdn.microsoft.com/visualc/rss.xml>
- Visual FoxPro — <http://msdn.microsoft.com/vfoxpro/rss.xml>
- Visual J# — <http://msdn.microsoft.com/vjsharp/rss.xml>
- Visual Studio — <http://msdn.microsoft.com/vstudio/rss.xml>
- Web Services — <http://msdn.microsoft.com/webservices/rss.xml>
- Windows Embedded — <http://msdn.microsoft.com/embedded/rss.xml>
- XML — <http://msdn.microsoft.com/embedded/xml.xml>



Рис. 3.4. Новостные RSS-фиды от MSDN



## Яндекс.Новости

Служба Яндекс открыла проект Яндекс.Новости (<http://news.yandex.ru>), к которому в настоящее время присоединилось свыше 500 Internet-изданий. Новости сортируются по десяти категориям, существует возможность поиска с указанием раздела и времени публикации новости. Поиск новостей возможен как по всем источникам, так и по заданным пользователем. Имеется также возможность поиска за произвольный период времени. Для сбора и экспорта новостей используется формат RSS 2.0.

Сегодня бесплатная служба синдикации новостного контента Яндекс представляет такие основные каналы:

- Главные новости — <http://news.yandex.ru/index.rss>
- Политика — <http://news.yandex.ru/politics.rss>
- В мире — <http://news.yandex.ru/world.rss>
- Общество — <http://news.yandex.ru/society.rss>
- Экономика — <http://news.yandex.ru/business.rss>
- Спорт — <http://news.yandex.ru/sport.rss>
- Происшествия — <http://news.yandex.ru/incident.rss>
- Культура — <http://news.yandex.ru/culture.rss>
- Здоровье — <http://news.yandex.ru/health.rss>
- Компьютеры — <http://news.yandex.ru/computers.rss>
- Internet — <http://news.yandex.ru/internet.rss>
- Авто — <http://news.yandex.ru/auto.rss>

## InfoStream

Разработанная в Информационном центре ЭЛВИСТИ система InfoStream® (<http://infostream.ua>) обеспечивает персонализацию интерфейса пользователей, работающих в режиме он-лайн, т.е. сохранение их постоянных запросов и организацию подписки, что реализуется на основе современной технологии RSS 0.91.

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

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

где в качестве значения *<запрос>* можно ввести слово или словосочетание на языке запросов информационно-поисковой системы InfoReS.

На основе технологии InfoStream созданы такие новостные каналы:

- Агропром — <http://uaport.net/cgi-bin/infostream.rss?rubr01>
- Банки — <http://uaport.net/cgi-bin/infostream.rss?rubr02>
- Экономика — <http://uaport.net/cgi-bin/infostream.rss?rubr03>
- Экономика Украины — <http://uaport.net/cgi-bin/infostream.rss?rubr04>
- Недвижимость — <http://uaport.net/cgi-bin/infostream.rss?rubr05>
- Биржи — <http://uaport.net/cgi-bin/infostream.rss?rubr06>

- Инвестиции — <http://uaoport.net/cgi-bin/infostream.rss?rubr07>
- Приватизация — <http://uaoport.net/cgi-bin/infostream.rss?rubr08>
- Нормативные акты — <http://uaoport.net/cgi-bin/infostream.rss?rubr09>
- Оборона, Конверсия — <http://uaoport.net/cgi-bin/infostream.rss?rubr10>
- Официальная хроника —  
<http://uaoport.net/cgi-bin/infostream.rss?rubr11>
- Криминал — <http://uaoport.net/cgi-bin/infostream.rss?rubr12>
- Обзоры прессы — <http://uaoport.net/cgi-bin/infostream.rss?rubr13>
- Связь — <http://uaoport.net/cgi-bin/infostream.rss?rubr14>
- Экология — <http://uaoport.net/cgi-bin/infostream.rss?rubr15>
- Энергетика — <http://uaoport.net/cgi-bin/infostream.rss?rubr16>
- Медицина — <http://uaoport.net/cgi-bin/infostream.rss?rubr17>
- Наука и техника — <http://uaoport.net/cgi-bin/infostream.rss?rubr18>
- Компьютеры — <http://uaoport.net/cgi-bin/infostream.rss?rubr19>
- Астрология — <http://uaoport.net/cgi-bin/infostream.rss?rubr20>
- Культура — <http://uaoport.net/cgi-bin/infostream.rss?rubr21>
- Катастрофы — <http://uaoport.net/cgi-bin/infostream.rss?rubr22>
- Образование — <http://uaoport.net/cgi-bin/infostream.rss?rubr23>
- Внешнеэкономическая деятельность — <http://uaoport.net/cgi-bin/infostream.rss?rubr25>
- Масс-медиа — <http://uaoport.net/cgi-bin/infostream.rss?rubr26>
- Калейдоскоп — <http://uaoport.net/cgi-bin/infostream.rss?rubr27>
- Религия — <http://uaoport.net/cgi-bin/infostream.rss?rubr28>
- Спорт — <http://uaoport.net/cgi-bin/infostream.rss?rubr29>
- Туризм — <http://uaoport.net/cgi-bin/infostream.rss?rubr30>
- Транспорт — <http://uaoport.net/cgi-bin/infostream.rss?rubr31>
- Автотранспорт — <http://uaoport.net/cgi-bin/infostream.rss?rubr32>
- Политика — <http://uaoport.net/cgi-bin/infostream.rss?rubr34>
- Страхование — <http://uaoport.net/cgi-bin/infostream.rss?rubr35>

### 3.9. Системы поиска RSS-фидов

Для нахождения RSS-фидов существуют многочисленные списки и каталоги, однако объемы существующих RSS-ресурсов таковы, что пользователям уже недостаточно десятка-другого рубрик первого уровня, имеющихся в каталогах. Как всегда, в подобных случаях на помощь приходят информационно-поисковые системы, которые позволяют находить как целые RSS-фиды, так и отдельные сообщения по ключевым словам. Поэтому в Internet появились поисковые сайты по RSS-фидам.

Одним из первых был создан сервис Feedster.com, который, кроме непосредственного поиска, позволяет подписаться на его результаты в формате RSS. В настоящее время Feedster обрабатывает 500 тысяч RSS-сообщений в сутки (рис. 3.5).

Еще одна поисковая система доступна на сайте <http://Assimilatethe.net>; она охватывает свыше 3500 RSS-ресурсов. Система ищет по заголовкам и описаниям RSS-сообщений. В базе данных системы Assimilatethe сейчас порядка 193 000 сообщений.

Как известно, RSS — самый распространенный формат для “живых журналов”, т.е. блогов (сокращение от слова Weblog). Для поиска по блогам также существуют сотни каталогов и поисковых систем. Среди основных поисковых систем по блогам можно назвать следующие:



Рис. 3.5. Результаты поиска в системе Feedster

- DayPop — <http://www.daypop.com>
- Blog Search Engine — <http://blogsearchengine.com>
- Feedster — <http://www.feedster.com>
- BlogStreet — <http://www.blogstreet.com>
- Blogarama — <http://blogarama.com/in.php?ID=2080>
- Globe of Blogs — <http://www.globeofblogs.com>

- BlogDex — <http://blogdex.media.mit.edu>
- Weblogs.com — <http://weblogs.com>
- BlogWise — <http://www.blogwise.com>
- BlogHop — <http://www.bloghop.com>
- BlogUniverse — <http://www.bloguniverse.com>

### 3.10. Агрегаторы

Пользователи, конечно же, могут читать RSS-файлы с помощью стандартных Web-браузеров, что, однако, сопряжено с просмотром XML-разметки и полным отсутствием всякого оформления. Но за это и боролись создатели формата RSS. А вот для интерпретации этого формата существует бесчисленное множество программ, созданных, в основном, за последние два-три года. Это означает, что пользователи могут получить доступ к данным в формате RSS с помощью специальных программ. Эти программы называются *RSS-агрегаторами* и в наглядном виде отображают содержание RSS-фидов.

Программа-агрегатор позволяет собирать RSS-файлы с Web-сайтов, одновременно следить за появлением на них новостей и читать содержание этих новостей. Программы-агрегаторы (их еще называют RSS-парсерами) выполняют синтаксический разбор данных, представленных в формате RSS, после чего могут реализовать любые действия по отношению к этим данным, — например, отсылать их по электронной почте либо отображать на определенном Web-сайте. Сегодня наиболее популярны агрегаторы, позволяющие собирать RSS-данные с разных Web-сайтов вместе.

#### Feedreader

Feedreader — это свободно распространяемая программа для Windows (<http://www.feedreader.com>), позволяющая читать данные в формате RSS версий 0.9, 0.91, 1.0, а также различную информацию от таких систем, как Dublin Core и Slashback (стандарты описания метаданных информационных ресурсов Сети). Утилита очень удобна в использовании, обеспечивает работу с информацией на русском и украинском языках и обладает широким кругом сервисных возможностей (рис. 3.6). FeedReader версии 2.5 можно загрузить по адресу [http://sourceforge.net/project/showfiles.php?group\\_id=70179](http://sourceforge.net/project/showfiles.php?group_id=70179); размер файла инсталляции — 1,2 Мбайт.

Feedreader — типичный RSS-агрегатор, интерфейс которого напоминает интерфейс почтовых программ. У пользователя, знакомого с почтовыми клиентами, работа с программой не вызывает затруднений. Остановимся подробнее на самых необходимых возможностях этой программы.

Для настройки подписки на RSS-фид пользователю следует активизировать опцию New и ввести следующую информацию:

- адрес RSS-фида;
- название фида (оно может быть определено пользователем);
- периодичность обращения к фиду на Web-сайте для обновления.



*Рис. 3.6. Интерфейс агрегатора Feedreader напоминает почтовую программу*

При этом имеется возможность изменить кодировку, размер шрифтов, поместить фид в отдельную папку и сгруппировать фиды.

Для управления подпиской существуют дополнительные опции, доступные в контекстном меню, отображаемом при щелчке правой кнопкой мыши 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 (на которое есть ссылка — “link”), заголовок и аннотация которого вызвали интерес, следует выполнить одно из следующих действий:

- дважды щелкнуть левой кнопкой мыши на заголовке сообщения;
- щелкнуть на ссылке Read on в поле аннотации;
- щелкнуть на соответствующей кнопке, расположенной перед заголовком сообщения;
- щелкнуть правой кнопкой мыши на заглавии сообщения — в результате можно открыть текст сообщения в новом окне браузера;
- щелкнуть на ссылке на первоисточник, что позволит перейти на соответствующий сайт в Internet.

## FeedDemon

Feed Demon представляет собой коммерческую программу ([www.feeddemon.com](http://www.feeddemon.com)), обеспечивающую удобную работу с RSS версии 2.0. Имеется возможность по-пробовать работу программы в “триальном” режиме. Утилита работает в среде Windows, корректно обращается с русской и украинской кодировками, обеспечивает поиск и фильтрацию информации фидов. Триал-версию FeedDemon 1.0 можно найти по адресу <http://www.feeddemon.com/download/downloadhandler.asp?file=feeddemon-trial.exe>; размер файла инсталляции — 2,3 Мбайт. Удобный интерфейс агрегатора позволяет легко отслеживать и читать свежие фиды. Feed Demon позволяет также представлять содержимое новостных лент в виде своеобразной газеты.

Приступить к использованию программы можно немедленно после инсталляции, так как пользователь сразу же начинает получать рассылки с сайтов Rollingstone.com, Scripting News, Sladshot, Wired, Yahoo! и др. Программа позволяет сохранять сообщения (News Bins) и отслеживать их по ключевым словам, запуская функцию Watches. Отдельные RSS-фиды можно перенаправлять в тематические списки или каналы. FeedDemon также позволяет проводить поиск и читать новости в автономном режиме.

Для подписки на фиды в программе следует ввести URL источника или импортировать файл OPML. Цена FeedDemon 1.0 составляет \$29,95.

## Abilon и ActiveRefresh

Это два агрегатора от одного производителя — компании ActiveRefresh (<http://www.activerefresh.com/download.php>). Бесплатная программа Abilon вполне подходит для среднего пользователя; она проста и надежна, отличается высокой скоростью и малой ресурсоемкостью (339 Кбайт). Она обладает возможностью закачки новых каналов с сайтов MoreOver, MyRss и NewsIsFree. Однако ей не хватает возможностей глобального поиска и сжатия информации.

В отличие от Abilon, ActiveRefresh — это платная программа, обеспечивающая полную реализацию концепций компании и позволяющая агрегировать обычные Web-сайты, импортировать с них новости, представленные в HTML, следить за почтовыми ящиками, проводить глобальный поиск и т.д.

## Syndirella0.9b

Программа Syndirella может как показывать информацию с обычных Web-страниц, так и отображать данные, представленные в формате RSS. Программа реализована на платформе .NET, функционирует в среде операционных систем Windows и требует установки Internet Explorer версии 5.0 или выше. Для работы программы необходимо установить библиотеку Microsoft .NET Framework runtime версии 1.0 (20 Мбайт). Однако если этот компонент уже установлен, то сама программа Syndirella займет всего 250 Кбайт. Адрес для загрузки: <http://www.yole.ru/projects/syndirella>.

## Прочие программы для среды Windows

Кроме перечисленных выше, сегодня большую популярность получили еще два агрегатора для работы под Windows — Awasu и Beaver. Особенность бесплатной программы Awasu (<http://www.awasu.com>) заключается в ее возможности объединять потоки множества новостных сайтов и блогов. Beaver (<http://www31.brinkster.com/toolmaker>) принимает фиды форматов RSS/RDF и имеет привычный интерфейс в стиле Outlook Express.

## **K.R.S.S. 2.6**

KDE's Rich Site Summary viewer — приложение для Linux, позволяющее отображать данные в формате RSS на экране в виде HTML-страниц. Есть возможность настраивать способ отображения информации с использованием технологии CSS (Cascading Style Sheets) и устанавливать специальные фильтры новостей. Адрес для загрузки программы: <http://krss.sourceforge.net/downloads.html>; размер файла — 394 Кбайт.

## **Liferea**

В последнее время для ОС Linux большую популярность приобрел агрегатор Liferea (<http://liferea.sourceforge.net>). Он поддерживает многочисленные форматы новостных фидов, основанные на XML, — такие как RSS, RDF, Atom, Echo, PIE, а также OCS и OPML для списков фидов. Эта программа распространяется с библиотекой GTK2.

## **Опера 7.5 и прочее**

Норвежская компания Opera Software (<http://www.opera.com>) выпустила новую версию браузера Opera 7.5, в которой появился встроенный RSS-агрегатор. Доступ к нему организован через интерфейс почтового клиента.

В настоящее время создаются (и уже созданы!) многочисленные инструментальные средства для разработки программ работы с RSS-данными. Например, для разработки программ-парсеров на языке Perl создан модуль XML::RSS, который загружается с сайта <http://search.cpan.org>.

Встраиваемые в Internet Explorer инструментальные панели от Dogpile (<http://www.dogpile.com/info.dogpl/tbar>) и HotBot Desktop (<http://www.hotbot.com/tools/desktop>) поддерживают технологии RSS и Atom. С помощью этих средств заголовки сайтов, поддерживающих RSS, просматриваются прямо в окне браузера.

Одна из самых заметных особенностей интерфейса будущей версии ОС Windows-Longhorn заключается в наличии многофункциональной боковой панели (Sidebar). На нее может быть помещена любая информация — от часов и списка контактов до новостей, импортируемых в формате RSS. При этом средства настройки панели включены в состав инструментария разработчиков и поддаются настройке с их стороны.

## **3.11. Новые подходы**

С помощью современной RSS-технологии пользователи Internet получили надежный и простой доступ к ресурсам оперативной информации с Web-сайтов Сети. Перспективность и популярность RSS как стандарта обусловлена, прежде всего, его доступностью и простотой. Сегодня практически все ведущие информационные сайты в мире и “живые журналы”, работающие в Internet, используют RSS как инструмент оперативного представления обновлений своих ресурсов.

Еще один аспект применения RSS-технологии стал актуален в связи с массовым распространением неостребованных рассылок по электронной почте — спама. Действительно, электронная почта привлекательна и для спамеров. Нередко списки электронных адресов подписчиков новостей на сайтах и порталах становятся добычей взломщиков, что делает подписку через e-mail достаточно рискованным занятием.

Поэтому можно предположить, что на смену рассылкам придет использование RSS-фидов. В отличие от рассылок по электронной почте, где доставка инициируется администраторами сайтов после того, как подписчик оставил им свой адрес, в случае использования RSS пользователь сам вводит адрес необходимого ему RSS-фида в программу-агрегатор. Эта программа периодически проверяет, не изменилось ли содержание RSS-фида, и при наличии изменений автоматически закачивает его содержимое. Главным преимуществом RSS-технологии здесь является то, что пользователь сам принимает решение о получении каждого конкретного сообщения.

Все большую популярность RSS-технология приобретает у владельцев Web-ресурсов (не только новостных, но и коммерческих) еще и благодаря своей экономичности — не требуется никаких средств для борьбы со спамом, нет необходимости в фильтрации писем и в управлении рассылкой. При этом все, кому это необходимо, получают желаемую информацию о важных событиях, корпоративных анонсах, обновлениях Web-сайтов и пр.

Индустрия рекламы также не осталась в стороне от использования технологии RSS. Хотя RSS последних версий допускает вставку гиперссылок и изображений, однако как рекламный носитель она несколько уступает электронным письмам в HTML-формате. В настоящее время в Сан-Франциско создается первая он-лайновая рекламная сеть RSSAds, которая базируется на внедрении текстовой рекламы в заголовки RSS версий 0.90, 0.91, 1.0, 2.0 и Atom. Основателям этой сети удалось разработать систему подсчета рекламных показов — как только RSS-клиент обращается на сервер с запросом, система учитывает это событие. RSSAds планирует продавать рекламу, используя разнообразные модели оплаты: за количество показов, за время показов, за количество кликов, а также за размещение рекламы в заголовках.

Системы синдикации Internet-новостей решают проблему нахождения необходимой информации, но оставляют без внимания такие задачи, как обобщение данных — их обработку и анализ. Одним из самых перспективных направлений обобщения информационных потоков в настоящее время является метод “глубинного анализа текстов” (Text Mining). Применительно к новостным потокам его идеологию можно сформулировать как постоянное, воспроизводимое во времени выполнение их содержательного анализа. Непрерывная аналитическая обработка сообщений является самой характерной особенностью этого метода, который позволяет формировать автоматические дайджесты, выявлять новые понятия и их взаимосвязи, рассчитывать разнообразные рейтинги. Именно системы такого типа смогут избавить пользователей от дублирующейся информации, информационного шума, позволят выявлять главные тенденции, находить коррелирующие события. По прогнозам аналитической компании IDC, спрос на подобные системы существенно возрастет в течение ближайших 4-5 лет. Ожидается, что в 2005 году прибыль от продажи таких систем составит 1,5 млрд долларов США, а в 2006 году они будут доминировать в сфере анализа информации.

## **3.12. Информационные ресурсы для мобильных устройств**

### **3.12.1. Wireless Application Protocol**

Wireless Application Protocol (WAP) — это открытый протокол и технический стандарт, разработанный по инициативе фирмы “Unwired Planet” (сегодня ее название “OpenWave”). WAP обеспечивает передачу информации из Internet на



дисплеи мобильных телефонов [26]. Широкое применение этого стандарта привело к тому, что в настоящее время можно подключиться и обмениваться информацией через Internet непосредственно с мобильных телефонов, без посредничества компьютера. WAP существенно расширяет набор таких услуг мобильной связи, как обыкновенные звонки и короткие текстовые сообщения (SMS), позволяя внедрять сервисы, аналогичные тем, которые предлагаются в World Wide Web. Internet на экране мобильного телефона чем-то напоминает телетекст на миниатюрном телевизоре, однако интерактивность, т.е. возможность не только получать, но и вводить информацию, существенно расширяет сферу применения WAP.

Первые мобильные телефоны с поддержкой WAP появились на рынке в середине 1999 года. С 2001 года практически все производители выпускают мобильные телефоны с поддержкой WAP-протокола. По прогнозам бюллетеня Analysys, число пользователей мобильных устройств с поддержкой протокола WAP к 2005 году превысит 370 млн человек. В настоящее время для доступа к беспроводным сетям 66% предприятий пользуются мобильными ПК, 24% — карманными ПК, 21% — мобильными телефонными аппаратами с поддержкой протокола WAP.

Сегодня происходит широкое внедрение нового стандарта скоростного обмена данными в сетях мобильной связи — General Packet Radio Service (GPRS), согласно которому данные передаются пакетами по каналам, свободным от голосового трафика. Сети GPRS могут поддерживать максимальную скорость 107,2 Кбит/с (в то время как GSM — всего 9,6 Кбит/с). Поэтому они обеспечивают эффективную высокоскоростную работу в Internet и постоянное пребывание на линии. Трафик здесь подсчитывается не по времени соединения, а по объему информации, что дополнительно стимулирует к работе с лаконичными, информационно насыщенными ресурсами.

Надо отметить, что по протоколу WAP на мобильный телефон содержание Web-ресурсов сети Internet непосредственно не передается. В Internet информация представлена в виде HTML-страниц, работа с которыми предполагает быстрые коммуникации, мощные процессоры, большие объемы памяти компьютеров, большие экраны. Мобильные же телефоны обладают медленными процессорами, небольшими объемами памяти и совсем маленькими дисплеями. Поэтому для WAP используется свой специализированный язык разметки — Wireless Markup Language (WML), более простой и строгий, чем HTML. В WAP также применяются язык сценариев — WMLScript (упрощенная версия JavaScript), а также собственный формат растровых изображений — Wireless Bitmap (WBMP).

### 3.12.2. WAP-ресурсы

WAP — это, прежде всего, мобильная технология доступа к информационным ресурсам. Сегодня происходит бурное насыщение ниши сетевых WAP-ресурсов. Количество русскоязычных WAP-сайтов составляет уже несколько тысяч и постоянно увеличивается. Для ориентации в этих ресурсах уже мало традиционных каталогов типа Wapper (<http://www.wapper.ru/catalog>), MOBIL (<http://www.MOBIL.ru/wap.php> — рис. 3.7) или WapGate (<http://www.wapgate.ru>). Как и для англоязычных (всемирных) ресурсов, возникла необходимость создания “WAP-поисковиков”, одним из лучших среди которых в настоящее время является “Яндекс” (<http://wap.yandex.ru>).

Сегодня многие банки стараются предоставлять в режиме он-лайн тот же набор услуг, что и в обычных отделениях (денежные переводы, пополнение карточного счета, погашение потребительского кредита). Одна из главных тенденций развития

Internet-банкинга — его использование “в одном флаконе” с другими возможностями удаленного доступа: call-центром (когда клиент, пользуясь Internet-банкингом, тут же консультируется с сотрудником банка по телефону) и Mobile-банкингом (банковские услуги через мобильный телефон при помощи технологии WAP).



Рис. 3.7. MOBIL.RU — каталог русскоязычных WAP-ресурсов

Наиболее полезны и удобны услуги WAP, связанные с доступом к электронной почте. Благодаря им пользователь может в любой момент просмотреть свежую корреспонденцию на дисплее мобильного телефона. В последнее время такая услуга появилась у крупнейших бесплатных почтовых серверов (wap.mail.ru, wap.imail.ru, wap.newmail.ru). Наряду с электронной почтой, сервер wap.beer.ru предлагает виртуальный ежедневник, в котором можно распланировать свой день, а затем получать соответствующие напоминания.

Для ориентации в мире WAP можно воспользоваться поисковыми серверами — например, wap.yandex.ru или www.wapgate.ru (второй ресурс является обычным Web-сайтом, с которого можно, кроме того, просматривать WAP-ресурсы). На WAP-сайтах сегодня находится не так уж мало информации по различным темам. Так, поклонников карманных ПК наверняка заинтересует сайт wap.handy.ru. Любители игр в режиме on-line оценят сервер www.wirelessgames.com/index.wml, зарегистрировавшись на котором, можно поиграть прямо с экрана мобильного телефона. В мире WAP не забыты и такие традиционные для Internet средства общения, как chat (wap.chat.ru) и ICQ (www.wapicq.com).

Ниже приведен список еще нескольких интересных ресурсов средств массовой информации в русско- и украиноязычном WAP-пространстве.

- РосБизнесКонсалтинг — <http://wap.rbc.ru>. Новости политики и культуры, фондовые индексы, котировки акций.
- Polit.RU — <http://wap.polit.ru>. Новости российской политики.
- RevKom — <http://wap.revkom.ru>. Новости мира связи, коммуникаций, цифровых устройств.
- RosWeb — <http://wap.rosweb.ru>. Сборник последних международных и российских новостей.
- UAport — <http://wap.uaport.net>. Украинский интегратор сетевых новостей Internet-холдинг UAport.
- UAtoday — <http://uatoday.net/wap>. Газета “Украина сегодня” в WAP-формате. Новости в реальном времени.
- IT-News — <http://wap.it-news.net.ua>. Новости информационных технологий.
- Апорт — <http://wap.aport.ru>. Новости прессы, электронных СМИ.
- БиЛайн — <http://wap.beeline.ru>. Новости.
- Med-news — <http://wap.bmed-news.kiev.ua>. Страница медицинских новостей.
- Финансовые новости — <http://www.audit-it.ru/wap>. Налоговые новости, курсы валют от ЦБ РФ.
- GALA.NET — <http://wap.gala.net>. Новости.
- BBC-news — <http://www.bbc.co.uk/mobile/mainmenu.wml>. Деловая информация и новости, а также данные о котировках.
- Экономические новости — <http://wap.mfd.ru>. Экономические новости для финансистов и аналитиков.
- МТС — <http://wap.mts.ru>. Финансовые и бизнес-новости, биржевые индексы, новости РБК и АФП.
- Сетевая Лаборатория — <http://wap.netlab.ru>. Компьютерные новости.
- “Новая газета” — <http://wap.novayagazeta.ru>. Обзор и анализ событий в России. WAP-ресурс.
- Nursat — <http://wap.nursat.kz>. Новости агентства “Интерфакс-Казахстан”.
- Аудит-ИТ — <http://www.audit-it.ru/wap>. Новости налогообложения, бухучета, аудита.

### 3.12.3. Реализация WAP-протокола

Схема работы WAP-сервисов включает три основных компонента: WAP-микробраузер, WAP-шлюз и WAP-сервер (рис. 3.8).

В качестве микробраузера может выступать мобильный терминал или программный эмулятор. WAP-шлюз взаимодействует с микробраузером, используя стек протоколов WAP. Шлюз переводит полученные от пользователя запросы, представлен-

ные в бинарном виде, в принятый в World Wide Web формат HTTP-сообщений. При этом провайдеры информации в качестве WAP-сервера могут использовать любой HTTP-сервер, например Apache, применяя все существующие наработки для создания сервисов и администрирования. Когда загрузка информационного блока с WAP-сервера завершается, WAP-шлюз компилирует элементы WML в компактную бинарную форму, что позволяет обеспечить большую скорость обмена информацией.



Рис. 3.8. Общая схема WAP-технологии

Кроме стандартных HTTP-серверов, для реализации WAP-серверов могут использоваться и специальные разработки, среди которых наиболее известен сервер компании Nokia, объединяющий функции сервера и шлюза.

В качестве коммерческих WAP-шлюзов наиболее известны продукты компаний Nokia и Ericsson. Nokia Artuse WAP Gateway, связывающий Internet (или intranet) и мобильные сети, обеспечивает для мобильной связи доступ к различным Internet-сервисам, а также позволяет применять устройства с поддержкой WAP для доступа к Web-приложениям.

Ericsson же предлагает два продукта — Ericsson WAP/Gateway Proxy и Jambala WAP Gateway. Оба пакета представляют собой комплексные решения для организации на базе операторов мобильной связи доступа к WAP-сервисам. Они включают сервер-шлюз с возможностью компиляции HTML-страниц в WML, интерфейс для WTA, поддержку SMS-шлюза.

Кроме продуктов от Nokia и Ericsson, существует несколько продуктов других производителей. Среди них наибольший интерес представляет проект Kannel, который был основан компанией Wapit Ltd. в 1999 году с целью разработки свободно распространяемого WAP-шлюза для UNIX-платформ. В данный момент на сайте Kannel (<http://www.kannel.org>) можно бесплатно получить полноценный WAP- и SMS-шлюз для Linux RedHat 6.1 или Debian с исходными кодами. Существует версия Kannel под Windows, адаптированная под эту ОС фирмой Warpe (<http://kannel.dev.warpe.net>).

### 3.12.4. WML и микробраузеры

Язык WML основан на модели описания языков XML (extensible Markup Language), поэтому первой строкой в любом файле должно быть указание на документ описания DTD (Document Type Definition) для данного языка. Официальная спецификация WML разработана и поддерживается WAP Forum, производственным консорциумом, основанным Nokia, Phone.com, Motorola и Ericsson. Эта спецификация определяет синтаксис, переменные и элементы, используемые в файлах формата WML. Исходя из этого, начало любого WML-документа должно выглядеть так:

```
<?xml version="1.0"?>
<!DOCTYPE wml PUBLIC "-//WAPFORUM//DTD WML 1.1//EN"
"http://www.wapforum.org/DTD/wml_1.1.xml">
```

Указанный DTD описывает все обязательные теги и элементы стандарта WML. Сам WML-код должен быть заключен в теги. Поскольку WML был разработан для устройств с низкой пропускной способностью и маленьким дисплеем, в качестве одной из составляющих используемого дизайна была применена концепция дек (колод) и карт. Отдельный WML-документ называется декой (deck), а интерактивное взаимодействие с пользователем осуществляется с помощью карт (card). Преимущество такой реализации заключается в том, что несколько экранов могут быть загружены в буфер мобильного телефона за один раз.

Карты можно рассматривать как прямую аналогию с HTML-страницами. Так же, как и HTML-страницы, карты имеют заголовок, определяемый параметром title, и внутренний идентификатор (id), по которому формируются гиперссылки на карту. На экране браузера может отображаться только одна карта, но поскольку она обычно достаточно мала по объему и на запрос ее уходит больше времени, чем на загрузку, близкие по смыслу карты объединяются в деки, что существенно ускоряет работу. Отдельный WML-файл представляет собой отдельную деку.

Описание текстовых элементов внутри карты начинается с парного тега "абзац". С этим тегом можно задавать следующие параметры: align="выравнивание" и mode="wrapmode". Параметр выравнивания может принимать значения right, center или left, а параметр mode определяет, будет ли текст автоматически переноситься на экране (значение wrap) или задействуется скролинг (значение nowrap). Для форматирования текста может использоваться несколько тегов, аналогичных HTML: <b> </b>, <i></i>, <u></u>.

Стандарт WML предусматривает собственный формат изображений для отображения в браузерах — WBMP. Это двухцветное изображение с специфическим алгоритмом сжатия. Для преобразования обычных цветных растровых файлов в WBMP существует несколько средств: например, по адресу [www.rcp.co.uk/distributed/Downloads](http://www.rcp.co.uk/distributed/Downloads) можно получить плагин Photoshop для экспорта в этот формат. На Web-странице [http://www.netec.de/downloads/wap\\_pictus.htm](http://www.netec.de/downloads/wap_pictus.htm) приведена программа WAP Pictus для конвертирования изображений в формат WBMP. Кроме того, конвертор изображений в формат WBMP содержится, например, в WAP-эмуляторе для ПК Deck-it WAP Previewer. Для присоединения изображения к карте служит тег img с параметрами, идентичными аналогичному HTML-тегу.

На сегодняшний день существует несколько микробраузеров, т.е. программ, встраиваемых в мобильные телефоны для работы по WAP-протоколу с WML-документами. Самый известный микробраузер создан компанией Unwired Planet (UP); он применяется в WAP-телефонах Motorola, Alcatel, Samsung, Nokia и Ericsson имеют собственные решения; кроме того, существуют модели телефонов (Benefon, Sony), использующие микробраузер корпорации Microsoft — Mobile Explorer.

Недавно представители компании Nokia заявили о том, что, крупнейший производитель модулей памяти для компьютеров, компания Samsung Electronics согласилась лицензировать исходный код браузера для мобильных устройств и технологию обмена сообщениями Smart Messaging. Оригинальная версия программы поддерживает WML 1.3, XHTML и WAP CSS и отвечает спецификациям WAP Forum и W3C. Использование реализованных в программном продукте технологий позволит удовлетворить всем требованиям заказчиков. Браузер для мобильных устройств от Nokia представляет собой платформу-независимый продукт, отвечающий требованиям OEM.

Существующие микробраузеры имеют свои особенности. Если говорить о параметрах, имеющих значение при разработке сервисов, то, прежде всего, для разных микробраузеров следует учитывать различия в максимальных размерах загружаемого за один сеанс блока информации (деки). Кроме того, не везде решены проблемы русификации (не говоря уже об украинизации) микробраузеров. Следует отметить, что в настоящее время в этой области повсеместно применяется кодировка символов utf-8, что, однако, не учитывается, например, во многих телефонах от Motorola. Для обеспечения работы с этими моделями информационным провайдером до сих пор приходится применять транслитерацию в WML-документах.

### 3.12.5. Эмуляторы WAP

Доступ к WAP-ресурсам возможен не только с мобильных телефонов. Оперативная и лаконичная информация WAP-сайтов доступна также с помощью программ-эмуляторов для ПК, например Deck-It ([www.pyweb.com](http://www.pyweb.com)), M3Gate ([www.numeric.ru/m3gate/r\\_index.htm](http://www.numeric.ru/m3gate/r_index.htm)), WinWAP ([www.winwap.org](http://www.winwap.org)). Эти программы позволяют просматривать WAP-серверы с обычных компьютеров, подключенных к Internet. Кроме того, просмотр WAP-серверов возможен и с помощью специализированных Web-сайтов сети Internet (как правило, для этого в специальное поле эмулятора достаточно ввести адрес WAP-сайта).

Универсальный Web-браузер Opera на данный момент — единственный продукт данного класса, поддерживающий протокол WAP и обеспечивающий просмотр страниц, написанных на языке WML.

Одной из лучших программ-эмуляторов WAP для ПК является Klondike WAP Browser Personal Edition 1.5 компании Apache Software (рис. 3.9). Этот эмулятор позволяет посещать WAP-сайты и сайты, автоматически перекодировываемые сервером с HTML в WML. С помощью Klondike WAP Browser можно также провести тестирование любого WAP-сайта в режиме офф-лайн.

Эмулятор M3Gate 0.5 компании Numeric Algorithm Laboratories поддерживает работу по стандарту WAP 1.1 с WML и WMLScript, изображениями WBMP и PNG, различными кодировками, позволяет отлаживать WAP-приложения, просматривать WML-код документов. M3Gate интегрируется с Internet Explorer или Netscape Navigator и автоматически запускается при попытке открыть WAP-ресурс. WinWAP — простой в использовании WAP-эмулятор, который работает через HTTP и WAP-шлюзы. С его помощью можно просматривать исходный код, переменные, Cookie. Включает в себя множество стандартных для Web-браузеров возможностей: печать, поддержку технологии drag and drop, использование закладок, изменение размера окна, поиск текста и др. Имеется версия для WinWAP Pocket PC.

Практически все производители мобильных телефонов имеют свои версии WAP-эмуляторов для персональных компьютеров.

Достаточно обширный список WAP-эмуляторов для ПК можно найти в Internet по адресу <http://www.wapgate.ru/soft/?s=2>.

Кроме программных WAP-эмуляторов (их еще называют WAP-браузерами), можно воспользоваться специальными Web-серверами, доступными с обычного компьютера, подключенного к Internet. Эти сайты (например, [www.wapgate.ru](http://www.wapgate.ru)) представляют возможность просмотра WAP-страниц путем указания WAP-адресов в специальных окнах ввода. Сегодня такие сайты открывают перед пользователями Internet новый срез информационных источников — кратких и оперативных, доступных к тому же с помощью привычного интерфейса.

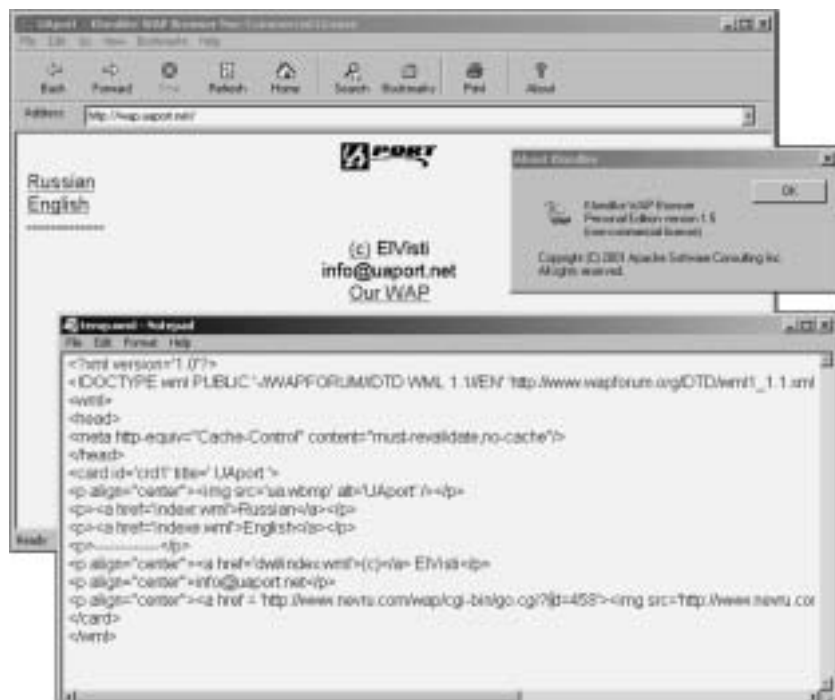


Рис. 3.9. Исходный текст WML, вызванный из эмулятора Klondike WAP Browser



Рис. 3.10. Wapsilon и его “скин” — телефон Nokia 7210

WAP-эмулятор Wapsilon v 2.4, который представлен на сайте wapsilon.com, ориентирован, прежде всего, на Nokia 7210 (рис. 3.10).

WAP-эмулятор Gelon (разработчик Gelon.net) обеспечивает отображение WML-страниц с помощью обширного набора “скинов”: Ericsson R320, Nokia 7110, Nokia 6210, Siemens C35, Siemens M35, Siemens S35, Motorola A6188, Motorola P7389 (рис. 3.11).



Рис. 3.11. Gelon — сайт-эмулятор WAP с множеством “скинов”

Еще один WAP-эмулятор для телефонов Nokia 3310, 6210, 7110, 9110, построенный на основе Java-апплетов, можно найти на украинском сайте “Мобильный портал”, находящемся по адресу <http://www.mobile-portal.kiev.ua/wap.phtml?t=emul>.

Эмулятор, расположенный по адресу TatTag.com, обладает способностью настраиваться на различные кодировки, что очень удобно при работе с многочисленными вариантами кириллических кодировок (рис. 3.12).

### 3.12.6. Проблемы и перспективы WAP

На сегодняшний день существует несколько проблем, ограничивающих распространение WAP-технологии. Прежде всего, такая связь с Internet остается дорогой и медленной. Кроме того, прогресс слабо коснулся возможностей отображения информации на небольших дисплеях мобильных устройств. Через год после рождения WAP стали появляться сомнения в наличии достаточного числа



потребителей, необходимого для окупаемости данной технологии. Абоненты сотовых сетей все еще редко используют свои телефоны для выхода в Internet. Компания Nielsen Norman Group провела в Лондоне “драйв-тест” технологии WAP, выдав добровольцам телефоны с WAP. По результатам исследования, Nielsen Norman рекомендовала операторам сотовой связи следующее: “Не тратьте деньги на внедрение тех услуг, которыми никто не будет пользоваться. Мы документально зафиксировали, что сфера применения WAP весьма узка”.



Рис. 3.12. Эмулятор TagTag.com

На смену телефонам с поддержкой WAP приходят другие компактные устройства, поддерживающие иные принципы работы в Internet. Например, компания Interactive Intelligence объявила о создании нового протокола передачи данных Mobilite 1.0, который по всем основным показателям опережает WAP и позволяет развивать и реализовать корпоративные решения с использованием карманных ПК, работающих под управлением Pocket PC, Palm OS и других систем. По словам Дональда Брауна, исполнительного директора Interactive Intelligence, протокол WAP прекрасно подходит лишь для решения относительно простых задач, поскольку обладает ограниченными функциональными возможностями. В отличие от WAP, программное обеспечение протокола Mobilite поддерживает все известные платформы, на которых работают карманные ПК.

Кроме того, во всем мире сейчас началось внедрение технологии скоростного обмена данными в сетях мобильной связи — General Packet Radio Service

(GPRS), с помощью которой можно работать в Internet, принимать почту, отправлять цифровые фотоснимки, играть в он-лайнные игры. Причем все это относительно недорого. Трафик в GPRS подсчитывается не в привычных для абонента минутах соединения, а в байтах — за объем переданной информации. Первым мобильным аппаратом, поддерживающим GPRS, была Motorola P7386i. Из удачных моделей стоило бы выделить Siemens S45, ME45, Nokia 8310 и несколько моделей от Ericsson.

Несмотря на критику WAP, этой технологии суждено стать стандартом “де-факто” — просто потому, что ею уже пользуется очень много людей. В принципе, у WAP-технологии неплохие перспективы, так как в комитет по разработке входят такие компании, как Alcatel, Matsushita, Swisscom Motorola, Nokia, Philips, Qualcomm, T-Mobil Samsung, Intel, NEC, Siemens, Fujitsu, IBM, Ericsson Psion Software, AT&T Wireless Services, BellSouth Cellular Corporation, Sonera, Telenor, Telstra, Vodafone, BT Cellnet, Sprint PCS, Telia Mobile. Большинство аналитиков сходятся во мнении, что в течение нескольких лет (где-то 2003–2005 гг.) число мобильных телефонов с доступом к Internet в мире превысит число подсоединенных к Internet ПК. Согласно прогнозу компании Strategy Analytics, уже в 2003 году 95% мобильных телефонов, выпущенных в США и Западной Европе, будут способны работать по протоколу WAP. Только в 2003 году в эфире было около миллиарда цифровых мобильных телефонов и новых WAP-оптимизированных устройств.

В настоящее время становится актуальной задача создания мобильных сетей третьего поколения (3G), которая заключается не только в разработке мобильных телефонов лучшего качества. Каналы связи 3G должны обеспечивать работу мультимедийных приложений, включая видеотелефонию (компания Orange уже работает над подобным проектом), видео по требованию и другие формы широкополосной связи. Менеджер компании Ericsson по связям с общественностью Питер Бодор (Peter Bodor) считает недостаточный успех WAP следствием того, что эта технология обманула ожидания пользователей, чего с 3G быть не должно. “Разочарование в WAP вызвано неспособностью индустрии оправдать ожидания, причем главной проблемой стала ее медлительность, — говорит он. — С 3G работать в Internet будет так же легко, как и из дома, к тому же локально-ориентированные услуги сделают эту работу более персонифицированной.”

Недавно более десятка ведущих коммуникационных компаний, включая AT&T Wireless, Cingular Wireless, mm02, NTT DoCoMo, Telefonica Moviles, Vodafone, Fujitsu, Matsushita, Mitsubishi Electric, Motorola, NEC, Nokia, Samsung, Sharp, Siemens, Sony Ericsson, Toshiba и Symbian, объявили об инициативе, направленной на создание максимально “однородного” рынка услуг мобильной связи. Данная инициатива предполагает, что ее участники будут разрабатывать свое ПО для мобильных систем следующих поколений (включая терминальные клиентские модули и серверные решения) строго в соответствии со спецификациями основных органов по стандартизации, таких как 3GPP.

Несмотря на прогнозы скептиков, предсказывавших уход WAP с рынка еще в 2003 году и рассматривающих эту технологию как переходную, на июнь 2004 года зафиксирован значительный рост числа WAP-сайтов (эти данные предоставила Mobile Data Association). Их количество выросло на 42% по сравнению с аналогичным периодом 2003 года и достигло 1,1 млн. Наибольшей популярностью сегодня пользуются WAP-сайты и порталы крупных мобильных операторов, основанные на WAP-технологии.

С помощью мобильных устройств с WAP-протоколом пользователи уже сегодня получают надежный и простой доступ к ресурсам сети Internet, справочной и оперативной информации, поступающей со всего мира.

### 3.12.7. Доступ к сетевому контенту с КПК

В течение нескольких последних лет во всем мире продолжается бум беспроводных (wireless) технологий. Естественно, не последнюю очередь в этом играют мобильные устройства [19]. На сегодня известно два основных типа устройств, применяемых в беспроводных технологиях. Во-первых, это мобильные телефоны, во-вторых — “наладонники”, они же КПК (карманные персональные компьютеры) или PDA (Personal Digital Assistant) двух конкурирующих типов — Palm и Pocket PC.

Широкое распространение уже в течение нескольких лет получили наладонники, представляющие собой, как правило, бесклавишные компьютеры карманного формата, которые обладают такими же потенциальными возможностями, как и обычные персональные компьютеры (ПК), если не принимать во внимание несколько непривычные и, откровенно говоря, весьма ограниченные возможности ввода данных, а также небольшие дисплеи этих устройств. Рабочие части экранов КПК имеют разрешение всего 160x160, 240x320 или, что пока очень редко, 320x320 точек (монохромных или цветных). Сенсорные экраны КПК позволяют осуществлять рукописный ввод информации. Например, у КПК Palm часть экрана занимает область, где можно рисовать символы, которые автоматически преобразуются в распознанные буквы. Практически для всех КПК разработаны и “крупногабаритные” клавиатуры, которые, однако, не входят в базовые поставки наладонников.

Что же касается конкуренции различных моделей, то можно отметить, что среди обычных пользователей наиболее популярны сегодня Palm и его клоны, а среди корпоративных пользователей все большую популярность получает модель Pocket PC, которая гарантирует полную совместимость с серверами, настольными компьютерами и корпоративными сетями. Обеспечивая завидное единообразие Pocket PC, корпорация Microsoft установила жесткий контроль над дизайном, процессорами, размерами и разрешением дисплеев.

В то же время американский журнал Business Week отмечает, что в 2003 году именно клоны Palm (КПК, работающие на базе операционной системы Palm OS) стали самыми новаторскими (Tapwave Zodiac, Sony CLIE PEG-UX50, Handspring Treo 600).

#### С КПК — в Сеть

Как известно, есть, как минимум, две основные причины, тормозящие широкое распространение доступа в Internet с мобильных устройств. Во-первых, скорость передачи данных не всегда высокая, и, во-вторых, это неудобства, возникающие при просмотре Web-страниц, созданных для больших экранов.

Известно, что для мобильных телефонов, работающих по протоколу WAP, разработчики соответствующих сайтов реализуют структурирование и сокращение объемов данных, которые приходится загружать на мобильные устройства. Этот же принцип используется и в методике Web-клиппинга (Web clipping) для наладонников.

Существуют два подхода к Web-клиппингу — серверный и локальный. В случае серверного клиппинга, который в настоящее время применяется реже, используется специализированный прокси-сервер, который принимает запросы со

стороны Palm, получает данные с Web-сайтов, а затем пересылает “карманнику” сжатый ответ. С помощью специального приложения пользователь генерирует локальный запрос в специальном формате PQA на получение данных из Сети. Этот запрос передается на прокси-сервер, который определяет, что именно требуется, переходит на соответствующие Web-сайты и извлекает необходимые данные. Прокси-сервер сжимает данные и пересылает их обратно на Palm, где ответ отображается с помощью приложения обработки запросов.

Локальный Web-клиппинг реализуется программными приложениями на самих Palm. Например, недавно выпущен программный пакет Mobile Internet Kit (MIK), позволяющий использовать PDA для работы с множеством ресурсов Сети. Он включает около 450 программ, в том числе реализующих Web-клиппинг для доступа с Palm к таким ресурсам, как Amazon.com, Britannica, MapQuest.com и многим другим. Кроме того, MIK включает приложения для полноценной работы с электронной почтой, SMS-сервисом, WAP-сайтами.

Корпорация Microsoft выработала собственный подход, предложив новое программное обеспечение для Pocket PC — Pocket Internet Explorer для Pocket PC. Это решение идеологически близко к локальному Web-клиппингу. Браузер Pocket Internet Explorer, входящий в состав ПО Pocket PC корпорации Microsoft, предлагает доступ к содержимому оригинальных страниц Web-сайтов, поскольку способен переформатировать страницы так, чтобы они наилучшим образом отображались на цветном экране размером 320x240 пикселей, которым оснащен Pocket PC.

### 3.12.8. Информационные ресурсы для КПК

С развитием технологии GPRS стали активно разрабатываться и КПК-версии самых разнообразных Web-сайтов Internet. При создании таких ресурсов всегда учитываются две основные особенности наладонников — ограниченные размеры экрана и скудные возможности клавиатурного ввода. Интерфейс типичного PDA-сайта, как

правило, представляет собой основное меню, перемещаясь по которому, посетитель выбирает ту позицию (раздел, тематику или группу товаров), которая его интересует. Несмотря на это, создатели как правило стремятся к тому, чтобы PDA-версии принципиально не отличались от Web-сайтов по своему наполнению (контенту).

В последнее время все основные мировые поисковые системы и интеграторы контента представили PDA-варианты своих Web-сайтов (например, Google, Yahoo, Яндекс, Lenta.ru, Newsru и многие другие).

PDA-поисковик от Google, оптимизированный под Palm, находится по адресу <http://www.google.com/palm>. Правила и условия работы с этой информационно-поисковой системой приведены по адресу <http://www.google.com/wireless/pda.html>.

Другой популярный поисковый портал Yahoo! также имеет PDA-версию, расположенную по адресу <http://wap.oa.yahoo.com>



Рис. 3.13. Мобильный портал от Yahoo!

(рис. 3.13). У пользователей имеется возможность персонализации, работы с электронной почтой, получения оперативных новостей, информации о финансах, спорте, погоде и кинофильмах.

PDA-вариант поисковика Яндекс находится по адресу <http://yandex.ru/yandsearch?useie5=1>. Основной интерфейс представляет собой окно для ввода запроса. В результате обработки запроса на PDA-сайте выводятся заголовки релевантных документов (по 10 документов на страницу), которые представляют собой гиперссылки на соответствующие Web-ресурсы.

Известный российский новостной сайт LENTA.RU имеет PDA-версию, размещенную по адресу <http://pda.lenta.ru>, газета “Известия” — <http://pda.izv.info>, питерский филиал “Московского комсомольца” — <http://www.mk-piter.ru/pda>, газета “Деловая панорама” — <http://pda.dpw.ru>.

Создаются и специализированные поисковые сайты, ориентированные на пользователей КПК, посвященные карманным компьютерам, программам для них, мобильным устройствам и технологиям. Так, в России поисковая система PDANewsCollector.ru дополняет поисковый сервис возможностями оперативного просмотра новостей большинства популярных российских и украинских сайтов, посвященных тематике мобильных устройств.

Обширный список русскоязычных ресурсов для КПК от Антона Носика находится по адресу <http://pda.lenta.ru/info/pdalinks.htm>. Приведем лишь некоторые, наиболее информативные из них.

- [pda.utro.ru](http://pda.utro.ru) — Утро.Ру, карманная версия.
- [palm.newsru.com](http://palm.newsru.com) — новости от NewsRu.Com.
- [pda.gismeteo.ru](http://pda.gismeteo.ru) — погода от GISmeteo.Ru.
- [r0.ru](http://r0.ru) — облегченный Рамблер.
- [pda.mail.ru](http://pda.mail.ru) — карманный Mail.Ru.
- [r0.ru/\\_lgmail.html](http://r0.ru/_lgmail.html) — карманная почта Рамблера.
- [www.google.com/palm](http://www.google.com/palm) — поиск в Google для PDA.
- [www.hpc.ru/pda/links](http://www.hpc.ru/pda/links) — каталог сайтов для КПК/PDA
- [www.77.ru/ppc.php](http://www.77.ru/ppc.php) — карта Москвы и адресный поиск.
- [www.pocketpcrussia.com/pda](http://www.pocketpcrussia.com/pda) — форум для владельцев КПК/PDA.
- [pda.translate.ru/?lang=ru](http://pda.translate.ru/?lang=ru) — он-лайнный переводчик.
- [pda.eda-server.ru](http://pda.eda-server.ru) — информационно-поисковый сервер по продуктам питания.
- [pda.lenta.ru](http://pda.lenta.ru) — версия новостного сайта Lenta.ru.
- [pda.izv.ru](http://pda.izv.ru) — газета “Известия”.
- [www.mk-piter.ru/pda](http://www.mk-piter.ru/pda) — питерский филиал газеты “Московский комсомолец”.
- [pda.dpw.ru](http://pda.dpw.ru) — газета “Деловая панорама”.

Еще один интересный ресурс, адаптированный под наладонники, — это он-лайнный переводчик, который находится по адресу <http://pda.translate.ru/?lang=ru>.



Рис. 3.14. Новостной сайт [pda.uaport.net](http://pda.uaport.net)

Появляются и Internet-магазины, ориентированные на “мобильных” клиентов. Так, например, недавно в России компания Торман.Ru открыла PDA-версию своего магазина — [www.pda.topman.ru](http://www.pda.topman.ru). В пресс-релизе, посвященном этому событию, особое внимание обращается на такой технологический момент: “клиенту, посещающему электронный магазин с помощью КПК, не нужно использовать прокрутку, усложнявшую процедуру ознакомления и выбора товара”.

Один из крупнейших русскоязычных новостных ресурсов для КПК размещен по адресу <http://pda.uaport.net> (рис. 3.14). Сайт предназначен для предоставления пользователям мобильных устройств оперативной информации и представляет собой PDA-версию новостного раздела поискового портала UAport (<http://infostream.com.ua/>

<http://infostream.com.ua/> rss). Оригинальная технология InfoStream-клиппинга обеспечивает просмотр с экрана мобильных устройств информации с сотен Web-сайтов сети Internet. Доступ к этому информационному PDA-сайту бесплатный и не требует регистрации.

### 3.12.9. Эмуляция мобильности

Информация сайтов, предназначенных для КПК, доступна как с помощью многочисленных программ-эмуляторов для ПК, так и непосредственно с экранов Web-браузеров. Действительно, бывает полезным поработать с компактной информацией, представленной на WAP- или PDA-сайтах, находясь в стационарных условиях у монитора мощного компьютера. Эмуляторы полезны разработчикам специализированных Internet-устройств, программистам КПК, а также потенциальным покупателям, выбирающим своего “карманного питомца”.

Для наладонников типа Palm существует ряд эмуляторов, однако имеется лишь один несомненный лидер среди программ данного класса — Palm OS Emulator (POSE), разработка которого в свое время патронировалась Palm Inc. (рис. 3.15). Программа свободно распространяется и пригодна для разнообразных Palm-платформ. Исходные коды Emulator Application 3.5 для Windows, Mac OS, Unix доступны на [sourceforge.net/projects/pose](http://sourceforge.net/projects/pose). (Скачать их также можно, например, с [palm.com.ua/?files&id=913](http://palm.com.ua/?files&id=913).) С помощью POSE легко моделируется несколько десятков разновидностей КПК, при этом для работы программе требуются “образы” прошивки (ROM) соответствующих КПК. Для эмуляции внешних образов наладонников предусмотрено использование разнообразных “скинов”. В процессе работы на виртуальном КПК можно устанавливать практически любые программные коды, предназначенные для работы на соответствующих Palm.

В последних версиях Web-браузера Opera под Windows и Linux (начиная с седьмой) реализована технология Small-Screen Rendering ([www.opera.com/products/smartphone/smallscreen](http://www.opera.com/products/smartphone/smallscreen)). При нажатии клавиш <Shift+F11> включается режим эмуляции экрана мобильного устройства. Как и в случае с WAP-сайтами, в этом браузере обеспечивается полноценная эмуляция экранов КПК. При этом исключается так называемая “горизонтальная прокрутка”, являющаяся настоящим бичем в случае просмотра Web-ресурсов с мобильных устройств.

В отличие от WAP-сайтов, информация которых адекватно не отображается основными браузерами настольных ПК, страницы сайтов для наладонников представлены в формате HTML (пусть и не всегда последней версии; например, большинство браузеров для Palm-устройств корректно интерпретирует лишь HTML версии 2). Браузер Pocket Internet Explorer настолько приближен к IE для ПК, что практически исключает потребность в использовании соответствующего эмулятора.



Рис. 3.15. Эмулятор POSE

### 3.12.10. RSS-формат на КПК

Технологию RSS News на сегодняшний день поддерживает все большее количество новостных Web-сайтов. Напомним, что формат RSS, являясь унифицированным подмножеством современного языка разметки XML, обеспечивает согласованный способ резюмирования содержимого Web-сайтов. Его применение предоставляет администраторам сайтов новостей, он-лайнных дневников (weblog), форумов и других часто обновляемых Web-ресурсов простой метод подачи информации о происходящих событиях.

RSS можно рассматривать и как формат, ориентированный, прежде всего, на публикацию и обеспечение экспорта новостей. После того как информация преобразована в формат RSS, ориентированная на него программа может загружать новые версии Web-сайтов и выполнять определенные действия, например автоматически обновлять список актуальных информационных сообщений. Такие программы называют RSS-агрегаторами. Они выполняют синтаксический разбор (парсинг) данных, представленных в формате RSS, после чего могут реализовать любые действия по отношению к этим данным, опираясь на полученные результаты. Данные в формате RSS представляют собой каналы или фиды (feed-файлы), в которых записывается новостная информация Web-сайта. Соответственно, если есть необходимость оперативно отслеживать изменения на Web-сайте, не посещая его, то можно подписаться с помощью программы-агрегатора на этот фид (конечно, если он существует).

Широчайшее распространение формата RSS (на одном только сайте NewsIsFree представлено свыше 6 тыс. каналов с различных, в том числе российских и украинских, новостных сайтов) обусловило появление множества программ-агрегаторов, ориентированных на КПК. Владельцы же КПК, установив на свои устройства RSS-агрегаторы, могут эффективно просматривать новостные файлы в RSS-формате.

Для платформы Palm OS наиболее популярной, пожалуй, является программа Hand RSS компании Stand Alone. С помощью этого “карманного агрегатора” владельцы устройств, работающих под управлением Palm OS, могут читать новости в формате RSS с новостных лент BBC News, CNet News, Fark.com, MSNBC, Salon.com, The Register, Wired и т.д. Скачать демоверсию программы или купить ее можно по адресу [http://www.standalone.com/cgi/prc\\_request.cgi](http://www.standalone.com/cgi/prc_request.cgi).

В качестве еще одного эффективного агрегатора можно назвать программу Quick Palm RSS Reader ([remus.manilasites.com](http://remus.manilasites.com)).



Рис. 3.16. Эмулятор MobileRSS

Из специализированных для Pocket PC можно назвать агрегатор новостей в RSS/RDF PocketFeed (<http://www.furrygoat.com/Software>). Пятнадцатидневную демоверсию еще одной программы для этой платформы (PocketPC 2002 и Windows Mobile 2003) — PocketRSS 1.3 — можно скачать на сайте <http://www.happyjackroad.com/AtomicDB/pocketpc/pocketRSS/pocketRSS.asp>.

Но вовсе необязательно устанавливать программу-агрегатор прямо в наладоннике. Как и для WAP-сайтов, в этом случае также существуют серверные решения, выполняющие всю работу по интерпретации RSS-фидов и преобразованию результатов в формат, пригодный для КПК. Один из лучших сайтов подобного назначения — MobileRSS ([mobilerss.net](http://mobilerss.net) — рис. 3.16).

Для работы с этим бесплатным сервером необходима формальная авторизация. Зарегистрированный клиент вводит и активизирует адреса необходимых ему RSS-фидов, после чего просматривает их в свободном режиме. Этот зарубежный сервис, помимо прочего, обеспечивает и корректную работу с кириллическими шрифтами.

### 3.12.11. Игрушка или рабочий инструмент

По многим причинам (прежде всего, благодаря внедрению технологии GPRS, обеспечивающей скоростной доступ в Internet) наладонники завоевывают внимание всех целевых групп — от детей и домохозяек до корпоративных клиентов. Так, число пользователей КПК в Европе составляет 4% от всего населения. Производители “карманников” за счет недополучения прибыли пытаются развить рынок КПК, захватить на нем лидирующие позиции. В минувшем году они поставили своей целью обеспечить продажу полнофункциональных карманных устройств по цене ниже \$300. Так, современная модель Palm Zire продается всего за 130 евро. В Palm Inc. полагают, что такой агрессивный маркетинг позволит удвоить число



европейцев, пользующихся КПК. Hewlett-Packard объявила о выпуске новых моделей iPAQ, работающих под управлением ОС Pocket PC производства Microsoft (рис. 3.17). Ожидается, что розничная цена iPAQ h1910 в США составит \$299. Чуть раньше о намерении выпустить свой наладонник на основе Pocket PC заявила компания Dell. Ее детище будет продаваться по демпинговой цене \$199.

Во многих странах наладонники широко применяются различными ведомствами. К примеру, в армии США стандартом являются КПК Palm. В московской милиции с помощью КПК обеспечивается доступ к различным информационным массивам, включающим текстовую и графическую информацию (“Розыск лиц”, “Паспорта”, “Оружие”, “Угон” и др.).

Сегодня стало действительно велением времени получение мобильного, оперативного и дешевого доступа к актуальной информации. С помощью PDA-устройств, таких как Palm или Pocket PC, пользователи получают простой и надежный доступ к самой актуальной информации из Internet/intranet-сетей в удобном, компактном виде.



Рис. 3.17. Pocket PC от Hewlett-Packard

### 3.13. Службы доставки новостей по электронной почте

Сегодня, когда информационное пространство World Wide Web развивается феноменальными темпами, что может быть реальной альтернативой онлайн-методам получения актуальной информации из Internet? Казалось бы, ставшие традиционными методы доставки информации по электронной почте утрачивают свою актуальность. Однако это не совсем так. Крупнейшие информационные службы рассматривают электронную почту как один из самых надежных каналов распространения своего контента.

#### 3.13.1. История сервиса

Электронная почта (e-mail), безусловно, была первым механизмом доставки информации в Internet. Это первый сервис Сети, возникшей свыше трех десятилетий тому назад. Сразу же, наряду с персональной перепиской, возникла необходимость информирования групп людей, которая нашла свое выражение в так называемых списках рассылок (Mail lists), реализующих рассылку сообщений одновременно по нескольким электронным адресам. До настоящего времени практически все ведущие информационные агентства мира используют технологию списков рассылки, поддерживаемую сотнями специализированных пакетов программ, интегрируемых с биллинговыми системами и с системами документооборота информационных служб. Пользователь, подписываясь на такую рассылку (как платную, так и бесплатную), точно знает, на какую тему он будет получать электронные письма. Другое дело, когда его “подписывают” несанкционированно, после чего следует получение сотен не востребуемых писем рекламного

характера. Такие письма называют спамом, и об этом явлении речь пойдет чуть ниже. Электронной почте, как средству доставки контента, посвящены многочисленные сетевые публикации. Пожалуй, наиболее полным, энциклопедическим источником по этой теме в Рунете можно назвать “Библиотеку электронной почты” (<http://www.mailinfo.ru> — рис. 3.18).

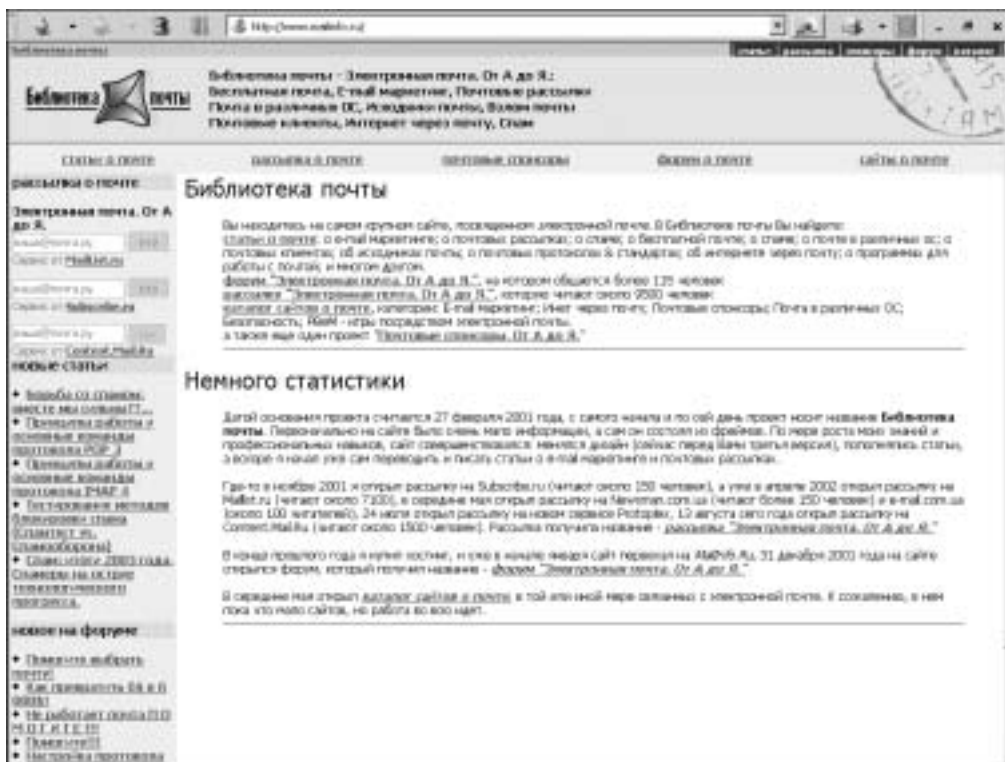


Рис. 3.18. Библиотека электронной почты

В Библиотеке электронной почты содержатся статьи о маркетинге e-mail, почтовых рассылках, спаме, бесплатной почте, почтовых клиентах, протоколах и стандартах и о многом другом. На сервере также размещен форум “Электронная почта. От А до Я”, на котором постоянно общаются свыше 100 пользователей. Рассылку с этого сайта читают около 10 000 подписчиков. На сервере также приведен каталог сайтов об электронной почте, содержащий такие категории, как маркетинг e-mail, Internet через почту, почта в различных ОС, безопасность, игры посредством электронной почты и др.

### 3.13.2. Система телеконференций Usenet

Практически одновременно с Internet появилась технология телеконференций Usenet, которая обеспечила возможность общения и информирования групп пользователей Сети, используя в качестве транспорта преимущественно электронную почту. Абонент может подписаться на рассылку информации из телеконференций, используя как для подписки, так и для получения информации

e-mail, а также в рамках технологии Usenet размещать собственную информацию в уже существующих телеконференциях или создавать собственные. В настоящее время в Сети на Usenet-серверах (их еще называют News-серверами) существуют десятки тысяч доступных абонентам телеконференций. Телеконференции Usenet размещаются на большинстве Internet-узлов посредством реализации механизма обмена данными, с помощью которого каждый подписчик имеет возможность обратиться к интересующей его телеконференции, получить поступающие новости или послать свою информацию средствами электронной почты. При этом Usenet в целом может рассматриваться как динамическая распределенная база данных.

Протокол как набор правил и команд, в соответствии с которым в сети Internet обеспечивается доступ к телеконференциям Usenet, называется NNTP (Network News Transport Protocol). При работе с электронной почтой в режиме off-line используются стандартные шлюзы между системой новостей и электронной почтой, а также серверами новостей. Эти серверы (шлюзы "mail-news") имеют свои электронные адреса, по которым им пишутся письма-запросы. Как правило, провайдер, к которому подключаются абоненты, предоставляет им такую услугу, как доступ к своему почтовому серверу новостей и NNTP-серверу. Но как получить доступ к информации телеконференций Usenet, которые не представлены на серверах конкретного провайдера? Для этого достаточно указать в настройках агента новостей имя другого сервера новостей (NNTP-сервера), к которому обеспечивается свободный доступ. Такие серверы, "открытые для всех" (или Public News-Server), существуют в Internet в большом количестве, а их актуальный перечень можно найти, например, на таких Web-страницах: <http://www.greenline.it/news.htm> или [http://www.newsservers.net/free\\_news\\_servers](http://www.newsservers.net/free_news_servers) (рис. 3.19).

Для поиска открытых серверов новостей существует поисковый механизм, находящийся по адресу <http://freenews.maxbaud.net>, где в качестве запроса вводится название группы новостей. В результате обработки запроса на этом сервере выводится список открытых серверов, поддерживающих запрашиваемую телеконференцию Usenet.

Средства Usenet оказались достаточно эффективным инструментом для распространения платной информации, поэтому в свое время получили широкое распространение телеконференции коммерческой службы ClariNet, распространяющей на платной основе информацию ведущих мировых информационных агентств.

В Украине уже десять лет существует электронная газета ElVisti.Info, которая содержит около 30 новостных телеконференций, доступных на платной основе пользователям украинских Internet-узлов (рис. 3.20). Разделы ElVisti.Info поступают в виде телеконференций, доступных с помощью различных серверов новостей в режиме полной подписки (feed). Объем поступающей на компьютер пользователя текстовой информации из ElVisti.Info составляет до 5 Мбайт в сутки.

Что же касается бесплатных, слабо модерлируемых телеконференций, то они в последнее время стремительно теряют свою актуальность, так как стали площадкой для спамеров, которые, с одной стороны, публикуют в них гигабайты коммерческой рекламы, а с другой стороны, пытаются отлавливать в них электронные адреса наивных пользователей, использующих телеконференции по их прямому назначению.

Несмотря на то, что технология Usenet, появившись значительно раньше Web-технологии, имеет собственные механизмы и средства доступа к информации, сегодня можно констатировать, что, наряду с электронной почтой, и Web-



Рис. 3.19. Список доступных NNTP-серверов на сайте NewsServers

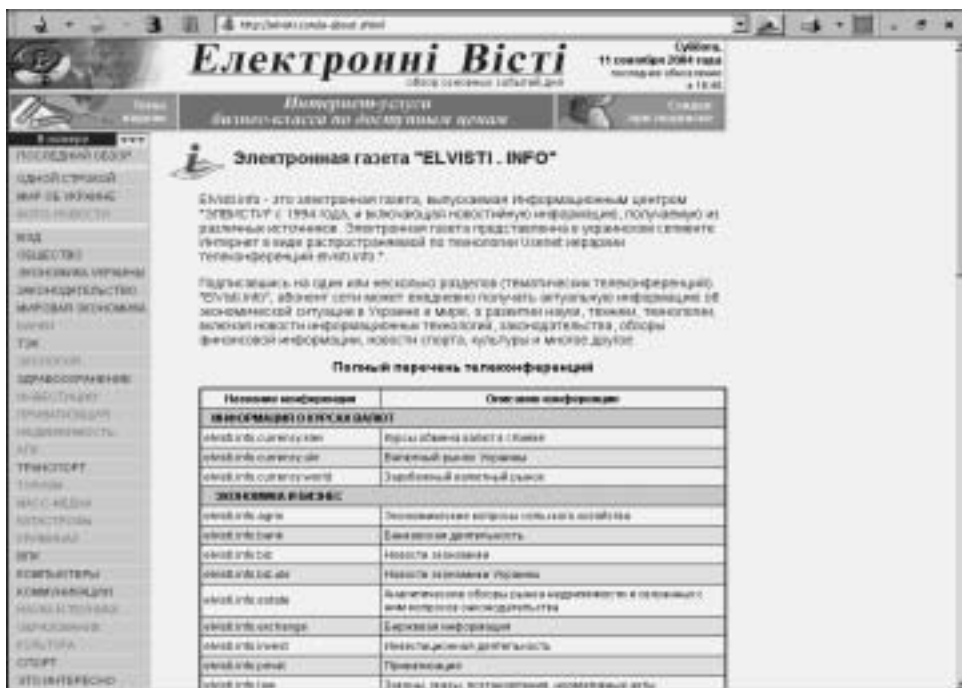


Рис. 3.20. Телеконференції ElVisti.Info

механизмы (engine) представляют доступ к ресурсам Usenet. Сегодня существуют тысячи серверов, обеспечивающих интеграцию этих двух технологий. Многие из таких серверов, кроме того, обеспечивают поисковые возможности, которых так не хватало системе Usenet на этапе ее становления. В качестве примера специализированного поискового сервера по телеконференциям Usenet можно назвать NewzBot (<http://www.newzbot.com> рис. 3.21).

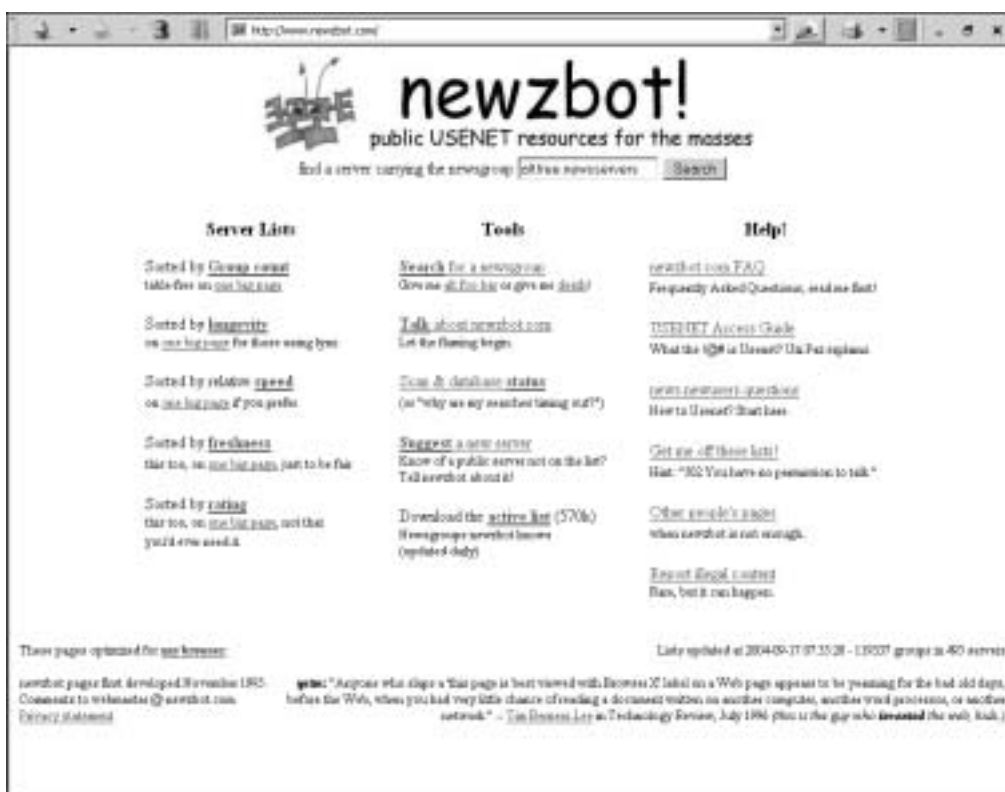


Рис. 3.21. Специализированная поисковая система NewzBot

Кроме того, многие универсальные сетевые поисковые службы обеспечивают полноценную работу с Usenet. В частности, крупнейшая поисковая служба Google со своего сайта “Google Groups” (<http://groups.google.com>) обеспечивает как просмотр списка и актуальности заполнения отдельных телеконференций, так и полноценный информационный поиск (рис. 3.22).

### 3.13.3. Доставка новостей с отдельных сайтов

Едва ли не стандартной функцией на современных Web-серверах и порталах является подписка на новости. Пользователь может зарегистрироваться на том или ином сервере и указать свой электронный адрес для получения необходимой, как ему кажется, рассылки. В распоряжении администраторов сайтов имеется широкий спектр программ рассылки; кроме того, они могут обратиться к специализированным службам рассылки, отдавая эту функцию на аутсорсинг.

Для пользователя подписка может привести к нескольким возможным “мелким неприятностям”. Во-первых, электронный адрес абонента может указать не он сам, а, например, “доброжелатель” (которым может оказаться и незлой администратор Web-сайта). Существует несколько подходов, позволяющих обойти такую принудительную подписку. Конечно же, стоит подписываться только на новости с авторитетных, надежных Web-сайтов. Администраторы корректных подписок обеспечивают возможность отказа от получения новостей как через Web-интерфейс, так и с помощью электронного письма. С другой стороны, корректная рассылка начинается всегда с обмена письмами, подтверждающими подписку.



Рис. 3.22. Сервер Google Groups

Во-вторых, даже на самых надежных Web-сайтах возможна утечка информации. Электронные адреса подписчиков могут попасть в руки спамеров, и тогда на подписчика могут обрушиться потоки невостребованной рекламы.

И, наконец, в-третьих, указанная в явном виде возможность отказа от рассылки не всегда срабатывает, то ли по техническим причинам, то ли по “забывчивости” администратора рассылки. Абонент может годами получать обновления с сайта мебельного магазина, хотя необходимое ему кресло уже давно было куплено. Грубо говоря, предоставление пользователю возможности отписки от новостей не входит непосредственно в область бизнес-интересов администраторов Web-сайтов, а относится лишь к элементарной сетевой этике.

## Пополнение баз данных

Механизмы доставки информации по электронной почте широко используются в технологиях поддержки баз данных, которые ведутся на локальных компьютерах пользователей и регулярно пополняются. Например, для обновления баз данных справочников и классификаторов Российского Информационного Комплекса (сервер “Ваш информационный мир офф-лайн” — [http://info.rosinfocom.ru/nws\\_distribution.php](http://info.rosinfocom.ru/nws_distribution.php)) при установленной программной оболочке достаточно подписаться на рассылку новостей.

Эта же технология используется для обновления законодательных баз данных. Например, в Украине таким образом пополняются нормативно-правовые базы данных компаний ЛИГА (<http://www.liga.kiev.ua>) или НАУ (<http://nau.kiev.ua>).

### 3.13.4. Специализированные службы рассылки новостей

Несмотря на бурное развитие Web-технологий, сегодня лучший способ бесплатно и без особых усилий найти целевую аудиторию в Internet — это создание собственной почтовой рассылки. В России и Украине есть несколько серверов служб бесплатных почтовых рассылок, где можно открыть рассылку, а также организовать для нее рекламу. Предоставляемыми службами рассылок могут воспользоваться все, кому есть чем делиться с широкой аудиторией, при соблюдении некоторых условий, в том числе обеспечивая периодичность выпусков.

Службы рассылок предоставляют такие сервисные возможности:

- как правило, удобные механизмы рассылок, которые постоянно развиваются;
- организация доставки писем подписчикам в различных форматах и кодировках, формирование архивов, ведение статистики, поддержка рассылок на различных языках;
- возможность для подписчиков управления всеми аспектами своей подписки.

Количество подписчиков — самый очевидный показатель популярности рассылки. Рейтинг различных рассылок определяется по признаку, называемому “он-лайн-активностью подписчиков” или просто “активностью”, которая рассчитывается как отношение количества зафиксированных чтений рассылки к тиражу рассылки.

В настоящее время в русскоязычном сегменте Сети существует три ведущие службы почтовых рассылок:

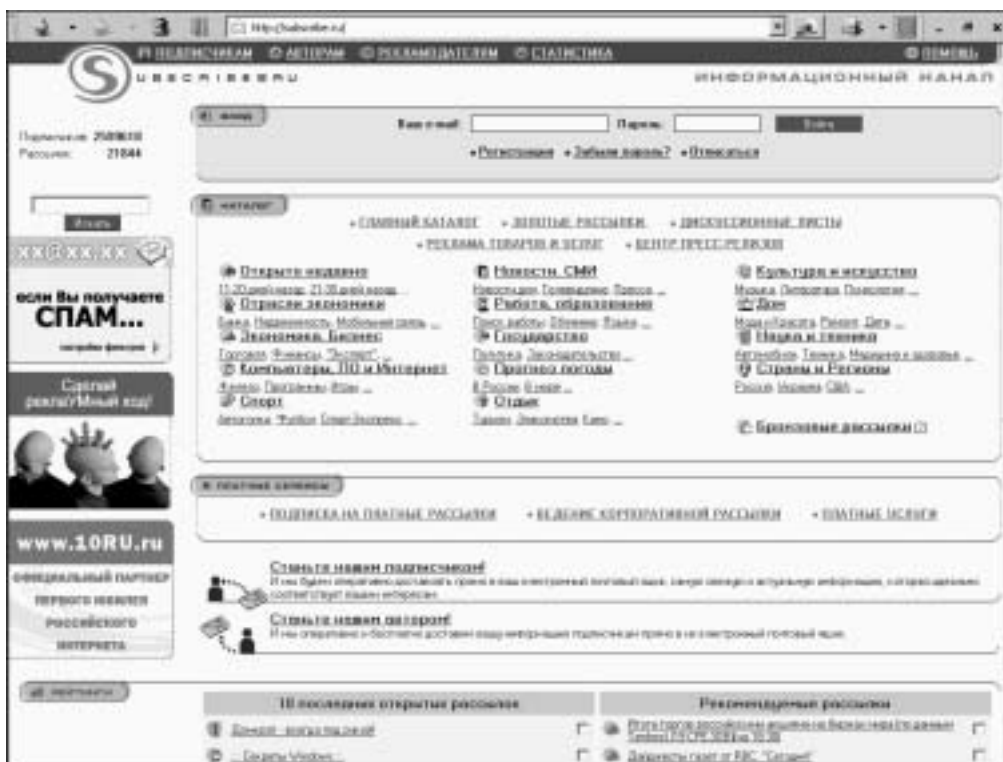
- <http://subscribe.ru>
- <http://content.ru>
- <http://maillist.ru>

Стоит упомянуть и об украинской службе <http://newsman.com.ua>. Некоторая статистика по названным службам приведена в табл. 3.1.

Без сомнения, рассылки Городского кота (<http://subscribe.ru>) — это самый популярный в Рунете сервис. Информационный канал Subscribe.Ru — старейший и один из крупнейших в русскоязычном сегменте Сети. Для продвижения новых рассылок, примерно раз в неделю служба рассылает новости тиражом около 500 тыс. писем с описаниями новых рассылок. Служба Subscribe.ru предоставляет форму для поиска в архивах рассылок, которые хранятся на самом сервере (рис. 3.23).

**Таблица 3.1. Статистические показатели важнейших почтовых рассылок России и Украины**

Служба почтовых рассылок	Subscribe.ru	Content.ru	MailList.ru	Newsman.com.ua
Дата открытия	1998	2000	1999	1999
Количество подписчиков	2 240 000	2 630 000	500 000	54 000
Количество рассылок	17 600	12 000	15 000	1700



*Рис. 3.23. Служба Subscribe.Ru*

Помимо бесплатных услуг, служба Subscribe.ru предоставляет ряд платных. Платные услуги подразделяются на собственные платные услуги (“Рассылки без рекламы”, “Персональная отправка”, “POP-ящик для подписки” и другие) и платные рассылки — информация от нескольких десятков информационных агентств, партнеров службы.

Очень несложно открыть рассылку и на <http://content.ru> — рис. 3.24. Количество рассылок на этом сервере превышает 7 тыс. Здесь отсутствует классификация рассылок, поэтому их ранжирование реализовано только в зависимости от популярности, определяемой тиражом.

На сервере компании Agava SoftWare (<http://maillist.ru> — рис. 3.25) рассылку также открыть достаточно просто, однако она будет зарегистрирована в каталоге только после согласования с администраторами сервера, которые самое большое значение придают названию рассылки. Действительно, удачное



название очень важно при создании собственной рассылки. Даже плохая рассылка может иметь более высокую популярность благодаря удачному названию. На MailList.ru возможно создание коммерческих рассылок, в рамках которых заказчик получает ряд преимуществ по сравнению с владельцем обычных рассылок: в коммерческих рассылках отсутствует реклама службы рассылок, заказчик может полностью управлять содержимым своей рассылки, самостоятельно устанавливать обратный адрес и изменять другие настройки.



Рис. 3.24. Служба Content.ru

Для рассылки тиража очень удобно использовать content.ru и maillist.ru. На этих серверах доступна статистика по количеству. Службы subscribe.ru и content.ru предоставляют специальную кнопку определения тиража для размещения на сайте.

Естественно, большинство из служб рассылки существуют за счет рекламы, размещаемой в некоммерческих рассылках их клиентов. Вместе с тем возможны и коммерческие рассылки, за которые платит клиент-рекламодатель, который вправе использовать в рассылках исключительно собственные рекламные баннеры.

### 3.13.5. Интеграция новостей с целью рассылки

В настоящее время в связи с развитием наполнения Web-пространства новостной информацией возникла необходимость интеграции динамического новостного контента с большого числа сайтов для обеспечения доступа к нему как через

Web-интерфейс, так и средствами электронной почты. Во всем мире стали появляться специализированные информационные службы, занимающиеся сбором и обработкой новостей с Internet-изданий с целью их дальнейшей рассылки.



Рис. 3.25. Служба MailList.ru

Например, известный российский поисковый портал Яндекс недавно открыл проект Яндекс.Новости (<http://news.yandex.ru> — рис. 3.26), к которому в настоящее время присоединились несколько сотен Internet-изданий. Главной особенностью Яндекс.Новости как открытого публичного сервиса является наличие тем, которые объединяют содержательно близкие новости с различных сайтов.

Посетитель Яндекс.Новостей может воспользоваться тематическими разделами (все полученные новости группируются по нескольким рубрикам), а также подписаться на новости определенной тематики или соответствующие конкретному поисковому запросу новости. Поиск новостей возможен как по всем источникам, так и по отдельным источникам, заданным пользователем. Имеется также возможность поиска за произвольный период времени, а также подписка на получение анонсов новостей по электронной почте.

Крупнейшее в России агентство по интеграции новостей Интегрум (<http://www.integrum.ru> — рис. 3.27) обеспечивает сбор в единый массив электронных версий коммерческих, статистических и новостных информационных продуктов. Контент-механизм службы является авторской разработкой агентства — это лингвистическая поисковая система Артефакт, построенная на использовании сложных морфологических алгоритмов. Сервис агентства “Персональная газета” обеспечивает создание и функционирование запросов-

роботов, осуществляющих автоматический поиск и доставку по электронной почте материалов подписчикам по заданным ими запросам — ключевым словам и логическим операторам. Сервис имеет развитую систему настроек отбора по контексту и источникам информации. Каждый запрос обрабатывается системой Артефакт от одного до трех раз в сутки. В результате выбираются предварительно загруженные в базу данных документы, соответствующие запросам, которые высылаются пользователям по электронной почте.

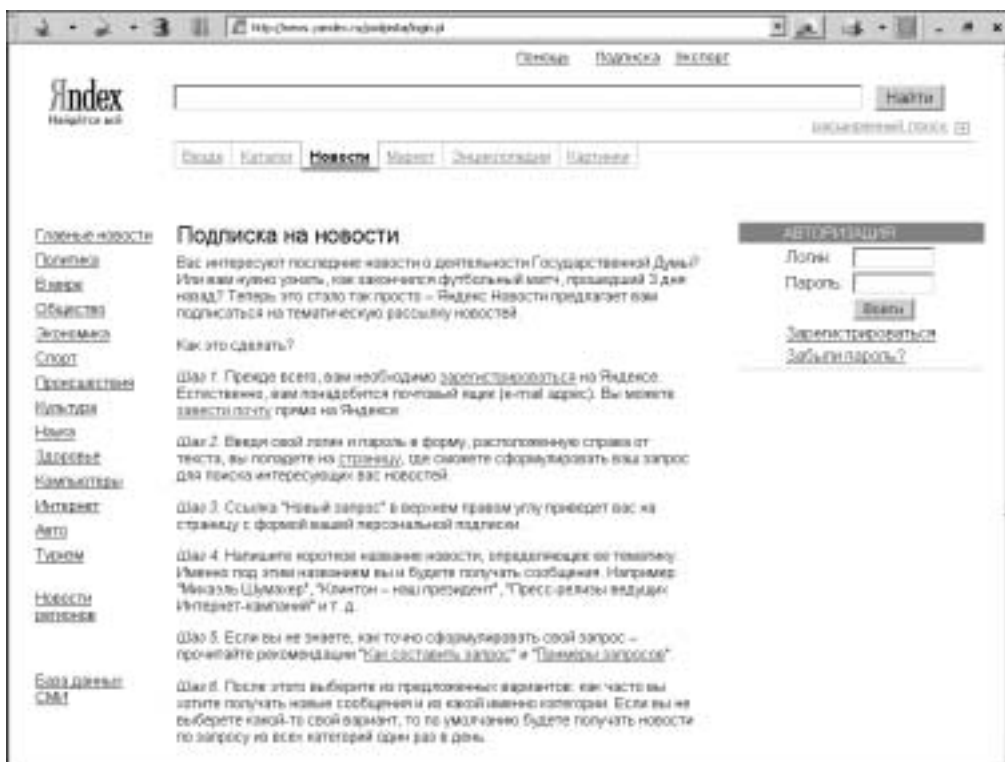


Рис. 3.26. Yandex.Новости

Система InfoStream® (<http://infostream.ua> — рис. 3.28), разработанная украинской компанией EIVisti, предназначена для автоматизированного сбора новостной информации с сайтов, ее обработки, систематизации и обеспечения доступа к ней. Если пользователь хочет получать новостную информацию в режиме избирательного распространения информации по интересующей его тематике (она определяется на языке запросов с помощью ключевых слов, логических операторов, операторов контекстной близости и скобок) по e-mail, SMS или встроить постоянную подборку в свою Web-страницу, то к его услугам сервис InfoStream Client.

Сегодня системой InfoStream охватывается ежедневно свыше 20 тыс. документов из более чем 600 информационных источников, перечень которых постоянно растет. Сведения о новых информационных источниках поступают как непосредственно от разработчика, так и от пользователей сервисов InfoStream. В результате реализуется эффективный механизм обратной связи между службой сопровождения системы и пользователями.



Рис. 3.27. Персональная газета службы Интегрум



Рис. 3.28. Служба InfoStream Client

### 3.13.6. Спам — альтернатива востребованной рассылке

Открытость и доступность электронной почты позволила наводнить Сеть множеством ненужных или малополезных посланий (рекламные сообщения, предложения о посещении тех или иных ресурсов и др.). Очень часто почтовые ящики доверчивых пользователей превращаются в мусорную яму, оказавшись заваленными невостребованными письмами — спамом [24]. Наиболее часто спам трактуется как непрошеное рекламное сообщение или информация, рассылаемые по электронной почте в личные почтовые ящики или телеконференции. Наряду с попытками четкого юридического определения, назовем несколько действий, которые пользователи Internet воспринимают как рассылку спама.

Во-первых, это рассылка (массовая или индивидуальная) почтовых сообщений без предварительного желания адресата получить подобную корреспонденцию. Во-вторых, это подписка пользователя Internet на список рассылки без его ведома. В-третьих, — помещение в телеконференции Usenet, дискуссионные листы, гостевые книги сообщений, не имеющих отношения к их тематикам (off-topic).

Пока спам способен приносить своим создателям хоть какие-то деньги, это явление в Internet будет процветать. Массовый характер несанкционированной рассылки информации сегодня представляет собой одну из самых серьезных проблем Сети, поскольку на обработку, просмотр и удаление непрошенных писем тратится все больше времени. Так, согласно исследованиям компании MessageLabs, британские фирмы из-за спама недополучают \$4,6 млрд прибыли. По их информации, каждый сотрудник компании ежедневно тратит 10 минут на просмотр и удаление спама. Это составляет 95% от общего времени просмотра почты. Борьба со спамом обходится недешево — работодатели теряют по \$472 на каждого работника в год.

### 3.13.7. Перспективы технологий доставки новостей

Электронная почта — это всего лишь один из способов доставки пользователю персонализированного контента. С одной стороны, этот способ не самый прогрессивный, но, безусловно, самый надежный. Сегодня доставку оперативной информации на компьютеры пользователей в режиме он-лайн обеспечивает новый класс программ — агенты новостей. Настройка взаимодействия агентов новостей с Сетью, ориентированная на интересы и потребности каждого конкретного пользователя, делает их очень востребованным продуктом, особенно сегодня, когда на повестке дня стоит персонализация Internet. Клиентские агенты новостей представляют собой приложения, устанавливаемые на компьютерах пользователей и облегчающие поиск и доставку необходимой информации. Они используют, как правило, целый набор традиционных поисковых систем и каталогов.

Пожалуй, наиболее известной утилитой этого класса является *Copernic 2000*. Программа позволяет искать информацию, “паразитируя” на таких поисковых машинах, как *AltaVista*, *DejaNews*, *Euroseek*, *Excite*, *HotBot*, *Infoseek*, *Lycos*, *Yahoo!*, одновременно используя более 30 информационных ресурсов. Некоторые агенты новостей предполагают наличие сразу двух частей — клиентской и серверной. На серверной части выполняется сбор и обработка информации, а на клиентской — настройка и запуск агента. Среди таких систем можно выделить известную российскую систему *News Alert*, которая постоянно сканирует новости из российских Web-ресурсов и обеспечивает их избирательное распространение в режиме реального времени. Клиентская программа *News Alert* имеет размер всего 50 Кбайт, она оповещает о новостях и показывает их анонсы по желанию пользователей.

В последние несколько лет стремительную популярность набирают технологии, ориентированные на использование формата RSS (Rich Site Summary), основное назначение которого — обеспечение однотипного обмена данными в такой сложной системе, как Internet. RSS обеспечивает согласованный способ резюмировать содержимое и обновления Web-сайтов. Именно возможность однотипного экспорта новостей обусловила появление нового типа новостных агентов — RSS-агрегаторов, представляющих собой клиентские программы доступа к данным в формате RSS. С помощью этих программ, например FeedReader (<http://www.feedreader.com>) или Syndirella (<http://www.yole.ru/projects/syndirella>), пользователи могут получать доступ к RSS-данным с помощью интерфейсов, которые удивительно напоминают интерфейсы почтовых программ. При этом аналогия интерфейсов подтверждается и теми преимуществами, которые традиционно присущи электронной почте, а именно:

- информация поступает по запросу пользователя после соединения с соответствующим сервером, а на сервере накапливается по мере поступления;
- информация из почтового ящика может накапливаться в архивах пользователя;
- информация в электронном почтовом ящике максимально персонализирована.